G C A T
T A C G
G C A T

Grand Celebration: 10th Anniversary of the Human Genome Project

Volume 3

Edited by

John Burn, James R. Lupski,
Karen E. Nelson and Pabulo H. Rampelotto

Printed Edition of the Special Issue Published in *Genes*

www.mdpi.com/journal/genes

John Burn, James R. Lupski, Karen E. Nelson and
Pabulo H. Rampelotto (Eds.)

Grand Celebration: 10th Anniversary of the Human Genome Project

Volume 3

MDPI

This book is a reprint of the special issue that appeared in the online open access journal *Genes* (ISSN 2073-4425) in 2014 (available at:
http://www.mdpi.com/journal/genes/special_issues/Human_Genome).

Guest Editors
John Burn
University of Newcastle
UK

James R. Lupski
Baylor College of Medicine
USA

Karen E. Nelson
J. Craig Venter Institute (JCVI)
USA

Pabulo H. Rampelotto
Federal University of Rio Grande do Sul
Brazil

Editorial Office
MDPI AG
Klybeckstrasse 64
Basel, Switzerland

Publisher
Shu-Kun Lin

Assistant Editor
Rongrong Leng

1. Edition 2016

MDPI • Basel • Beijing • Wuhan

ISBN 978-3-03842-123-8 complete edition (Hbk)
ISBN 978-3-03842-169-6 complete edition (PDF)

ISBN 978-3-03842-124-5 Volume 1 (Hbk) ISBN 978-3-03842-170-2 Volume 1 (PDF)
ISBN 978-3-03842-125-2 Volume 2 (Hbk) ISBN 978-3-03842-171-9 Volume 2 (PDF)
ISBN 978-3-03842-126-9 Volume 3 (Hbk) ISBN 978-3-03842-172-6 Volume 3 (PDF)

Table of Contents

List of Contributors

Bjorn T. Adalsteinsson: Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK.

Mauricio Arcos-Burgos: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Euan A. Ashley: Department of Medicine, Stanford University, Stanford, CA 94305, USA.

Graeme C. Black: Manchester Centre for Genomic Medicine, Central Manchester University Hospitals, NHS Foundation Trust, Manchester, M13 9WL, UK.

Cesar L. Boguszewski: Endocrine Division (SEMPR), Department of Internal Medicine, Federal University of Parana, Avenida Agostinho Leão Junior, 285-Alto da Glória. CEP 80030-110, Curitiba-PR, Brazil.

Margaret C.S. Boguszewski: Endocrine Division (SEMPR), Department of Internal Medicine, Federal University of Parana, Avenida Agostinho Leão Junior, 285-Alto da Glória. CEP 80030-110, Curitiba-PR, Brazil.

Guillaume Bourque: McGill University and Genome Quebec Innovation Center, Montréal, QC, H3A 1A4, Canada.

Scott D. Boyd: Department of Pathology, Stanford University, Stanford, CA 94305, USA.

Matthew Brown: University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Road, Woolloongabba, Brisbane 4102, Queensland, Australia.

Margit Burmeister: Molecular & Behavioral Neuroscience Institute/Departments of Human Genetics/Department of Psychiatry/Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

Grégory Caignard: Complex Traits Group/Department of Human Genetics, McGill University, Montréal, QC H3G 0B1, Canada.

Aaron Chuah: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Jeffrey M. Craig: Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, Victoria 3052, Australia; Department of Paediatrics, The University of Melbourne, Parkville, Victoria 3052, Australia.

Christiaan de Leeuw: Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; Institute for Computing and Information Sciences, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.

Emma Duncan: University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Road, Woolloongabba, Brisbane 4102, Queensland, Australia; Department of Endocrinology and Diabetes, Royal Brisbane and Women's Hospital, Herston 4029, Queensland, Australia.

Megan M. Eva: Department of Medicine/Complex Traits Group/Department of Human Genetics, McGill University, Montréal, QC H3G 0B1, Canada.

Robert Eveleigh: McGill University and Genome Quebec Innovation Center, Montréal, QC, H3A 1A4, Canada.

Anne C. Ferguson-Smith: Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK.

Donald Freed: Program in Biochemistry, Cellular and Molecular Biology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA; Department of Neurology, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205, USA.

Stephen J. Galli: Department of Microbiology and Immunology/Department of Pathology, Stanford University, Stanford, CA 94305, USA.

Richard A. Gibbs: Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.

Philippe Gros: Department of Biochemistry/Complex Traits Group, McGill University, Montréal, QC H3G 0B1, Canada.

Anke R. Hammerschlag: Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.

Gavin A. Huttley: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Angad S. Johar: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

David P. Kelsell: Centre for Cutaneous Research, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK.

Julio Licinio: Mind and Brain Theme, South Australian Health and Medical Research Institute, and Department of Psychiatry, School of Medicine, Flinders University, PO Box 11060 Adelaide SA 5001, Adelaide, Australia.

Danielle Malo: Department of Medicine/Complex Traits Group/Department of Human Genetics, McGill University, Montréal, QC H3G 0B1, Canada.

Thiviyani Maruthappu: Centre for Cutaneous Research, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK.

Claudio A. Mastronardi: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Amy L. McGuire: Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA.

Jason D. Merker: Department of Pathology, Stanford University, Stanford, CA 94305, USA.

Thomas Mikeska: Genetic Technologies Ltd., Fitzroy, Victoria 3065, Australia.

William G. Newman: Manchester Centre for Genomic Medicine, University of Manchester, Manchester, M13 9WL, UK.

Hardip R. Patel: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Gilberto Paz-Filho: Genome Biology Department, The John Curtin School of Medical Research, The Australian National University, Garran Rd, building 131, Acton, Canberra, ACT 0200, Australia.

Stacey Pereira: Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA.

Jonathan Pevsner: Department of Psychiatry and Behavioral Sciences/Program in Biochemistry, Cellular and Molecular Biology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA; Department of Neurology, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205, USA.

Tinca J. C. Polderman: Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.

Danielle Posthuma: Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; Department of Child and Adolescent Psychiatry, Erasmus University Medical Center and Sophia Children's Hospital, P.O. Box 2060, 3000 CB Rotterdam, The Netherlands; Department of Clinical Genetics, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands.

Erin Sandford: Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA.

Iris Schrijver: Department of Pediatrics, Stanford University, Stanford, CA 94305, USA.

Claire A. Scott: Centre for Cutaneous Research, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK.

Eileen M. Shore: Department of Genetics, Perelman School of Medicine/Center for Research in FOP and Related Disorders, Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, 3450 Hamilton Walk, Philadelphia, PA 19104, USA.

August B. Smit: Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.

Eric L. Stevens: Department of Neurology, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205, USA; Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA; Present address: CFSAN Division of Microbiology, Food and Drug Administration, College Park, MD 20740, USA.

Henning Tiemeier: Department of Child and Adolescent Psychiatry, Erasmus University Medical Center and Sophia Children's Hospital, P.O. Box 2060, 3000 CB Rotterdam, The Netherlands.

Rebekah van Bruggen: Department of Biochemistry/Complex Traits Group, McGill University, Montréal, QC H3G 0B1, Canada.

Matthijs Verhage: Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; Department of Clinical Genetics, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands.

Silvia M. Vidal: Complex Traits Group/Department of Human Genetics, McGill University, Montréal, QC H3G 0B1, Canada; McGill Life Sciences Complex, Bellini Building, 3649 Sir William Osler Promenade, Room 367, Montreal, QC, H3G 0B1, Canada.

Tonya White: Department of Child and Adolescent Psychiatry, Erasmus University Medical Center and Sophia Children's Hospital, P.O. Box 2060, 3000 CB Rotterdam, The Netherlands.

Ma-Li Wong: Mind and Brain Theme, South Australian Health and Medical Research Institute, and Department of Psychiatry, School of Medicine, Flinders University, PO Box 11060 Adelaide SA 5001, Adelaide, Australia.

James L. Zehnder: Department of Medicine/ Department of Pathology, Stanford University, Stanford, CA 94305, USA.

Preface

In 1990, scientists began working together on one of the largest biological research projects ever proposed. The project proposed to sequence the three billion nucleotides in the human genome. The Human Genome Project took 13 years and was completed in April 2003, at a cost of approximately three billion dollars. It was a major scientific achievement that forever changed the understanding of our own nature. The sequencing of the human genome was in many ways a triumph for technology as much as it was for science. From the Human Genome Project, powerful technologies have been developed (e.g., microarrays and next generation sequencing) and new branches of science have emerged (e.g., functional genomics and pharmacogenomics), paving new ways for advancing genomic research and medical applications of genomics in the 21st century. The investigations have provided new tests and drug targets, as well as insights into the basis of human development and diagnosis/treatment of cancer and several mysterious humans diseases. This genomic revolution is prompting a new era in medicine, which brings both challenges and opportunities. Parallel to the promising advances over the last decade, the study of the human genome has also revealed how complicated human biology is, and how much remains to be understood. The legacy of the understanding of our genome has just begun. To celebrate the 10th anniversary of the essential completion of the Human Genome Project, in April 2013 *Genes* launched this Special Issue, which highlights the recent scientific breakthroughs in human genomics, with a collection of papers written by authors who are leading experts in the field.

John Burn, James R. Lupski,
Karen E. Nelson and Pabulo H. Rampelotto
Guest Editors

Genes and Genetic Testing in Hereditary Ataxias

Erin Sandford and Margit Burmeister

Abstract: Ataxia is a neurological cerebellar disorder characterized by loss of coordination during muscle movements affecting walking, vision, and speech. Genetic ataxias are very heterogeneous, with causative variants reported in over 50 genes, which can be inherited in classical dominant, recessive, X-linked, or mitochondrial fashion. A common mechanism of dominant ataxias is repeat expansions, where increasing lengths of repeated DNA sequences result in non-functional proteins that accumulate in the body causing disease. Greater understanding of all ataxia genes has helped identify several different pathways, such as DNA repair, ubiquitination, and ion transport, which can be used to help further identify new genes and potential treatments. Testing for the most common mutations in these genes is now clinically routine to help with prognosis and treatment decisions, but next generation sequencing will revolutionize how genetic testing will be done. Despite the large number of known ataxia causing genes, however, many individuals with ataxia are unable to obtain a genetic diagnosis, suggesting that more genes need to be discovered. Utilization of next generation sequencing technologies, expression studies, and increased knowledge of ataxia pathways will aid in the identification of new ataxia genes.

Reprinted from *Genes*. Cite as: Sandford, E.; Burmeister, M. Genes and Genetic Testing in Hereditary Ataxias. *Genes* **2014**, *5*, 586-603.

1. Introduction

Ataxia is a neurological sign that involves a lack of coordinated muscle movement, which impacts walking, speech, and vision. Ataxia can present as an isolated symptom, or present as one of many symptoms of a more complex disease. Acquired ataxias may be temporary or permanent, and can be caused by environmental factors, such as alcohol, trauma, or exposure to toxins, or by other underlying medical conditions such as stroke, infection, tumors, or vitamin deficiencies. However, many ataxias have an underlying genetic cause. Hereditary ataxias are a group of highly heterogeneous diseases, but each usually follows a typical Mendelian dominant, recessive, or X-linked inheritance. The prevalence of hereditary ataxias varies by population and has been estimated at 1–9 per 100,000 people [1–4]. Many hereditary diseases also present with ataxia as one symptom of a more complex phenotype. This review will focus on disorders classified primarily as ataxia, along with those ataxias that result in other symptoms like intellectual disability, with known genetic association.

Early work on the genetic origins of ataxia began in 1993 with the discovery of a CAG repeat responsible for spinocerebellar ataxia (SCA) type 1 [5]. Continued screening for CAG repeat expansions identified several additional dominant SCAs that are caused by the same mechanism [6–10]. With the advancement of next generation sequencing technology, genome and exome sequencing have become an affordable option for screening for disease genes. Exome sequencing for Mendelian diseases first gained prominence in 2010 with the discovery of the disease gene for

Miller syndrome and since then, mutations in several new ataxia genes have been identified utilizing exome sequencing, including *ATP2B3*, *KCND3*, *DNMT1*, *UCHL1*, and *TPP1*, illustrating the utility of the technology [11–17]. While mutations in many of these new genes were found in only one family ("private" mutations), thus far, mutations in *KCND3* were found in multiple different families on several continents [13,14]. Despite these advances, it is estimated that up to 40% of those with ataxia do not know the genetic cause, illustrating the need to continue research into the identification of ataxia genes in order to provide a diagnosis and potentially a treatment [18].

2. Phenotypes of Hereditary Ataxias

Hereditary ataxias exhibit a wide range of phenotypes, in both clinical features and age of onset. Some ataxias are described as "pure cerebellar", where symptoms are all related to cerebellar control of muscle movement. This can include ataxic gait and movement of body and limbs, along with nystagmus, dysarthia, and hypotonia. Many of these features are easily observed by external examination. Magnetic resonance imaging often provides the clearest explanation for the ataxia through the identification of cerebellar atrophy, but may appear normal in some cases [19,20]. Other ataxias can present with more extensive additional neurological symptoms, such as Parkinsonism, epilepsy, dementia, and neuropathy. Multisystem involvement can include symptoms such as deafness and intellectual disability. These symptoms may be progressive, gradually becoming more severe over time, or non-progressive, where the symptoms are stable.

The age of symptom onset in affected individuals can vary dramatically, both within and across different ataxias, with symptoms present from birth through onset in the 7th and 8th decades of life. Late onset ataxias are more commonly progressive and can result in patients becoming wheelchair bound or even experience a reduced lifespan. Congenital ataxias display symptoms within the first year of life and are often non-progressive, however many congenital ataxias more often present as multisystem diseases. These children may display muscular hypotonia prior to onset of ataxia symptoms, resulting in "floppy baby syndrome".

A common phenomenon in the dominantly inherited ataxias is anticipation, where the younger generation exhibits symptoms at an earlier age. The rate of anticipation can vary, depending on genetic and environmental factors, but differences in age of onset, up to 20 years, have been reported. Much, but not all, of anticipation can be explained by increasing repeat length of the CAG expansions (see Section 3.1.1). Anticipation can be difficult for clinicians to correctly diagnose, as younger individuals with a family history of ataxia may describe more psychosomatic symptoms in the expectation of developing symptoms later in life.

3. Ataxia Genetics

Hereditary ataxias are genetically and phenotypically heterogeneous. Similar phenotypes may be caused by mutations in many different genes, and several genes cause different types of ataxia depending upon the mutation. While many ataxias appear worldwide, such as Friedreich's ataxia or SCA3, others are more common in one population. Dentatorubral-pallidoluysian atrophy (DRPLA) is most common in Japan and SCA2 is prevalent in Cuba. Other ataxias may be completely restricted to certain populations, such as Cayman ataxia on the Cayman Islands. Knowledge of a patient's ethnic origin can, therefore, be helpful, along with phenotype and family history. In most newly diagnosed cases with ataxia, screening panels for many ataxia genes is recommended. As ataxia can be misdiagnosed as multiple sclerosis or Parkinson's, finding a genetic cause often solidifies a diagnosis, not only for an individual, but for the whole family.

3.1. Autosomal Dominant

Many dominant ataxias have been classified as SCAs or episodic ataxias (EA). At least 34 different SCAs and seven EAs have been described clinically, with 28 having known associated genetic mutations. Dominant ataxias tend to have an onset later in life and be slowly progressive. SCAs, particularly those caused by repeat expansions, can exhibit a larger range of symptom onset and a faster rate of progression. A detailed review of the clinical characteristics of SCAs was published in 2009 [21]. Individuals with EA experience episodes of ataxia that can range from minutes to hours in duration and are triggered by environmental stimuli such as stress, alcohol, or exercise [22–24].

In some cases, the causative gene was identical in several previously reported SCAs, so SCA15, SCA16, and SCA29 all are caused by mutations in *ITPR1* and SCA19 and SCA22 are caused by mutations in *KCND3* [13,14,25–27]. Repeat expansion in *CACNA1A* results in SCA6 while single nucleotide variants (SNV), insertions, and deletions result in EA2. Known autosomal dominant (AD) ataxia genes are reported in Table S1.

3.1.1. Repeat Expansions

The most common forms of dominant ataxias are caused by repeat expansion. Short repeats, typically three to six bases long, appear at variable repeat number within many genes. Occasionally these repeat regions become unstable during replication, leading to either deletions of repeats, which rarely causes problems, or to expansion of the number of repeats. Typically within a repeat region, there are instances of non-repeated bases, such as a CAA in a string of CAG. Mutations that convert these imperfections in the repeat region to match the surrounding repeats result in an unstable sequence and increased likelihood of expansion. In ataxias, the number of repeats may increase anywhere from less than 2 to over 100 fold, depending on the gene. The most common repeat expansions are CAG expansions. As CAG encodes glutamine, these are also referred to as a polyglutamine or polyQ repeats, as these repeats form strings of glutamines (Q) in the coding region. There are currently seven known AD ataxias caused by CAG polyglutamine expansions: SCA1, SCA2, SCA3 (also known as Machado Joseph disease or MJD), SCA6, SCA7, SCA17, and DRPLA. In addition, repeat expansions outside the coding region, in introns or the untranslated regions of the gene, also can cause ataxia without causing polyglutamine disease, but rather by interfering with the regulation of the gene: SCA8 (CTG), SCA10 (ATTCT), SCA12 (CAG), SCA31 (TGGAA), and SCA36 (GGCCTG).

The most common SCAs reported are SCA1, SCA2, SCA3, SCA6, and SCA7. Rates for each vary by population; the National Ataxia Foundation reports that SCA6 is responsible for up the

30% of dominant ataxia cases in Japan, but only 15% in the U.S. and 2% in Italy. Together these five SCA make up about 60% of the reported dominant cases of ataxia [21]. With the high frequency of these SCAs, it is not surprising that they were the first genetic mutations responsible for SCA that were identified.

Age of onset is highly variable with repeat expansion disorders, ranging from early childhood to the later decades of adulthood. There is an inverse correlation between repeat length and age of onset, with longer repeats resulting in symptoms at a younger age. Expansions often increase in length in each subsequent generation, leading to a phenomenon called anticipation, where the next generation starts exhibiting symptoms at an earlier age than the previous. Reduction of repeat length has been reported but this occurs more rarely. Many individuals have repeats at an intermediate length, resulting in incomplete penetrance of the disease, but these are more likely to expand in future generations. Expansion of repeat regions can therefore appear as sporadic cases when the repeat is newly expanded, as this individual may have no other affected family members.

Repeat expansions cause disease through toxic gain of function. This gain of function can allow expanded proteins to avoid degradation, exhibit changes in expression, and influence function of other interacting genes [28–30]. Recently, it has been demonstrated in SCA8 and FXTAS, an X-linked ataxia, that RNA can be translated independent of a traditional ATG start site, in a process referred to as repeat-associated non-ATG translation, contributing to the harmful effects of aberrant proteins [31–33].

3.1.2. Other Mutations in AD Ataxias

Several of the other more recently discovered dominant ataxias are caused by conventional mutations: SNVs, insertions, and deletions. Conventional mutations are much less common in AD ataxias than repeat expansions. Several mutations have only been reported in select populations or families, while others appear to exist worldwide. SCA28, for example, causes 1.5% of AD ataxia in Europeans, but has not been detected in other large populations such as Chinese [34,35]. Dominant mutations can result in disease through haploinsufficiency due to gene deletion or disruption of functionally important residues, or by dominant negative mechanisms. Although more rare than in repeat expansions, anticipation has been documented in cases of indel or SNV mutations. The mechanism behind anticipation in ataxia due to indel or SNV mutations is unknown.

The variety of mutation types present in dominant ataxias illustrates the need for careful attention to molecular assays used to screen for new mutations. *ITPR1* was initially discarded as a candidate gene but later reassessment of the same samples detected the disease-causing deletion [25,36]. The confirmation of *ITPR1* as an ataxia causing gene in humans led to the careful screening and discovery of mutations in SCA16 and SCA29 patients [26,27].

3.2. Autosomal Recessive

Autosomal recessive (AR) ataxias occur more frequently than AD ataxias. Known AR ataxia genes are reported in Table S2. Despite the greater frequency of AR ataxias, many of these cases go genetically undiagnosed. Often, only one individual in a family presents with recessive ataxia.

These cases may appear sporadic or idiopathic, making it difficult to distinguish AR from a *de novo* AD mutation or a new expansion event. In addition, the number of genes causing AR ataxia is large, and often mutations are family-specific or private variants, which appear most frequently under conditions of a suspected founder effect or consanguineous union. Recessive ataxias, more often than dominant, have symptom onset from birth or in early childhood, but this may be due to ascertainment, and later onset recessive ataxias certainly also exist. Unlike AD, early onset AR are typically non-progressive in their symptoms, with more multisystem involvement leading to other symptoms such as intellectual disability [37,38].

The most common autosomal recessive ataxia, and the most common early onset ataxia, is Friedreich's ataxia (FRDA). FRDA is estimated to have a prevalence of 1 in 20–50,000. In certain regions of the world, carrier rates have been estimated to be as high as 1 in 11 [39]. It is most commonly seen in individuals of European ancestry but is present worldwide. FRDA is primarily caused by a GAA intronic repeat expansion of the frataxin gene, with rare conventional mutations also reported [40,41]. The intronic expansion interferes with transcription and results in suppression of gene expression [42,43]. FRDA is a prime example where understanding the cellular pathology has guided research towards treatment, with several groups exploring methods to therapeutically increase the expression of frataxin [44], some of which are in or nearing clinical trials

Ataxia telangiectasia (A-T) is an early onset ataxia affecting 1 in 40–100,000. As mutations disrupt DNA repair, individuals with A-T are susceptible to radiation and oxidative stress. Heterozygous carriers for mutated *ATM* gene have a greater susceptibility to developing cancer. Mutations in *ATM* are highly variable, with over 600 unique variants reported.

There are several other ataxias that exhibit clinical features and molecular pathology similar to A-T. A-T like disorder is caused by mutations in *MRE11A*, another DNA repair gene. Individuals with A-T like disorder share the same neurological defects, along with oculomotor apraxia, but lack the telangiectasia and other features. Four other genes have been identified to cause ataxia with oculomotor apraxia (AOA). AOA2 is caused by mutations in *SETX* and is predicted to be responsible for 8% of non-Friedreich recessive ataxias [45]. It is prevalent among French-Canadians, but also present in other populations [45]. AOA1 is common among Japanese and Portuguese, where it was additionally characterized with features of low serum albumin and high cholesterol levels [46]. Mutations in *GRID2* and *PIK3R5*, which cause AOA and AOA3, are much less common.

3.3. X-Linked

In contrast to AD and AR ataxias, there are comparatively few known X-linked ataxias. The most common X-linked ataxia is fragile X-associated tremor ataxia syndrome (FXTAS). FXTAS is caused by a CGG repeat expansion in the 5' untranslated region of the *FMR1* gene [47]. This ataxia-associated expansion is often referred to as a fragile X "premutation". The normal length of the *FMR1* repeat is less than 39 repeats, whereas 55 to 200 repeats are considered to be a premutation. Males with greater than 200 repeats have the full expansion mutation, which causes fragile X syndrome, a severe disease caused by expansion of the same repeat [48]. In the U.S., carrier rates for the *FMR1* premutation are estimated at 1 in 209 for females and 1 in 430 for males [49]. A study in a population from Quebec estimated premutation rates at 1 in 259 for females and 1 in 813 for males [50,51].

FXTAS is characterized by tremor and ataxia with late onset, usually past the fifth decade. As the gene is X-linked, males are far more commonly affected than females [52]. In female carriers, an estimated 20% experience symptoms of premature ovarian insufficiency, with onset of menopause before 40 and/or fertility issues [53]. Males with fragile X display a very different phenotype from FXTAS, with prominent intellectual disability and abnormal facial features. *FMR1* is a prime example of how subtle differences in mutations within the same gene can greatly impact the phenotype.

Other X-linked ataxias are rare, often restricted to a single family. A mutation in *ATP2B3*, also known as *PMCA3*, was recently associated with spinocerebellar ataxia in an Italian family. Researchers found a single point mutation disrupted calcium transport in the cell, resulting in a "pure cerebellar" phenotype, with congenital onset ataxia, cerebellar atrophy, hypotonia, and slow eye movements [12,54]. Although the phenotype reported is similar to that seen in other families, this is the only reported *ATP2B3* ataxia mutation to date. Sideroblastic anemia with ataxia (ASAT) is caused by mutations in *ABCB7*.

3.4. Mitochondrial

Mitochondrial DNA is maternally transmitted through mitochondria in the oocyte. Mutations in mitochondrial DNA genes tend to result in more multisystem diseases that can contain ataxia as a symptom. Neuropathy, ataxia, and retinitis pigmentosa (NARP) is caused by mutations in the mitochondrial DNA gene *MTATP6* [55].

Mutations in nuclear genes that function primarily in the mitochondria can also cause ataxia. Despite their association to the mitochondria, these mutations are inherited in an AR pattern. Mutations in *POLG*, which is a subunit of mitochondrial DNA polymerase, are responsible for ataxia and other multisystem features [56]. *C10orf2*, or twinkle, is necessary for proper mtDNA replication and is responsible for a variety of neurological phenotypes including infantile onset ataxia [57–60].

3.5. Multiple Systems Atrophy and other Multisystem Diseases that Include Ataxia

Multisystem diseases can be more difficult to diagnose due to the variability in presentation. More diverse neurological phenotypes, such as seizures and myopathy, and non-neurological symptoms such as hearing loss, cardiac problems, and diabetes can complicate these disorders. Multiple system atrophy (MSA) is a progressive neurodegenerative disorder. Individuals may initially present with Parkinsonism or ataxia, and progress to more severe cerebellar atrophy and nervous system dysfunctions. Mutations in *COQ2* shown to be responsible for MSA have been shown to be more common in the Japanese population [61]. Refsum disease can also cause cerebellar ataxia but ataxia is not present in all Refsum patients. Several members of the peroxisome biogenesis factor family are responsible for several peroxisome biogenesis disorders that can appear similar to Refsum. These diseases range in severity from resulting in early death to survival and functional ability in adulthood. The broad phenotypes displayed in these diseases can make them difficult to diagnose and classify.

4. Mutations in Conserved Pathways Cause Ataxia

Despite the great advances made in sequencing technology and the discovery of new genes, there is still a gap in the full understanding of the function of these gene products. Many of the functions of ataxia genes have yet to be discovered, despite overwhelming evidence that they are responsible for causing disease. Discovery of genetic pathways involved in ataxia genes is important to our understanding of disease pathogenesis, and may also impact some treatments. Expression studies and protein interaction assays focused on known ataxia genes have helped identify pathways and protein interactions [62]. Researching expression and pathways can be difficult, primarily due to the low availability of relevant tissue, as brain donation and biopsy are delicate topics for those with ataxia and their families. Knowledge of these pathways will not only be important for efforts for treatment development but aid in the discovery of new ataxia genes through the identification of common pathways and interactions. A success story for this approach is the identification of a new EA candidate gene, *UBR4*, which was selected as a candidate gene due to its role in ubiquitination and localization with another ataxia gene, *ITPR1* [63].

4.1. DNA Repair

The ability of a cell to repair damage to DNA is important in order to maintain proper function and avoid deleterious mutations. DNA damage can result in cell death by apoptosis or the formation of cancerous cells. Several ataxia genes have roles in DNA repair, with many involved in ataxia with oculomotor apraxia. *MRE11* acts in a complex to locate damaged DNA, where it recruits *ATM* to phosphorylate p53 and induce DNA repair [64]. In individuals with ataxia-causing *MRE11* mutations, *MRE11* fails to effectively form a complex and recruit *ATM* [65]. Mutations in *SETX*, responsible for ataxia with oculomotor apraxia, greatly decrease the ability of cells to repair double strand breaks caused by oxidative stress [66]. Single strand repair mechanisms are impaired by mutations in *TDP1* and *APTX* [67–69].

4.2. Channelopathies

Mutations in genes responsible for the transport of ions in and out of the cell result in channelopathies. Channelopathies have received the most attention as a common pathway in neurological disease, with several reviews focused on the role of channel genes in disease and neurological disorders. Defects in ion channel genes usually result in dominant negative mechanisms, as they can alter the current and exchange of ions across cell membranes, affecting cell signaling or causing intracellular accumulation. Ion voltage channels help to regulate the action potential of neurons and release neurotransmitters. EA1, EA2, and EA5 are all caused by mutations in channel genes, a potassium voltage gated channel and two calcium voltage dependent channels [24,70,71]. Two other potassium voltage gated channel genes, *KCNC3* and *KCND3*, are responsible for SCA13 and SCA19/22 [13,14,72]. Inositol 1,4,5 triphosphate binding to *ITPR1* mediates the release of calcium from intracellular stores in the endoplasmic reticulum [73,74]. Deletions in *ITPR1* are hypothesized to cause adult onset ataxia through haploinsufficiency, and mutations in conserved domains affect channel function resulting in congenital ataxia [25–27].

4.3. Ubiquitination

Ubiquitination serves multiple roles within the cell, including targeting proteins for degradation. Many ataxias result from a mutant protein escaping this degradation system. Disruption of ubiquitination systems can cause failures in many cellular processes, such as protein degradation pathways, membrane tracking, apoptosis, and immune system processes.

Several ataxia genes are in or interact with ubiquitination system proteins. Mutations in *RNF170*, an E3 ubiquitin ligase, are responsible for AD sensory ataxia [75,76]. *RNF170* was shown to associate with inositol 1,4,5-triphosphate receptors (*IP3*), while mutations in its receptor, *IP3R1* (*ITPR1*) also cause ataxia, providing a link between ubiquitination and ion channel signaling [76]. *ATXN3* is a de-ubiquitinating enzyme that interacts with parkin, an E3 ubiquitin ligase, resulting in more de-ubiquitinated parkin in the presence of *ATXN3* repeat expansion mutants [77]. Recently mutations in two different E3 ligases have been associated with ataxia and hypogonadism: *RNF216* and *STUB1* [78,79].

4.4. Transcription/Translation

The ability to control gene expression and protein abundance is important for proper function in the cell and organism. Failure in proper transcriptional mechanism and regulation can result in a variety of diseases including cancer, autoimmune, and neurological [80] disorders. *ATXN1* forms a complex with the transcriptional repressor capicua and may interact with the transcription factor *RORα* [29,81]. Nemo-like kinase (*NLK*) has been shown to interact with *ATXN1* transcriptional complex, and decreased expression of *NLK* positively modulates the phenotype in SCA1 models, providing another biological target for future treatments [82]. The transcription factor *RORα* exhibits decreased levels in SCA1 and SCA3, with null and mutant mice for *Rora* showing cerebellar defects and ataxia [81,83,84]. Along with DNA repair, mutations in *SETX* also interfere with transcription, highlighting interactions between senataxin and proteins involved in transcription and RNA processing [85].

5. Genetic Testing of Ataxias and Personalized Medicine

5.1. Is Genetic Testing of Ataxia Useful?

Rare forms of ataxia respond to Vitamin E or Coenzyme Q10 (AVED and SCAR9/ARCA2), but for most ataxias, only symptomatic treatment is available. Genetic testing for diseases with no treatment is controversial and in the U.S., is usually considered a personal choice, whereas developing countries do not routinely offer testing. Thus, what are the reasons to test a patient with ataxia for known genes? Indeed, some of those with ataxia have commented upon finding out that they had e.g., SCA2, so now what? There is little difference in treatment or prognosis, so why all the expense of testing?

One reason given for genetic testing is family planning. This most often arises in dominant ataxias where the disease is seen in prior generations, but can apply to recessive ataxias, especially in isolated populations where disease alleles may be at a higher frequency. Individuals who are carriers of a disease mutation may make different reproductive choices to avoid passing the disease on to the next generation. They may choose to avoid passing on genetic material by not having children, choosing adoption, or use of egg/sperm donors. Those wanting biological children may utilize pre-implantation screening with *in vitro* fertilization or termination of pregnancy after prenatal diagnosis [86]. Pre-implantation testing and testing of children have resulted in a new ethical conundrum of “genetic ignorance”, where parents may decide to remain ignorant about their own results, but wish to test offspring or embryos, possibly unnecessarily [87].

A more complex situation is that of FXTAS. This diagnosis in an older male with ataxia implies a very high risk of the more severe fragile X syndrome in any grandsons or nephews through his female relatives. Since the *FMR1* expansion is on the X chromosome, females can be asymptomatic carriers. For example, a female with a male relative with FXTAS may be a carrier, and hence be at risk of having a son with fragile X syndrome. A survey on genetic screening for *FMR1* mutations showed that while individuals are concerned about finding out they are carriers, and the emotional stress that may accompany that, many note the value in being able to make informed reproductive choices and possible benefits to other family members [88].

Genetic testing also will allow pre-symptomatic testing. One man, after finding out the genetic cause of his mother’s ataxia, decided to get tested himself before purchasing a house—he reasoned that if he had the mutation, the house should accommodate his future potential disability needs such as walker and wheel chair accessibility. Pre-symptomatic testing can result in unintended consequences for those tested, which may explain why for some neurodegenerative diseases, like Huntington’s, a minority of those at risk are tested [89,90]. There are reports of greater instances of depression in individuals with positive test results, possible stigmatization by peers or family members, or having difficulty obtaining life insurance.

Genetic testing offers a definitive diagnostic confirmation for patients. Some individuals with ataxia are first diagnosed as having amyotrophic lateral sclerosis, multiple sclerosis, MSA, or Parkinson’s. Genetic testing will become an increasingly important part of differential diagnosis in these individuals. Desire for knowledge and closure about what is causing their symptoms can be comforting to affected individuals and family members. Genetic testing also has clear clinical ramifications for prognosis. As there are clinical differences in progression rates between different forms of ataxia [21], genetic testing can help patients and their physicians understand their own prognosis. For example, SCA7 leads to severe vision problems or blindness, and SCA6 also leads to some vision problems, whereas SCA2 symptoms also include neuropathy, tremors and cramps. Rarely, Vitamin-E responsive ataxia may be confused clinically with Friedreich’s ataxia, and hence this diagnosis will open up a new treatment with large doses of vitamin E. In turn, sometimes spastic paraplegias are misdiagnosed as ataxias, and treating the spasticity aspect may bring relief [91].

5.2. Genetic Testing Now and in the Future Era of Cheap Sequencing

Currently, genetic testing is performed by a number of academic and commercial laboratories. For recessive ataxias, usually Friedreich's ataxia is tested first, although depending on the symptoms, other ataxias are included. For dominant ataxias, the five most common SCAs are often tested first [92]. If these are negative, comprehensive panels for most known dominant or recessive ataxia genes, or all, are available from commercial sources [93]. Comprehensive panels are helpful when the mode of inheritance is not clear—e.g., if a parent died before symptom onset, or has some neurological symptoms that are not identical, or had been diagnosed with a different disorder.

Next generation exome sequencing has shown some success in clinical environments, demonstrating that it may be more efficient than testing for mutations in ataxia genes individually [94–98]. Several academic clinical laboratories offer targeted sequencing of hundreds of genes, whereas others, such as Baylor and University of Chicago, offer exome sequencing, but may specifically evaluate ataxia genes [99,100]. At what point it is best to move from sequencing genes one at a time to large panels, and when from large panels to whole exome, is currently not clear. In addition, whole exome or genome sequencing can identify mutations in genes unrelated to ataxia, such as genes associated with early onset breast cancer. This is not unique to ataxia, and how to deal with such secondary or "incidental" findings is currently actively debated by ethicists and clinicians.

In addition, there are current limitations to next generation sequencing technology. Currently large repeat expansions cannot be accurately sequenced or mapped to identify the common repeat expansion mutations. This is a limitation of the short reads captured by the sequencing technology; reads comprised of entirely repeats cannot be aligned to accurately determine placement or length. New computational methods are being developed in an attempt to tackle this problem. It is important for clinicians and genetic counselors to consider that next generation sequencing does not guarantee a diagnosis and should address this point with patients desiring sequencing.

Hence, a fully comprehensive genome analysis that covers all ataxia gene mutations is not currently available. Given the heterogeneity of ataxias, and the large number of genes still being detected each year, a comprehensive genetic test will be a challenge for researchers and clinicians.

6. Conclusions

The variability in phenotypic symptoms and genetic causes provide a challenge for clinicians and geneticists in studying ataxia. Advancements in sequencing technology have greatly increased our rate of discovery of new ataxia genes and ability to screen for known genes. With the price of whole exome sequencing and soon whole genome sequencing falling below $1000 a sample, it has become the cost effective approach to screen multiple genes at the same time. Continuing expression studies and investigation into the role of genes will help identify shared pathways and functions. A challenge in this movement towards next generation sequencing technology is the discovery of new repeat expansions, which are difficult to detect using this new technology. The number of known genes mutations responsible for ataxia keeps growing every year; however we still do not have well defined functions or pathways for many of these genes. With greater understanding of the pathways these genes are involved in and how each mutation causes disease, we may be able to generate more targeted and effective treatments in the future.

Acknowledgments

We would like to acknowledge the National Ataxia Foundation (NAF) for their assistance in recruitment and opportunity to meet and discuss with those affected by ataxia. We would like to thank our many clinical colleagues for recruitment and discussions, and all of the families that have taken time to donate samples and personal stories. Our ataxia research is funded by the National Institutes of Health (NIH) grant R01-NS078560.

Author Contributions

All authors contributed to and wrote this review.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Durr, A. Autosomal dominant cerebellar ataxias: Polyglutamine expansions and beyond. *Lancet Neurol.* **2010**, *9*, 885–894.
2. Joo, B.-E.; Lee, C.-N.; Park, K.-W. Prevalence rate and functional status of cerebellar ataxia in Korea. *Cerebellum Lond. Engl.* **2012**, *11*, 733–738.
3. Coutinho, P.; Ruano, L.; Loureiro, J.L.; Cruz, V.T.; Barros, J.; Tuna, A.; Barbot, C.; Guimarães, J.; Alonso, I.; Silveira, I.; *et al.* Hereditary ataxia and spastic paraplegia in Portugal: A population-based prevalence study. *JAMA Neurol.* **2013**, *70*, 746–755.
4. Koht, J.; Tallaksen, C.M.E. Cerebellar ataxia in the eastern and southern parts of Norway. *Acta Neurol. Scand. Suppl.* **2007**, *187*, 76–79.
5. Orr, H.T.; Chung, M.Y.; Banfi, S.; Kwiatkowski, T.J., Jr.; Servadio, A.; Beaudet, A.L.; McCall, A.E.; Duvick, L.A.; Ranum, L.P.; Zoghbi, H.Y. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **1993**, *4*, 221–226.
6. Sanpei, K.; Takano, H.; Igarashi, S.; Sato, T.; Oyake, M.; Sasaki, H.; Wakisaka, A.; Tashiro, K.; Ishida, Y.; Ikeuchi, T.; *et al.* Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat. Genet.* **1996**, *14*, 277–284.
7. Pulst, S.M.; Nechiporuk, A.; Nechiporuk, T.; Gispert, S.; Chen, X.N.; Lopes-Cendes, I.; Pearlman, S.; Starkman, S.; Orozco-Diaz, G.; Lunkes, A.; *et al.* Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.* **1996**, *14*, 269–276.

8. Kawaguchi, Y.; Okamoto, T.; Taniwaki, M.; Aizawa, M.; Inoue, M.; Katayama, S.; Kawakami, H.; Nakamura, S.; Nishimura, M.; Akiguchi, I. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **1994**, *8*, 221–228.
9. Haberhausen, G.; Damian, M.S.; Leweke, F.; Müller, U. Spinocerebellar ataxia, type 3 (SCA3) is genetically identical to Machado-Joseph disease (MJD). *J. Neurol. Sci.* **1995**, *132*, 71–75.
10. Zhuchenko, O.; Bailey, J.; Bonnen, P.; Ashizawa, T.; Stockton, D.W.; Amos, C.; Dobyns, W.B.; Subramony, S.H.; Zoghbi, H.Y.; Lee, C.C. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat. Genet.* **1997**, *15*, 62–69.
11. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; *et al*. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35.
12. Zanni, G.; Calì, T.; Kalscheuer, V.M.; Ottolini, D.; Barresi, S.; Lebrun, N.; Montecchi-Palazzi, L.; Hu, H.; Chelly, J.; Bertini, E.; *et al*. Mutation of plasma membrane Ca2+ ATPase isoform 3 in a family with X-linked congenital cerebellar ataxia impairs Ca2+ homeostasis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14514–14519.
13. Lee, Y.-C.; Durr, A.; Majczenko, K.; Huang, Y.-H.; Liu, Y.-C.; Lien, C.-C.; Tsai, P.-C.; Ichikawa, Y.; Goto, J.; Monin, M.-L.; *et al*. Mutations in KCND3 cause spinocerebellar ataxia type 22. *Ann. Neurol.* **2012**, *72*, 859–869.
14. Duarri, A.; Jezierska, J.; Fokkens, M.; Meijer, M.; Schelhaas, H.J.; den Dunnen, W.F.A.; van Dijk, F.; Verschuuren-Bemelmans, C.; Hageman, G.; van de Vlies, P.; *et al*. Mutations in potassium channel kcnd3 cause spinocerebellar ataxia type 19. *Ann. Neurol.* **2012**, *72*, 870–880.
15. Winkelmann, J.; Lin, L.; Schormair, B.; Kornum, B.R.; Faraco, J.; Plazzi, G.; Melberg, A.; Cornelio, F.; Urban, A.E.; Pizza, F.; *et al*. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* **2012**, *21*, 2205–2210.
16. Bilguvar, K.; Tyagi, N.K.; Ozkara, C.; Tuysuz, B.; Bakircioglu, M.; Choi, M.; Delil, S.; Caglayan, A.O.; Baranoski, J.F.; Erturk, O.; *et al*. Recessive loss of function of the neuronal ubiquitin hydrolase UCHL1 leads to early-onset progressive neurodegeneration. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3489–3494.
17. Sun, Y.; Almomani, R.; Breedveld, G.J.; Santen, G.W.E.; Aten, E.; Lefeber, D.J.; Hoff, J.I.; Brusse, E.; Verheijen, F.W.; Verdijk, R.M.; *et al*. Autosomal recessive spinocerebellar ataxia 7 (SCAR7) is caused by variants in TPP1, the gene involved in classic late-infantile neuronal ceroid lipofuscinosis 2 disease (CLN2 disease). *Hum. Mutat.* **2013**, *34*, 706–713.
18. Sailer, A.; Houlden, H. Recent advances in the genetics of cerebellar ataxias. *Curr. Neurol. Neurosci. Rep.* **2012**, *12*, 227–236.
19. Lhatoo, S.D.; Rao, D.G.; Kane, N.M.; Ormerod, I.E. Very late onset Friedreich's presenting as spastic tetraparesis without ataxia or neuropathy. *Neurology* **2001**, *56*, 1776–1777.
20. Castelnovo, G.; Biolsi, B.; Barbaud, A.; Labauge, P.; Schmitt, M. Isolated spastic paraparesis leading to diagnosis of Friedreich's ataxia. *J. Neurol. Neurosurg. Psychiatry* **2000**, *69*, 693.
21. Paulson, H.L. The spinocerebellar ataxias. *J. Neuroophthalmol.* **2009**, *29*, 227–237.

22. Baloh, R.W. Episodic ataxias 1 and 2. *Handb. Clin. Neurol.* **2012**, *103*, 595–602.

23. Jen, J.C.; Wan, J.; Palos, T.P.; Howard, B.D.; Baloh, R.W. Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. *Neurology* **2005**, *65*, 529–534.

24. Escayg, A.; de Waard, M.; Lee, D.D.; Bichet, D.; Wolf, P.; Mayer, T.; Johnston, J.; Baloh, R.; Sander, T.; Meisler, M.H. Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. *Am. J. Hum. Genet.* **2000**, *66*, 1531–1539.

25. Van de Leemput, J.; Chandran, J.; Knight, M.A.; Holtzclaw, L.A.; Scholz, S.; Cookson, M.R.; Houlden, H.; Gwinn-Hardy, K.; Fung, H.-C.; Lin, X.; *et al.* Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genet.* **2007**, *3*, e108.

26. Iwaki, A.; Kawano, Y.; Miura, S.; Shibata, H.; Matsuse, D.; Li, W.; Furuya, H.; Ohyagi, Y.; Taniwaki, T.; Kira, J.; *et al.* Heterozygous deletion of ITPR1, but not SUMF1, in spinocerebellar ataxia type 16. *J. Med. Genet.* **2008**, *45*, 32–35.

27. Huang, L.; Chardon, J.W.; Carter, M.T.; Friend, K.L.; Dudding, T.E.; Schwartzentruber, J.; Zou, R.; Schofield, P.W.; Douglas, S.; Bulman, D.E.; *et al.* Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *Orphanet J. Rare Dis.* **2012**, *7*, 67.

28. Cummings, C.J.; Reinstein, E.; Sun, Y.; Antalffy, B.; Jiang, Y.; Ciechanover, A.; Orr, H.T.; Beaudet, A.L.; Zoghbi, H.Y. Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. *Neuron* **1999**, *24*, 879–892.

29. Lam, Y.C.; Bowman, A.B.; Jafar-Nejad, P.; Lim, J.; Richman, R.; Fryer, J.D.; Hyun, E.D.; Duvick, L.A.; Orr, H.T.; Botas, J.; *et al.* ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* **2006**, *127*, 1335–1347.

30. Suzuki, K.; Zhou, J.; Sato, T.; Takao, K.; Miyagawa, T.; Oyake, M.; Yamada, M.; Takahashi, H.; Takahashi, Y.; Goto, J.; *et al.* DRPLA transgenic mouse substrains carrying single copy of full-length mutant human DRPLA gene with variable sizes of expanded CAG repeats exhibit CAG repeat length- and age-dependent changes in behavioral abnormalities and gene expression profiles. *Neurobiol. Dis.* **2012**, *46*, 336–350.

31. Zu, T.; Gibbens, B.; Doty, N.S.; Gomes-Pereira, M.; Huguet, A.; Stone, M.D.; Margolis, J.; Peterson, M.; Markowski, T.W.; Ingram, M.A.C.; *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 260–265.

32. Cleary, J.D.; Ranum, L.P.W. Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum. Mol. Genet.* **2013**, *22*, R45–R51.

33. Todd, P.K.; Oh, S.Y.; Krans, A.; He, F.; Sellier, C.; Frazer, M.; Renoux, A.J.; Chen, K.; Scaglione, K.M.; Basrur, V.; *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **2013**, *78*, 440–455.

34. Cagnoli, C.; Stevanin, G.; Brussino, A.; Barberis, M.; Mancini, C.; Margolis, R.L.; Holmes, S.E.; Nobili, M.; Forlani, S.; Padovan, S.; *et al.* Missense mutations in the AFG3L2 proteolytic domain account for ~1.5% of European autosomal dominant cerebellar ataxias. *Hum. Mutat.* **2010**, *31*, 1117–1124.

35. Jia, D.; Tang, B.; Chen, Z.; Shi, Y.; Sun, Z.; Zhang, L.; Wang, J.; Xia, K.; Jiang, H. Spinocerebellar ataxia type 28 (SCA28) is an uncommon cause of dominant ataxia among Chinese kindreds. *Int. J. Neurosci.* **2012**, *122*, 560–562.

36. Knight, M.A.; Kennerson, M.L.; Anney, R.J.; Matsuura, T.; Nicholson, G.A.; Salimi-Tari, P.; Gardner, R.J.M.; Storey, E.; Forrest, S.M. Spinocerebellar ataxia type 15 (sca15) maps to 3p24.2-3pter: Exclusion of the ITPR1 gene, the human orthologue of an ataxic mouse mutant. *Neurobiol. Dis.* **2003**, *13*, 147–157.

37. Embiruçu, E.K.; Martyn, M.L.; Schlesinger, D.; Kok, F. Autosomal recessive ataxias: 20 types, and counting. *Arq. Neuropsiquiatr.* **2009**, *67*, 1143–1156.

38. Fogel, B.L.; Perlman, S. Clinical features and molecular genetics of autosomal recessive cerebellar ataxias. *Lancet Neurol.* **2007**, *6*, 245–257.

39. Zamba-Papanicolaou, E.; Koutsou, P.; Daiou, C.; Gaglia, E.; Georghiou, A.; Christodoulou, K. High frequency of Friedreich's ataxia carriers in the Paphos district of Cyprus. *Acta Myol.* **2009**, *28*, 24–26.

40. Campuzano, V.; Montermini, L.; Moltò, M.D.; Pianese, L.; Cossée, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A.; *et al.* Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **1996**, *271*, 1423–1427.

41. Bidichandani, S.I.; Ashizawa, T.; Patel, P.I. Atypical Friedreich ataxia caused by compound heterozygosity for a novel missense mutation and the GAA triplet-repeat expansion. *Am. J. Hum. Genet.* **1997**, *60*, 1251–1256.

42. Bidichandani, S.I.; Ashizawa, T.; Patel, P.I. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* **1998**, *62*, 111–121.

43. Campuzano, V.; Montermini, L.; Lutz, Y.; Cova, L.; Hindelang, C.; Jiralerspong, S.; Trottier, Y.; Kish, S.J.; Faucheux, B.; Trouillas, P.; *et al.* Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum. Mol. Genet.* **1997**, *6*, 1771–1780.

44. Chapdelaine, P.; Coulombe, Z.; Chikh, A.; Gerard, C.; Tremblay, J.P. A Potential New Therapeutic Approach for Friedreich Ataxia: Induction of Frataxin Expression with TALE Proteins. *Mol. Ther. Nucleic Acids* **2013**, *2*, e119.

45. Le Ber, I.; Bouslam, N.; Rivaud-Péchoux, S.; Guimarães, J.; Benomar, A.; Chamayou, C.; Goizet, C.; Moreira, M.-C.; Klur, S.; Yahyaoui, M.; *et al.* Frequency and phenotypic spectrum of ataxia with oculomotor apraxia 2: A clinical and genetic study in 18 patients. *Brain J. Neurol.* **2004**, *127*, 759–767.

46. Moreira, M.C.; Barbot, C.; Tachi, N.; Kozuka, N.; Uchida, E.; Gibson, T.; Mendonça, P.; Costa, M.; Barros, J.; Yanagisawa, T.; *et al.* The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nat. Genet.* **2001**, *29*, 189–193.

47. Hagerman, R.J.; Leehey, M.; Heinrichs, W.; Tassone, F.; Wilson, R.; Hills, J.; Grigsby, J.; Gage, B.; Hagerman, P.J. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **2001**, *57*, 127–130.
48. Fu, Y.H.; Kuhl, D.P.; Pizzuti, A.; Pieretti, M.; Sutcliffe, J.S.; Richards, S.; Verkerk, A.J.; Holden, J.J.; Fenwick, R.G., Jr.; Warren, S.T. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **1991**, *67*, 1047–1058.
49. Tassone, F.; Iong, K.P.; Tong, T.-H.; Lo, J.; Gane, L.W.; Berry-Kravis, E.; Nguyen, D.; Mu, L.Y.; Laffin, J.; Bailey, D.B.; *et al*. FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med.* **2012**, *4*, 100.
50. Rousseau, F.; Rouillard, P.; Morel, M.L.; Khandjian, E.W.; Morgan, K. Prevalence of carriers of premutation-size alleles of the FMRI gene—And implications for the population genetics of the fragile X syndrome. *Am. J. Hum. Genet.* **1995**, *57*, 1006–1018.
51. Dombrowski, C.; Lévesque, S.; Morel, M.L.; Rouillard, P.; Morgan, K.; Rousseau, F. Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: Loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* **2002**, *11*, 371–378.
52. Rodriguez-Revenga, L.; Madrigal, I.; Pagonabarraga, J.; Xunclà, M.; Badenas, C.; Kulisevsky, J.; Gomez, B.; Milà, M. Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet. EJHG* **2009**, *17*, 1359–1362.
53. Schwartz, C.E.; Dean, J.; Howard-Peebles, P.N.; Bugge, M.; Mikkelsen, M.; Tommerup, N.; Hull, C.; Hagerman, R.; Holden, J.J.; Stevenson, R.E. Obstetrical and gynecological complications in fragile X carriers: a multicenter study. *Am. J. Med. Genet.* **1994**, *51*, 400–402.
54. Bertini, E.; des Portes, V.; Zanni, G.; Santorelli, F.; Dionisi-Vici, C.; Vicari, S.; Fariello, G.; Chelly, J. X-linked congenital ataxia: A clinical and genetic study. *Am. J. Med. Genet.* **2000**, *92*, 53–56.
55. Holt, I.J.; Harding, A.E.; Petty, R.K.; Morgan-Hughes, J.A. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am. J. Hum. Genet.* **1990**, *46*, 428–433.
56. Synofzik, M.; Srulijes, K.; Godau, J.; Berg, D.; Schöls, L. Characterizing POLG ataxia: Clinics, electrophysiology and imaging. *Cerebellum Lond. Engl.* **2012**, *11*, 1002–1011.
57. Faruq, M.; Narang, A.; Kumari, R.; Pandey, R.; Garg, A.; Behari, M.; Dash, D.; Srivastava, A.; Mukerji, M. Novel mutations in typical and atypical genetic loci through exome sequencing in autosomal recessive cerebellar ataxia families. *Clin. Genet.* **2013**, doi:10.1111/cge.12279.
58. Nikali, K.; Suomalainen, A.; Saharinen, J.; Kuokkanen, M.; Spelbrink, J.N.; Lönnqvist, T.; Peltonen, L. Infantile onset spinocerebellar ataxia is caused by recessive mutations in mitochondrial proteins Twinkle and Twinky. *Hum. Mol. Genet.* **2005**, *14*, 2981–2990.
59. Hartley, J.N.; Booth, F.A.; del Bigio, M.R.; Mhanni, A.A. Novel Autosomal Recessive c10orf2 Mutations Causing Infantile-Onset Spinocerebellar Ataxia. *Case Rep. Pediatr.* **2012**, *2012*, 303096.
60. Wanrooij, S.; Falkenberg, M. The human mitochondrial replication fork in health and disease. *Biochim. Biophys. Acta* **2010**, *1797*, 1378–1388.

61. Multiple-System Atrophy Research Collaboration Mutations in COQ2 in familial and sporadic multiple-system atrophy. *N. Engl. J. Med.* **2013**, *369*, 233–244.
62. Lim, J.; Hao, T.; Shaw, C.; Patel, A.J.; Szabó, G.; Rual, J.-F.; Fisk, C.J.; Li, N.; Smolyar, A.; Hill, D.E.; *et al.* A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **2006**, *125*, 801–814.
63. Conroy, J.; McGettigan, P.; Murphy, R.; Webb, D.; Murphy, S.M.; McCoy, B.; Albertyn, C.; McCreary, D.; McDonagh, C.; Walsh, O.; *et al.* A novel locus for episodic ataxia:UBR4 the likely candidate. *Eur. J. Hum. Genet.* **2014**, *22*, 505–510.
64. Lee, J.-H.; Paull, T.T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **2005**, *308*, 551–554.
65. Regal, J.A.; Festerling, T.A.; Buis, J.M.; Ferguson, D.O. Disease-associated MRE11 mutants impact ATM/ATR DNA damage signaling by distinct mechanisms. *Hum. Mol. Genet.* **2013**, *22*, 5146–5159.
66. Suraweera, A.; Becherel, O.J.; Chen, P.; Rundle, N.; Woods, R.; Nakamura, J.; Gatei, M.; Criscuolo, C.; Filla, A.; Chessa, L.; *et al.* Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *J. Cell Biol.* **2007**, *177*, 969–979.
67. Takahashi, T.; Tada, M.; Igarashi, S.; Koyama, A.; Date, H.; Yokoseki, A.; Shiga, A.; Yoshida, Y.; Tsuji, S.; Nishizawa, M.; *et al.* Aprataxin, causative gene product for EAOH/AOA1, repairs DNA single-strand breaks with damaged 3'-phosphate and 3'-phosphoglycolate ends. *Nucleic Acids Res.* **2007**, *35*, 3797–3809.
68. Zhou, T.; Lee, J.W.; Tatavarthi, H.; Lupski, J.R.; Valerie, K.; Povirk, L.F. Deficiency in 3'-phosphoglycolate processing in human cells with a hereditary mutation in tyrosyl-DNA phosphodiesterase (TDP1). *Nucleic Acids Res.* **2005**, *33*, 289–297.
69. El-Khamisy, S.F.; Saifi, G.M.; Weinfeld, M.; Johansson, F.; Helleday, T.; Lupski, J.R.; Caldecott, K.W. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* **2005**, *434*, 108–113.
70. Browne, D.L.; Gancher, S.T.; Nutt, J.G.; Brunt, E.R.; Smith, E.A.; Kramer, P.; Litt, M. Episodic ataxia/myokymia syndrome is associated with point mutations in the human potassium channel gene, KCNA1. *Nat. Genet.* **1994**, *8*, 136–140.
71. Ophoff, R.A.; Terwindt, G.M.; Vergouwe, M.N.; van Eijk, R.; Oefner, P.J.; Hoffman, S.M.; Lamerdin, J.E.; Mohrenweiser, H.W.; Bulman, D.E.; Ferrari, M.; *et al.* Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca^{2+} channel gene CACNL1A4. *Cell* **1996**, *87*, 543–552.
72. Waters, M.F.; Minassian, N.A.; Stevanin, G.; Figueroa, K.P.; Bannister, J.P.A.; Nolte, D.; Mock, A.F.; Evidente, V.G.H.; Fee, D.B.; Müller, U.; *et al.* Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes. *Nat. Genet.* **2006**, *38*, 447–451.
73. Streb, H.; Irvine, R.F.; Berridge, M.J.; Schulz, I. Release of Ca^{2+} from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* **1983**, *306*, 67–69.

74. Berridge, M.J. Inositol trisphosphate and calcium signalling mechanisms. *Biochim. Biophys. Acta* **2009**, *1793*, 933–940.

75. Valdmanis, P.N.; Dupré, N.; Lachance, M.; Stochmanski, S.J.; Belzil, V.V.; Dion, P.A.; Thiffault, I.; Brais, B.; Weston, L.; Saint-Amant, L.; *et al.* A mutation in the RNF170 gene causes autosomal dominant sensory ataxia. *Brain J. Neurol.* **2011**, *134*, 602–607.

76. Lu, J.P.; Wang, Y.; Sliter, D.A.; Pearce, M.M.P.; Wojcikiewicz, R.J.H. RNF170 protein, an endoplasmic reticulum membrane ubiquitin ligase, mediates inositol 1,4,5-trisphosphate receptor ubiquitination and degradation. *J. Biol. Chem.* **2011**, *286*, 24426–24433.

77. Durcan, T.M.; Kontogiannea, M.; Thorarinsdottir, T.; Fallon, L.; Williams, A.J.; Djarmati, A.; Fantaneanu, T.; Paulson, H.L.; Fon, E.A. The Machado-Joseph disease-associated mutant form of ataxin-3 regulates parkin ubiquitination and stability. *Hum. Mol. Genet.* **2011**, *20*, 141–154.

78. Margolin, D.H.; Kousi, M.; Chan, Y.-M.; Lim, E.T.; Schmahmann, J.D.; Hadjivassiliou, M.; Hall, J.E.; Adam, I.; Dwyer, A.; Plummer, L.; *et al.* Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. *N. Engl. J. Med.* **2013**, *368*, 1992–2003.

79. Shi, C.-H.; Schisler, J.C.; Rubel, C.E.; Tan, S.; Song, B.; McDonough, H.; Xu, L.; Portbury, A.L.; Mao, C.-Y.; True, C.; *et al.* Ataxia and hypogonadism caused by the loss of ubiquitin ligase activity of the U box protein CHIP. *Hum. Mol. Genet.* **2014**, *23*, 1013–1024.

80. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251.

81. Serra, H.G.; Duvick, L.; Zu, T.; Carlson, K.; Stevens, S.; Jorgensen, N.; Lysholm, A.; Burright, E.; Zoghbi, H.Y.; Clark, H.B.; *et al.* RORalpha-mediated Purkinje cell development determines disease severity in adult SCA1 mice. *Cell* **2006**, *127*, 697–708.

82. Ju, H.; Kokubu, H.; Todd, T.W.; Kahle, J.J.; Kim, S.; Richman, R.; Chirala, K.; Orr, H.T.; Zoghbi, H.Y.; Lim, J. Polyglutamine disease toxicity is regulated by Nemo-like kinase in spinocerebellar ataxia type 1. *J. Neurosci.* **2013**, *33*, 9328–9336.

83. Konno, A.; Shuvaev, A.N.; Miyake, N.; Miyake, K.; Iizuka, A.; Matsuura, S.; Huda, F.; Nakamura, K.; Yanagi, S.; Shimada, T.; *et al.* Mutant Ataxin-3 with an Abnormally Expanded Polyglutamine Chain Disrupts Dendritic Development and Metabotropic Glutamate Receptor Signaling in Mouse Cerebellar Purkinje Cells. *Cerebellum* **2013**, *13*, 29–41.

84. Dussault, I.; Fawcett, D.; Matthyssen, A.; Bader, J.A.; Giguère, V. Orphan nuclear receptor ROR alpha-deficient mice display the cerebellar defects of staggerer. *Mech. Dev.* **1998**, *70*, 147–153.

85. Suraweera, A.; Lim, Y.; Woods, R.; Birrell, G.W.; Nasim, T.; Becherel, O.J.; Lavin, M.F. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum. Mol. Genet.* **2009**, *18*, 3384–3396.

86. Cruz-Mariño, T.; Velázquez-Pérez, L.; González-Zaldivar, Y.; Aguilera-Rodríguez, R.; Velázquez-Santos, M.; Vázquez-Mojena, Y.; Estupiñán-Rodríguez, A.; Reynaldo-Armiñán, R.; Almaguer-Mederos, L.E.; Laffita-Mesa, J.M.; *et al.* Couples at risk for spinocerebellar ataxia type 2: the Cuban prenatal diagnosis experience. *J. Community Genet.* **2013**, *4*, 451–460.

87. Erez, A.; Plunkett, K.; Sutton, V.R.; McGuire, A.L. The right to ignore genetic status of late onset genetic disease in the genomic era; Prenatal testing for Huntington disease as a paradigm. *Am. J. Med. Genet. A* **2010**, *152A*, 1774–1780.

88. Archibald, A.D.; Hickerton, C.L.; Jaques, A.M.; Wake, S.; Cohen, J.; Metcalfe, S.A. "It's about having the choice": Stakeholder perceptions of population-based genetic carrier screening for fragile X syndrome. *Am. J. Med. Genet. A* **2013**, *161A*, 48–58.
89. Andersson, P.L.; Juth, N.; Petersén, Å.; Graff, C.; Edberg, A.-K. Ethical aspects of undergoing a predictive genetic testing for Huntington's disease. *Nurs. Ethics* **2013**, *20*, 189–199.
90. Tanaka, K.; Sekijima, Y.; Yoshida, K.; Tamai, M.; Kosho, T.; Sakurai, A.; Wakui, K.; Ikeda, S.; Fukushima, Y. Follow-up nationwide survey on predictive genetic testing for late-onset hereditary neurological diseases in Japan. *J. Hum. Genet.* **2013**, *58*, 560–563.
91. Doi, H.; Ohba, C.; Tsurusaki, Y.; Miyatake, S.; Miyake, N.; Saitsu, H.; Kawamoto, Y.; Yoshida, T.; Koyano, S.; Suzuki, Y.; *et al.* Identification of a novel homozygous SPG7 mutation in a Japanese patient with spastic ataxia: making an efficient diagnosis using exome sequencing for autosomal recessive cerebellar ataxia and spastic paraplegia. *Intern. Med. Tokyo Jpn.* **2013**, *52*, 1629–1633.
92. UW Laboratory Medicine Clinical Test Information. Available online: http://web.labmed.washington.edu/tests/genetics/SCAPN/ (accessed on 18 April 2014).
93. Athena Diagnostics: Test Catalog. Available online: http://www.athenadiagnostics.com/content/test-catalog/ (accessed on 18 April 2014).
94. Yang, Y.; Muzny, D.M.; Reid, J.G.; Bainbridge, M.N.; Willis, A.; Ward, P.A.; Braxton, A.; Beuten, J.; Xia, F.; Niu, Z.; *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **2013**, *369*, 1502–1511.
95. Hammer, M.B.; Eleuch-Fayache, G.; Gibbs, J.R.; Arepalli, S.K.; Chong, S.B.; Sassi, C.; Bouhlal, Y.; Hentati, F.; Amouri, R.; Singleton, A.B. Exome sequencing: an efficient diagnostic tool for complex neurodegenerative disorders. *Eur. J. Neurol. Off. J. Eur. Fed. Neurol. Soc.* **2013**, *20*, 486–492.
96. Sawyer, S.L.; Schwartzentruber, J.; Beaulieu, C.L.; Dyment, D.; Smith, A.; Warman Chardon, J.; Yoon, G.; Rouleau, G.A.; Suchowersky, O.; Siu, V.; *et al.* Exome sequencing as a diagnostic tool for pediatric-onset ataxia. *Hum. Mutat.* **2014**, *35*, 45–49.
97. Németh, A.H.; Kwasniewska, A.C.; Lise, S.; Parolin Schnekenberg, R.; Becker, E.B.E.; Bera, K.D.; Shanks, M.E.; Gregory, L.; Buck, D.; Zameel Cader, M.; *et al.* Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain J. Neurol.* **2013**, *136*, 3106–3118.
98. Ohba, C.; Osaka, H.; Iai, M.; Yamashita, S.; Suzuki, Y.; Aida, N.; Shimozawa, N.; Takamura, A.; Doi, H.; Tomita-Katsumoto, A.; *et al.* Diagnostic utility of whole exome sequencing in patients showing cerebellar and/or vermis atrophy in childhood. *Neurogenetics* **2013**, *14*, 225–232.
99. Baylor College of Medicine: Medical Genetics Laboratories. Available online: https://www.bcm.edu/research/medical-genetics-labs/test_detail.cfm?testcode=1500/ (accessed on 18 April 2014).
100. The University of Chicago: Ataxia Exome Panel. Available online: http://dnatesting.uchicago.edu/tests/676/ (accessed on 18 April 2014).

Functional Gene-Set Analysis Does Not Support a Major Role for Synaptic Function in Attention Deficit/Hyperactivity Disorder (ADHD)

Anke R. Hammerschlag, Tinca J. C. Polderman, Christiaan de Leeuw, Henning Tiemeier, Tonya White, August B. Smit, Matthijs Verhage and Danielle Posthuma

Abstract: Attention Deficit/Hyperactivity Disorder (ADHD) is one of the most common childhood-onset neuropsychiatric disorders. Despite high heritability estimates, genome-wide association studies (GWAS) have failed to find significant genetic associations, likely due to the polygenic character of ADHD. Nevertheless, genetic studies suggested the involvement of several processes important for synaptic function. Therefore, we applied a functional gene-set analysis to formally test whether synaptic functions are associated with ADHD. Gene-set analysis tests the joint effect of multiple genetic variants in groups of functionally related genes. This method provides increased statistical power compared to conventional GWAS. We used data from the Psychiatric Genomics Consortium including 896 ADHD cases and 2455 controls, and 2064 parent-affected offspring trios, providing sufficient statistical power to detect gene sets representing a genotype relative risk of at least 1.17. Although all synaptic genes together showed a significant association with ADHD, this association was not stronger than that of randomly generated gene sets matched for same number of genes. Further analyses showed no association of specific synaptic function categories with ADHD after correction for multiple testing. Given current sample size and gene sets based on current knowledge of genes related to synaptic function, our results do not support a major role for common genetic variants in synaptic genes in the etiology of ADHD.

Reprinted from *Genes*. Cite as: Hammerschlag, A.R.; Polderman, T.J.C.; de Leeuw, C.; Tiemeier, H.; White, T.; Smit, A.B.; Verhage, M.; Posthuma, D. Functional Gene-Set Analysis Does Not Support a Major Role for Synaptic Function in Attention Deficit/Hyperactivity Disorder (ADHD). *Genes* **2014**, *5*, 604-614.

1. Introduction

Attention Deficit/Hyperactivity Disorder (ADHD) is one of the most common childhood-onset neuropsychiatric disorders. The worldwide prevalence is estimated at ~5% [1], and remained relatively stable across the last three decades [2]. ADHD is characterized by a persistent pattern of inattention and/or impulsiveness and hyperactivity. Despite high heritability estimates for ADHD, averaging 70% [3], the identification of genes has been difficult. Most likely this is mainly due to the polygenic character of ADHD, similar to that of other complex traits, meaning that many genetic variants with small effects contribute to ADHD risk [4].

Genome-wide association studies (GWAS) of ADHD have yielded no significant single nucleotide polymorphism (SNP) associations thus far [5]. However, it has been reported that the top hits of GWAS point to the involvement of synaptic processes such as neurotransmission, cell-cell communication systems, potassium channel subunits and regulators, and more basic processes like neuronal migration, neurite outgrowth, spine formation, neuronal plasticity, cell division, and adhesion [6–8]. Furthermore, many genes previously implicated in ADHD [9] are expressed in the synapse (*i.e.*, *DBH*, *SLC6A2*, *ADRA2A*, *HTR1B*, *HTR2A*, *TPH1/2*, *MAOA*, *CHRNA4*, *SNAP25*, and *BDNF*), suggesting the involvement of synaptic function in the etiology of ADHD.

In addition to common genetic variants, rare variants may contribute to ADHD risk. Increased structural variation burden has been reported, particularly in subjects with intellectual disability [10–13]. Interestingly, biological pathways enriched for GWAS SNP associations with low *p*-values overlap with pathways enriched for rare structural variants, including pathways important for synaptic function [12]. Of special interest are SNPs and duplications spanning the *CHRNA7* gene, which is primarily involved in modulation of rapid synaptic transmission and which has been associated with other neuropsychiatric phenotypes in addition to ADHD [12,13]. Furthermore, strong associations have been reported for structural variation affecting metabotropic glutamate receptor genes and genes that interact with them. Several of these genes are important modulators of synaptic transmission and neurogenesis [11].

Given the polygenic nature of ADHD, it is likely that non-random combinations of genetic variants are involved in the etiology of ADHD. Genes do not work in isolation; rather, they form complex molecular networks and cellular pathways. Therefore, it is plausible that the numerous genetic variants of small effect aggregate in genes that share a similar cellular function. Evaluating the joint effect of multiple SNPs in functionally related genes increases the statistical power to detect associations with ADHD compared to single SNP methods, as it reduces multiple testing. Moreover, single SNP associations do not necessarily lead to knowledge about underlying biological mechanisms, while a set of genes with the same function could result in more insight in the molecular or cellular mechanisms of ADHD [14].

Prior studies that tested the joint effect of genetic variants generally grouped genes based on biological pathways. However, grouping genes based on cellular function (“horizontal grouping”) instead of biological pathways (“vertical grouping”) may be especially powerful in synaptic protein networks [15,16]. Many different pathways regulate synaptic function, but act not independent, as many proteins act across pathways. For example, different neuromodulator pathways (e.g., dopamine or serotonin) include receptors that are activated by the specific neuromodulators, but are functionally and often structurally similar to each other. It may well be that genetic variants influencing complex traits like ADHD concentrate at similar cellular function, by which they influence different pathways leading to similar consequences in synaptic function.

The majority of gene-set analyses that have been conducted have used publicly available gene sets. However, currently available public gene sets are generally incomplete and neither error-free nor unbiased, especially with regard to genes active in the brain [17,18]. Fortunately, expert-curated sets of genes are increasingly becoming available, such as the mir-137 gene set [19], specific synaptic gene sets [15], and gene sets for glial function [20].

As the results of previous GWAS and genes affected by structural variation suggested involvement of synaptic function, we hypothesized that synaptic processes play a role in the etiology of ADHD. Collective testing of genetic variants in genes grouped according to similar synaptic functions may be the most optimal way to test this. Therefore, we applied a functional gene-set analysis for ADHD using 18 previously published, expert-curated pre- and postsynaptic gene sets [15]. To our knowledge, this is the first study to conduct hypothesis-driven gene-set analysis for ADHD by grouping synaptic genes according to cellular function. We used ADHD data from the Psychiatric Genomics Consortium (PGC) [5].

2. Methods

2.1. Sample

We used GWAS summary statistics from the currently largest publicly available ADHD data set, as provided by the PGC [5]. Details on the data set have been described previously [5]. In short, the data set consisted of four projects: the Children's Hospital of Philadelphia (CHOP), phase I of the International Multicenter ADHD Genetics Project (IMAGE), phase II of IMAGE (IMAGE II), and the Pfizer-funded study from the University of California, Los Angeles, Washington University, and Massachusetts General Hospital (PUWMa). The total sample consisted of 896 unrelated cases and 2455 controls, and 2064 trio samples (alleles transmitted to offspring were considered as "trio cases", and non-transmitted alleles as "pseudo-controls"). All samples were of European ancestry and met diagnostic criteria for ADHD as defined by the DSM-IV. All samples underwent the same quality control and analysis steps. The strongest single SNP association with ADHD in this data set was $p = 1.10 \times 10^{-6}$ [5].

2.2. Defining Functional Gene Sets

Generation of the synaptic gene sets has been described previously [15]. Briefly, synaptic gene grouping was based on cellular function as determined by previous synaptic protein identification experiments and data mining for synaptic genes and gene function. This resulted in the inclusion of 1028 genes, expressed in either the pre- or postsynapse or in both, divided over 17 synaptic gene sets with a specific synaptic function, and one synaptic gene set with unassigned cellular function. The gene sets with gene IDs are available at the Complex Trait Genetics webpage [21].

2.3. Power Analysis

The Genetic Power Calculator (GPC) [22,23] was used to define the minimal genotype relative risk that could reliably be detected for a gene set given the current sample size. Because the PGC data set consists of both case-control samples and trio samples, power was calculated using the weighted mean of the noncentrality parameters of the samples. To use the GPC for gene-set analysis, we assumed that the risk allele frequency represents the average allele frequency of all contributing risk variants in a gene set, and that the relative risk is representing the global effect of the gene set. We further used a disease prevalence of 5% (as estimated by Polanczyk *et al.* [1]), and a multiplicative model (power calculation based on the allelic test). Tests were corrected for the number of gene sets ($\alpha = 0.05/18 = 2.8 \times 10^{-3}$).

2.4. Gene-Set Analysis

Gene-set analysis was conducted using JAG [24]. To test the hypothesis that synaptic function was associated with ADHD, we conducted self-contained tests for each gene set and one overall test including all synaptic gene sets. For each gene set, the test statistic was defined as the sum over the $-\log_{10}$ of SNP *p*-values annotated to genes in that gene set. These SNP *p*-values were taken from the PGC association results. To allow for unbiased interpretation of the test statistic, 10,000 permutations were conducted in which any relation between a genetic variant and affection status was disconnected. As such, linkage disequilibrium (LD), and number of SNPs and genes within each gene set stayed intact. For each permutation of the data set, the test statistics of the gene sets were computed. The self-contained *p*-value was calculated as the proportion of test statistics in the permuted data sets that was higher than the original test statistic. Bonferroni correction was applied to account for multiple testing with a corrected significance threshold of $\alpha = 0.05/18 = 2.8 \times 10^{-3}$.

For the permutations of the data set, we used the genotype data of the European ancestry samples from the 1000 Genomes project [25] with a simulated binary phenotype (as we had no access to raw data of the PGC). Using this as reference data, we could appropriately account for LD effects on correlations in SNP *p*-values in the PGC association data. For the test statistics of the original gene sets, only SNPs that were also available in the 1000 Genomes genotype data were used.

Competitive tests were performed for gene sets found to be significant in the self-contained test. While self-contained tests evaluate whether a gene set is associated with ADHD under the null hypothesis of no association, a competitive test shows whether the observed (self-contained) association is stronger than expected by chance for gene sets with the same number of genes. To this end, 150 random gene sets were generated, matching for the same number of genes. JAG calculated a self-contained *p*-value for each of these random gene sets. The competitive *p*-value was then computed as the proportion of random gene sets with self-contained *p*-values lower than the self-contained *p*-value for the gene set itself. Only gene sets with a competitive *p*-value < 0.05 were considered to be significant.

3. Results

3.1. Power Analysis

Power analyses showed that for gene sets containing on average SNPs with a risk allele frequency (RAF) of at least 0.1, our sample had sufficient power (≥0.80) to detect gene sets with a genotype relative risk (GRR) of 1.23 (Figure 1). For gene sets containing a mean RAF of at least 0.2, we had sufficient power to detect gene sets with a GRR of 1.17.

3.2. Gene-Set Analysis

A total number of 1,206,461 SNPs were available for gene-set analysis. Of these, 61,413 SNPs mapped to 956 genes (out of 1028) within our gene sets. All 956 synaptic genes together were significantly associated with ADHD in the self-contained test (Table 1). However, the competitive test showed that the synaptic genes were not more strongly associated with ADHD than randomly generated gene sets matched for same number of genes, suggesting that the self-contained *p*-value was significant merely due to a large number of SNPs being evaluated, which did not particularly aggregate in genes involved in synaptic function.

Figure 1. Statistical power to detect gene sets in the Psychiatric Genomics Consortium (PGC) Attention Deficit/Hyperactivity Disorder (ADHD) sample. Power is displayed for different genotype relative risks (GRR), and risk allele frequencies (RAF) of 0.1 and 0.2. The weighted mean of the noncentrality parameters of the case-control sample (896 cases and 2455 controls) and trio sample (2064 trios) was used to calculate power. Power analyses assume a disease prevalence of 5% and a multiplicative model. We assumed that gene sets behave as individual single nucleotide polymorphisms (SNPs). Tests are corrected for number of gene sets ($\alpha = 2.8 \times 10^{-3}$). Dotted horizontal line represents power of 0.80.

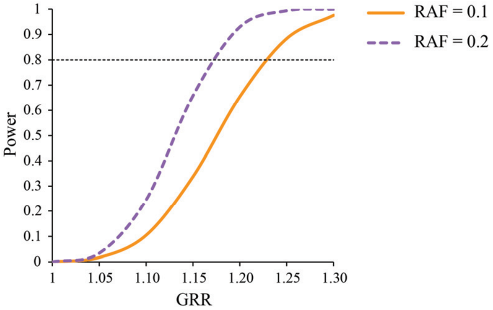


Table 1. Association findings between synaptic gene sets and ADHD.

Gene Set	Number of Genes in Original Set	Number of Genes Present in GWAS Data	Number of SNPs Present in GWAS Data	Self-Contained *p*-Value ($\alpha = 2.8 \times 10^{-3}$)	Competitive *p*-Value ($\alpha = 0.05$)
All synaptic genes	1028	956	61413	0.0393 *	0.1733
Ion balance/transport	43	40	1454	0.0118	NA
Cell metabolism	57	51	1059	0.0429	NA
Endocytosis	26	26	1075	0.0554	NA
Cell adhesion and trans-synaptic signaling	81	76	13550	0.0709	NA
Exocytosis	87	83	4855	0.0962	NA
Protein cluster	47	42	4182	0.1491	NA
Peptide/neurotrophin signals	28	25	1742	0.1659	NA
Structural plasticity	98	90	4655	0.1764	NA
Tyrosine kinase signaling	7	7	1281	0.2030	NA
Neurotransmitter metabolism	29	27	1059	0.2959	NA
RNA and protein synthesis, folding and breakdown	71	64	1152	0.4994	NA
Ligand-gated ion channel signaling	36	32	2935	0.6500	NA
G-protein-coupled receptor signaling	41	40	3129	0.6578	NA
Unassigned	61	53	2258	0.6644	NA
Intracellular signal transduction	150	145	9563	0.7001	NA
G-protein relay	27	25	946	0.7047	NA
Intracellular trafficking	80	75	2024	0.7334	NA
Excitability	59	56	4508	0.7914	NA

* $\alpha = 0.05$.

Self-contained tests for the specific synaptic gene sets showed associations at nominal significance levels for the involvement of *ion balance/transport* and *cell metabolism* in ADHD (Table 1). However, these associations did not survive Bonferroni correction. All other self-contained *p*-values were >0.05. We thus conclude that no significant associations were found between any of the specific synaptic gene sets and ADHD. Consequently, no subsequent competitive tests were performed for the synaptic gene sets of specific functions.

4. Discussion

Results from previous GWAS have led to the conclusion that ADHD is a heritable, yet polygenic disorder influenced by many genetic variants of small effect. Top hits from previous studies have suggested a role for synaptic processes in the etiology of ADHD. In the current study, we tested the hypothesis that genetic variants that influence the risk for ADHD cluster in synaptic gene sets. We used expert-curated gene sets of pre- and postsynaptic genes. Using the largest public ADHD GWAS sample currently available, our study had sufficient statistical power to detect gene sets representing a GRR of at least 1.17 (or 1.23 for less common alleles) for the liability to develop ADHD. The self-contained test of all synaptic genes together showed a significant association with ADHD. However, for complex traits that are polygenic, any large group of genes is likely to be associated due to background polygenic effects. The competitive test showed that the association was not stronger compared to that of randomly generated gene sets with the same number of genes. This suggests that the association was not a result of the selection of synaptic genes, but merely because of the large number of genes. Hence, our results support the idea that ADHD is a polygenic disorder, and suggest that overall synaptic function does not play a major role in the etiology of ADHD, given current synaptic genes.

In addition, no specific synaptic function categories were associated with ADHD after correction for multiple testing. These results suggest that if common genetic variants in the current synaptic gene sets with a specific function play a role in the etiology of ADHD, their effect is modest at most, even when considering the joint effect of multiple genetic variants.

Although previous analyses suggested involvement of several synaptic processes in ADHD [6,7,11–13], it should be kept in mind that the majority of previous results reported non-significant, suggestive results, and hence no strong conclusions could be drawn regarding the impact of those processes on ADHD. For example, a recent study used a different type of categorization of gene sets: they constructed gene sets based on pathways and candidate genes, and did report significant associations of dopamine/norepinephrine and serotonin pathways, and genes involved in neuritic outgrowth, with the hyperactive/impulsive component of ADHD [26]. However, in this study competitive tests to investigate if reported associations were stronger than can be expected by the polygenic nature of ADHD were not performed. Consequently, it remains unclear whether the reported associations are due to the background polygenic effects like our apparent association of synaptic genes with ADHD.

Synaptic function has been implicated and confirmed for other psychiatric disorders, especially schizophrenia [19,24] and bipolar disorder [27,28]. For example, gene sets of *cell adhesion and trans-synaptic signaling* and *excitability* showed replicated associations with schizophrenia [19,24].

Recent cross-disorder analyses by the PGC reported overlap in genetic liability between psychiatric disorders (schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, and ADHD) [29,30]. However, of all five psychiatric disorders, ADHD showed the weakest genetic overlap with other psychiatric disorders, having only a moderate genetic correlation with major depressive disorder, and showing no overlap with schizophrenia, bipolar disorder, and autism spectrum disorder. Our current findings fit into this overall picture of a separate genetic etiology of ADHD, by showing no evidence for an association with common variants in the current curated list of synaptic genes.

The list of genes involved in synaptic function is however a dynamic list: it depends on available experimental data and expert curation. When more experimental data is generated more genes may be included, which may have been missed in the current analyses. However, if genetic variants with an effect on ADHD risk aggregate in genes that are active in the synapse, it is expected that many genes within this gene set play a role in ADHD. Thus, an indication of association should be present if any of our current gene sets has a strong effect on ADHD risk, even when the current gene sets are not complete. Our results do not show any clear trends of association between the gene sets and ADHD.

An alternative explanation for the lack of association in our study could be the heterogeneous nature of ADHD. It is known that ADHD is characterized by a heterogeneous manifestation of symptoms, possibly reflecting genetic heterogeneity [31]. Genetic heterogeneity makes it more challenging to detect genetic variation that plays a role in the etiology of ADHD, as the heterogeneity results in an apparent reduction of the effect sizes of true genetic variants. The current lack of association of synaptic functions with ADHD diagnosis together with previous reports that implicate a role of synaptic function based on smaller scaled samples, may reflect the involvement of synaptic function in only very specific sub-populations of ADHD symptoms. Future studies focusing on ADHD symptom profiles are needed to detect such specific associations between synaptic function and ADHD subtypes.

5. Conclusions

We find no evidence for involvement of specific synaptic functions in the etiology of ADHD, given current sample size and gene sets based on current knowledge of genes related to synaptic function. Our results suggest that if common genetic variants in the current synaptic gene sets play a role in the etiology of ADHD, their effect is modest at most, even when considering the joint effect of multiple genetic variants.

Acknowledgments

This work is supported by the Netherlands Organization for Scientific Research (NWO Brain & Cognition 433-09-228, and NWO Complexity Project 645-000-003) and the European Union Seventh Framework Program (“SynSys” project) (HEALTHF2-2009-242167). The data set used for the analysis described in this paper was obtained from the Psychiatric Genomics Consortium (PGC) webpage [32]. Statistical analyses were carried out on the Genetic Cluster Computer [33] hosted by

SURFsara and financially supported by the Netherlands Organization for Scientific Research (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. Scripts to modify JAG to be able to work with SNP *p*-values instead of raw data as input can be obtained through the first (Anke R. Hammerschlag) or third author (Christiaan de Leeuw) of the current paper.

Author Contributions

Danielle Posthuma conceived the project. Anke R. Hammerschlag performed the analyses and drafted the first version of the manuscript. Anke R. Hammerschlag, Tinca J. C. Polderman and Danielle Posthuma wrote the final version. Matthijs Verhage and August B. Smit constructed the synaptic gene sets. Christiaan de Leeuw wrote scripts to modify JAG to be able to work with SNP *p*-values instead of raw data as input. All authors provided a critical review of the manuscript and approved the final version.

Conflicts of Interest

Matthijs Verhage and August B. Smit are advisors of Sylics BV, which has no conflict of interest with the contents of this study.

References

1. Polanczyk, G.; de Lima, M.S.; Horta, B.L.; Biederman, J.; Rohde, L.A. The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *Am. J. Psychiatry* **2007**, *164*, 942–948.
2. Polanczyk, G.V.; Willcutt, E.G.; Salum, G.A.; Kieling, C.; Rohde, L.A. ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* **2014**, *43*, 434–442.
3. Posthuma, D.; Polderman, T.J.C. What have we learned from recent twin studies about the etiology of neurodevelopmental disorders? *Curr. Opin. Neurol.* **2013**, *26*, 111–121.
4. Visscher, P.M.; Brown, M.A.; McCarthy, M.I.; Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **2012**, *90*, 7–24.
5. Neale, B.M.; Medland, S.E.; Ripke, S.; Asherson, P.; Franke, B.; Lesch, K.P.; Faraone, S.V.; Nguyen, T.T.; Schäfer, H.; Holmans, P.; *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **2010**, *49*, 884–897.
6. Lesch, K.P.; Timmesfeld, N.; Renner, T.J.; Halperin, R.; Röser, C.; Nguyen, T.T.; Craig, D.W.; Romanos, J.; Heine, M.; Meyer, J.; *et al.* Molecular genetics of adult ADHD: Converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.* **2008**, *115*, 1573–1585.
7. Franke, B.; Neale, B.M.; Faraone, S.V. Genome-wide association studies in ADHD. *Hum. Genet.* **2009**, *126*, 13–50.

8. Poelmans, G.; Pauls, D.L.; Buitelaar, J.K.; Franke, B. Integrated genome-wide association study findings: Identification of a neurodevelopmental network for attention deficit hyperactivity disorder. *Am. J. Psychiatry* **2011**, *168*, 365–377.
9. Gizer, I.R.; Ficks, C.; Waldman, I.D. Candidate gene studies of ADHD: A meta-analytic review. *Hum. Genet.* **2009**, *126*, 51–90.
10. Williams, N.M.; Zaharieva, I.; Martin, A.; Langley, K.; Mantripragada, K.; Fossdal, R.; Stefansson, H.; Stefansson, K.; Magnusson, P.; Gudmundsson, O.O.; *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: A genome-wide analysis. *Lancet* **2010**, *376*, 1401–1408.
11. Elia, J.; Glessner, J.T.; Wang, K.; Takahashi, N.; Shtir, C.J.; Hadley, D.; Sleiman, P.M.; Zhang, H.; Kim, C.E.; Robison, R.; *et al.* Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat. Genet.* **2012**, *44*, 78–84.
12. Stergiakouli, E.; Hamshere, M.; Holmans, P.; Langley, K.; Zaharieva, I.; Hawi, Z.; Kent, L.; Gill, M.; Williams, N.; Owen, M.J.; *et al.* Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *Am. J. Psychiatry* **2012**, *169*, 186–194.
13. Williams, N.M.; Franke, B.; Mick, E.; Anney, R.J.L.; Freitag, C.M.; Gill, M.; Thapar, A.; O'Donovan, M.C.; Owen, M.J.; Holmans, P.; *et al.* Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: The role of rare variants and duplications at 15q13.3. *Am. J. Psychiatry* **2012**, *169*, 195–204.
14. Schadt, E.E. Molecular networks as sensors and drivers of common human diseases. *Nature* **2009**, *461*, 218–223.
15. Ruano, D.; Abecasis, G.R.; Glaser, B.; Lips, E.S.; Cornelisse, L.N.; de Jong, A.P.H.; Evans, D.M.; Davey Smith, G.; Timpson, N.J.; Smit, A.B.; *et al.* Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am. J. Hum. Genet.* **2010**, *86*, 113–125.
16. De Jong, A.P.H.; Verhage, M. Presynaptic signal transduction pathways that modulate synaptic transmission. *Curr. Opin. Neurobiol.* **2009**, *19*, 245–253.
17. Feldman, I.; Rzhetsky, A.; Vitkup, D. Network properties of genes harboring inherited disease mutations. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4323–4328.
18. Rossin, E.J.; Lage, K.; Raychaudhuri, S.; Xavier, R.J.; Tatar, D.; Benita, Y.; International Inflammatory Bowel Disease Genetics Constortium; Cotsapas, C.; Daly, M.J. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **2011**, *7*, e1001273.
19. Ripke, S.; O'Dushlaine, C.; Chambert, K.; Moran, J.L.; Kähler, A.K.; Akterin, S.; Bergen, S.E.; Collins, A.L.; Crowley, J.J.; Fromer, M.; *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **2013**, *45*, 1150–1159.
20. Goudriaan, A.; de Leeuw, C.; Ripke, S.; Hultman, C.M.; Sklar, P.; Sullivan, P.F.; Smit, A.B.; Posthuma, D.; Verheijen, M.H.G. Specific glial functions contribute to schizophrenia susceptibility. *Schizophr. Bull.* **2014**, *40*, 925–935.

21. Complex Trait Genetics. Available online: http://ctglab.nl/software/ (accessed on 4 April 2014).
22. Purcell, S.; Cherny, S.S.; Sham, P.C. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **2003**, *19*, 149–150.
23. Purcell, S.; Sham, P.C. Genetic Power Calculator. Available online: http://pngu.mgh.harvard.edu/~purcell/gpc/ (accessed on 3 March 2014).
24. Lips, E.S.; Cornelisse, L.N.; Toonen, R.F.; Min, J.L.; Hultman, C.M.; International Scizophrenia Consortium; Holmans, P.A.; O'Donovan, M.C.; Purcell, S.M.; Smit, A.B.; *et al*. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol. Psychiatry* **2012**, *17*, 996–1006.
25. 1000 Genomes Project Consortium; Abecasis, G.R.; Auton, A.; Brooks, L.D.; DePristo, M.A.; Durbin, R.M.; Handsaker, R.E.; Kang, H.M.; Marth, G.T.; McVean, G.A. An integrated map of genetic variation from 1092 human genomes. *Nature* **2012**, *491*, 56–65.
26. Bralten, J.; Franke, B.; Waldman, I.; Rommelse, N.; Hartman, C.; Asherson, P.; Banaschewski, T.; Ebstein, R.P.; Gill, M.; Miranda, A.; *et al.* Candidate genetic pathways for attention-deficit/hyperactivity disorder (ADHD) show association to hyperactive/impulsive symptoms in children with ADHD. *J. Am. Acad. Child Adolesc. Psychiatry* **2013**, *52*, 1204.e1–1212.e1.
27. Ferreira, M.A.R.; O'Donovan, M.C.; Meng, Y.A.; Jones, I.R.; Ruderfer, D.M.; Jones, L.; Fan, J.; Kirov, G.; Perlis, R.H.; Green, E.K.; *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.* **2008**, *40*, 1056–1058.
28. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **2011**, *43*, 977–983.
29. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **2013**, *381*, 1371–1379.
30. Cross-Disorder Group of the Psychiatric Genomics Consortium; Lee, S.H.; Ripke, S.; Neale, B.M.; Faraone, S.V.; Purcell, S.M.; Perlis, R.H.; Mowry, B.J.; Thapar, A.; Goddard, M.E.; *et al*. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **2013**, *45*, 984–994.
31. Faraone, S.V. Genetics of childhood disorders: XX. ADHD, Part 4: is ADHD genetically heterogeneous? *J. Am. Acad. Child Adolesc. Psychiatry* **2000**, *39*, 1455–1457.
32. Psychiatric Genomics Consortium. Available online: https://pgc.unc.edu/Sharing.php#SharingOpp/ (accessed on 23 October 2013).
33. Genetic Cluster Computer. Available online: http://www.geneticcluster.org/ (accessed on 4 April 2014).

Discovery in Genetic Skin Disease: The Impact of High Throughput Genetic Technologies

Thiviyani Maruthappu, Claire A. Scott and David P. Kelsell

Abstract: The last decade has seen considerable advances in our understanding of the genetic basis of skin disease, as a consequence of high throughput sequencing technologies including next generation sequencing and whole exome sequencing. We have now determined the genes underlying several monogenic diseases, such as harlequin ichthyosis, Olmsted syndrome, and exfoliative ichthyosis, which have provided unique insights into the structure and function of the skin. In addition, through genome wide association studies we now have an understanding of how low penetrance variants contribute to inflammatory skin diseases such as psoriasis vulgaris and atopic dermatitis, and how they contribute to underlying pathophysiological disease processes. In this review we discuss strategies used to unravel the genes underlying both monogenic and complex trait skin diseases in the last 10 years and the implications on mechanistic studies, diagnostics, and therapeutics.

Reprinted from *Genes*. Cite as: Maruthappu, T.; Scott, C.A.; Kelsell, D.P. Discovery in Genetic Skin Disease: The Impact of High Throughput Genetic Technologies. *Genes* **2014**, *5*, 615-634.

1. Introduction

The advent of high throughput single nucleotide polymorphism (SNP) genotyping and latterly, next generation sequencing (NGS) technology including whole exome sequencing (WES) have revolutionised our approach to genetic diagnostics and novel gene discovery in the genodermatoses—a group of inherited skin disorders.

Prior to this, technologies including linkage analysis using genome wide microsatellite panels in combination with candidate gene screening by PCR and Sanger sequencing have been the primary method for discerning new skin disease-associated loci. Successes with this approach include Hailey-Hailey Disease (OMIM #169600) [1], Netherton Syndrome (OMIM #256500) [2], Darier-Disease (OMIM #124200) [3], and Dyschromatosis symmetrica hereditaria (OMIM #127400) [4]. Candidate gene screening approaches have also yielded success, particularly in deciphering the keratin disorders [5]. However, clinical and likely genetic heterogeneity of skin diseases and the availability of DNA from probands only, or from small families, have hindered disease gene discovery for many disorders [6]. This can now be surmounted with high-density SNP homozygosity mapping for consanguineous recessive disorders, and in particular NGS and WES for dominant and recessive disorders, which has facilitated our understanding of some of the genetic make up of common diseases.

Skin diseases are ideal for determining genotype-phenotype correlations because of the relative ease with which clinical and histological examination can be made. In addition, inflammatory pathways involved in the pathogenesis of skin diseases such as psoriasis vulgaris (PV) are relevant to a number of other immune-mediated diseases including inflammatory bowel disease and rheumatoid arthritis [7].

The genetic bases of many monogenic skin diseases have been unravelled and in this review we focus on examples of discoveries in cutaneous genetics, applying different strategies such as SNP microarray, microsatellite linkage analysis, targeted NGS and WES. Equally, it has also been informative in understanding the significance of *de novo* mutations including the unusual phenomenon of revertant mosaicism in the skin, where spontaneous correction of a disease-causing mutation in a somatic cell occurs [8]. We have also gained insights into complex trait diseases and will explore what contributions these have made to mechanistic insights, diagnosis and treatment of common skin diseases including psoriasis, atopic dermatitis (AD) (eczema), and skin cancer.

2. Harlequin Ichthyosis

The discovery that *ABCA12* gene mutations are associated with the skin disease harlequin ichthyosis (HI) is an example of where SNP microarray technology was used successfully to elucidate the genetic locus associated with this disease [9].

The inherited ichthyoses are a heterogeneous group of disorders characterised by skin scaling, often of the whole surface, and hyperkeratosis [10]. Syndromic (affecting multiple tissues) as well as nonsyndromic forms of ichthyosis exist and mutations in multiple genes are associated with disease including *TGM1* (OMIM *190195), *NIPAL4* (OMIM *609383), *STS* (OMIM *300747), *ALOX12B* (OMIM *603741), *ALOXE3* (OMIM *607206), *CYP4F22* (OMIM *611495), and *FLG* (OMIM *135940) amongst others (reviewed in [10]). Autosomal recessive congenital ichthyosis (ARCI) is comprised of three main groups: congenital ichthyosiform erythroderma (CIE), lamellar ichthyosis (LI) and HI [10]. HI (OMIM #242500) is the most severe form of ichthyosis and has a high perinatal mortality, with babies presenting at birth with hard scale plates with deep fissures, eclabium, and bilateral ectropion (reviewed in [9,11]).

The discovery of the genetic cause of HI was hampered by availability of DNA from only affected family members or from small families due to the severity of the condition, thus genetic linkage studies were unfeasible [9]. To investigate the genetic basis of HI, Kelsell *et al.* (2005) [9] used a SNP microarray to map a block of homozygosity on chromosome 2q35 and to identify a minimal region between HI patients from consanguineous parents, which contained the *ABCA12* gene. ABCA12 belongs to the ATP-binding cassette (ABC) A family of transporters, some members of which have been implicated in lipid transport (reviewed in [12]).

ABCA12 was a promising gene candidate for HI because patient skin displayed aberrant lipid distribution [9] and missense mutations in *ABCA12* were already known to be associated with another form of ARCI, LI [13]. PCR and Sanger sequencing of the *ABCA12* gene in HI patients confirmed that recessive mutations were associated with HI [9,14]. Mutations in *ABCA12* are now known to be associated with all three forms of ARCI (reviewed in [10]). However, unlike for LI and CIE, in which largely missense *ABCA12* mutations are associated with disease [13,15,16], HI is usually associated with loss of function gene mutations including nonsense, frameshift, and splice site mutations, which severely disrupt the cellular functions of ABCA12 [9,17–19]. However, there are reports of patients who have ABCA12 missense mutations [9,11,18,20–22]. HI patients with homozygous loss of function mutations have an increased risk of mortality, indicating a survival advantage for patients with compound heterozygous mutations [11].

ABCA12 is thought to transport lipids via lamellar granules where they are processed and released to form lipid lamellae constituting the stratum corneum in the epidermis [14,23]. A reduction in the number, and structural abnormalities, of lamellar granules has been observed in HI patient skin [14,24,25]. In addition, characterisation of HI patient skin has shown a loss of nonpolar lipids [26] and abnormal glucosylceramide localization [14], and experiments with patient-derived keratinocytes showed aberrant glucosylceramide accumulation in lamellar granules [27], which is indicative of a lipid transport defect as a result of loss of ABCA12 function [14,26,27].

Similarly, *Abca12* knockout mice models [28–30] and an *abca12* knockout zebrafish model [31] showed features of aberrant lipid transport compared to controls (reviewed in [32]). HI skin also shows features of premature terminal differentiation and a decreased expression of certain proteases, which suggests that loss of ABCA12 disrupts keratinocyte differentiation and epidermal desquamation, resulting in the formation of an aberrant epidermal barrier [26].

Prior to the discovery of the genetic cause of HI, prenatal diagnostic investigations depended on obtaining a foetal biopsy for analysis by electron microscopy, and on sonography [33,34] (reviewed in [35]). The discovery of the genetic cause of different ichthyoses, including HI, represents a major milestone in the ability to perform genetic diagnosis, carrier screening, genetic counselling, and prenatal diagnosis.

Current approaches to genetic screening for HI can involve screening specific exons, as there are some recurrent ethnic group mutations in *ABCA12* [18,19] and using WES, circumventing the need for performing PCR and Sanger sequencing of all 53 coding exons of the *ABCA12* gene.

3. Exfoliative Ichthyosis

The discovery of cystatin A (*CSTA*) gene mutations in association with exfoliative ichthyosis [36] is an example of the successful implementation of combining SNP microarray analysis with targeted NGS to determine the genetic cause of disease.

Autosomal recessive exfoliative ichthyosis (OMIM #607936) is characterised by palmoplantar skin peeling and dry scaly skin, with trauma and moisture aggravating the condition [36]. Microsatellite linkage analysis of two related Bedouin families initially suggested linkage of the disease to chromosome 12q13, which contains the type II keratin cluster [37].

Blaydon *et al.* (2011) [36] revisited this family and applied whole genome homozygosity mapping which revealed a common block of homozygosity between affected Bedouin patients on chromosome 3q21 as the likely disease gene location. Sequence capture and NGS of this region was then performed and revealed a splice site mutation in *CSTA*, which was found to segregate with exfoliative ichthyosis in the Bedouin family. This locus was missed in the microsatellite genome scans performed by Hatsell *et al.* (2003) [37] due to markers for this region being uninformative. Sanger sequencing of *CSTA* in a different family with exfoliative ichthyosis revealed a homozygous nonsense mutation which also segregated with disease [36]. In a subsequent study, WES revealed a novel homozygous nonsense mutation in *CSTA* in a large family with acral peeling skin syndrome [38] with similar clinical features to the patients reported in Blaydon *et al.* (2011) [36].

Cystatins are cysteine protease inhibitors which are thought to have a protective function against endogenous and external proteases, and to potentially modulate the degradation of intra- and extracellular proteins (reviewed in [39]). CSTA has been identified as a constituent of the cornified envelope [40] and is expressed in the suprabasal layers of the epidermis, the highest expression of which is in the granular layer [36,41]. CSTA is secreted by keratinocytes *in vitro* and has also been found in sweat, and is believed to have a protective role by inhibiting the proteolytic activity of dust mite allergens Der p 1 and Der f 1 [42]. CSTA levels have also been implicated as prognostic markers in different cancers [43–45].

Characterisation of skin from exfoliative ichthyosis patients with *CSTA* mutations revealed widened intercellular gaps in the lower epidermis, whereas the upper epidermal layers appeared normal with no evident barrier defect [36]. Experiments using an *in vitro* keratinocyte cell knockdown model showed an adhesion defect in response to mechanical stress, and an organotypic *CSTA* knockdown model showed similar abnormalities to the patient skin [36]. This finding is indicative of CSTA having a key role in keratinocyte adhesion in the basal epidermal layers and that loss of CSTA causes a predisposition to epidermal splitting. There were no obvious abnormalities in a murine model with a chromosomal deletion, which included the *Csta* gene [46], although investigation of a skin phenotype was not described.

4. Olmsted Syndrome

The genetic basis of various skin diseases (Table 1) has been determined using exome sequencing technology. One example where WES enabled the identification of the underlying causative genes is Olmsted syndrome (OS) [47,48]. OS (OMIM #614594) is a rare disorder characterised by mutilating palmoplantar keratoderma and periorificial keratosis. Additional clinical features include constriction of the digits, dystrophy of the nails, diffuse alopecia and a predisposition to infection and development of squamous cell carcinoma on keratotic lesions [47]. Different modes of inheritance have been hypothesised [47–50].

Table 1. Examples of genes associated with skin disease discovered using exome sequencing technology.

Gene	Disease	Mode of Inheritance	Reference
AAGAB	Punctate palmoplantar keratoderma Type I	AD	[51,52]
ADAM10	Reticulate acropigmentation of Kitamura	AD	[53]
AQP5	Nonepidermolytic palmoplantar keratoderma	AD	[54]
ENPP1	Cole disease	AD	[55]
EXPH5	Inherited skin fragility	AR	[56]
HOXC13	Pure hair and nail ectodermal dysplasia	AR	[57]
KANK2	Palmoplantar keratoderma and woolly hair	AR	[58]
MBTPS2	Olmsted syndrome	XLR	[48]
POFUT1	Dowling-Degos disease	AD	[59]
POGLUT1	Dowling-Degos disease	AD	[60]
SERPINB7	Nagashima-type palmoplantar keratosis	AR	[61]
TRPV3	Olmsted syndrome	AD/AR	[47]/[62]

AD: autosomal dominant; AR: autosomal recessive; XLR: X-linked recessive.

WES was used successfully to identify mutations in the Transient Receptor Potential Cation Channel, Subfamily V, Member 3 (*TRPV3*) gene [47], and the Membrane-Bound Transcription Factor Protease, Site 2 (*MBTPS2*) gene [48] to be associated with OS.

Lin *et al.* (2012) performed WES of an OS patient and her unaffected parents and identified a novel *de novo* heterozygous mutation p.G573S in *TRPV3* [47]. Screening for *TRPV3* mutations in five other OS patients revealed that three were heterozygous for p.G573S, one heterozygous for p.G573C and one heterozygous for p.W692G [47].

TRPV3 is a member of the TRPV cation channel family, and is known to be expressed in various tissue types including skin and hair follicles [63–65]. The murine *TRPV3* mutants p.G573S and p.G573C were discovered in spontaneous hairless rodent strains that develop dermatitis, a trait inherited in an autosomal dominant manner [66]. *Trpv3* knockout mice display wavy hair, curly whiskers and a defective skin barrier, and it is believed that TRPV3 associates with TGF-α/EGFR in a signalling pathway to modulate keratinocyte differentiation and hair morphogenesis [67].

In vitro functional studies with the three OS-associated TRPV3 mutants indicated that they are gain of function mutants, creating constitutively open channels and causing increased cell death of cells expressing the mutants [47]. Similar results were obtained in *in vitro* expression studies with the murine TRPV3 mutants p.G573S and p.G573C [68]. It has been hypothesised that *in vivo* the mutants may cause apoptosis and subsequent keratoderma in patients, and could contribute to their pruritis [47].

A subsequent study using WES revealed the recurrent *TRPV3* mutation p.G573S in sporadic OS [69]. Screening by Sanger sequencing has also revealed a homozygous mutation in an OS patient, indicating recessive inheritance [62]. Both recessive [70] and sporadic [71] *TRPV3* mutations have been associated with atypical OS with erythromelalgia.

Exome sequencing of two affected males reported previously in a consanguineous pedigree [72] in which OS followed a suggested X-linked recessive inheritance pattern, revealed a novel *MBTPS2* gene mutation which segregated with disease in the family [48]. This discovery expands the number of disorders attributed to *MBTPS2* gene mutations, as other mutations in this gene are associated with ichthyosis follicularis with atrichia and photophobia (IFAP) syndrome [73–75], BRESEK/BRESHECK syndrome [76], and keratosis follicularis spinulosa decalvans (KFSD) [77].

MBTPS1 and MBTPS2 are involved in activating signalling proteins such as the transcription factors SREBPS, enabling cells to respond to sterols [78,79] and in the processing of ATF6, which is a component of the unfolded protein response (UPR) [80]. *In vitro* functional studies with IFAP and KFSD MBTPS2 mutants revealed decreased sterol responsiveness compared to wild-type [73,77], and mutants which caused the greatest impairment of enzyme activity seemed to be associated with increased disease severity in patients [73].

5. Complex Traits of the Skin

In the last 10 years there have been landmark discoveries in our understanding of the genetic basis and pathophysiology of inflammatory skin diseases, most notably PV and AD. Both are common, complex diseases, in which a host of environmental factors can trigger disease in genetically susceptible individuals [81,82]. Inflammatory dermatoses are associated with both a significant burden on healthcare resources and patients' quality of life [83,84].

Identification of susceptibility loci for PV and AD have resulted from developments in genome wide association studies (GWAS), which have been applied to all common disorders. Information has been generated by the HapMap and 1000 Genomes projects, in parallel with the technology to genotype multiple individual DNA samples at one million or more loci, allowing SNPs to be reviewed and enabling comparisons of allele frequency between large numbers of cases and controls to identify those which confer risk of disease [85]. The development of DNA microarray based genotyping allows up to a million SNPs to be tested simultaneously.

6. Psoriasis

PV is a common and chronic inflammatory disease, which can affect the skin, nails and joints. It is characterised by immune-mediated epidermal hyperproliferation [86]. It is a highly heritable disease, with increased concordance in monozygotic *versus* dizygotic twins (65%–72% *versus* 15%–30% respectively) [87]. During the last 10 years, almost 40 GWAS-identified novel psoriasis-susceptibility loci have been identified and more recently, the genes within these loci and their significance to the pathophysiology of PV are becoming clearer [88]. Interestingly, several show clustering to a distinct segment of the inflammatory cascade [89]. Psoriasis susceptibility locus 1 (*PSORS1*), located on the MHC region on chromosome 6p21, has been most consistently identified in GWAS with a significant odds ratio of 3.0 [90]. Genes implicated within this 250 kb interval include *HLA-C* (human leukocyte antigen C), *CCHCR1* (coiled-coil α-helical rod protein 1), and *CDSN* (corneodesmosin). These were considered as potential disease-associated genes due to their function and the presence of disease-associated SNPs within their coding sequence [91]. Identification of the causal disease susceptibility allele was extremely challenging, ultimately Nair *et al.* (2006) sequenced the entire *PSORS1* region in individuals bearing different *HLA-C* alleles to identify SNPs unique to the *PSORS1* haplotype. They indicated that *HLA-Cw6* was the major *PSORS1* disease allele [92], reflecting the importance of antigen presentation in the pathophysiology of PV.

Identification of susceptibility loci has contributed to our understanding of PV pathogenesis, which appears to involve the innate and adaptive immune responses. Pathways that have been identified in various studies include IL12/IL17 axis activation (*IL23R*, *IL12B*, *IL23A*, and *TRAF31P2*), type 1 interferon induction (*IFIH1*, *RNF114*, and *TYK2*) and NF-κB signaling (*CARD14*, *REL*, *NFKBIA*, *TNFAIP3*, and *TNF1P*) [89,90,92–97]. Of particular interest is the Th1-Th17 axis, involving the recently described subset of IL17 expressing T cells (Th17) [98] which is thought to play a major role in the development and maintenance of psoriatic plaques [97].

IL12 and IL23 are cytokines that induce naïve CD4^{+} lymphocytes to differentiate into type 1 helper cells and type 17 helper cells, both of which are key mediators of PV [97]. IL12 and IL23 share a common p40 subunit encoded by the *IL12B* gene. In mice, injection of IL23 results in epidermal hyperplasia, which is mediated by IL22 produced by Th17 cells. This shows similarities to phenomena observed in humans [99]. GWAS have identified three SNPs with strong evidence of association with PV mapping near *IL12B*, *IL23A* (encoding the p19 subunit of IL23) and *IL23R* (encoding a subunit of the IL23 receptor) [94] raising the possibility that dysregulated IL23 signaling could lead to chronic immune responses within epithelial cells. Ustekinumab (Stelara$^{®}$) is a human IgG1κ monoclonal antibody against the p40 subunit of the IL12 and IL23 cytokines that has demonstrated significant improvement in outcome measures for the treatment of PV in Phase III clinical trials [100]. A significant proportion of patients had at least 90% improvement in their psoriasis area-and-severity index (PASI) score, with a proportion experiencing complete clearance by 12 weeks [100]. These findings also establish a central role for the IL12/IL23 p40 cytokines in the pathophysiology of PV.

Another approach to utilise the discoveries gained from GWAS studies is personalised medicine. For example, patients with PV who carry risk variants in *IL12B* may benefit preferably from a monoclonal antibody targeting its p40 subunit, e.g., Ustekinumab. Studies using molecular profiling of PV and clinical phenotyping to predict treatment response have shown promise [101] and larger studies are underway. This is one example of how PV has been used as a paradigm for autoimmune disease and for proof-of-principle studies of targeted biologic therapies, because of the ease of accessing the skin and objectively measuring disease severity and responses to treatment.

Rare variants with large effect have been observed in families where PV segregates as an apparent Mendelian trait. The psoriasis susceptibility locus 2 (*PSORS2*) was first mapped in 1994 to human chromosomal region 17q25-qter in a large family of European ancestry [102]. More recently, it has been shown that *PSORS2* is due to gain of function mutations in the caspase recruitment domain family member 14 (*CARD14*) [96] using linkage analysis, targeted and exome capture in combination with NGS. On the basis of these findings, further work has uncovered rare missense variations in *CARD14* linked to PV using a large case-control study [95]. *CARD14* encodes a NF-κB activator within the skin epidermis. The mutations identified lie within the coiled-coiled domain of *CARD14* and result in enhanced NF-κB activity compared with wild-type CARD14 [95].

Generalised pustular psoriasis (GPP) can present with an acute, widespread and life-threatening eruption associated with fever and leukocytosis. It has long been considered a variant of PV. Mutations in *IL36RN*, which encodes the IL36 receptor antagonist and abrogates downstream activation of NF-κB signaling, have been shown to underlie GPP in consanguineous pedigrees of North African origin [103]. This mutation results in enhanced production of IL1, IL6, and IL8 inflammatory cytokines, which may contribute to the profound systemic inflammatory response seen clinically in these patients [103]. Similar recessive mutations in *IL36RN* have not been observed in patients with PV alone [104]. Genetic studies suggest that in fact, PV and GPP are etiologically distinct clinical entities, which consequently have important therapeutic implications [105].

7. Atopic Dermatitis (Eczema)

AD is a chronic inflammatory skin disease characterised by disturbed skin barrier function and dry, itchy skin. Its prevalence worldwide is increasing and in some countries affects almost 20% of children [106]. Like PV, concordance is observed in twin studies with rates of 0.72–0.86 in monozygotic and 0.21–0.23 in dizygotic twin pairs [107]. A complex interplay between environmental, genetic and immunological factors, as for many common disorders, all contribute to susceptibility and severity.

The filaggrin story is central to our understanding of AD and ichthyosis vulgaris (IV). It exemplifies how the study of a monogenic disorder can translate to a complex trait disease. In 2006, null mutations in the filaggrin gene *FLG* were first identified in Irish families with IV, which often causes dry, scaly skin and is also a strong genetic risk factor for AD [108]. Histological evidence for the possible lack of filaggrin in IV dates back to 1985 [109] however these preliminary studies were hindered by the daunting size and repetitive nature of *FLG*, particularly exon 3. The McLean group developed a successful strategy to analyse this locus with the use of long range PCR to amplify exon 3 in combination with short specific PCRs to amplify remaining overlapping fragments that were then used to reconstruct the repetitive sequence [108]. Further research has identified significant associations of *FLG* mutations with atopic asthma, allergic rhinitis and peanut allergy [110], as well as early onset and increased severity of AD [111]. These studies have been reproduced in a variety of geographical populations, including European, Japanese, Taiwanese, Chinese, and Korean [112–114]. Indeed, the correlation between *FLG* mutations and AD is considered one of the most robust examples of genotype-phenotype relationship in complex trait disease with an odds ratio of up to 13.4 [115].

Filaggrin plays a key role in epidermal barrier function. Briefly, its degradation products act as “natural moisturising factors” in the skin and assist the formation of a flattened granular cell layer upon keratinocyte terminal differentiation [116]. Studies describing murine models of filaggrin haploinsufficiency have shown skin barrier impairment and enhanced sensitisation to percutaneous allergens [117,118]. The significant effect of *FLG* mutations on AD risk highlights the role of impaired skin barrier function in the pathogenesis of atopic diseases. Filaggrin replacement therapies could prove significant in the management of AD. Recently, Otsuka *et al.* (2014) [119] identified a novel compound JTC801, with potential therapeutic applicability. This has been shown to increase expression of filaggrin in both human and murine keratinocytes and, when administered orally, it can hinder the development of AD-like inflammation in the NC/nga AD mouse model [119].

Although the AD spotlight has focused largely on filaggrin, several other genes have been implicated in the pathogenesis of this disorder. To date, a total of 19 genome-wide significant ($p < 5 \times 10^{-8}$) susceptibility loci have been identified through GWAS [120]. The first GWAS data was published in 2009 and included 939 cases and 975 controls in addition to 270 complete nuclear families with two affected siblings [121]. It identified a novel susceptibility locus in 11q13.5, located 38 kb downstream of *C11orf30*. The peak association was observed 68 kb upstream of the leucine rich repeat containing 32 gene (*LRRC32*) which has been shown to be expressed in activated human regulatory T cells [122]. Carriers have a risk of developing AD that is 1.47 times that of controls [121]. A 2011 Meta analysis of GWAS for AD included 5606 cases and 20565 controls and an additional 5419 cases and 19833 controls in a validation study [114]. Three novel risk loci reached genome-wide significance: rs479844 upstream of ovo-like zinc finger 1 (*OVOL1*), rs2164983 near actin-like 9 (*ACTL9*) and rs2897442 in kinesin family member 3A (*KIF3A*). They also confirmed association with the *FLG* locus. *OVOL1* disruption in mice leads to keratinocyte hyperproliferation and hair shaft abnormalities [93]. It is thought to play a role in regulating epidermal proliferation and loricrin expression, impairing premature terminal differentiation [123]. *KIF3A* associated SNPs map within a cluster of cytokine and immune mediated genes including Th2 cytokine genes: *IL13* and *IL4*. These cytokines have been implicated in other autoimmune and inflammatory diseases including PV [124], Crohn's Disease [125] and asthma [125]. Increased levels of Th2 cytokines such as these have been reported in AD as well as greater levels of mRNA expression in acute skin lesions compared with unaffected skin in patients [126–128]. These GWAS findings highlight the role of skin barrier function (*FLG*), epidermal proliferation and differentiation (*OVOL1*) and the adaptive immune system response (*IL13-RAD50*, *LRRC32*) in the pathophysiology of AD.

Despite these promising discoveries, less than 20% of disease variance has been explained [129]. The phenomenon of "missing heritability" has been observed across other complex diseases and suggests that unmapped common and rare variants with small effect size in GWAS as well as genetic interactions may contribute to the remaining heritability [129]. Epigenetic studies focusing on the contribution of DNA and chromatin methylation may also explain the role that they play in the formation and progression of complex diseases by regulating gene expression [130]. Future work integrating GWAS and epigenetic data may provide insights into our understanding of complex trait disease. In summary, GWAS data reinforces the concept that multiple low risk variants are most likely to contribute to AD and PV, but that larger sample sizes may be necessary to identify them.

8. Conclusions

The post-Human Genome Project era has seen remarkable advances in our understanding of genes underlying both rare and common skin disease. Such insights have proved significant beyond the field of dermatology because of shared mechanisms of disease for example, PV and inflammatory bowel disease. The wider relevance of skin disease is highlighted by the fact that skin is frequently a marker of internal disease. For example, mutations in *ADAM17* not only cause inflammatory skin and bowel disease but increased susceptibility to infection and cardiomyopathy [131]. Similarly, the study of tylosis with oesophageal cancer, an autosomal dominant cancer syndrome that presents with skin thickening of the palms and soles, has brought to light the role of the inactive rhomboid family member iRHOM2 in cancer pathophysiology [132] and wound healing [133]. This also highlights that mechanistic studies are facilitated by the relative ease with which patient material can be obtained by skin biopsy to derive cell lines for functional studies.

Skin disease is particularly remarkable for its intragenic heterogeneity, for example distinct dominant and recessive mutations in the desmosomal Desmoplakin gene *DSP* can result in a spectrum of disease phenotypes ranging from arrhythmogenic right ventricular cardiomyopathy (ARVC) and striate palmoplantar keratoderma to palmoplantar keratoderma with woolly hair and ARVC (reviewed in [134]).

GWAS, WES and whole genome sequencing (WGS) involving increasingly larger cohorts of ethnically diverse populations may also identify additional low and high penetrance variants that contribute to phenotypic variability. WGS is becoming increasingly affordable and offers scope to become the most cost-effective method for genetic diagnostics. In parallel, advances in bioinformatics and statistics are necessary to analyse the vast quantity of data generated by these studies, and distinguish significant findings. We may also see a move towards re-classification of skin diseases and malignancies based on genome sequence and subsequently, a targeted therapeutic approach to optimise treatment outcome.

Acknowledgments

TM is funded by a Medical Research Council Clinical Fellowship and CAS is funded by a Barts and The London Charity project grant (awarded to DPK).

Author Contributions

Thiviyani Maruthappu, Claire A. Scott and David P. Kelsell wrote the paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Sudbrak, R.; Brown, J.; Dobson-Stone, C.; Carter, S.; Ramser, J.; White, J.; Healy, E.; Dissanayake, M.; Larregue, M.; Perrussel, M.; *et al.* Hailey-hailey disease is caused by mutations in ATP2C1 encoding a novel Ca^{2+} pump. *Hum. Mol. Genet.* **2000**, *9*, 1131–1140.
2. Chavanas, S.; Bodemer, C.; Rochat, A.; Hamel-Teillac, D.; Ali, M.; Irvine, A.D.; Bonafe, J.L.; Wilkinson, J.; Taieb, A.; Barrandon, Y.; *et al.* Mutations in spink5, encoding a serine protease inhibitor, cause netherton syndrome. *Nat. Genet.* **2000**, *25*, 141–142.
3. Sakuntabhai, A.; Ruiz-Perez, V.; Carter, S.; Jacobsen, N.; Burge, S.; Monk, S.; Smith, M.; Munro, C.S.; O'Donovan, M.; Craddock, N.; *et al.* Mutations in ATP2A2, encoding a Ca^{2+} pump, cause darier disease. *Nat. Genet.* **1999**, *21*, 271–277.
4. Miyamura, Y.; Suzuki, T.; Kono, M.; Inagaki, K.; Ito, S.; Suzuki, N.; Tomita, Y. Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. *Am. J. Hum. Genet.* **2003**, *73*, 693–699.
5. Irvine, A.D.; McLean, W.H. The molecular genetics of the genodermatoses: Progress to date and future directions. *Br. J. Dermatol.* **2003**, *148*, 1–13.
6. Mardis, E.R. Next-generation DNA sequencing methods. *Ann. Rev. Genomics Hum. Genet.* **2008**, *9*, 387–402.
7. Ainsworth, C. Immunology: A many layered thing. *Nature* **2012**, *492*, S52–S54.
8. Choate, K.A.; Lu, Y.; Zhou, J.; Choi, M.; Elias, P.M.; Farhi, A.; Nelson-Williams, C.; Crumrine, D.; Williams, M.L.; Nopper, A.J.; *et al.* Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. *Science* **2010**, *330*, 94–97.

9. Kelsell, D.P.; Norgett, E.E.; Unsworth, H.; Teh, M.T.; Cullup, T.; Mein, C.A.; Dopping-Hepenstal, P.J.; Dale, B.A.; Tadini, G.; Fleckman, P.; *et al.* Mutations in ABCA12 underlie the severe congenital skin disease harlequin ichthyosis. *Am. J. Hum. Genet.* **2005**, *76*, 794–803.
10. Oji, V.; Tadini, G.; Akiyama, M.; Blanchet Bardon, C.; Bodemer, C.; Bourrat, E.; Coudiere, P.; DiGiovanna, J.J.; Elias, P.; Fischer, J.; *et al.* Revised nomenclature and classification of inherited ichthyoses: Results of the first ichthyosis consensus conference in soreze 2009. *J. Am. Acad. Dermatol.* **2010**, *63*, 607–641.
11. Rajpopat, S.; Moss, C.; Mellerio, J.; Vahlquist, A.; Ganemo, A.; Hellstrom-Pigg, M.; Ilchyshyn, A.; Burrows, N.; Lestringant, G.; Taylor, A.; *et al.* Harlequin ichthyosis: A review of clinical and molecular findings in 45 cases. *Arch. Dermatol.* **2011**, *147*, 681–686.
12. Dean, M.; Rzhetsky, A.; Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **2001**, *11*, 1156–1166.
13. Lefevre, C.; Audebert, S.; Jobard, F.; Bouadjar, B.; Lakhdar, H.; Boughdene-Stambouli, O.; Blanchet-Bardon, C.; Heilig, R.; Foglio, M.; Weissenbach, J.; *et al.* Mutations in the transporter ABCA12 are associated with lamellar ichthyosis type 2. *Hum. Mol. Genet.* **2003**, *12*, 2369–2378.
14. Akiyama, M.; Sugiyama-Nakagiri, Y.; Sakai, K.; McMillan, J.R.; Goto, M.; Arita, K.; Tsuji-Abe, Y.; Tabata, N.; Matsuoka, K.; Sasaki, R.; *et al.* Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer. *J. Clin. Investig.* **2005**, *115*, 1777–1784.
15. Natsuga, K.; Akiyama, M.; Kato, N.; Sakai, K.; Sugiyama-Nakagiri, Y.; Nishimura, M.; Hata, H.; Abe, M.; Arita, K.; Tsuji-Abe, Y.; *et al.* Novel ABCA12 mutations identified in two cases of non-bullous congenital ichthyosiform erythroderma associated with multiple skin malignant neoplasia. *J. Investig. Dermatol.* **2007**, *127*, 2669–2673.
16. Sakai, K.; Akiyama, M.; Yanagi, T.; McMillan, J.R.; Suzuki, T.; Tsukamoto, K.; Sugiyama, H.; Hatano, Y.; Hayashitani, M.; Takamori, K.; *et al.* ABCA12 is a major causative gene for non-bullous congenital ichthyosiform erythroderma. *J. Investig. Dermatol.* **2009**, *129*, 2306–2309.
17. Akiyama, M. ABCA12 mutations and autosomal recessive congenital ichthyosis: A review of genotype/phenotype correlations and of pathogenetic concepts. *Hum. Mutat.* **2010**, *31*, 1090–1096.
18. Thomas, A.C.; Cullup, T.; Norgett, E.E.; Hill, T.; Barton, S.; Dale, B.A.; Sprecher, E.; Sheridan, E.; Taylor, A.E.; Wilroy, R.S.; *et al.* ABCA12 is the major harlequin ichthyosis gene. *J. Investig. Dermatol.* **2006**, *126*, 2408–2413.
19. Thomas, A.C.; Sinclair, C.; Mahmud, N.; Cullup, T.; Mellerio, J.E.; Harper, J.; Dale, B.A.; Turc-Carel, C.; Hohl, D.; McGrath, J.A.; *et al.* Novel and recurring ABCA12 mutations associated with harlequin ichthyosis: Implications for prenatal diagnosis. *Br. J. Dermatol.* **2008**, *158*, 611–613.

20. Akiyama, M.; Sakai, K.; Sugiyama-Nakagiri, Y.; Yamanaka, Y.; McMillan, J.R.; Sawamura, D.; Niizeki, H.; Miyagawa, S.; Shimizu, H. Compound heterozygous mutations including a *de novo* missense mutation in ABCA12 led to a case of harlequin ichthyosis with moderate clinical severity. *J. Investig. Dermatol.* **2006**, *126*, 1518–1523.
21. Umemoto, H.; Akiyama, M.; Yanagi, T.; Sakai, K.; Aoyama, Y.; Oizumi, A.; Suga, Y.; Kitagawa, Y.; Shimizu, H. New insight into genotype/phenotype correlations in ABCA12 mutations in harlequin ichthyosis. *J. Dermatol. Sci.* **2011**, *61*, 136–139.
22. Scott, C.A.; Plagnol, V.; Nitoiu, D.; Bland, P.J.; Blaydon, D.C.; Chronnell, C.M.; Poon, D.S.; Bourn, D.; Gardos, L.; Csaszar, A.; *et al.* Targeted sequence capture and high-throughput sequencing in the molecular diagnosis of ichthyosis and other skin diseases. *J. Investig. Dermatol.* **2013**, *133*, 573–576.
23. Sakai, K.; Akiyama, M.; Sugiyama-Nakagiri, Y.; McMillan, J.R.; Sawamura, D.; Shimizu, H. Localization of ABCA12 from Golgi apparatus to lamellar granules in human upper epidermal keratinocytes. *Exp. Dermatol.* **2007**, *16*, 920–926.
24. Dale, B.A.; Holbrook, K.A.; Fleckman, P.; Kimball, J.R.; Brumbaugh, S.; Sybert, V.P. Heterogeneity in harlequin ichthyosis, an inborn error of epidermal keratinization: Variable morphology and structural protein expression and a defect in lamellar granules. *J. Investig. Dermatol.* **1990**, *94*, 6–18.
25. Milner, M.E.; O'Guin, W.M.; Holbrook, K.A.; Dale, B.A. Abnormal lamellar granules in harlequin ichthyosis. *J. Investig. Dermatol.* **1992**, *99*, 824–829.
26. Thomas, A.C.; Tattersall, D.; Norgett, E.E.; O'Toole, E.A.; Kelsell, D.P. Premature terminal differentiation and a reduction in specific proteases associated with loss of ABCA12 in harlequin ichthyosis. *Am. J. Pathol.* **2009**, *174*, 970–978.
27. Mitsutake, S.; Suzuki, C.; Akiyama, M.; Tsuji, K.; Yanagi, T.; Shimizu, H.; Igarashi, Y. ABCA12 dysfunction causes a disorder in glucosylceramide accumulation during keratinocyte differentiation. *J. Dermatol. Sci.* **2010**, *60*, 128–129.
28. Yanagi, T.; Akiyama, M.; Nishihara, H.; Ishikawa, J.; Sakai, K.; Miyamura, Y.; Naoe, A.; Kitahara, T.; Tanaka, S.; Shimizu, H. Self-improvement of keratinocyte differentiation defects during skin maturation in ABCA12-deficient harlequin ichthyosis model mice. *Am. J. Pathol.* **2010**, *177*, 106–118.
29. Zuo, Y.; Zhuang, D.Z.; Han, R.; Isaac, G.; Tobin, J.J.; McKee, M.; Welti, R.; Brissette, J.L.; Fitzgerald, M.L.; Freeman, M.W. ABCA12 maintains the epidermal lipid permeability barrier by facilitating formation of ceramide linoleic esters. *J. Biol. Chem.* **2008**, *283*, 36624–36635.
30. Smyth, I.; Hacking, D.F.; Hilton, A.A.; Mukhamedova, N.; Meikle, P.J.; Ellis, S.; Satterley, K.; Collinge, J.E.; de Graaf, C.A.; Bahlo, M.; *et al.* A mouse model of harlequin ichthyosis delineates a key role for ABCA12 in lipid homeostasis. *PLoS Genet.* **2008**, *4*, e1000192.
31. Li, Q.; Frank, M.; Akiyama, M.; Shimizu, H.; Ho, S.Y.; Thisse, C.; Thisse, B.; Sprecher, E.; Uitto, J. Abca12-mediated lipid transport and Snap29-dependent trafficking of lamellar granules are crucial for epidermal morphogenesis in a zebrafish model of ichthyosis. *Dis. Model. Mech.* **2011**, *4*, 777–785.

32. Scott, C.A.; Rajpopat, S.; Di, W.L. Harlequin ichthyosis: ABCA12 mutations underlie defective lipid transport, reduced protease regulation and skin-barrier dysfunction. *Cell Tissue Res.* **2013**, *351*, 281–288.
33. Akiyama, M.; Suzumori, K.; Shimizu, H. Prenatal diagnosis of harlequin ichthyosis by the examination of keratinized hair canals and amniotic fluid cells at 19 weeks' estimated gestational age. *Prenat. Diagn.* **1999**, *19*, 167–171.
34. Vohra, N.; Rochelson, B.; Smith-Levitin, M. Three-dimensional sonographic findings in congenital (harlequin) ichthyosis. *J. Ultrasound Med.* **2003**, *22*, 737–739.
35. Akiyama, M. Harlequin ichthyosis and other autosomal recessive congenital ichthyoses: The underlying genetic defects and pathomechanisms. *J. Dermatol. Sci.* **2006**, *42*, 83–89.
36. Blaydon, D.C.; Nitoiu, D.; Eckl, K.M.; Cabral, R.M.; Bland, P.; Hausser, I.; van Heel, D.A.; Rajpopat, S.; Fischer, J.; Oji, V.; *et al.* Mutations in CSTA, encoding cystatin a, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. *Am. J. Hum. Genet.* **2011**, *89*, 564–571.
37. Hatsell, S.J.; Stevens, H.; Jackson, A.P.; Kelsell, D.P.; Zvulunov, A. An autosomal recessive exfoliative ichthyosis with linkage to chromosome 12q13. *Br. J. Dermatol.* **2003**, *149*, 174–180.
38. Krunic, A.L.; Stone, K.L.; Simpson, M.A.; McGrath, J.A. Acral peeling skin syndrome resulting from a homozygous nonsense mutation in the CSTA gene encoding cystatin A. *Pediatr. Dermatol.* **2013**, *30*, e87–e88.
39. Turk, V.; Bode, W. The cystatins: Protein inhibitors of cysteine proteinases. *FEBS Lett.* **1991**, *285*, 213–219.
40. Steven, A.C.; Steinert, P.M. Protein composition of cornified cell envelopes of epidermal keratinocytes. *J. Cell Sci.* **1994**, *107*, 693–700.
41. Palungwachira, P.; Kakuta, M.; Yamazaki, M.; Yaguchi, H.; Tsuboi, R.; Takamori, K.; Ogawa, H. Immunohistochemical localization of cathepsin l and cystatin A in normal skin and skin tumors. *J. Dermatol.* **2002**, *29*, 573–579.
42. Kato, T.; Takai, T.; Mitsuishi, K.; Okumura, K.; Ogawa, H. Cystatin a inhibits IL-8 production by keratinocytes stimulated with Der p 1 and Der f 1: Biochemical skin barrier against mite cysteine proteases. *J. Allergy Clin. Immunol.* **2005**, *116*, 169–176.
43. Li, C.; Chen, L.; Wang, J.; Zhang, L.; Tang, P.; Zhai, S.; Guo, W.; Yu, N.; Zhao, L.; Liu, M.; *et al.* Expression and clinical significance of cathepsin B and stefin a in laryngeal cancer. *Oncol. Rep.* **2011**, *26*, 869–875.
44. Strojan, P.; Budihna, M.; Smid, L.; Svetic, B.; Vrhovec, I.; Skrk, J. Cathepsin B and L and stefin A and B levels as serum tumor markers in squamous cell carcinoma of the head and neck. *Neoplasma* **2001**, *48*, 66–71.
45. Parker, B.S.; Ciocca, D.R.; Bidwell, B.N.; Gago, F.E.; Fanelli, M.A.; George, J.; Slavin, J.L.; Moller, A.; Steel, R.; Pouliot, N.; *et al.* Primary tumour expression of the cysteine cathepsin inhibitor stefin A inhibits distant metastasis in breast cancer. *J. Pathol.* **2008**, *214*, 337–346.

46. Bilodeau, M.; MacRae, T.; Gaboury, L.; Laverdure, J.P.; Hardy, M.P.; Mayotte, N.; Paradis, V.; Harton, S.; Perreault, C.; Sauvageau, G. Analysis of blood stem cell activity and cystatin gene expression in a mouse model presenting a chromosomal deletion encompassing Csta and Stfa2l1. *PLoS One* **2009**, *4*, e7500.
47. Lin, Z.; Chen, Q.; Lee, M.; Cao, X.; Zhang, J.; Ma, D.; Chen, L.; Hu, X.; Wang, H.; Wang, X.; *et al.* Exome sequencing reveals mutations in TRPV3 as a cause of olmsted syndrome. *Am. J. Hum. Genet.* **2012**, *90*, 558–564.
48. Haghighi, A.; Scott, C.A.; Poon, D.S.; Yaghoobi, R.; Saleh-Gohari, N.; Plagnol, V.; Kelsell, D.P. A missense mutation in the MBTPS2 gene underlies the X-linked form of olmsted syndrome. *J. Investig. Dermatol.* **2013**, *133*, 571–573.
49. Cambiaghi, S.; Tadini, G.; Barbareschi, M.; Caputo, R. Olmsted syndrome in twins. *Arch. Dermatol.* **1995**, *131*, 738–739.
50. Larregue, M.; Callot, V.; Kanitakis, J.; Suau, A.M.; Foret, M. Olmsted syndrome: Report of two new cases and literature review. *J. Dermatol.* **2000**, *27*, 557–568.
51. Giehl, K.A.; Eckstein, G.N.; Pasternack, S.M.; Praetzel-Wunder, S.; Ruzicka, T.; Lichtner, P.; Seidl, K.; Rogers, M.*; Graf, E.; Langbein,* L.**;** *et al.* Nonsense mutations in AAGAB cause punctate palmoplantar keratoderma type buschke-fischer-brauer. *Am. J. Hum. Genet.* **2012**, *91*, 754–759.
52. Pohler, E.; Mamai, O.; Hirst, J.; Zamiri, M.; Horn, H.; Nomura, T.; Irvine, A.D.; Moran, B.; Wilson, N.J.; Smith, F.J.; *et al.* Haploinsufficiency for AAGAB causes clinically heterogeneous forms of punctate palmoplantar keratoderma. *Nat. Genet.* **2012**, *44*, 1272–1276.
53. Kono, M.; Sugiura, K.; Suganuma, M.; Hayashi, M.; Takama, H.; Suzuki, T.; Matsunaga, K.; Tomita, Y.; Akiyama, M. Whole-exome sequencing identifies ADAM10 mutations as a cause of reticulate acropigmentation of kitamura, a clinical entity distinct from dowling-degos disease. *Hum. Mol. Genet.* **2013**, *22*, 3524–3533.
54. Blaydon, D.C.; Lind, L.K.; Plagnol, V.; Linton, K.J.; Smith, F.J.; Wilson, N.J.; McLean, W.H.; Munro, C.S.; South, A.P.; Leigh, I.M.; *et al.* Mutations in *AQP5*, encoding a water-channel protein, cause autosomal-dominant diffuse nonepidermolytic palmoplantar keratoderma. *Am. J. Hum. Genet.* **2013**, *93*, 330–335.
55. Eytan, O.; Morice-Picard, F.; Sarig, O.; Ezzedine, K.; Isakov, O.; Li, Q.; Ishida-Yamamoto, A.; Shomron, N.; Goldsmith, T.; Fuchs-Telem, D.; *et al.* Cole disease results from mutations in ENPP1. *Am. J. Hum. Genet.* **2013**, *93*, 752–757.
56. McGrath, J.A.; Stone, K.L.; Begum, R.; Simpson, M.A.; Dopping-Hepenstal, P.J.; Liu, L.; McMillan, J.R.; South, A.P.; Pourreyron, C.; McLean, W.H.; *et al.* Germline mutation in EXPH5 implicates the Rab27B effector protein Slac2-b in inherited skin fragility. *Am. J. Hum. Genet.* **2012**, *91*, 1115–1121.
57. Lin, Z.; Chen, Q.; Shi, L.; Lee, M.; Giehl, K.A.; Tang, Z.; Wang, H.; Zhang, J.; Yin, J.; Wu, L.; *et al.* Loss-of-function mutations in HOXC13 cause pure hair and nail ectodermal dysplasia. *Am. J. Hum. Genet.* **2012**, *91*, 906–911.

58. Ramot, Y.; Molho-Pessach, V.; Meir, T.; Alper-Pinus, R.; Siam, I.; Tams, S.; Babay, S.; Zlotogorski, A. Mutation in KANK2, encoding a sequestering protein for steroid receptor coactivators, causes keratoderma and woolly hair. *J. Med. Genet.* **2014**, *51*, 388–394.

59. Li, M.; Cheng, R.; Liang, J.; Yan, H.; Zhang, H.; Yang, L.; Li, C.; Jiao, Q.; Lu, Z.; He, J.; *et al.* Mutations in POFUT1, encoding protein *O*-fucosyltransferase 1, cause generalized dowling-degos disease. *Am. J. Hum. Genet.* **2013**, *92*, 895–903.

60. Basmanav, F.B.; Oprisoreanu, A.M.; Pasternack, S.M.; Thiele, H.; Fritz, G.; Wenzel, J.; Grosser, L.; Wehner, M.; Wolf, S.; Fagerberg, C.; *et al.* Mutations in POGLUT1, encoding protein O-glucosyltransferase 1, cause autosomal-dominant dowling-degos disease. *Am. J. Hum. Genet.* **2014**, *94*, 135–143.

61. Kubo, A.; Shiohama, A.; Sasaki, T.; Nakabayashi, K.; Kawasaki, H.; Atsugi, T.; Sato, S.; Shimizu, A.; Mikami, S.; Tanizaki, H.; *et al. Mutatio*ns in SERPINB7, encoding a member of the serine protease inhibitor superfamily, cause nagashima-type palmoplantar *keratosis. Am. J. Hum. Genet.* **2013**, *93*, 945–956.

62. Eytan, O.; Fuchs-Telem, D.; Mevorach, B.; Indelman, M.; Bergman, R.; Sarig, O.; Goldberg, I.; Adir, N.; Sprecher, E. Olmsted syndrome caused by a homozygous recessive mutation in TRPV3. *J. Investig. Dermatol.* **2014**, *134*, 1752–1754.

63. Smith, G.D.; Gunthorpe, M.J.; Kelsell, R.E.; Hayes, P.D.; Reilly, P.; Facer, P.; Wright, J.E.; Jerman, J.C.; Walhin, J.P.; Ooi, L.; *et al.* TRPV3 is a temperature-sensitive vanilloid receptor-like protein. *Nature* **2002**, *418*, 186–190.

64. Peier, A.M.; Reeve, A.J.; Andersson, D.A.; Moqrich, A.; Earley, T.J.; Hergarden, A.C.; Story, G.M.; Colley, S.; Hogenesch, J.B.; McIntyre, P.; *et al.* A heat-sensitive TRP channel expressed in keratinocytes. *Science* **2002**, *296*, 2046–2049.

65. Xu, H.; Ramsey, I.S.; Kotecha, S.A.; Moran, M.M.; Chong, J.A.; Lawson, D.; Ge, P.; Lilly, J.; Silos-Santiago, I.; Xie, Y.; *et al.* TRPV3 is a calcium-permeable temperature-sensitive cation channel. *Nature* **2002**, *418*, 181–186.

66. Asakawa, M.; Yoshioka, T.; Matsutani, T.; Hikita, I.; Suzuki, M.; Oshima, I.; Tsukahara, K.; Arimura, A.; Horikawa, T.; Hirasawa, T.; *et al.* Association of a mutation in TRPV3 with defective hair growth in rodents. *J. Investig. Dermatol.* **2006**, *126*, 2664–2672.

67. Cheng, X.; Jin, J.; Hu, L.; Shen, D.; Dong, X.P.; Samie, M.A.; Knoff, J.; Eisinger, B.; Liu, M.L.; Huang, S.M.; *et al.* Trp channel regulates EGFR signaling in hair morphogenesis and skin barrier formation. *Cell* **2010**, *141*, 331–343.

68. Xiao, R.; Tian, J.; Tang, J.; Zhu, M.X. The TRPV3 mutation associated with the hairless phenotype in rodents is constitutively active. *Cell Calcium* **2008**, *43*, 334–343.

69. Lai-Cheong, J.E.; Sethuraman, G.; Ramam, M.; Stone, K.; Simpson, M.A.; McGrath, J.A. Recurrent heterozygous missense mutation, p.Gly573ser, in the TRPV3 gene in an Indian boy with sporadic olmsted syndrome. *Br. J. Dermatol.* **2012**, *167*, 440–442.

70. Duchatelet, S.; Guibbal, L.; de Veer, S.; Fraitag, S.; Nitschke, P.; Zarhrate, M.; Bodemer, C.; Hovnanian, A. Olmsted syndrome with erythromelalgia caused by recessive TRPV3 mutations. *Br. J. Dermatol.* **2014**, doi:10.1111/bjd.12951.

71. Duchatelet, S.; Pruvost, S.; de Veer, S.; Fraitag, S.; Nitschke, P.; Bole-Feysot, C.; Bodemer, C.; Hovnanian, A. A new TRPV3 missense mutation in a patient with olmsted syndrome and erythromelalgia. *JAMA Dermatol.* **2014**, *150*, 303–306.
72. Yaghoobi, R.; Omidian, M.; Sina, N.; Abtahian, S.A.; Panahi-Bazaz, M.R. Olmsted syndrome in an Iranian family: Report of two new cases. *Arch. Iran. Med.* **2007**, *10*, 246–249.
73. Oeffner, F.; Fischer, G.; Happle, R.; Konig, A.; Betz, R.C.; Bornholdt, D.; Neidel, U.; Boente Mdel, C.; Redler, S.; Romero-Gomez, J.; *et al.* IFAP syndrome is caused by deficiency in MBTPS2, an intramembrane zinc metalloprotease essential for cholesterol homeostasis and ER stress response. *Am. J. Hum. Genet.* **2009**, *84*, 459–467.
74. Ding, Y.G.; Wang, J.Y.; Qiao, J.J.; Mao, X.H.; Cai, S.Q. A novel mutation in MBTPS2 causes ichthyosis follicularis, alopecia and phot*ophobia* (IFAP) syndrome in a chinese family. *Br. J. Dermatol.* **2010**, *163*, 886–889.
75. Tang, L.; Liang, J.; Wang, W.; Yu, L.; Yao, Z. A novel mutation in MBTPS2 *causes a broad phe*n**otyp**ic *s*pectrum of ichthyosis follicularis, atrichia, and photophobia syndrome in a large chinese family. *J. Am. Acad. Dermatol.* **2011**, *64*, 716–722.
76. Naiki, M.; Mizuno, S.; Yamada, K.; Yamada, Y.; Kimura, R.; Oshiro, M.; Okamoto, N.; Makita, Y.; Seishima, M.; Wakamatsu, N. MBTPS2 mutation causes bresek/bresheck syndrome. *Am. J. Med. Genet. Part A* **2012**, *158A*, 97–102.
77. Aten, E.; Brasz, L.C.; Bornholdt, D.; Hooijkaas, I.B.; Porteous, M.E.; Sybert, V.P.; Vermeer, M.H.; Vossen, R.H.; van der Wielen, M.J.; Bakker, E.; *et al.* Keratosis follicularis spinulosa decalvans is caused by mutations in MBTPS2. *Hum. Mutat.* **2010**, *31*, 1125–1133.
78. Sakai, J.; Nohturfft, A.; Goldstein, J.L.; Brown, M.S. Cleavage of sterol regulatory element-binding proteins (srebps) at site-1 requires interaction with SREBP cleavage-activating protein. Evidence from *in vivo* competition studies. *J. Biol. Chem.* **1998**, *273*, 5785–5793.
79. Rawson, R.B.; Zelenski, N.G.; Nijhawan, D.; Ye, J.; Sakai, J.; Hasan, M.T.; Chang, T.Y.; Brown, M.S.; Goldstein, J.L. Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Mol. Cell* **1997**, *1*, 47–57.
80. Ye, J.; Rawson, R.B.; Komuro, R.; Chen, X.; Dave, U.P.; Prywes, R.; Brown, M.S.; Goldstein, J.L. ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol. Cell* **2000**, *6*, 1355–1364.
81. Bisgaard, H.; Simpson, A.; Palmer, C.N.; Bonnelykke, K.; McLean, I.; Mukhopadhyay, S.; Pipper, C.B.; Halkjaer, L.B.; Lipworth, B.; Hankinson, J.; *et al.* Gene-environment interaction in the onset of eczema in infancy: Filaggrin loss-of-function mutations enhanced by neonatal cat exposure. *PLoS Med.* **2008**, *5*, e131.
82. Enamandram, M.; Kimball, A.B. Psoriasis epidemiology: The interplay of genes and the environment. *J. Investig. Dermatol.* **2013**, *133*, 287–289.
83. Carroll, C.L.; Balkrishnan, R.; Feldman, S.R.; Fleischer, A.B., Jr.; Manuel, J.C. The burden of atopic dermatitis: Impact on the patient, family, and society. *Pediatr. Dermatol.* **2005**, *22*, 192–199.

84. Baker, C.S.; Foley, P.A.; Braue, A. Psoriasis uncovered—Measuring burden of disease impact in a survey of australians with psoriasis. *Australas. J. Dermatol.* **2013**, *54*, 1–6.

85. Stranger, B.E.; Stahl, E.A.; Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **2011**, *187*, 367–383.

86. Weinstein, G.D.; Frost, P. Abnormal cell proliferation in psoriasis. *J. Investig. Dermatol.* **1968**, *50*, 254–259.

87. Wuepper, K.D.; Coulter, S.N.; Haberman, A. Psoriasis vulgaris: A genetic approach. *J. Investig. Dermatol.* **1990**, *95*, 2S–4S.

88. Capon, F.; Barker, J.N. The quest for psoriasis susceptibility genes in the postgenome-wide association studies era: Charting the road ahead. *Br. J. Dermatol.* **2012**, *166*, 1173–1175.

89. Capon, F.; Burden, A.D.; Trembath, R.C.; Barker, J.N. Psoriasis and other complex trait dermatoses: From loci to functional pathways. *J. Investig. Dermatol.* **2012**, *132*, 915–922.

90. Nair, R.P.; Stuart, P.; Henseler, T.; Jenisch, S.; Chia, N.V.; Westphal, E.; Schork, N.J.; Kim, J.; Lim, H.W.; Christophers, E.; *et al.* Localization of psoriasis-susceptibility locus PSORS1 to a 60-kb interval telomeric to HLA-C. *Am. J. Hum. Genet.* **2000**, *66*, 1833–1844.

91. Capon, F.; Munro, M.; Barker, J.; Trembath, R. Searching for the major histocompatibility complex psoriasis susceptibility gene. *J. Investig. Dermatol.* **2002**, *118*, 745–751.

92. Nair, R.P.; Stuart, P.E.; Nistor, I.; Hiremagalore, R.; Chia, N.V.; Jenisch, S.; Weichenthal, M.; Abecasis, G.R.; Lim, H.W.; Christophers, E.; *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. J. Hum. Genet.* **2006**, *78*, 827–851.

93. Nair, M.; Teng, A.; Bilanchone, V.; Agrawal, A.; Li, B.; Dai, X. Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. *J. Cell Biol.* **2006**, *173*, 253–264.

94. Nair, R.P.; Duffin, K.C.; Helms, C.; Ding, J.; Stuart, P.E.; Goldgar, D.; Gudjonsson, J.E.; Li, Y.; Tejasvi, T.; Feng, B.J.; *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappab pathways. *Nat. Genet.* **2009**, *41*, 199–204.

95. Jordan, C.T.; Cao, L.; Roberson, E.D.; Duan, S.; Helms, C.A.; Nair, R.P.; Duffin, K.C.; Stuart, P.E.; Goldgar, D.; Hayashi, G.; *et al.* Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappab, in psoriasis. *Am. J. Hum. Genet.* **2012**, *90*, 796–808.

96. Jordan, C.T.; Cao, L.; Roberson, E.D.; Pierson, K.C.; Yang, C.F.; Joyce, C.E.; Ryan, C.; Duan, S.; Helms, C.A.; Liu, Y.; *et al.* PSORS2 is due to mutations in CARD14. *Am. J. Hum. Genet.* **2012**, *90*, 784–795.

97. Di Cesare, A.; di Meglio, P.; Nestle, F.O. The IL-23/Th17 axis in the immunopathogenesis of psoriasis. *J. Investig. Dermatol.* **2009**, *129*, 1339–1350.

98. Weaver, C.T.; Hatton, R.D.; Mangan, P.R.; Harrington, L.E. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Ann. Rev. Immunol.* **2007**, *25*, 821–852.

99. Zheng, Y.; Danilenko, D.M.; Valdez, P.; Kasman, I.; Eastham-Anderson, J.; Wu, J.; Ouyang, W. Interleukin-22, a T(h)17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **2007**, *445*, 648–651.

100. Krueger, G.G.; Langley, R.G.; Leonardi, C.; Yeilding, N.; Guzzo, C.; Wang, Y.; Dooley, L.T.; Lebwohl, M.; Group, C.P.S. A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. *N. Engl. J. Med.* **2007**, *356*, 580–592.
101. Suarez-Farinas, M.; Shah, K.R.; Haider, A.S.; Krueger, J.G.; Lowes, M.A. Personalized medicine in psoriasis: Developing a genomic classifier to predict histological response to alefacept. *BMC Dermatol.* **2010**, *10*, doi:10.1186/1471-5945-10-1.
102. Tomfohrde, J.; Silverman, A.; Barnes, R.; Fernandez-Vina, M.A.; Young, M.; Lory, D.; Morris, L.; Wuepper, K.D.; Stastny, P.; Menter, A.; *et al.* Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. *Science* **1994**, *264*, 1141–1145.
103. Marrakchi, S.; Guigue, P.; Renshaw, B.R.; Puel, A.; Pei, X.Y.; Fraitag, S.; Zribi, J.; Bal, E.; Cluzeau, C.; Chrabieh, M.; *et al.* Interleukin-36-receptor antagonist deficiency and generalized pustular psoriasis. *N. Engl. J. Med.* **2011**, *365*, 620–628.
104. Berki, D.M.; Mahil, S.K.; Burden, A.D.; Trembath, R.C.; Smith, C.H.; Capon, F.; Barker, J.N. Loss of IL36RN function does not confer susceptibility to psoriasis vulgaris. *J. Investig. Dermatol.* **2014**, *134*, 271–273.
105. Capon, F. IL36RN mutations in generalized pustular psoriasis: Just the tip of the iceberg? *J. Investig. Dermatol.* **2013**, *133*, 2503–2504.
106. Flohr, C.; Mann, J. New insights into the epidemiology of childhood atopic dermatitis. *Allergy* **2014**, *69*, 3–16.
107. Larsen, F.S.; Holm, N.V.; Henningsen, K. Atopic dermatitis: A genetic-epidemiologic study in a population-based twin sample. *J. Am. Acad. Dermatol.* **1986**, *15*, 487–494.
108. Smith, F.J.; Irvine, A.D.; Terron-Kwiatkowski, A.; Sandilands, A.; Campbell, L.E.; Zhao, Y.; Liao, H.; Evans, A.T.; Goudie, D.R.; Lewis-Jones, S.; *et al.* Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat. Genet.* **2006**, *38*, 337–342.
109. Sybert, V.P.; Dale, B.A.; Holbrook, K.A. Ichthyosis vulgaris: Identification of a defect in synthesis of filaggrin correlated with an absence of keratohyaline granules. *J. Investig. Dermatol.* **1985**, *84*, 191–194.
110. Brown, S.J.; Asai, Y.; Cordell, H.J.; Campbell, L.E.; Zhao, Y.; Liao, H.; Northstone, K.; Henderson, J.; Alizadehfar, R.; Ben-Shoshan, M.; *et al.* Loss-of-function variants in the filaggrin gene are a significant risk factor for peanut allergy. *J. Allergy Clin. Immunol.* **2011**, *127*, 661–667.
111. Brown, S.J.; Relton, C.L.; Liao, H.; Zhao, Y.; Sandilands, A.; McLean, W.H.; Cordell, H.J.; Reynolds, N.J. Filaggrin haploinsufficiency is highly penetrant and is associated with increased severity of eczema: Further delineation of the skin phenotype in a prospective epidemiological study of 792 school children. *Br. J. Dermatol.* **2009**, *161*, 884–889.
112. Hirota, T.; Takahashi, A.; Kubo, M.; Tsunoda, T.; Tomita, K.; Sakashita, M.; Yamada, T.; Fujieda, S.; Tanaka, S.; Doi, S.; *et al.* Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **2012**, *44*, 1222–1226.

113. Sun, L.D.; Xiao, F.L.; Li, Y.; Zhou, W.M.; Tang, H.Y.; Tang, X.F.; Zhang, H.; Schaarschmidt, H.; Zuo, X.B.; Foelster-Holst, R.; *et al.* Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese han population. *Nat. Genet.* **2011**, *43*, 690–694.

114. Paternoster, L.; Standl, M.; Chen, C.M.; Ramasamy, A.; Bonnelykke, K.; Duijts, L.; Ferreira, M.A.; Alves, A.C.; Thyssen, J.P.; Albrecht, E.; *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.* **2012**, *44*, 187–192.

115. Palmer, C.N.; Irvine, A.D.; Terron-Kwiatkowski, A.; Zhao, Y.; Liao, H.; Lee, S.P.; Goudie, D.R.; Sandilands, A.; Campbell, L.E.; Smith, F.J.; *et al.* Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat. Genet.* **2006**, *38*, 441–446.

116. Sandilands, A.; Sutherland, C.; Irvine, A.D.; McLean, W.H. Filaggrin in the frontline: Role in skin barrier function and disease. *J. Cell Sci.* **2009**, *122*, 1285–1294.

117. Fallon, P.G.; Sasaki, T.; Sandilands, A.; Campbell, L.E.; Saunders, S.P.; Mangan, N.E.; Callanan, J.J.; Kawasaki, H.; Shiohama, A.; Kubo, A.; *et al.* A homozygous frameshift mutation in the mouse flg gene facilitates enhanced percutaneous allergen priming. *Nat. Genet.* **2009**, *41*, 602–608.

118. Oyoshi, M.K.; Murphy, G.F.; Geha, R.S. Filaggrin-deficient mice exhibit Th17-dominated skin inflammation and permissiveness to epicutaneous sensitization with protein antigen. *J. Allergy Clin. Immunol.* **2009**, *124*, 485–493

119. Otsuka, A.; Doi, H.; Egawa, G.; Maekawa, A.; Fujita, T.; Nakamizo, S.; Nakashima, C.; Nakajima, S.; Watanabe, T.; Miyachi, Y.; *et al.* Possible new therapeutic strategy to regulate atopic dermatitis through upregulating filaggrin expression. *J. Allergy Clin. Immunol.* **2014**, *133*, 139–146.

120. Tamari, M.; Hirota, T. Genome-wide association studies of atopic dermatitis. *J. Dermatol.* **2014**, *41*, 213–220.

121. Esparza-Gordillo, J.; Weidinger, S.; Folster-Holst, R.; Bauerfeind, A.; Ruschendorf, F.; Patone, G.; Rohde, K.; Marenholz, I.; Schulz, F.; Kerscher, T.; *et al.* A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat. Genet.* **2009**, *41*, 596–601.

122. Wang, R.; Wan, Q.; Kozhaya, L.; Fujii, H.; Unutmaz, D. Identification of a regulatory T cell specific cell surface molecule that mediates suppressive signals and induces Foxp3 expression. *PLoS One* **2008**, *3*, e2705.

123. Buschke, S.; Stark, H.J.; Cerezo, A.; Pratzel-Wunder, S.; Boehnke, K.; Kollar, J.; Langbein, L.; Heldin, C.H.; Boukamp, P. A decisive function of transforming growth factor-beta/smad signaling in tissue morphogenesis and differentiation of human HaCat keratinocytes. *Mol. Biol. Cell* **2011**, *22*, 782–794.

124. Chang, M.; Li, Y.; Yan, C.; Callis-Duffin, K.P.; Matsunami, N.; Garcia, V.E.; Cargill, M.; Civello, D.; Bui, N.; Catanese, J.J.; *et al.* Variants in the 5q31 cytokine gene cluster are associated with psoriasis. *Genes Immun.* **2008**, *9*, 176–181.

125. Li, Y.; Chang, M.; Schrodi, S.J.; Callis-Duffin, K.P.; Matsunami, N.; Civello, D.; Bui, N.; Catanese, J.J.; Leppert, M.F.; Krueger, G.G.; *et al.* The 5q31 variants associated with psoriasis and crohn's disease are distinct. *Hum. Mol. Genet.* **2008**, *17*, 2978–2985.

126. Leung, D.Y. New insights into atopic dermatitis: Role of skin barrier and immune dysregulation. *Allergol. Int.* **2013**, *62*, 151–161.

127. Hamid, Q.; Naseer, T.; Minshall, E.M.; Song, Y.L.; Boguniewicz, M.; Leung, D.Y. *In vivo* expression of IL-12 and IL-13 in atopic dermatitis. *J. Allergy Clin. Immunol.* **1996**, *98*, 225–231.

128. Hamid, Q.; Boguniewicz, M.; Leung, D.Y. Differential *in situ* cytokine gene expression in acute *versus* chronic atopic dermatitis. *J. Clin. Investig.* **1994**, *94*, 870–876.

129. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **2008**, *456*, 18–21.

130. Gudjonsson, J.E.; Krueger, G. A role for epigenetics in psoriasis: Methylated cytosine-guanine sites differentiate lesional from nonlesional skin and from normal skin. *J. Investig. Dermatol.* **2012**, *132*, 506–508.

131. Blaydon, D.C.; Biancheri, P.; Di, W.L.; Plagnol, V.; Cabral, R.M.; Brooke, M.A.; van Heel, D.A.; Ruschendorf, F.; Toynbee, M.; Walne, A.; *et al.* Inflammatory skin and bowel disease linked to ADAM17 deletion. *N. Engl. J. Med.* **2011**, *365*, 1502–1508.

132. Blaydon, D.C.; Etheridge, S.L.; Risk, J.M.; Hennies, H.C.; Gay, L.J.; Carroll, R.; Plagnol, V.; McRonald, F.E.; Stevens, H.P.; Spurr, N.K.; *et al.* RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome. *Am. J. Hum. Genet.* **2012**, *90*, 340–346.

133. Brooke, M.A.; Etheridge, S.L.; Kaplan, N.; Simpson, C.; O'Toole, E.A.; Ishida-Yamamoto, A.; Marches, O.; Getsios, S.; Kelsell, D.P. iRHOM*2*-dependent regulation of ADAM17 in cutaneous disease and epidermal barrier function. *Hum. Mol. Genet.* **2014**, *23*, 4064–4076.

134. Brooke, M.A.; Nitoiu, D.; Kelsell, D.P. Cell-cell connectivity: Desmosomes and disease. *J. Pathol.* **2012**, *226*, 158–171.

Epigenetic Control of the Genome—Lessons from Genomic Imprinting

Bjorn T. Adalsteinsson and Anne C. Ferguson-Smith

Abstract: Epigenetic mechanisms modulate genome function by writing, reading and erasing chromatin structural features. These have an impact on gene expression, contributing to the establishment, maintenance and dynamic changes in cellular properties in normal and abnormal situations. Great effort has recently been undertaken to catalogue the genome-wide patterns of epigenetic marks—creating reference epigenomes—which will deepen our understanding of their contributions to genome regulation and function with the promise of revealing further insights into disease etiology. The foundation for these global studies is the smaller scale experimentally-derived observations and questions that have arisen through the study of epigenetic mechanisms in model systems. One such system is genomic imprinting, a process causing the mono-allelic expression of genes in a parental-origin specific manner controlled by a hierarchy of epigenetic events that have taught us much about the dynamic interplay between key regulators of epigenetic control. Here, we summarize some of the most noteworthy lessons that studies on imprinting have revealed about epigenetic control on a wider scale. Specifically, we will consider what these studies have revealed about: the variety of relationships between DNA methylation and transcriptional control; the regulation of important protein-DNA interactions by DNA methylation; the interplay between DNA methylation and histone modifications; and the regulation and functions of long non-coding RNAs.

Reprinted from *Genes*. Cite as: Adalsteinsson, B.T.; Ferguson-Smith, A.C. Epigenetic Control of the Genome—Lessons from Genomic Imprinting. *Genes* **2014**, *5*, 635-655.

1. A Primer on Epigenetics, DNA Methylation and Histone Modifications

Epigenetic modifications perform three main functions in mammalian cells: they contribute to the control of chromosome architecture ensuring stability and appropriate segregation of chromosomes during mitosis; they contribute to regulation of the silencing and inaccessibility of repetitive elements and endogenous retroelements; and they can initiate and maintain the activity and repression of individual genes or clusters of genes. Here we focus on the role of epigenetic modifications in the control of mammalian transcription and the contribution of genomic imprinting studies to our understanding of epigenetic mechanisms.

In mammals, the different cells that make up an organism generally contain the same DNA yet their cellular morphology and function can vary greatly. This is largely a result of differential gene expression, which is developmentally regulated and can then be maintained after repeated cell divisions. The maintenance of expression states/levels requires heritable information to be passed through cell division to ensure propagation in each daughter cell, and it is this information that has been termed epigenetic. Further, cells are subject to dynamic changes in gene expression, dependent, for example, on intrinsic and extrinsic cues, which can be mediated through epigenetic processes. Epigenetic mechanisms include DNA methylation and post translational modifications to core histones. Other related components have been proposed as epigenetic such as non-coding RNAs (ncRNAs) and nucleosomal positioning, however these might also be considered mediators and/or facilitators of epigenetic states. The characterization and mapping of genome-wide epigenetic modifications represent an ever increasing field of research. These studies are revealing genome-wide patterns of epigenetic regulation that not only have confirmed many of the conclusions suggested from more traditional experimental approaches in model systems but also allow for the generation of new hypotheses that await experimental testing. One model system that contributed a foundation for these studies is the process of genomic imprinting.

DNA methylation is a process whereby a methyl (CH_3) group is added most commonly to a cytosine in DNA. In mammals it is generally found at CpG dinucleotides and can be correlated with gene repression in a variety of ways (discussed in more detail below). CpG sites are generally depleted in the genome, apart from stretches of DNA called CpG islands where CpG density is high. CpG islands can be concentrated at gene promoters and are generally unmethylated. CpG sites outside CpG islands are generally methylated (reviewed in [1,2])—resulting in a genome-wide methylation pattern that can be described as roughly bimodal. Acquisition of DNA methylation is catalyzed by a family of DNA methyltransferases (DNMTs, reviewed in [3]). DNMT1 has affinity for hemi-methylated DNA and is responsible for maintaining methylation after DNA replication and DNMT3A and DNMT3B catalyze *de novo* DNA methylation while the DNA methyltransferase homologue, DNMT3L acts as a cofactor and has no methyltransferase activity.

Waves of DNA methylation loss and acquisition are orchestrated during embryonic development. After fertilization the two parental genomes are mostly stripped of their epigenetic marks, a process that presumably "resets" the genome to a naive state applicable for pluripotency (DNA methylation at certain sequences in imprinted loci are among few genomic regions to "escape" this demethylation, see details in Section 2). Around blastocyst implantation *de novo* methylation then occurs and, to our knowledge, no further genome wide erasure/acquisition waves occur in somatic cells. Another wave of genome-wide reprogramming occurs in primordial germ cells (this time DNA methylation at imprinted loci is also lost, see details in Section 2); erasure of DNA methylation commences in the embryonic germline after embryonic day 7.5 (E7.5) in the mouse and progressive *de novo* methylation follows at E12.5 in prospermatogonia of male embryos, but occurs after birth in oocytes of female embryos (reviewed in [4]). This germline epigenetic reprogramming is required for generating functional germ cells and failure to do this appropriately usually results in infertility or developmentally abnormal embryos that die during gestation [5,6].

Covalent post-translational modifications to core histones (histone modifications henceforth) can impact the conformation of the nucleosome-nucleosome architecture within chromatin and influence its function such that some modifications are associated with an active chromatin state and others with a repressive state (for extensive review refer to [7]). The full repertoire of histone modifications is unknown, but is complex, with some specific amino acid residues influencing the ability of others to be modified, and some sites having the potential to be modified in multiple different ways. It is currently unclear whether many of the modifications truly are epigenetically heritable in a replication-dependent manner, like DNA methylation. Lysine methylation and lysine acetylation are among the best characterized histone modifications whose correlations with gene activity and repression have been extensively studied. Furthermore, enzymes involved in "writing" and "erasing" these epigenetic marks have been identified and characterized; histone lysine methyltransferases deposit methyl groups to lysine, and histone lysine demethylases remove them. Histone acetyltransferases (KATs) and histone deacetylases (HDACs) deposit and remove acetyl groups, respectively. Generally, regions with acetylated histones are associated with gene activity and regions devoid of acetylated histones are repressed, while associations between histone methylation and gene transcription are more site specific; histone 3 lysine 4 (H3K4) and H3K36 methylation are for example found on expressed genes while H3K9 methylation is associated with repressed genes. Their distribution in the genome can be associated with certain genomic motifs, e.g., gene regions such as promoters or open reading frames (ORFs), or intergenic regions such as repeats. For example, H3K4me3 (me3 denotes tri-methylated) is found at the promoters of active genes, whilst H3K4me1 is associated with enhancers, H3K20me3 is found at repressed repeat regions, and H3K9me3 at promoters of repressed genes, retroelements, imprinted loci and at pericentromeric repeat regions.

2. Genomic Imprinting and Targeting DNA Methylation

Genomic imprinting is a process causing the mono-allelic expression of a specific subset of mammalian genes in a parental origin specific manner (reviewed in [8,9])—*i.e.*, genes that are expressed either from the paternally inherited chromosome or from the maternally inherited chromosome (paternal allele and maternal allele henceforth) are imprinted. The non-equivalence of parental genomes in mammals was discovered in 1984 [10,11], and individual imprinted genes were first discovered in 1991 (reviewed in [8]). Today, over 100 imprinted genes have been identified, most of which are organized in clusters and are regulated in a coordinated manner by a single imprinting control region (ICR) [9]. Most clusters contain at least one non coding gene and multiple protein coding genes, whose functions regulate embryonic development, placentation and a range of post-natal processes.

Epigenetic mechanisms allow the transcriptional machinery of the cell to distinguish the two parental chromosomes at imprinted loci and hence provide an important paradigm for understanding epigenetic control of gene activity and repression. Specifically, the discovery of differences in DNA methylation in the same place on the two parental chromosomes suggested the importance of epigenetic mechanisms in regulating imprinting [12,13] and the potential for epigenetic control in a wider context. The identification of imprinting control regions and their validation genetically as functional elements essential for the imprinting of multiple genes in *cis*, elucidated imprinting control. The loss of imprinting after targeted deletion of DNMT1 proved that DNA methylation was required for imprinting [14]. Importantly, in the absence of DNMT1, some imprinted genes were activated but others became repressed, an indication that methylation could impact activity as well as repression.

The acquisition of methylation at ICRs occurs in the germ line *de novo* by DNMT3A and DNMT3L with a small number of ICRs becoming methylated in sperm cells, and the majority acquiring methylation in oocytes—paternal and maternal ICRs, respectively. It is of interest that paternal ICRs are always located in intergenic regions while maternal ICRs are located at promoter sequences. Importantly, erasure of imprints occurs in the wave of demethylation that occurs in the primordial germ cells. However, in order to retain the memory of the parental origin that is subsequently established after that reprogramming, imprints must be retained during the post-fertilization epigenetic reprogramming phase [4]. Interestingly, other regions of the genome seem refractory to zygotic reprogramming [15] though these are not necessarily parent-specific or retained like imprints during development. The relationship, if any, of these regions to ICRs remains unclear. In addition to the ICR, other differentially methylated regions (DMRs) are located at some imprinted clusters, but a notable difference between ICRs and these DMRs is that differential methylation of the latter is not germline established, but rather is acquired post-fertilization. In all cases, these so-called secondary DMRs—to distinguish them from regions such as ICRs that acquire differential methylation in the germline—require the ICR for their establishment. The mechanisms through which ICRs control gene expression in their respective clusters are diverse and remain the subject of active research, including analysis of regulation by ncRNAs and of the relationships between DNA methylation and histone and non-histone proteins.

Both in imprinted and non-imprinted contexts, little is known about why certain DNA sequences become methylated and not others, or how this may change dynamically within a sequence such as a particular CpG island at a gene promoter. Most likely, it is a process that must be targeted in some manner. Targeting of the DNA methylation machinery has received much attention and efforts made to identify intrinsic sequence specificities of DNMTs and their cofactors. It has thus generally been assumed that the acquisition of methylation represents the "active" process in establishing differential methylation. However, recent studies on DMRs in the germlines and their propagation after fertilization suggest it might also be protection from DNA methylation and maintenance at methylated regions that determine differential methylation (Figure 1A, reviewed in [16]): Rather than appearing as discrete methylated sequences in otherwise unmethylated regions, maternal ICRs (which represent the vast majority of ICRs) are surrounded by methylation at both flanks. In contrast, these ICRs are unmethylated in sperm but are also flanked by methylation at surrounding sequences, suggesting that DNA methylation may be the "default" state and that it is protection from methylation at the ICRs, and perhaps other non-imprinted sequences as well, that establishes their differential methylation. Furthermore, in the germline, far more sequences are differentially methylated between oocytes and sperm than the ICRs; recent genome-wide studies suggest they are in the counts of thousands in oocytes and hundreds in sperm [15,17,18]. In contrast to ICRs these sequences generally lose methylation after fertilization, suggesting targeted maintenance of DNA methylation at specific sequences is essential for the germline-derived differential methylation of imprinted loci. Hence perhaps, loss of maintenance, in addition to active removal of DNA methylation at non-imprinted loci, contributes to the mechanism through which demethylation occurs in somatic cells. KRAB zinc finger proteins (ZFP) represent a family of over 350 tetrapod-specific genes whose functions remain poorly understood. They bind DNA and have previously been shown to recruit the repressive chromatin machinery in a site-specific manner. One of these KRAB-ZFPs, ZFP57, has been shown to be required to maintain the DNA methylation memory at imprints during post-fertilization reprogramming when the bulk of the genome is changing its epigenetic state [19]. ZFP57 binds methylated DNA and is thought to recruit methyltransferases to imprinting control regions hence preventing them from loss of their imprints.

3. DNA Methylation and Gene Repression—The Chicken or the Egg?

3.1. DNA Methylation Correlates with Repression

The correlation between DNA methylation and gene repression was noted in several experiments assaying viral and endogenous gene expression in mammalian, frog and sea urchin cells in the late 1970s and early 80s [20–30]. Experiments were conducted to determine whether the observed relationship was purely correlational, or whether DNA methylation functionally regulated gene expression. This was, however, challenging, but the strong evidence in many different contexts, showing that hypomethylated regions were associated with activity and hypermethylated regions refractory to transcription, suggested that absence of DNA methylation may be necessary though not sufficient for transcription. Vardimon *et al*. injected bacterial plasmids containing *in vitro* methylated or unmethylated DNA encoding a viral gene into frog oocyte nuclei [31]. They observed maintenance of the respective methylation states over a 24 h period, and expression of the gene in oocytes that were injected with unmethylated DNA but not in those that were injected with methylated DNA [31]. In a similar experiment, Stein *et al*. transfected *in vitro* methylated or unmethylated plasmids containing the *Aprt* (adenine phosphoribosiltransferase) gene into cultured *Aprt* null mouse cells. They observed maintenance of the respective *Aprt* methylation states after integration into the endogenous genome over several cell divisions for both unmethylated and methylated plasmids, and that integration of the unmethylated but not the methylated gene rescued the *Aprt* null phenotype, suggesting methylation of the gene was associated with inhibition of its transcription [32].

Correlations between gene expression and DNA methylation have been assessed at CpG sites across whole chromosomes or the whole genome. Consistent with the earlier studies, DNA methylation of promoter sequences, though rare at CpG island promoters, was observed to correlate with gene repression [33–35]. The functional role of DNA methylation in repressing gene expression is further suggested by results from studies in which the genes encoding the DNA methyltransferases are deleted conditionally in various cell lineages. Generally, the loss of DNMTs results in dysregulation of multiple genes, with a trend towards gene activation rather than silencing, again suggesting that DNA methylation represses gene expression (reviewed in [36]). Furthermore, treatment of cells *in vivo* with the DNA methyltransferase inhibitor 5-Azacytidine was shown to result in gene activation in several experiments in the 1980s, with concomitant loss of DNA methylation (reviewed in [37]). Together all these findings have led to the general assumption that loss and acquisition of DNA methylation at a gene promoter results in gene activation and silencing, respectively, but none actually proved that the acquisition of DNA methylation itself causes the gene silencing in all contexts.

3.2. DNA Methylation as a Consequence of Transcriptional Silencing

Studies of the temporal onset of mono-allelic expression of imprinted genes and the acquisition of differential methylation at secondary DMRs during mouse development indicate that DNA methylation can be acquired *after* gene repression (Figure 1B). The imprinted genes *Gtl2*, *Cdkn1C*, *H19* and *Igf2r* each contain a secondary DMR in their promoters, which become differentially methylated days *after* their mono-allelic expression is observed (summarized in [38]). Generally, mono-allelic expression of these genes is initiated around the morula or blastocyst stage (E3.5-4.5), while differential methylation of the respective secondary DMR occurs after E6.5 [13,38–42]. In the most extreme case, *Igf2r* is mono-allelically expressed from the maternal allele from E6.5 onward but the silent paternal allele only becomes methylated at or after E15.5 [13,42]. It is reasonable to assume that this temporal relationship, where methylation is acquired as a consequence of gene repression, also applies to non-imprinted genes (Figure 1B). In particular, is has recently been shown that DNA methylation levels are secondary to the binding of transcription factors; Stadler *et al.* [43] identified multiple clusters of CpG sites that have low to intermediate levels of methylation, 10%–50%, in mouse embryonic stem (ES) cells. These low methylated regions (LMRs) are likely distal regulatory regions, and are bound by various transcription factors. Scrambling binding sites for the insulator protein CTCF or knocking out the transcription factor REST led to increased methylation at the LMRs. Furthermore, reintroduction of REST into the *REST*$^{-/-}$ cells reverted the methylation status of the LMRs to the normal low levels [43]. These findings suggest DNA methylation may not have a direct role in silencing gene expression in all situations. In such cases DNA methylation might rather be acquired after gene silencing to maintain the repressed state or as a secondary readout of other mechanisms of genome control. Nonetheless, there are situations where acquisition of DNA methylation unquestionably does regulate gene expression, notably at the germline DMRs of imprinted genes [1,8,9,14–16].

4. How Does DNA Methylation Confer Effects on Gene Expression?

4.1. Proteins Attracted and Repelled

In situations where DNA methylation does indeed direct gene repression there are currently two model mechanisms that are generally acknowledged [1,44]: First, DNA methylation can attract proteins that bring about gene repression through recruitment of chromatin modifiers. A group of proteins, collectively referred to as methyl binding proteins (MBPs) have been characterized and shown to specifically bind to methylated, but not unmethylated, DNA [44–49]. MBPs are known to interact with histone modifiers such as HDACs, e.g., in forming complexes, such as the nucleosome remodeling deacetylase (NuRD) complex, which through their histone deacetylase activity and subsequent chromatin condensation bring about gene repression [50–55]. Secondly, certain proteins may interact with DNA in a methylation dependent manner. Here, DNA methylation may be refractory to the binding of proteins, such as transcription factors or other regulatory proteins [56–58], that are necessary for gene expression (Figure 1C). For this latter model, the best characterized example is the regulation of CTCF binding at the imprinted H19/Igf2 cluster *via* differential DNA methylation on the two parental alleles (reviewed in [1]).

4.2. Regulation of CTCF Binding at the H19/Igf2 Imprinted Cluster; the Insulator Mechanism

In the H19/Igf2 imprinted cluster, the protein coding gene *Igf2* is expressed from the paternally inherited allele [59]. This expression pattern is dependent on the regional ICR [60], on its differential methylation [12,14,61] and on the insulator protein CTCF binding to the ICR. On the unmethylated maternal allele, CTCF can bind, while its binding is inhibited on the methylated paternally inherited chromosome [62–65]—thus CTCF binding to DNA is methylation-sensitive (Figure 1C). *Igf2* and a downstream non-coding RNA gene, *H19,* share enhancers that are located at the 3' end of *H19* [66,67] and the parental specific expression of *Igf2* and *H19* are ultimately determined by interaction with these sequences; on the paternally inherited chromosome, *Igf2*-enchancer interaction is possible and the gene is expressed. On the maternally inherited chromosome this contact is blocked by CTCF binding to the ICR and this facilitates enhancer interaction with a now active *H19* instead, and also results in *Igf2* repression.

What is the mechanism of CTCF’s enhancer blocking activity? The current model (reviewed in [68]) suggests that in the H19/Igf2 cluster, chromatin loop formation on the maternal allele spatially inhibits enhancer interaction with *Igf2.* The process appears to depend on three elements; dimerization, CTCF binding to more than one region and physical contact between these neighboring sites via CTCF interaction [69–74]. The model suggests that on the unmethylated maternally inherited chromosome, CTCF binds to the ICR and also to an upstream somatic DMR located 5' of *Igf2*. Binding does not occur at the paternal allele where methylation inhibits the binding. On the maternal allele ICR-DMR contact is made possible by CTCF dimerization bringing together the two distinct loci, and because they flank *Igf2*, the gene is ‘looped out’ (Figure 1D). Further chromatin contacts within the cluster, some facilitated by CTCF, then result in physical separation between the *Igf2* loop and the enhancers. Recently cohesins have been shown to bind to over half of CTCF binding sites in the genome, including in the H19/Igf2 cluster [75]. Given the ability of cohesins to tether DNA strands (*i.e.*, sister chromatids after cell’s S-phase) it is possible that cohesins contribute mechanistically to these chromatin contacts on the maternal H19/Igf2 locus. On the paternal allele, where CTCF cannot bind, long-range chromatin interactions are not observed within the cluster, suggesting a state that allows interaction between the 90 kb distant enhancers and *Igf2* (Figure 1D) [73]. Similar interactions involving CTCF have been noted at other loci (Figure 1D).

5. Relationship between DNA Methylation and Histone Modifications

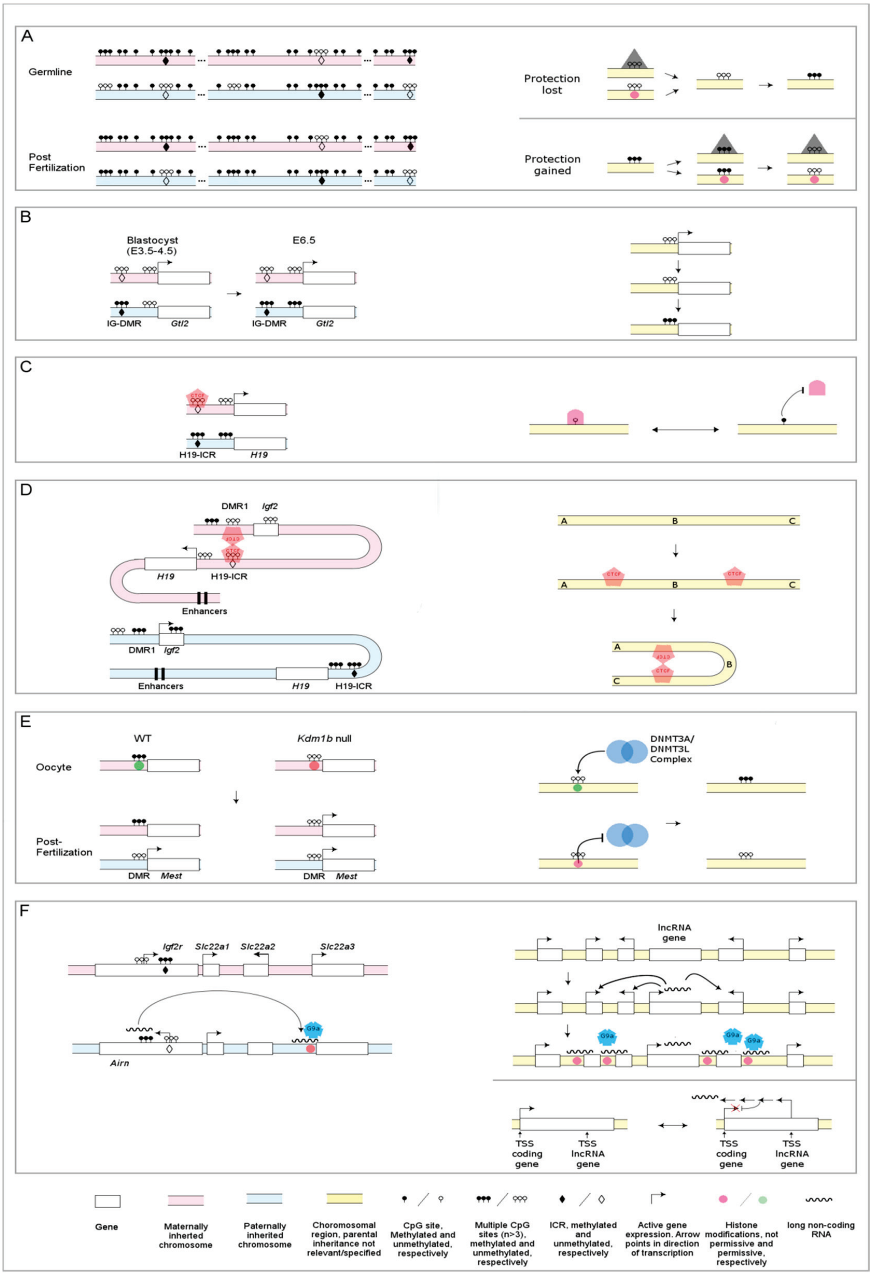
Similar to DNA methylation, correlation between multiple histone modifications in various genomic elements, including promoters, have been associated with gene activity and repression, and early studies illustrating this indeed investigated the relationship in the context of imprinted loci [76–82]. A functional relationship may therefore potentially exist between DNA methylation and histone modifications whereby the acquisition of one may be dependent on, or mutually exclusive, with the other. Indeed, as noted above, MBPs can recruit histone modification enzymes. Well-defined examples of histone modifications that regulate *de novo* DNA methylation are however scarce [83,84]. One very compelling example again comes from the study of genomic imprinting, as discussed below.

DNMT3L lacks a DNA methyltransferase activity, but it is necessary for methylation of DNA in certain situations [85,86] because it forms a complex with DNMT3A and DNMT3B, impacts their activity and contributes to their structural interaction with chromatin [87–90]. The ability is likely a result of a recently discovered affinity of DNMT3L to histone H3 [90] and this interaction is dependent on the methylation state of H3 at lysine K4—the binding only occurs when the histone is unmethylated hence H3 methylation might shield from DNA methylation [90]. A functional role for H3K4 methylation in modulating DNA methylation came from an imprinting study where Ciccone *et al.* showed that this interaction has important regulatory implications. The group generated mice deficient for a H3K4 demethylase enzyme, KDM1B, which resulted in increased H3K4 methylation in oocytes, where KDM1B is almost exclusively expressed. Consistent with inhibition of the DNMT3L-DNMT3A complex binding to methylated histone H3, DMRs at four imprinted regions that normally acquire DNA methylation in the female germ line were unmethylated in the *Kdm1b* null oocytes and imprinted expression of the corresponding genes was lost in embryos from *Kdm1b* null females (Figure 1E) [91]. These results strongly suggest a functional link between loss of H3K4 methylation and acquisition of DNA methylation, at least at imprinted regions (Figure 1E).

Cedar and Bergman take this further proposing a model of how the bimodal methylation pattern of mammalian genomes may be dependent on this same relationship. They suggest that *de novo* DNA methylation at the blastocyst stage is prevented at particular loci by deposition of H3K4 methylation. They further suggest that H3K4 methyltransferases may be targeted to CpG islands by RNA polymerase II and as a consequence, the DNA methyltransferase machinery containing DNMT3L, cannot access CpG sites in regulatory regions that are CpG islands [84].

H3K9 di- and trimethylation is associated with repressive DNA. DNA methylation is often found at such regions. Furthermore, DNA is globally hypomethylated in mouse ES cells carrying deletion of a H3K9 methyltransferase, G9a [92]. In this case the loss of DNA methylation is not a result of the aberrantly low levels of histone methylation, but rather due to loss of the histone methyltransferase enzyme itself; the DNA methyltransferase machinery interacts with G9a, and this interaction is mediated through a protein domain that is independent of the histone methyltransferase catalytic activity by a SET protein domain. Therefore, in *G9a*$^{-/-}$ mouse ES cells carrying *G9a* transgenes that lack histone methyltransferase activity, e.g., due to a point mutation in the SET domain, DNA methylation levels are partially rescued [93,94]. Regulation of DNA methylation through interaction of the DNMTs with histone modifiers, rather than with the histone modifications themselves, seems to be common and is observed for multiple mammalian histone methyltransferases [95–97], as well as in plant [98] and fungal systems [99]. Interestingly, in *G9a*$^{-/-}$ ES cells DNA methylation is lost at some imprinted loci [94,100], but where tested this is not observed in embryos [100,101]. This behaviour at imprints may suggest that ES cell culture is not a faithful model for assessing a requirement for histone modifying enzymes in DNA methylation, but equally might also reflect different properties of imprint-specific maintenance in ES cells compared to *in vivo*.

Figure 1. Regulatory epigenetic phenomena at imprinted loci. On the left are examples of various epigenetic mechanisms as observed in imprinted loci, and on the right models are presented of how those principles may apply more generally.


A
Germline
Post Fertilization
Protection lost
Protection gained
B
Blastocyst (E3.5-4.5)
E6.5
IG-DMR
Gtl2
C
CTCF
H19-ICR
H19
D
DMR1
Igf2
CTCF
H19
H19-ICR
Enhancers
A
B
C
E
WT
Kdm1b null
Oocyte
Post-Fertilization
DMR
Mest
DNMT3A/DNMT3L Complex
F
Igf2r
Slc22a1
Slc22a2
Slc22a3
G9a
Airn
lncRNA gene
TSS coding gene
TSS lncRNA gene
Gene
Maternally inherited chromosome
Paternally inherited chromosome
Choromosomal region, parental inheritance not relevant/specified
CpG site, Methylated and unmethylated, respectively
Multiple CpG sites (n>3), methylated and unmethylated, respectively
ICR, methylated and unmethylated, respectively
Active gene expression. Arrow points in direction of transcription
Histone modifications, not permissive and permissive, respectively
long non-coding RNA

(**A**) left: In the male germline (sperm), CpG dense regions are generally unmethylated and less dense regions are methylated. In the female germline (oocytes) CpG rich regions are more frequently methylated. This results in multiple differentially methylated regions between the male and female germlines. After fertilization only a small subset of these regions retain differential methylation. Retention of differential methylation at imprinting control regions (diamonds) post-fertilization may therefore be a targeted protection from either demethylation or *de novo* methylation; right: Model; Changes in DNA methylation may be mediated through loss and gain of such protection—when protection is lost (e.g., upper—as a result of factor (black triangle) binding or a histone modification that is non-permissive (red circle) for DNMT binding) CpGs become methylated by the methyltransferase machinery. If protection is gained (lower) the machinery cannot access the CpG sites to maintain methylation and after cell divisions methylation is therefore lost; (**B**) left: The secondary DMR located in the promoter region of the imprinted gene *Gtl2* becomes methylated on the paternal allele after expression is silenced; right: Model; DNA methylation at CpG sites in promoter regions of non-imprinted genes may therefore, at least in some cases, occur after gene silencing; (**C**) left: In the H19/Igf2 imprinted locus, CTCF (red pentagon) binds the H19-ICR, on the unmethylated maternal allele, not the methylated paternal allele; right: Model; Methylation of CpG sites can inhibit protein binding (purple) to DNA; (**D**) left: In the H19-Igf2 imprinted locus CTCF (red pentagon) binds to regions flanking *Igf2* and dimerizes, looping the gene and physically inhibiting its interaction with distal enhancers. On the paternally inherited allele, CTCF does not bind and enhancers are in contact with *Igf2* and the gene is expressed; right: Model; Looping of DNA sequences through the action of CTCF (red pentagon) can separate regions or bring them into contact; (**E**) left: In *Kdm1b*$^{-/-}$ (histone methyltransferase) mouse oocytes, imprints are not established at multiple ICRs due to the inhibitory effect of H3K4 methylation on DNMT3L. Histone modification states in WT and *Kdm1b*$^{-/-}$ mice are depicted as green and red circles to signify permissive and non-permissiveness to *de novo* DNA methylation, respectively. In embryos from *Kdm1b*$^{-/-}$ mothers, imprinted expression is lost, and genes are biallelically expressed (*Mest,* depicted) or repressed; right: Model; Histone modifications (red and green circles) can regulate DNA methylation; (**F**) left: In the Igf2r imprinted locus *Igf2r* expression is inhibited by transcriptional interference from the *Airn* lncRNA transcript on the paternal allele. The lncRNA recruits histone modifiers such as G9a (blue) to proximal imprinted genes that contribute to silencing of the imprinted *Slc22a3* in a lineage specific manner, e.g., through deposition of histone marks that are non-permissive for transcription (red circle); upper right: Model; lncRNAs may exert their effects in *trans* at proximal genes. As illustrated, a lncRNA is expressed and silences proximal genes, but not the more distal genes; lower right: Model; In the example provided a lncRNA and a coding gene are expressed from within the same ORF. Transcription of the lncRNA inhibits expression of the coding gene.

6. lncRNAs

6.1. lncRNAs, Definition, Characterization and Potential Functions

In recent years the roles of long noncoding RNAs (lncRNAs) in regulating genome function have received considerable attention, and are now emerging as a large group of genes with potential functions of fundamental importance for cell biology (for review see [2,102–104]). lncRNAs are defined as noncoding RNA transcripts of >200 bp [104]. Transcription of lncRNAs resembles that of mRNA genes; they are transcribed by the same transcriptional machinery and by RNA polymerase II, the transcripts are 5' capped and can be spliced and shuttled to the cytoplasm [102]. The lack of an open reading frame and their size are therefore the only criteria that currently define lncRNAs as a group [102,104]. On basis of high-throughput RNA sequencing experiments, the numbers of lncRNA transcripts have been suggested to range between 5000–15,000 [105,106]. With higher sensitivity, targeted capture experiments have identified lncRNAs that are undetectable by high-throughput technology, suggesting that this range is an underestimate [107]. However, as a result of their loosely defined criteria, lncRNAs as a group may be very heterogeneous. Therefore, the functional roles discussed below may only apply to a subset of their estimated numbers.

Despite the current excitement surrounding "new" roles for lncRNAs, they were shown to regulate genomic imprinting over a decade ago. Multiple potential functions of lncRNAs have been proposed whereby lncRNAs either exert their effects by acting in *trans* or by the act of their transcription in *cis* (transcriptional interference). Both effects have been shown to act at imprinted loci. Some *trans* acting lncRNAs, such as HOTAIR, have been suggested to exert their effects throughout the genome [108], while others, including most imprinted lncRNAs defined to date, act over a limited area surrounding or close to their transcriptional origin. Some lncRNAs may utilize both *cis* and *trans* acting mechanisms. An example is the imprinted Airn lncRNA whose transcription on the paternal chromosome represses *Igf2r* expression in *cis* by transcriptional interference [109], while the Airn RNA molecule itself is also necessary for regulating other genes in the cluster in a *trans*-targeted manner (see below, and Figure 1F [110]). Transcriptional interference is proposed to occur as a result of a collision between the transcriptional machineries of two adjacent or overlapping transcripts which might result in termination of one or both transcriptional events. Alternatively it may occur by promoter occlusion via inhibition of formation of a transcriptional initiation complex due to existing transcription of one transcript through the promoter of another [111].

Functions of *trans* acting lncRNAs have been proposed to fall into the following categories [112]: (1) Decoys: lncRNAs that bind to DNA binding proteins and prevent their interaction with DNA; (2) Scaffolds: lncRNA that function to join two or more proteins into an lncRNA-RNP (ribonucleoprotein) complex; (3) Guides: lncRNAs that bind proteins to guide them to certain genomic locations, e.g., by lending them specificity and/or binding capacity to certain DNA sequences or chromatin states.

6.2. lncRNAs in the Epigenetic Control of Genome Function—Lessons from Imprinting

Every cluster of imprinted genes contains at least one lncRNA and these lncRNAs are regulated by DNA methylation. This was demonstrated in experiments where the genes encoding DNA methyltransferases were deleted in mice to gauge effects on imprinting regulation. Promoters for the *Airn*, *Nespas/Gnasxl*, *Snrpn* and *Kcnq1ot1* lncRNA genes lie within the ICR for their respective region and are differentially methylated on the two parental chromosomes. Upon loss of DNMT1, the maintenance methyltransferase, methylation is lost at these ICRs in E10.5 embryos (the genetic manipulation is lethal at later embryonic stages) and *Airn, Nespas/Gnasxl*, *Snrpn* and *Kcnq1ot1* are biallelically expressed, with effects on neighboring imprinted protein coding genes, some of which may lose imprinting as a result of the lncRNA dysregulation [14,113,114]. *Kcnq1ot1* and *Airn* promoters are located in the ICRs, exhibit differential methylation, and, importantly, are located within genes running antisense to them, hence these provide examples of critical regulatory DNA methylation at genomic regions considered by some to have little or no consequence, *i.e.*, intragenic. The existence of other epigenetically regulated elements within genes and acting in this way to potentially regulate lncRNAs, may have very widespread effects on genome function.

Furthermore, imprinted lncRNAs have been demonstrated to be necessary for epigenetic control of genome function, to guide chromatin modifying enzymes in *trans* to specific sites in the genome. This is thought to mediate changes in histone modifications and be associated with changes in transcriptional activity. Although challenging to address experimentally, this function for lncRNAs is currently the topic of much attention. It was studies on imprinted gene regulation at the Igf2r/Airn and Kcnq1/Kcnq1ot1 imprinted clusters that provided examples of this type of regulation [104]. *Kcnq1ot1* and *Airn* are estimated as greater than 100 kb lncRNA transcripts, transcribed in an antisense orientation from within protein coding genes; *Kcnq1ot1* from *Kcnq1* in a 1 Mb imprinted cluster that contains eight maternally expressed protein coding genes, and *Airn* from *Igf2r* in a 400 kb long imprinted cluster that contains three maternally expressed protein coding genes. Both transcripts generate unspliced lncRNAs that are localized in the nucleus [103,115,116]. The ICRs of both genes are methylated on the maternal, but not paternally inherited chromosome, and determine monoallelic expression of the lncRNAs from the paternal allele. In mouse genetic mutants, where the promoters of *Airn* and *Kcnq1ot1* are deleted or their transcripts truncated by insertion of premature polyA sequence into the endogenous genes, biallelic expression of the imprinted protein coding genes occurs within their respective clusters [110,117–119]. These results suggested that the lncRNAs or the act of their transcription is necessary for silencing of genes in *cis* (Figure 1F). In addition, several lines of evidence further indicate that lncRNAs guide chromatin modifying enzymes in *trans* to establish repressive histone marks and gene silencing on the paternal allele (Figure 1F): In the Kcnq1 imprinted cluster *Osbpl5*, *Cd81*, *Ascl2* and *Tscc4* are imprinted exclusively in the placenta [40] and so are *Slc22a2* and *Slc22a3* in the Igf2r/Airn cluster [120]. The paternal chromosomes are bound by the histone methyltransferases G9a and/or Ezh-Eed2 in the extraembryonic lineage [40,110], and both Airn and Kcnq1ot1 lncRNAs associate with G9a histone methyltransferase in a lineage specific manner—in placenta but not embryo [110,115]. These results showed that lncRNAs may be a contributing factor for targeting epigenetic marks (Figure 1F) with genetic models being used alongside biochemical approaches to generate a more tractable and comparable experimental paradigm for added robustness. These studies have paved the way for explorations of the roles of multiple other lncRNAs which are found in association with different chromatin modifying enzymes [121,122]. Most recently, the imprinted lncRNA Gtl2/Meg3 has been shown to function in *trans* to target polycomb regulatory complexes in mouse and human stem cells in culture [123].

7. Conclusions

The robust genetic approaches applied to the regulation of imprinting have allowed it to be an excellent hypothesis-driven model to investigate and understand the epigenetic control of genome regulation. One of its greatest strengths as a model is that it allows the comparison of differentially expressed alleles of the two inherited copies of a gene with identical sequence within the same cell. Because these two parentally inherited alleles have well-defined different epigenetic states the contributions of these to gene expression can be determined. Since imprinted clusters employ multiple different epigenetic mechanisms, acting through various different mediators (long non-coding RNA, CTCF, *etc*.), this has enabled investigators to explore their hierarchical interactions and relationships with one another. As evidenced by the examples presented here, imprinting has provided insight into some of the most fundamental aspects of a range of epigenetic phenomena and their mediators.

Nevertheless, many important aspects of imprinting and epigenetic control remain to be elucidated. These include: what allows epigenetic marks to be *de novo* targeted differently in the male and female germlines; whether they are modulated by extrinsic or intrinsic signals, for example in the context of development and disease; and how DNA methylation is actively removed during reprogramming and perhaps at other times in development. The mechanisms regulating some of these processes are beginning to emerge where the context of imprinting has contributed; the DNA binding proteins ZFP57 and PGC/Stella have been shown to target and maintain DNA methylation at imprinted clusters during postfertilisation epigenetic reprogramming [19,124] and selective loss of imprinting is necessary for stem cell regulation in the neurogenic niche of the developing mouse [125]. Whether we can apply more generally what we learn from these mechanisms—for example about the general targeting of epigenetic states or the dynamic changes in epigenetic state in specific cellular niches—remains to be determined. It is likely that future studies, addressing these and other similarly fundamental questions in the context of imprinting will continue to add new layers to our understanding of genome regulation and the epigenetic control of genome function more widely.

Acknowledgments

We thank previous and current members of the Ferguson-Smith lab for their contribution to the ideas presented here and to Carol Edwards and Lisa Neumann for critical reading of the manuscript. Bjorn Adalsteinsson is funded by the Peoples Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement No.: 290123.

Author Contributions

Bjorn Thor Adalsteinsson wrote the manuscript. Anne C. Ferguson-Smith conceived, wrote and edited the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Miranda, T.B.; Jones, P.A. DNA methylation: The nuts and bolts of repression. *J. Cell. Physiol.* **2007**, *213*, 384–390.
2. Gibney, E.R.; Nolan, C.M. Epigenetics and gene expression. *Heredity (Edinb).* **2010**, *105*, 4–13.
3. Goll, M.G.; Bestor, T.H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **2005**, *74*, 481–514.
4. Smallwood, S.A.; Kelsey, G. *De novo* DNA methylation: A germ cell perspective. *Trends Genet.* **2012**, *28*, 33–42.
5. Kaneda, M.; Okano, M.; Hata, K.; Sado, T.; Tsujimoto, N.; Li, E.; Sasaki, H. Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **2004**, *429*, 900–903.
6. Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.; *et al*. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **2011**, *477*, 606–610.
7. Berger, S.L. The complex language of chromatin regulation during transcription. *Nature* **2007**, *447*, 407–412.
8. Ferguson-Smith, A.C. Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **2011**, *12*, 565–575.
9. Edwards, C.A.; Ferguson-Smith, A.C. Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **2007**, *19*, 281–289.
10. McGrath, J.; Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **1984**, *37*, 179–183.
11. Surani, M.A.H.; Barton, S.C.; Norris, M.L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **1984**, *308*, 548–550.
12. Ferguson-Smith, A.C.; Sasaki, H.; Cattanach, B.M.; Surani, M.A. Parental-origin-specific epigenetic modification of the mouse *H19* gene. *Nature* **1993**, *362*, 751–755.
13. Stöger, R.; Kubicka, P.; Liu, C.-G.; Kafri, T.; Razin, A.; Cedar, H.; Barlow, D.P. Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* **1993**, *73*, 61–71.
14. Li, E.; Beard, C.; Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **1993**, *366*, 362–365.
15. Smallwood, S.A.; Tomizawa, S.; Krueger, F.; Ruf, N.; Carli, N.; Segonds-Pichon, A.; Sato, S.; Hata, K.; Andrews, S.R.; Kelsey, G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **2011**, *43*, 811–814.
16. Kelsey, G.; Feil, R. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Phil. Trans. R. Soc. B* **2013**, *368*, doi:10.1098/rstb.2011.0336.
17. Kobayashi, H.; Sakurai, T.; Imai, M.; Takahashi, N.; Fukuda, A.; Yayoi, O.; Sato, S.; Nakabayashi, K.; Hata, K.; Sotomaru, Y.; *et al*. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **2012**, *8*, e1002440.

18. Smith, Z.D.; Chan, M.M.; Mikkelsen, T.S.; Gu, H.; Gnirke, A.; Regev, A.; Meissner, A. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **2012**, *484*, 339–344.
19. Li, X.; Ito, M.; Zhou, F.; Youngson, N.; Zuo, X.; Leder, P.; Ferguson-Smith, A.C. A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* **2008**, *15*, 547–557.
20. Groudine, M.; Eisenman, R.; Weintraub, H. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. *Nature* **1981**, *292*, 311–317.
21. Sutter, D.; Doerfler, W. Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 253–256.
22. Desrosiers, R.C.; Mulder, C.; Fleckenstein, B. Methylation of *Herpesvirus saimiri* DNA in lymphoid tumor cell lines. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3839–3843.
23. Cohen, J.C. Methylation of milk-borne and genetically transmitted mouse mammary tumor virus proviral DNA. *Cell* **1980**, *19*, 653–662.
24. Guntaka, R.V.; Rao, P.Y.; Mitsialis, S.A.; Katz, R. Modification of avian sarcoma proviral DNA sequences in nonpermissive XC cells but not in permissive chicken cells. *J. Virol.* **1980**, *34*, 569–572.
25. Van der Ploeg, L.H.T.; Flavell, R.A. DNA methylation in the human globin locus in erythroid and nonerythroid tissues. *Cell* **1980**, *19*, 947–958.
26. McGhee, J.D.; Ginder, G.D. Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* **1979**, *280*, 419–420.
27. Kuo, M.T.; Mandel, J.L.; Chambon, P. DNA methylation: Correlation with DNase I sensitivity of chicken ovalbumun and conalbumin chromatin. *Nucleic Acids Res.* **1979**, *7*, 2105–2113.
28. Mandel, J.L.; Chambon, P. DNA methylation: Organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucleic Acids Res.* **1979**, *7*, 2081–2103.
29. Bird, A.P.; Taggart, M.H.; Smith, B.A. Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* **1979**, *17*, 889–901.
30. Bird, A.; Taggart, M.; Macleod, D. Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of *X. laevis*. *Cell* **1981**, *26*, 381–390.
31. Vardimon, L.; Kressmann, A.; Cedar, H.; Maechler, M.; Doerfler, W. Expression of a cloned adenovirus gene is inhibited by *in vitro* methylation. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 1073–1077.
32. Stein, R.; Razin, A.; Cedar, H. *In vitro* methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3418–3422.
33. Ball, M.P.; Li, J.B.; Gao, Y.; Lee, J.-H.; LeProust, E.M.; Park, I.-H.; Xie, B.; Daley, G.Q.; Church, G.M. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **2009**, *27*, 361–368.

34. Deng, J.; Shoemaker, R.; Xie, B.; Gore, A.; LeProust, E.M.; Antosiewicz-Bourget, J.; Egli, D.; Maherali, N.; Park, I.-H.; Yu, J.; *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **2009**, *27*, 353–360.
35. Rauch, T.A.; Wu, X.; Zhong, X.; Riggs, A.D.; Pfeifer, G.P. A human B cell methylome at 100-base pair resolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 671–678.
36. Trowbridge, J.J.; Orkin, S.H. DNA methylation in adult stem cells. *Epigenetics* **2010**, *5*, 189–193.
37. Razin, A.; Cedar, H. DNA methylation and gene expression. *Microbiol. Rev.* **1991**, *55*, 451–458.
38. Sato, S.; Yoshida, W.; Soejima, H.; Nakabayashi, K.; Hata, K. Methylation dynamics of IG-DMR and *Gtl2*-DMR during murine embryonic and placental development. *Genomics* **2011**, *98*, 120–127.
39. Bhogal, B.; Arnaudo, A.; Dymkowski, A.; Best, A.; Davis, T.L. Methylation at mouse *Cdkn1c* is acquired during postimplantation development and functions to maintain imprinted expression. *Genomics* **2004**, *84*, 961–970.
40. Umlauf, D.; Goto, Y.; Cao, R.; Cerqueira, F.; Wagschal, A.; Zhang, Y.; Feil, R. Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat. Genet.* **2004**, *36*, 1296–1300.
41. Sasaki, H.; Ferguson-Smith, A.C.; Shum, A.S.W.; Barton, S.C.; Surani, M.A. Temporal and spatial regulation of H19 imprinting in normal and uniparental mouse embryos. *Development* **1995**, *121*, 4195–4202.
42. Lerchner, W.; Barlow, D.P. Paternal repression of the imprinted mouse *Igf2r* locus occurs during implantation and is stable in all tissues of the post-implantation mouse embryo. *Mech. Dev.* **1997**, *61*, 141–149.
43. Stadler, M.B.; Murr, R.; Burger, L.; Ivanek, R.; Lienert, F.; Schöler, A.; van Nimwegen, E.; Wirbelauer, C.; Oakeley, E.J.; Gaidatzis, D.; *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **2011**, *480*, 490–495.
44. Klose, R.J.; Bird, A.P. Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **2006**, *31*, 89–97.
45. Meehan, R.R.; Lewis, J.D.; McKay, S.; Kleiner, E.L.; Bird, A.P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **1989**, *58*, 499–507.
46. Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bird, A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **1992**, *69*, 905–914.
47. Hendrich, B.; Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **1998**, *18*, 6538–6547.
48. Hendrich, B.; Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **2003**, *19*, 269–277.
49. Bird, A.P.; Wolffe, A.P. Methylation-induced repression—Belts, braces, and chromatin. *Cell* **1999**, *99*, 451–454.

50. Jones, P.L.; Veenstra, G.J.C.; Wade, P.A.; Vermaak, D.; Kass, S.U.; Landsberger, N.; Strouboulis, J.; Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **1998**, *19*, 187–191.
51. Nan, X.; Ng, H.-H.; Johnson, C.A.; Laherty, C.D.; Turner, B.M.; Eisenman, R.N.; Bird, A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **1998**, *393*, 386–389.
52. Ng, H.-H.; Zhang, Y.; Hendrich, B.; Johnson, C.A.; Turner, B.M.; Erdjument-Bromage, H.; Tempst, P.; Reinberg, D.; Bird, A. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* **1999**, *23*, 58–61.
53. Wade, P.A.; Gegonne, A.; Jones, P.L.; Ballestar, E.; Aubry, F.; Wolffe, A.P. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat. Genet.* **1999**, *23*, 62–66.
54. Zhang, Y.; Ng, H.-H.; Erdjument-Bromage, H.; Tempst, P.; Bird, A.; Reinberg, D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **1999**, *13*, 1924–1935.
55. Sarraf, S.A.; Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **2004**, *15*, 595–605.
56. Prendergast, G.C.; Lawe, D.; Ziff, E.B. Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell* **1991**, *65*, 395–407.
57. Watt, F.; Molloy, P.L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **1988**, *2*, 1136–1143.
58. Comb, M.; Goodman, H.M. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res.* **1990**, *18*, 3975–3982.
59. DeChiara, T.M.; Robertson, E.J.; Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **1991**, *64*, 849–859.
60. Thorvaldsen, J.L.; Duran, K.L.; Bartolomei, M.S. Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* **1998**, *12*, 3693–3702.
61. Tremblay, K.D.; Duran, K.L.; Bartolomei, M.S. A 5' 2-kilobase-pair region of the imprinted mouse *H19* gene exhibits exclusive paternal methylation throughout development. *Mol. Cell. Biol.* **1997**, *17*, 4322–4329.
62. Szabó, P.E.; Tang, S.-H.E.; Rentsendorj, A.; Pfeifer, G.P.; Mann, J.R. Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function. *Curr. Biol.* **2000**, *10*, 607–610.
63. Kanduri, C.; Pant, V.; Loukinov, D.; Pugacheva, E.; Qi, C.-F.; Wolffe, A.; Ohlsson, R.; Lobanenkov, V.V. Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* **2000**, *10*, 853–856.
64. Hark, A.T.; Schoenherr, C.J.; Katz, D.J.; Ingram, R.S.; Levorse, J.M.; Tilghman, S.M. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **2000**, *405*, 486–489.

65. Bell, A.C.; Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **2000**, *405*, 482–485.

66. Webber, A.L.; Ingram, R.S.; Levorse, J.M.; Tilghman, S.M. Location of enhancers is essential for the imprinting of *H19* and *Igf2* genes. *Nature* **1998**, *391*, 711–715.

67. Leighton, P.A.; Saam, J.R.; Ingram, R.S.; Stewart, C.L.; Tilghman, S.M. An enhancer deletion affects both *H19* and *Igf2* expression. *Genes Dev.* **1995**, *9*, 2079–2089.

68. Phillips, J.E.; Corces, V.G. CTCF: Master weaver of the genome. *Cell* **2009**, *137*, 1194–1211.

69. Pant, V.; Kurukuti, S.; Pugacheva, E.; Shamsuddin, S.; Mariano, P.; Renkawitz, R.; Klenova, E.; Lobanenkov, V.; Ohlsson, R. Mutation of a single CTCF target site within the *H19* imprinting control region leads to loss of *Igf2* imprinting and complex patterns of *de novo* methylation upon maternal inheritance. *Mol. Cell. Biol.* **2004**, *24*, 3497–3504.

70. Yusufzai, T.M.; Tagami, H.; Nakatani, Y.; Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **2004**, *13*, 291–298.

71. Yoon, Y.S.; Jeong, S.; Rong, Q.; Park, K.-Y.; Chung, J.H.; Pfeifer, K. Analysis of the *H19ICR* insulator. *Mol. Cell. Biol.* **2007**, *27*, 3499–3510.

72. Li, T.; Hu, J.-F.; Qiu, X.; Ling, J.; Chen, H.; Wang, S.; Hou, A.; Vu, T.H.; Hoffman, A.R. CTCF regulates allelic expression of *Igf2* by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.* **2008**, *28*, 6473–6482.

73. Kurukuti, S.; Tiwari, V.K.; Tavoosidana, G.; Pugacheva, E.; Murrell, A.; Zhao, Z.; Lobanenkov, V.; Reik, W.; Ohlsson, R. CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10684–10689.

74. Murrell, A.; Heeson, S.; Reik, W. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat. Genet.* **2004**, *36*, 889–893.

75. Wendt, K.S.; Yoshida, K.; Itoh, T.; Bando, M.; Koch, B.; Schirghuber, E.; Tsutsumi, S.; Nagae, G.; Ishihara, K.; Mishiro, T.; *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **2008**, *451*, 796–801.

76. Hon, G.C.; Hawkins, R.D.; Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **2009**, *18*, R195–R201.

77. Ernst, J.; Kheradpour, P.; Mikkelsen, T.S.; Shoresh, N.; Ward, L.D.; Epstein, C.B.; Zhang, X.; Wang, L.; Issner, R.; Coyne, M.; *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **2011**, *473*, 43–49.

78. Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell* **2007**, *128*, 707–719.

79. Hon, G.; Wang, W.; Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **2009**, *5*, e1000566.

80. Zhou, V.W.; Goren, A.; Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **2011**, *12*, 7–18.

81. Grandjean, V.; O'Neill, L.; Sado, T.; Turner, B.; Ferguson-Smith, A. Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted *Igf2-H19* domain. *FEBS Lett.* **2001**, *488*, 165–169.

82. Pedone, P.V.; Pikaart, M.J.; Cerrato, F.; Vernucci, M.; Ungaro, P.; Bruni, C.B.; Riccio, A. Role of histone acetylation and DNA methylation in the maintenance of the imprinted expression of the *H19* and *Igf2* genes. *FEBS Lett.* **1999**, *458*, 45–50.
83. Chen, T. Mechanistic and functional links between histone methylation and DNA methylation. *Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 335–348.
84. Cedar, H.; Bergman, Y. Linking DNA methylation and histone modification: Patterns and paradigms. *Nat. Rev. Genet.* **2009**, *10*, 295–304.
85. Bourc'his, D.; Xu, G.-L.; Lin, C.-S.; Bollman, B.; Bestor, T.H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **2001**, *294*, 2536–2539.
86. Bourc'his, D.; Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **2004**, *431*, 96–99.
87. Jia, D.; Jurkowska, R.Z.; Zhang, X.; Jeltsch, A.; Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* **2007**, *449*, 248–251.
88. Margot, J.B.; Ehrenhofer-Murray, A.E.; Leonhardt, H. Interactions within the mammalian DNA methyltransferase family. *BMC Mol. Biol.* **2003**, *4*, doi:10.1186/1471-2199-4-7.
89. Suetake, I.; Shinozaki, F.; Miyagawa, J.; Takeshima, H.; Tajima, S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* **2004**, *279*, 27816–27823.
90. Ooi, S.K.T.; Qiu, C.; Bernstein, E.; Li, K.; Jia, D.; Yang, Z.; Erdjument-Bromage, H.; Tempst, P.; Lin, S.-P.; Allis, C.D.; *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **2007**, *448*, 714–717.
91. Ciccone, D.N.; Su, H.; Hevi, S.; Gay, F.; Lei, H.; Bajko, J.; Xu, G.; Li, E.; Chen, T. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **2009**, *461*, 415–418.
92. Ikegami, K.; Iwatani, M.; Suzuki, M.; Tachibana, M.; Shinkai, Y.; Tanaka, S.; Greally, J.M.; Yagi, S.; Hattori, N.; Shiota, K. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells* **2007**, *12*, 1–11.
93. Tachibana, M.; Matsumura, Y.; Fukuda, M.; Kimura, H.; Shinkai, Y. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J.* **2008**, *27*, 2681–2690.
94. Dong, K.B.; Maksakova, I.A.; Mohn, F.; Leung, D.; Appanah, R.; Lee, S.; Yang, H.W.; Lam, L.L.; Mager, D.L.; Schübeler, D.; *et al.* DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* **2008**, *27*, 2691–2701.
95. Lehnertz, B.; Ueda, Y.; Derijck, A.A.H.A.; Braunschweig, U.; Perez-Burgos, L.; Kubicek, S.; Chen, T.; Li, E.; Jenuwein, T.; Peters, A.H.F.M. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **2003**, *13*, 1192–1200.
96. Viré, E.; Brenner, C.; Deplus, R.; Blanchon, L.; Fraga, M.; Didelot, C.; Morey, L.; van Eynde, A.; Bernard, D.; Vanderwinden, J.-M.; *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **2006**, *439*, 871–874.

97. Li, H.; Rauch, T.; Chen, Z.-X.; Szabó, P.E.; Riggs, A.D.; Pfeifer, G.P. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J. Biol. Chem.* **2006**, *281*, 19489–19500.
98. Jackson, J.P.; Lindroth, A.M.; Cao, X.; Jacobsen, S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **2002**, *416*, 556–560.
99. Freitag, M.; Hickey, P.C.; Khlafallah, T.K.; Read, N.D.; Selker, E.U. HP1 is essential for DNA methylation in *Neurospora*. *Mol. Cell* **2004**, *13*, 427–434.
100. Xin, Z.; Tachibana, M.; Guggiari, M.; Heard, E.; Shinkai, Y.; Wagstaff, J. Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. *J. Biol. Chem.* **2003**, *278*, 14996–15000.
101. Wagschal, A.; Sutherland, H.G.; Woodfine, K.; Henckel, A.; Chebli, K.; Schulz, R.; Oakey, R.J.; Bickmore, W.A.; Feil, R. G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol. Cell. Biol.* **2008**, *28*, 1104–1113.
102. Mercer, T.R.; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300–307.
103. Mohammad, F.; Mondal, T.; Kanduri, C. Epigenetics of imprinted long noncoding RNAs. *Epigenetics* **2009**, *4*, 277–286.
104. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641.
105. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789.
106. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927.
107. Mercer, T.R.; Gerhardt, D.J.; Dinger, M.E.; Crawford, J.; Trapnell, C.; Jeddeloh, J.A.; Mattick, J.S.; Rinn, J.L. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **2012**, *30*, 99–104.
108. Chu, C.; Qu, K.; Zhong, F.L.; Artandi, S.E.; Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **2011**, *44*, 667–678.
109. Latos, P.A.; Pauler, F.M.; Koerner, M.V; Senergin, H.B.; Hudson, Q.J.; Stocsits, R.R.; Allhoff, W.; Stricker, S.H.; Klement, R.M.; Warczok, K.E.; *et al. Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **2012**, *338*, 1469–1472.
110. Nagano, T.; Mitchell, J.A.; Sanz, L.A.; Pauler, F.M.; Ferguson-Smith, A.C.; Feil, R.; Frase, P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **2008**, *322*, 1717–1720.
111. Osato, N.; Suzuki, Y.; Ikeo, K.; Gojobori, T. Transcriptional interferences in *cis* natural antisense transcripts of humans and mice. *Genetics* **2007**, *176*, 1299–1306.
112. Rinn, J.L.; Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166.

113. Howell, C.Y.; Bestor, T.H.; Ding, F.; Latham, K.E.; Mertineit, C.; Trasler, J.M.; Chaillet, J.R. Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* **2001**, *104*, 829–838.

114. Lewis, A.; Mitsuya, K.; Umlauf, D.; Smith, P.; Dean, W.; Walter, J.; Higgins, M.; Feil, R.; Reik, W. Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat. Genet.* **2004**, *36*, 1291–1295.

115. Pandey, R.R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-DiNardo, D.; Kanduri, C. *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **2008**, *32*, 232–246.

116. Seidl, C.I.M.; Stricker, S.H.; Barlow, D.P. The imprinted *Air* ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.* **2006**, *25*, 3565–3575.

117. Shin, J.-Y.; Fitzpatrick, G.V.; Higgins, M.J. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *EMBO J.* **2008**, *27*, 168–178.

118. Mancini-DiNardo, D.; Steele, S.J.S.; Levorse, J.M.; Ingram, R.S.; Tilghman, S.M. Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **2006**, *20*, 1268–1282.

119. Sleutels, F.; Zwart, R.; Barlow, D.P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **2002**, *415*, 810–813.

120. Zwart, R.; Sleutels, F.; Wutz, A.; Schinkel, A.H.; Barlow, D.P. Bidirectional action of the *Igf2r* imprint control element on upstream and downstream imprinted genes. *Genes Dev.* **2001**, *15*, 2361–2366.

121. Khalil, A.M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Morales, R.D.; Thomas, K.; Presser, A.; Bernstein, B.E.; van Oudenaarden, A.; *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11667–11672.

122. Guttman, M.; Donaghey, J.; Carey, B.W.; Garber, M.; Grenier, J.K.; Munson, G.; Young, G.; Lucas, A.B.; Ach, R.; Bruhn, L.; *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **2011**, *477*, 295–300.

123. Kaneko, S.; Bonasio, R.; Saldaña-Meyer, R.; Yoshida, T.; Son, J.; Nishino, K.; Umezawa, A.; Reinberg, D. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* **2014**, *53*, 1–11.

124. Nakamura, T.; Arai, Y.; Umehara, H.; Masuhara, M.; Kimura, T.; Taniguchi, H.; Sekimoto, T.; Ikawa, M.; Yoneda, Y.; Okabe, M.; *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **2007**, *9*, 64–71.

125. Ferrón, S.R.; Charalambous, M.; Radford, E.; McEwen, K.; Wildner, H.; Hind, E.; Morante-Redolat, J.M.; Laborda, J.; Guillemot, F.; Bauer, S.R.; *et al.* Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* **2011**, *475*, 381–385.

Whole Exome Sequencing of Extreme Morbid Obesity Patients: Translational Implications for Obesity and Related Disorders

Gilberto Paz-Filho, Margaret C.S. Boguszewski,Claudio A. Mastronardi, Hardip R. Patel, Angad S. Johar, Aaron Chuah, Gavin A. Huttley, Cesar L. Boguszewski, Ma-Li Wong, Mauricio Arcos-Burgos and Julio Licinio

Abstract: Whole-exome sequencing (WES) is a new tool that allows the rapid, inexpensive and accurate exploration of Mendelian and complex diseases, such as obesity. To identify sequence variants associated with obesity, we performed WES of family trios of one male teenager and one female child with severe early-onset obesity. Additionally, the teenager patient had hypopituitarism and hyperprolactinaemia. A comprehensive bioinformatics analysis found *de novo* and compound heterozygote sequence variants with a damaging effect on genes previously associated with obesity in mice (*LRP2*) and humans (*UCP2*), among other intriguing mutations affecting ciliary function (*DNAAF1*). A gene ontology and pathway analysis of genes harbouring mutations resulted in the significant identification of overrepresented pathways related to ATP/ITP (adenosine/inosine triphosphate) metabolism and, in general, to the regulation of lipid metabolism. We discuss the clinical and physiological consequences of these mutations and the importance of these findings for either the clinical assessment or eventual treatment of morbid obesity.

Reprinted from *Genes*. Cite as:Paz-Filho, G.; Boguszewski, M.C.S.; Mastronardi, C.A.; Patel, H.R.; Johar, A.S.; Chuah, A.; Huttley, G.A.; Boguszewski, C.L.; Wong, M.; Arcos-Burgos, M.; Licinio, J. Whole Exome Sequencing of Extreme Morbid Obesity Patients: Translational Implications for Obesity and Related Disorders. *Genes* **2014**, *5*, 709-725.

1. Introduction

Obesity is a global epidemic: the World Health Organization (WHO) estimates that half a billion people over the age of twenty worldwide are obese [1]. Global projections estimate that worldwide, 1.2 billion individuals will be obese by 2030 [2].

Obesity and its related traits have high estimates of heritability (h^2 between 40% and 70%) [3]. The investigation of candidate genes and genome-wide association studies have identified more than 60 obesity susceptibility genes that predispose to increased body weight, waist circumference, waist-hip ratio, body mass index (BMI) and fat percentage or fat mass. However, mutations in these genes account for a very small fraction of the obesity phenotypic variance [4,5]. It has been estimated that at least 7% of children with severe early-onset obesity (defined by an onset before the age of 10 years and BMI over three standard deviations (SD) above normal) have a single locus sequence variant determining obesity [6]. Nine susceptibility genes, determinants of non-syndromic Mendelian forms of human obesity, are involved in the hypothalamic control of energy balance via the leptin-melanocortin pathway and/or in neural development [4]: brain-derived neurotrophic factor (BDNF), leptin (LEP), leptin receptor (LEPR), melanocortin-4 receptor (MC4R), neurotrophic tyrosine kinase receptor type 2 (NTRK2), prohormone convertase 1 (PCSK1), proopiomelanocortin (POMC), single-minded homolog 1 (SIM1) and, more recently, melanocortin 2 receptor accessory protein 2 (MRAP2) [7].

Exome sequencing is rapidly becoming the first-line approach for monogenic disorders [8]. The use of whole-exome capture and the complete sequencing of the coding genome of parent-child trios is a highly effective approach for identifying homozygous, compound heterozygous and *de novo* coding sequence variants, as multiple *de novo* sequence variants occurring within a specific gene (or within a gene family or pathway) are extremely implausible events [8]. Its rationale is based on the fact that gene variants located in exons are more likely to be pathogenic than those located in introns or between genes. The power of this strategy has increased with the access to large numbers of publicly available exome sequences that allow the controlled comparison of frequencies, as well as the identification of *de novo* variants and stratification by ethnicity. This strategy has been used to identify candidate genes for several Mendelian and complex traits [9–12]. In the assessment of obesity, whole-exome sequencing has identified sequence variants in the leptin receptor gene [13], in the *ADCY3* gene [14] and in the *BBIP1* gene in patients with Bardet–Biedl syndrome [15]. However, no novel pathogenic genes or pathways associated with obesity have been identified through this approach yet.

In this study, by employing whole-exome capture and sequencing in the assessment of two patients with severe early-onset obesity (and their parents), we identified *de novo* mutations and the compound heterozygous status of several damaging variants. Intriguingly, some of these variants were harboured in genes involved in the pathophysiology of obesity (such as *LRP2* and *UCP2*), providing the foundation for future research in this field. Thus far, we emphasize that the networking of clinical case-reports and genetic analyses would be crucial to finding the major loci underpinning complex disorders.

2. Experimental

Two family trios, the probands of which had severe early-onset obesity (onset before the age of 10 years and BMI over three SD above normal) were included in this study. All parents and capable patients provided written informed consent for the genetic research studies, which were performed in accordance with the study protocol approved by the Australian National University Human Research Ethics Committee (Protocol 2011/108, approved on the 6 May, 2011) and in concordance with the Helsinki Declaration of 1975, as revised in 2008. DNA was extracted from peripheral blood from the patients and parents for genetic analysis.

2.1. DNA Library Preparation, Exome Capture and Sequencing Protocol

Libraries were constructed from 1 µg of genomic DNA using an Illumina TruSeq genomic DNA library kit (Illumina Inc., San Diego, CA, USA). Libraries were multiplexed with 6 samples pooled together (500 ng of each library). Exons were enriched from the pooled 3 µg of library DNA using an Illumina TruSeq Exome enrichment kit (Illumina Inc.). Each exome-enriched pool was run on a 100-base-pair paired-end run on an Illumina HiSeq 2000 sequencer (Illumina Inc.). We surveyed

201,071 genomic regions in total using the exome capture platform. Ninety percent of the bases in approximately 197,000 of these targeted regions had at least one read coverage. All regions were sampled at approximately 50× coverage.

2.2. Sequence Read Processing, Alignment, Bioinformatics and Genetic Analyses

The sequencing image data were processed in real time using Illumina Real Time Analysis (RTA) software (Illumina Inc.) and converted to fastq files containing DNA base calls (A, C, G and T) and quality scores using the Illumina CASAVA pipeline (a software program that converts raw image data into sequences). The resulting fastq files were further processed for variant analysis.

The entire workflow of data curation and analysis for variant-calling was developed by the Genome Discovery Unit (GDU) at The Australian National University. Key components of the workflow include: (i) quality assessment; (ii) read alignment; (iii) local realignment around the known and novel indel regions to refine indel boundaries; (iv) recalibration of base qualities; (v) variant calling; and (vi) assigning quality scores to variants (detailed workflow information is in the Supplemental Material).

Subsequently, we included a filtering phase (using information from dbSNP and the 1K Exome Project), with the following sequential steps: (1) identification of rare or *de novo* variants (a lower minor allele frequency cut-off (MAF) in the window of 0.1%–1.0%); (2) filtering of variants to include those that are potentially pathogenic or are specific variants associated with disease susceptibility using several tools, namely, SIFT, PolyPhen2, Mutation Taster, Mutation Assessor and Functional Analysis through Hidden Markov Models (FATHMM), as implemented by the DNA-seq Analysis Package (SVS7.7.6, Golden Helix, Bozeman, MT, USA) (variants were not excluded if classified as potentially damaging by at least one of these filtering tools); (3) filtering of damaging variants based on genes known to be associated with human disease; and (4) independent confirmation of selected variants by Sanger sequencing (Supplemental Material). The definition of *de novo* sequence variants, compound heterozygous polymorphisms and rare recessive homozygous polymorphisms was performed with different modules of the DNA-seq Analysis Package (SVS7.7.6, Golden Helix, Bozeman, MT, USA).

To identify potential enriched endocrine-physiological pathways, a genetic ontology pathway analysis was performed. For constructing the pathways, variants with potential functional changes detected by the *de novo* and the compound heterozygous analysis were examined with the set of algorithms implemented in MetaCore (Thomson Reuters, New York, NY, USA) for the heuristic interpretation of maps, networks and rich ontologies for diseases.

3. Clinical Reports

Patient 1: A Brazilian male teenager with a history of excessive weight gain starting at age 3, decelerated growth since age 11 and delayed puberty was first evaluated at age 14 y, 3 m. His body weight was 105.0 kg (+5.32 SD score), his height 152.5 cm (−1.45 SD score) and his BMI 45.2 kg/m^2 (+11.03 SD score) (Figure 1). The patient had no complaints of hearing deficits or vision loss. Testicular volumes were <2 mL bilaterally, and he was at Tanner pubertal stage P2–P3.

He had cubitus valgus and round facies. During physical examination, profuse sweating was noted. Other physical signs were unremarkable.

Obesity was a common finding in his family, but all individuals had normal height. His father is obese (BMI 36.7 kg/m^2), as are his paternal grandfather (BMI 44.6 kg/m^2) and paternal uncle (BMI 50.5 kg/m^2). His mother's BMI is 30.1 kg/m^2, and two maternal aunts are also obese (BMI 31.6 and 30.1 kg/m^2). His older brother is overweight (BMI 29 kg/m^2). None of his family members have a history of severe early-onset obesity. There was no history of consanguinity in the family. Pregnancy was uneventful, and size at birth was 4.2 kg and 51.5 cm.

Figure 1. Height, weight and BMI of Patient 1, from birth to age 15 y, 10 m.

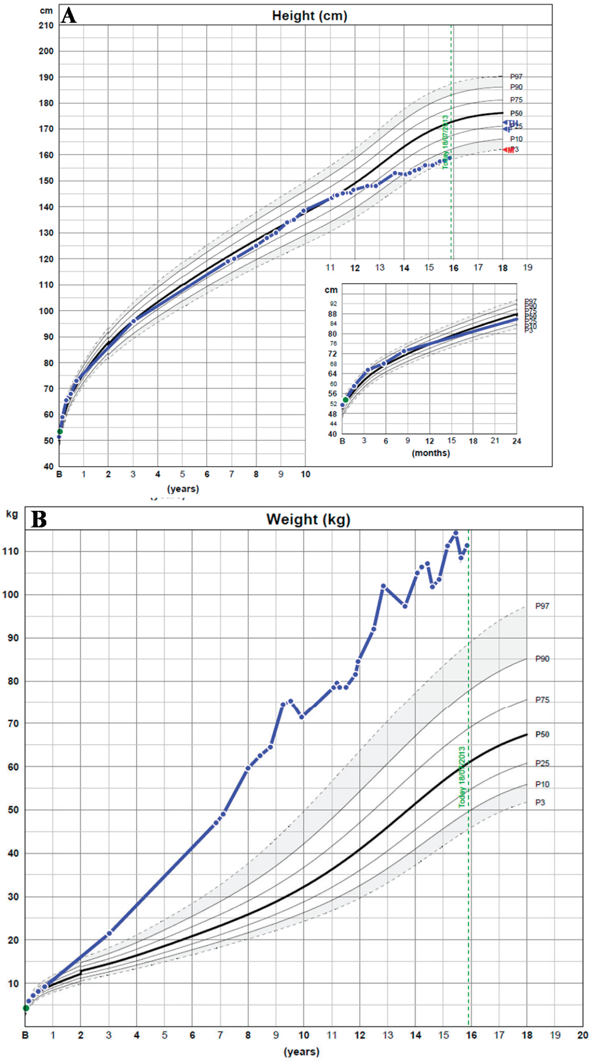
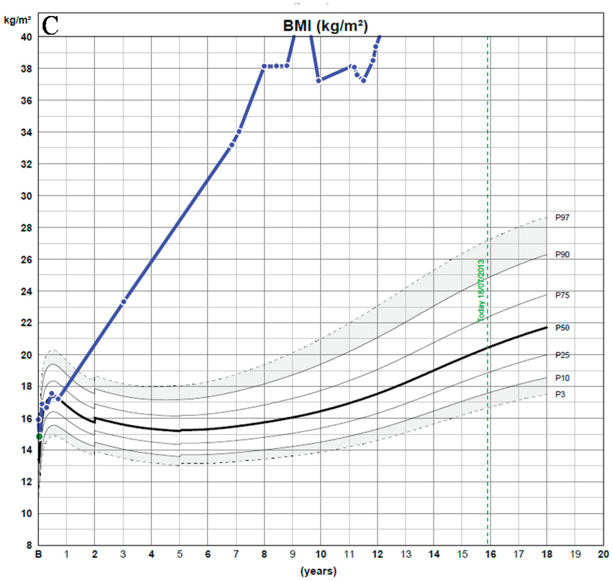


Figure 1. *Cont.*



Growth charts illustrating changes in height (**A**), weight (**B**) and BMI (**C**). M, maternal height; P, paternal height; TH, target height. From WHO Child Growth Standard: Methods and Developments.

He had been previously diagnosed with central hypothyroidism at age 9, with thyroid-stimulating hormone (TSH) of 7.4 mU/L and free T4 of 9.5 pmol/L, with undetectable titres of antithyroglobulin and antithyroperoxidase antibodies and a normal thyroid ultrasound. He also had elevated serum prolactin levels of 2,908 pmol/L, measured for the first time at age 9. Magnetic resonance enhanced by the contrast gadolinium showed a pituitary gland of normal volume, with no evidence of pituitary adenoma and without any structural abnormalities in the brain. The search for macroprolactin was negative (71% recovery after polyethylene glycol precipitation). He had been on treatment with levothyroxine 88 μg/day since age 9 y, 5 m, and cabergoline 0.5 mg every 10–15 days since age 13 y, 2 m, which had normalized his serum TSH and prolactin levels.

During the previous five years, he was treated with hypocaloric mixed diets, frequent physical activity, sibutramine 10 mg/day and orlistat 120 mg after meals. This approach resulted in only a 6-kg noncontinuous weight loss.

The patient's serum triglycerides and total cholesterol were elevated (3.11 and 4.84 mmol/L, respectively), with low HDL-cholesterol of 0.85 mmol/L and normal calculated LDL-cholesterol of 1.99 mmol/L. Fasting plasma glucose and insulin were 5.33 mmol/L and 13 μU/mL, respectively, with the homeostasis model assessment-estimated insulin resistance (HOMA-IR) index equal to 3.07. His serum insulin-like growth factor 1 (IGF-1) level was below the age reference range (71.25 nmol/L; reference value 130–563 nmol/L). Growth hormone (GH) secretion was evaluated during a standard insulin provocative test, with no development of hypoglycaemia during the test (lowest glucose level of 3.55 mmol/L and respective GH level of 0.06 μg/L). Morning cortisol and adrenocorticotropic hormone (ACTH) levels were normal (0.68 μmol/L and 1.60 pmol/L, respectively). His serum leptin levels were 8.1 and 18.0 µg/L at age 9 and appropriately elevated at age 13 (69.7 µg/L). At age 13, his total testosterone levels were pre-pubertal (0.90 nmol/L), and follicle-stimulating hormone (FSH) and luteinizing hormone (LH) were undetectable. At age 13 y, 9 m, his bone age was 15.6 years. Table 1 summarizes the laboratory test results and their respective reference range values.

Table 1. Laboratory test results for Patient 1.

Test	Value	Normal reference range
Thyroid-stimulating hormone (TSH) *	7.4 mU/L	0.3–5.0 mU/L
Free T4 *	9.5 pmol/L	10.3–25.7 pmol/L
Antithyroglobulin (ATG) and antithyroperoxidase (ATPO) antibodies *	Both negative	<9.0 IU/mL (ATG) <116 IU/mL (ATPO)
Prolactin *	2908 pmol/L	82–504 pmol/L
Macroprolactin *	Negative (71% recovery)	>50% recovery
Total cholesterol $	4.84 mmol/L	4.4 mmol/L
HDL cholesterol $	0.85 mmol/L	>1.16 mmol/L
LDL cholesterol $	1.99 mmol/L	<2.84 mmol/L
Triglycerides $	3.11 mmol/L	<1.02 mmol/L
Fasting plasma glucose $	5.33 mmol/L	3.89–5.5 mmol/L
Fasting insulin $	13 µU/mL	1.8–4.6 µU/mL
Insulin-like growth factor 1 (IGF-1) &	71.25 nmol/L	130–563 nmol/L
Growth hormone (GH)/glucose &#	0.06 µg/L/3.55 mmol/L	>5 µg/L/<1.94 mmol/L
Adrenocorticotropic hormone (ACTH) (morning)	1.60 pmol/L	2.2–13.2 pmol/L
Cortisol (morning)	0.68 µmol/L	0.14–0.70 µmol/L
Leptin	8.1 * and 69.7 µg/L &	Detectable
Total testosterone &	0.90 nmol/L	3.47–41.60 nmol/L
Follicle-stimulating hormone (FSH) &	Undetectable	0.5–10.5 IU/L
Luteinizing hormone (LH) &	Undetectable	0.5–7.9 IU/L
Total calcium ^	2.62 mmol/L	2.40–2.64 mmol/L
Inorganic phosphate ^	173.4 mmol/L	108.4–164.2 mmol/L
Magnesium ^	1.1 mmol/L	0.7–0.9 mmol/L
Alkaline phosphatise ^	114 U/L	66–571 U/L
25-hydroxy vitamin D ^	85 mmol/L	>75 mmol/L
Parathyroid hormone (PTH) ^	2.6 pmol/L	1.0–5.5 pmol/L
Selenium @	0.03 µmol/L	0.25–2.4 µmol/L
Total urinary protein @	0.08 g/24 hours	<0.15 g/24 hours

* Measured at age 9, not treated with levothyroxine and cabergoline; $ measured at age 12; # GH and lowest glucose level measured during a standard insulin provocative test; & measured at age 13; ^ measured at age 14; @ measured at age 16.

Radiographic studies showed lumbar scoliosis convex to the right, mild reduction of intervertebral spaces at L4-S1, mild shortening of L1, as well as a short fourth metacarpal. In order to exclude the diagnosis of Albright's hereditary osteodystrophy, serum electrolytes, alkaline phosphatase, vitamin D and parathyroid hormone (PTH) were measured, which were all unremarkable for Albright's hereditary osteodystrophy (total calcium 2.62 mmol/L; inorganic phosphate 173.4 mmol/L; magnesium 1.1 mmol/L, alkaline phosphatase 114 U/L, 25-hydroxy vitamin D 85 mmol/L and PTH 2.6 pmol/L). His serum levels of selenium were 0.03 µmol/L, 10-fold lower than the lowest limit of the normal range (reference values 0.25–2.4 µmol/L). Total urinary protein levels were 0.08 g/24 hours.

Intramuscular injections of testosterone esters (70 mg every four weeks) were initiated for puberty induction. Biosynthetic GH was initiated in a dose of 0.33 mg/day and increased to 0.66 mg/day according to IGF-1 levels. Four months after starting GH therapy, he developed episodes of fever of unexplained origin, which resolved spontaneously after six months. During that period, investigation for infection disease was negative, including a normal PET-scan, with leucocytosis as the only observed abnormality.

Patient 2: A two year-old Brazilian girl was evaluated for severe early-onset obesity. Her body weight was 23 kg (+4.79 SD score), her height was 93 cm (+2.32 SD score) and her BMI was 26.6 kg/m^2 (+4.49 SD score). She was born with 2.9 kg and 46 cm, from an uneventful pregnancy. Excessive weight gain was noted upon a few weeks after birth. Neurologic development was normal, with no evidence of Prader–Willi or Bardet–Biedl syndromes. She had normal serum leptin levels of 18 µg/L. A history of recurrent bacterial and viral respiratory tract infections was noted, which warranted the need for antibiotic therapy almost every month. There was no significant familial history of obesity or consanguinity. The physical examination was unremarkable.

4. Results

We called a total of 455,342 variants, 336,652 of them polymorphic, 21,613 matched at the dbNSFP, 12,286 with potential pathogenic effects and 2,291 with a minor allele frequency <1% when compared to the 1 kG phase 1. Our filtering approach by *de novo* functional mutation screening reported three *de novo* sequence variants with a potential damaging effect (Table 2). These sequence variants were found in Patient 1 (*UCP2*) and in Patient 2 (*AICDA* and *FAM71E2*).

The compound heterozygous analysis identified 20 variants with potential functional effects associated with eight genes, namely: *LRP2*, *AMPD3*, *OR8U8-OR9G1* and *SLC22A6* in Patient 1 and *TTN*, *APEH*, *DNAAF1* and *KIR3DL3* in Patient 2 (Table 3). From a literature search on the compound heterozygous variants that were classified as damaging, we identified two sequence variants in the *LRP2* gene that may potentially affect LRP2 protein function in Patient 1: a genomic variant G→T (NC_000002.11:g.170009391G>T), resulting in a nonsynonymous substitution on codon 4127 (NM_004525.2:c.12379C>A; NP_004516.2:p.Arg4127Ser), and a genomic variant C→T (NC_000002.11:g.170030506C>T), resulting in a nonsynonymous substitution on codon 3646 (NM_004525.2:c.10937G>A; (NP_004516.2:p.Arg3646His). The *LRP2* gene encodes a multi-ligand endocytic receptor (also known as megalin or glycoprotein 330) involved in the regulation of the leptin-melanocortin pathway.

Table 2. *De novo* damaging sequence variants.

Patient	Variant	Chr	Position	Ref All	Alt All	Identifier	Classification	Gene	Transcript	Exon	HGVS Coding	HGVS Protein
1	11:73689104-SNV	11	73,689,104	G	A	rs660339	Nonsyn SNV	UCP2	NM_003355	4	c.164C>T	p.Ala55Val
2	12:8757523-Ins	12	8,757,523	-	A	rs5796316	Splicing	AICDA	NM_020661	4	c.428-5_428-4insT	
2	19:55873642-SNV	19	55,873,642	C	T	rs4252574	Nonsyn SNV	FAM71E2	NM_001145402	3	c.535G>A	p.Glu179Lys

Chr: chromosome; Ref All: reference allele; Alt All: alternate allele; HGVS: Human Genome Variation Society nomenclature.

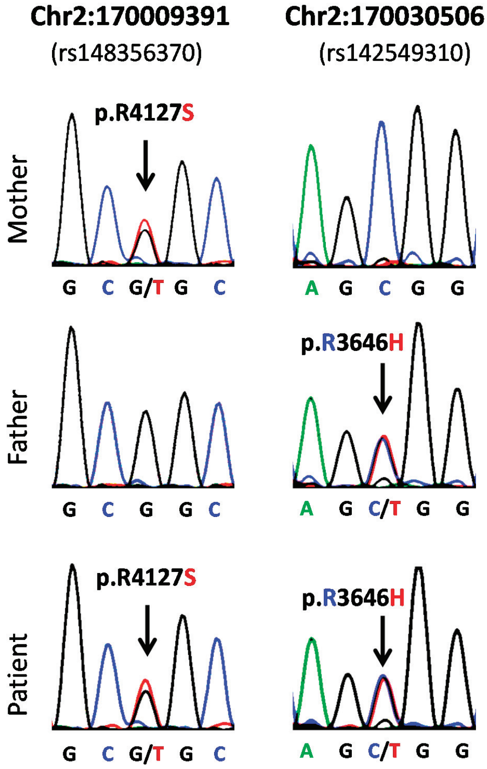
Table 3. Sequence rare variants (allele frequency in the general population <1%) with potential damaging effect by compound heterozygous analysis.

Patient	Single Nucleotide Variant	Chromosome	Position	Identifier	Gene	dbSNP MAF Frequency	Alleles	Reference Allele	Reference Aminoacid	Altered Aminoacid	HGVS Protein
1	2:170009391-SNV	2	170009391	rs148356370	LRP2	0.005	G/T	G	R	S	p.R4127S
1	2:170030506-SNV	2	170030506	rs142549310	LRP2	0.002	C/T	C	R	H	p.R3646H
1	11:10518373-SNV	11	10518373	rs144107914	AMPD3	0.001	C/T	C	S	L	p.S323L
1	11:10527316-SNV	11	10527316	N/A	AMPD3	***	A/G	G	R	Q	p.R571Q
1	11:56468198-SNV	11	56468198	rs4990194	OR8U8-OR9G1	0.069	A/G	A	Y	C	p.Y112C
1	11:56468212-SNV	11	56468212	rs591369	OR8U8-OR9G1	**	A/G	G	V	M	p.V117M
1	11:56468554-SNV	11	56468554	rs12420076	OR8U8-OR9G1	0.061	A/C	A	K	Q	p.K231Q
1	11:56468560-SNV	11	56468560	rs10896516	OR8U8-OR9G1	**	C/T	T	Y	H	p.Y233H
1	11:56468561-SNV	11	56468561	rs10896517	OR8U8-OR9G1	0.047	A/G	A	Y	C	p.Y233C
1	11:62748503-SNV	11	62748503	rs150409056	SLC22A6	*	G/T	G	R	S	p.R331S
1	11:62749384-SNV	11	62749384	rs200609617	SLC22A6	***	C/T	C	A	T	p.A243T
2	2:179507021-SNV	2	179507021	N/A	TTN	***	G/A	G	R	C	p.R4436C
2	2:179577628-SNV	2	179577628	N/A	TTN	***	C/T	C	V	I	p.V7798I
2	2:179634421-SNV	2	179634421	rs200875815	TTN	***	T/G	T	T	P	c.8749A>C
2	3:49716372-SNV	3	49716372	N/A	APEH	***	A/G	G	R	H	p.R383H
2	3:49720698-SNV	3	49720698	N/A	APEH	***	A/G	G	A	T	p.A708T
2	16:84203467-SNV	16	84203467	rs143322223	DNAAF1	***	C/G	C	E	Q	p.E345Q
2	16:84208329-SNV	16	84208329	rs139519641	DNAAF1	*	A/G	A	?	?	Splicing
2	19:55239223-SNV	19	55239223	rs117372288	KIR3DL3	***	A/G	A	V	I	p.V168I
2	19:55241240-SNV	19	55241240	rs111516669	KIR3DL3	**	A/G	A	V	I	p.V313I

Note: The first two rows show the sequence variants that affect function, harboured at the *LRP2* gene (NC_000002.11:g.170009391G>T and NC_000002.11:g.170030506C>T). N/A: not available. * Monomorphic in available data from dbSNP database; ** MAF only available for 2 chromosomes; *** There is no frequency data.

Sanger sequencing of the two amplicons containing the two *LRP2* sequence variants validated the findings obtained from the whole-exome capture and sequencing analysis (Figure 2). These results confirmed that each parent is heterozygous for one of the sequence variants (the mother is heterozygous for the mentioned sequence variant (g/t) (NC_000002.11:g.170009391G>T), and the father is heterozygous for the sequence variant (c/t) (NC_000002.11:g.170030506C>T), whereas the patient is heterozygous for both *LRP2* sequence variants. Therefore, the patient is compound heterozygous for the *LRP2* gene as initially described in the whole-exome capture and sequencing analysis (Figure 2).

Figure 2. Sanger sequencing results from Patient 1 and parents.



Sanger sequencing of the *LRP2* gene shows that both parents are heterozygous for one of the mutations. The mother is heterozygous for the Chr2:170009391 sequence variant (g/t) (NC_000002.11:g.170009391G>T), and the father is heterozygous for the Chr2: 170030506 variant (c/t) (NC_000002.11:g.170030506C>T). The patient is compound heterozygous and has both sequence variants.

MetaCore analysis, including those genes harbouring functional mutations, namely *UCP2*, *AICDA*, *FAM71E2*, *TTN*, *APEH*, *DNAAF1*, *KIR3DL3*, *LRP2*, *AMPD3*, *OR8U8-OR9G1* and *SLC22A6*, defined six pathways significantly overrepresented (after false discovery correction, FDR), e.g.,: (1) ATP/ITP (adenosine/inosine triphosphate) metabolism; (2) regulation of lipid metabolism/peroxisome proliferator-activated receptor (PPAR) regulation of lipid metabolism; (3) development of insulin, IGF-1 and TNF-alpha in brown adipocyte differentiation; (4) mitochondrial dysfunction in neurodegenerative diseases; (5) oxidative stress role of Sirtuin1 and PGC1 alpha in the activation of the defence system; and (6) CTP/UTP (cytidine/uridine triphosphate) metabolism (Table 4).

Table 4. Significant pathways from MetaCore analysis of candidate morbid obesity genes.

Pathways of Candidate Morbid Obesity Genes	Genes from Input List in Pathway	*p*-Value	FDR
ATP, ITP metabolism	AMPD3	1.107e-3	6.640e-3
Regulation of lipid metabolism PPAR regulation of lipid metabolism	UCP2	1.851e-2	3.164e-2
Development of insulin, IGF-1 and TNF-alpha in brown adipocyte differentiation	UCP2	2.332e-2	3.164e-2
Mitochondrial dysfunction in neurodegenerative diseases	UCP2	2.594e-2	3.164e-2
Oxidative stress role of Sirtuin1 and PGC1 alpha in the activation of the defence system	UCP2	2.637e-2	3.164e-2
CTP UTP metabolism	AICDA	4.709e-2	4.709e-2

Note: *p*-values and false discovery (FDR) correction adjusting for multiple comparisons, representing the probability that these pathways generated from our candidate gene list would appear by coincidental chance from feeding a random set of genes.

5. Conclusions

In this study, by employing whole-exome capture and sequencing, we identified novel sequence variants in the *LRP2* gene that might be associated with the phenotype of severe early-onset obesity in Patient 1. Similar to the other genes associated with monogenic forms of obesity, *LRP2* is also involved in the regulation of the leptin-melanocortin pathway [16]. In addition, in Patient 1, we found a *de novo* sequence variant in the *UCP2* gene, a transporter protein present in the mitochondrial inner membrane that is a key regulator of energy balance, the variants of which have already been associated with obesity [17]. In Patient 2, we identified sequence variants in the *DNAAF1* gene, which might be related to the second patient's phenotype. The *DNAAF1* gene is required for the stability of the ciliary architecture, and it has been demonstrated that ciliary dysfunction is associated with the pathogenesis of obesity [18,19].

Recently, LRP2 has been nominated as a novel appetite regulator responsible for generating satiety signals in hypothalamic neurons [16]. It is a multiligand endocytic receptor, a member of the low density lipoprotein receptor gene family, which binds a large variety of ligands. Leptin is one of its ligands; LRP2 mediates its reuptake in renal tubules [20] and promotes leptin transport across the choroid plexus [21]. LRP2 also binds to the long-form leptin receptor (LepRb), forming a complex that is co-localized and subjected to endocytosis in the hypothalamic neurons. Subsequently, the endocytosis of this co-localized complex leads to the activation of signal transducer and activator of transcription 3 (STAT3) signalling in hypothalamic neurons, including proopiomelanocortin (POMC)- and LepRb-expressing neurons. As a consequence, food intake and body weight are decreased. In the absence of functional LRP2, STAT3 signalling is decreased. Therefore, hunger is stimulated and satiety decreased [16].

The binding of LRP2 and LepRb is enhanced by clusterin, a sulphated glycoprotein widely expressed in hypothalamic areas involved in the regulation of food intake and energy metabolism [22]. Chronic intracerebroventricular (icv) administration of clusterin causes the reduction of food intake, body weight and epididymal fat mass [16]. LRP2 is expressed in rodent hypothalamus, and its inhibition by small interfering RNA significantly blunts the effects of clusterin icv injections on food intake and Stat3 activation [16]. Therefore, LRP2 acts as a key mediator of the food intake-suppressing effects of clusterin, and its absence can cause obesity in rodents (as previously demonstrated) [16].

Sequence variants in the *LRP2* gene have been previously associated with Donnai–Barrow/ facio-oculo-acoustico-renal (DB/FOAR) syndrome [23]. The phenotype of this syndrome includes agenesis of the corpus callosum, developmental delay, enlarged anterior fontanelle, high myopia, hypertelorism, proteinuria and sensorineural hearing loss, but not obesity. In a review by Pober *et al.*, sensorineural hearing loss, high myopia and proteinuria were present in 100% of DB/FOAR syndrome cases. None of those features were present in Patient 1; therefore, we ruled this diagnosis out. Serum selenium levels in patients with DB/FOAR syndrome have not been reported, but it has been shown that LRP2 mediates the reuptake of selenoproteins in the kidney and that LRP2-mutant mice have low selenium serum levels due to the increased urinary excretion of selenoproteins [24]. Since the patient's serum levels of selenium were 10-fold lower than those of the reference range, this finding supports the impairment of the biological function of LRP2.

Besides suffering from severe early-onset obesity, Patient 1 also had pituitary dysfunction characterized by GH deficiency, central hypothyroidism and hypogonadotropic hypogonadism, with concomitant hyperprolactinaemia. It is unclear whether sequence variants in the *LRP2* gene can directly affect pituitary development and function. LRP2 is essential for brain development [25], as knock-out mice exhibit holoprosencephaly [26] and *Lrp2*-mutant mice have abnormal cortical axon development [27]. Humans with DB/FOAR syndrome have structural brain abnormalities, mainly agenesis of the corpus callosum and, in one reported case, empty sella turcica [28]. These support the hypothesis that *LRP2* sequence variants might lead to abnormalities in pituitary development and hypopituitarism.

In addition, Patient 1 had a *de novo* variant in the *UCP2* gene. As a transporter protein that is expressed in the mitochondrial inner membrane, UCP2 decreases mitochondrial ATP production by mediating H^+ leak across the inner membrane [17]. Patient 1 has a G to A substitution at rs660339, resulting in an Ala55Val substitution, which has been associated with obesity in diverse settings [29,30]. Moreover, polymorphisms at rs660339 may also affect metabolic efficiency in terms of energy expenditure [31]. It is noteworthy to mention that this *de novo* variant has a reported minor allele frequency of about 0.5% and that it is a site that mutates recurrently.

The other *de novo* sequence variants that we found in Patient 2 (AICDA, Activation-Induced Cytidine Deaminase; and FAM71E2, Family with Sequence Similarity 71, Member E2) are protein coding genes for a RNA-editing deaminase and for a protein of unknown function, respectively. The variant rs5796316 in AICDA has been reported in one previous study. Whereas AICDA is possibly not implicated in the pathophysiology of obesity, the function of FAM71E2 is unknown to date. However, rs4252574 in FAM71E2 is quite common, and the variant A allele is the major allele reported at about 70%.

In Patient 2, we did not find gene variants that would strongly explain the phenotype, as we did for Patient 1. However, we identified sequence variants harboured in the *DNAAF1* that are possibly implicated in the patient's history of recurrent infections and severe obesity. That gene is responsible for encoding a protein that is cilium-specific and is required for the stability of the ciliary architecture. *DNAFF1* is one of the 21 genes in which mutations are associated with primary ciliary dyskinesia (PCD) [32]. It is unlikely that Patient 2 has PCD, given the absence of its clinical manifestations (bronchiectasis, defects in body situs and, later in life, infertility). However, it is possible that identified *DNAAF1* sequence variants are causing a milder form of PCD with recurrent airway infections and severe obesity. Ciliary dysfunction has been associated with severe early-onset obesity, and currently, it is known that two obesity syndromes are caused by mutations in genes regulating ciliary function: Bardet–Biedl syndrome and Alström syndrome [19]. It has been demonstrated that ciliary dysfunction leads to the development of obesity in animal models, due to diverse alterations in central and peripheral pathways regulating energy metabolism [18,33–38]. For Patient 2, electron microscopy would be useful in the assessment of the effect of the DNAAF1 gene variant on ciliary structure. Furthermore, as we have not confirmed the DNAAF1 variant by Sanger sequencing, we cannot confirm that it is in *trans* in the patient (*i.e.*, each parent contributing one of the variants).

It is unlikely that the other gene variants with potential damaging effect (Table 3) also play a role in the pathogenesis of obesity, since they are not related to the regulation of the leptin-melanocortin pathway. Particularly, *TTN* is a large gene, the variants of which are frequently unrelated to disease; and *OR8U8-OR9G1,* the variants of whichmost likely represent artefacts, possibly due to alignment problems.

By performing MetaCore pathway analysis of the genes harbouring functional mutations, we observed that six pathways are significantly overrepresented, all of them involving energy or ATP/ITP/CTP/UCP metabolism. The importance of UCP2 on energy metabolism was further strengthened by the observation that four of these pathways were centred on the *UCP2* gene.

In whole-genome association studies (GWAS), the *LRP2* and the *DNAAF1* genes were previously associated with increased BMI in a British population, without reaching a level of significance that is relevant for GWAS [39]. Curiously, a higher significance level for a single nucleotide polymorphism (SNP) within the *LRP2* gene ($p = 8.68 \times 10^{-6}$) was found in a GWAS of patients with anorexia nervosa [40]. Although variants at *UCP2* rs660339 have been associated with increased BMI in Europeans [41], this finding was not replicated in a recent meta-analysis in populations representing four ethnicities [42].

Whole-exome capture and sequencing can be used with family-based phenotype ascertainment strategies (nuclear and extended families) to exploit parent-child transmission and relative-relative sharing/not sharing patterns, as well as with arbitrary strategies of phenotype dichotomization to increase efficiency. In an extreme phenotype study design, individuals who are at both ends of a phenotype distribution are selected for sequencing. It is assumed that alleles contributing to the trait in individuals who are at both ends of the phenotype distribution are enriched, and sequencing even a modest sample size can potentially identify novel candidate alleles. The same consideration is also applicable to additional alternatives aimed to identify *de novo* variants that involve the sequencing of parent-offspring trios in which only the offspring is affected. The clinical use of whole-exome capture and sequencing is promising, as demonstrated in a recent study that evaluated 250 patients with undiagnosed diseases: the success rate in obtaining a genetic diagnosis was as high as 25% in that study [12].

Recently, guidelines for investigating and reporting the causality of sequence variants in human disease have been published, to avoid an acceleration of false-positive reports of causality [43]. In our study, we comply with those guidelines, but we acknowledge that our findings are not sufficient to implicate those gene variants as determinants of the obese phenotype. The significance of our results can be limited due to the fact that the number of probands is very small. However, a similar approach has already been validated and published by other studies, such as the Finding of Rare Disease Genes (FORGE) Canada Consortium [44]. In addition, WES has been applied for the diagnosis of several diseases, as observed in many case reports with very small sample sizes [45]. Furthermore, despite the fact that our results have not been replicated yet in other obese individuals, they are important to raise awareness of the *LRP2* gene as a possible candidate as a novel monogenic cause of obesity. In addition, our results lack confirmation through functional data, which should be pursued in future studies.

Whole-exome capture and sequencing analysis is a time- and resource-intense endeavour. Currently, we employ software that allows rapid selection of any genetic variant according to variant type, novelty (via screening public and private databases) and predicted protein effect. However, linking these results to phenotypic manifestations in a particular person is currently performed by a mixture of manual analysis using a number of additional databases (e.g., Human Genome Mutation Database, Online Mendelian Inheritance in Man (OMIM), PubMed and UCSC, among others). We built on existing analytic tools in order to rapidly detect and annotate genomic variants associated with human disease. We are aware that analytical criteria for filtering need to be flexible and up-to-date; therefore, we undertook a systematic upgrade and iterative processes of database evaluation by considering each filter.

In conclusion, by employing a novel and unique strategy for whole-exome capture and sequencing analysis of two trios comprised of patients with severe early-onset obesity, we have identified sequence variants in the *LRP2* and in the *UCP2* genes that might explain the phenotype of a patient with severe early-onset obesity, central hypothyroidism, hypogonadotropic hypogonadism, growth hormone deficiency and idiopathic hyperprolactinaemia. In addition, we identified a sequence variant in the *DNAAF1* gene that might be implicated in the development of severe obesity associated with ciliary dysfunction. Whereas *de novo* variants in the *UCP2* gene have already been associated with obesity, the role of *LRP2* and *DNAAF1* sequence variants in human obesity needs to be further investigated by functional studies, and the frequency and distribution of those sequence variants need to be evaluated in a larger number of obese individuals.

Acknowledgments

This work was funded by the Australian National University and the John James Foundation.

Author Contributions

Obtained clinical data and described the phenotypes: Margaret C.S. Boguszewski, Cesar L. Boguszewski, Gilberto Paz-Filho. Supervised the exome sequencing procedures and performed bioinformatics analyses: Oscar M. Arcos-Burgos, Gavin A. Huttley, Hardip R. Patel, Angad Johar, Aaron Chuah. Conceptualized the study and experimental design: Julio Licinio, Ma-Li Wong, Oscar M. Arcos-Burgos, Gilberto Paz-Filho. Performed Sanger sequencing and analysed data: Claudio A. Mastronardi. Wrote the manuscript: Margaret C.S. Boguszewski, Claudio A. Mastronardi, Gilberto Paz-Filho, Oscar M. Arcos-Burgos, Gavin A. Huttley, Julio Licinio, Ma-Li Wong.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. WHO. Obesity and Overweight. Available online: http://www.who.int/mediacentre/factsheets/fs311/en/index.html (accessed on 23 July 2014).
2. Malik, V.S.; Willett, W.C.; Hu, F.B. Global obesity: Trends, risk factors and policy implications. *Nat. Rev. Endocrinol.* **2013**, *9*, 13–27.
3. Maes, H.H.; Neale, M.C.; Eaves, L.J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **1997**, *27*, 325–351.
4. El-Sayed Moustafa, J.S.; Froguel, P. From obesity genetics to the future of personalized obesity therapy. *Nat. Rev. Endocrinol.* **2013**, *9*, 402–413.
5. Loos, R.J. Genetic determinants of common obesity and their value in prediction. *Best Pract. Res. Clin. Endocrinol. Metab.* **2012**, *26*, 211–226.
6. Farooqi, S.; O'Rahilly, S. Genetics of obesity in humans. *Endocr. Rev.* **2006**, *27*, 710–718.
7. Asai, M.; Ramachandrappa, S.; Joachim, M.; Shen, Y.; Zhang, R.; Nuthalapati, N.; Ramanathan, V.; Strochlic, D.E.; Ferket, P.; Linhart, K.; *et al.* Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science* **2013**, *341*, 275–278.
8. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for mendelian disease gene discovery. *Nat. Rev. Genet.* **2011**, *12*, 745–755.
9. Liew, W.K.; Ben-Omran, T.; Darras, B.T.; Prabhu, S.P.; de Vivo, D.C.; Vatta, M.; Yang, Y.; Eng, C.M.; Chung, W.K. Clinical application of whole-exome sequencing: A novel autosomal recessive spastic ataxia of charlevoix-saguenay sequence variation in a child with ataxia. *JAMA Neurol.* **2013**, *70*, 788–791.

10. Need, A.C.; Shashi, V.; Hitomi, Y.; Schoch, K.; Shianna, K.V.; McDonald, M.T.; Meisler, M.H.; Goldstein, D.B. Clinical application of exome sequencing in undiagnosed genetic conditions. *J. Med. Genet.* **2012**, *49*, 353–361.
11. Veltman, J.A.; Brunner, H.G. *De novo* mutations in human genetic disease. *Nat. Rev. Genet.* **2012**, *13*, 565–575.
12. Yang, Y.; Muzny, D.M.; Reid, J.G.; Bainbridge, M.N.; Willis, A.; Ward, P.A.; Braxton, A.; Beuten, J.; Xia, F.; Niu, Z.; *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **2013**, *369*, 1502–1511.
13. Gill, R.; Him Cheung, Y.; Shen, Y.; Lanzano, P.; Mirza, N.M.; Ten, S.; Maclaren, N.K.; Motaghedi, R.; Han, J.C.; Yanovski, J.A.; *et al.* Whole-exome sequencing identifies novel lepr mutations in individuals with severe early onset obesity. *Obesity* **2013**, *22*, 576–584.
14. Hendricks, A.E. *Whole Exome Sequencing Case-Control Using 1000 Severe Obesity Cases Identifies Putative New Loci and Replicates Previously Established Loci*; American Society of Human Genetics: Boston, MA, USA, 2013.
15. Scheidecker, S.; Etard, C.; Pierce, N.W.; Geoffroy, V.; Schaefer, E.; Muller, J.; Chennen, K.; Flori, E.; Pelletier, V.; Poch, O.; *et al.* Exome sequencing of bardet-biedl syndrome patient identifies a null mutation in the bbsome subunit bbip1 (bbs18). *J. Med. Genet.* **2014**, *51*, 132–136.
16. Gil, S.Y.; Youn, B.S.; Byun, K.; Huang, H.; Namkoong, C.; Jang, P.G.; Lee, J.Y.; Jo, Y.H.; Kang, G.M.; Kim, H.K.; *et al.* Clusterin and lrp2 are critical components of the hypothalamic feeding regulatory pathway. *Nat. Commun.* **2013**, *4*, 1862.
17. Fleury, C.; Neverova, M.; Collins, S.; Raimbault, S.; Champigny, O.; Levi-Meyrueis, C.; Bouillaud, F.; Seldin, M.F.; Surwit, R.S.; Ricquier, D.; *et al.* Uncoupling protein-2: A novel gene linked to obesity and hyperinsulinemia. *Nat. Genet.* **1997**, *15*, 269–272.
18. Berbari, N.F.; Pasek, R.C.; Malarkey, E.B.; Yazdi, S.M.; McNair, A.D.; Lewis, W.R.; Nagy, T.R.; Kesterson, R.A.; Yoder, B.K. Leptin resistance is a secondary consequence of the obesity in ciliopathy mutant mice. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7796–7801.
19. Mok, C.A.; Heon, E.; Zhen, M. Ciliary dysfunction and obesity. *Clin. Genet.* **2010**, *77*, 18–27.
20. Hama, H.; Saito, A.; Takeda, T.; Tanuma, A.; Xie, Y.; Sato, K.; Kazama, J.J.; Gejyo, F. Evidence indicating that renal tubular metabolism of leptin is mediated by megalin but not by the leptin receptors. *Endocrinology* **2004**, *145*, 3935–3940.
21. Dietrich, M.O.; Spuch, C.; Antequera, D.; Rodal, I.; de Yebenes, J.G.; Molina, J.A.; Bermejo, F.; Carro, E. Megalin mediates the transport of leptin across the blood-csf barrier. *Neurobiol. Aging* **2008**, *29*, 902–912.
22. Aronow, B.J.; Lund, S.D.; Brown, T.L.; Harmony, J.A.; Witte, D.P. Apolipoprotein j expression at fluid-tissue interfaces: Potential role in barrier cytoprotection. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 725–729.
23. Pober, B.R.; Longoni, M.; Noonan, K.M. A review of donnai-barrow and facio-oculo-acoustico-renal (db/foar) syndrome: Clinical features and differential diagnosis. *Birth Defects Res. Part A Clin. Mol. Teratol.* **2009**, *85*, 76–81.

24. Chiu-Ugalde, J.; Theilig, F.; Behrends, T.; Drebes, J.; Sieland, C.; Subbarayal, P.; Kohrle, J.; Hammes, A.; Schomburg, L.; Schweizer, U. Mutation of megalin leads to urinary loss of selenoprotein p and selenium deficiency in serum, liver, kidneys and brain. *Biochem. J.* **2010**, *431*, 103–111.
25. May, P.; Woldt, E.; Matz, R.L.; Boucher, P. The ldl receptor-related protein (lrp) family: An old family of proteins with new physiological functions. *Ann. Med.* **2007**, *39*, 219–228.
26. Willnow, T.E.; Hilpert, J.; Armstrong, S.A.; Rohlmann, A.; Hammer, R.E.; Burns, D.K.; Herz, J. Defective forebrain development in mice lacking gp330/megalin. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8460–8464.
27. Ha, S.; Stottmann, R.W.; Furley, A.J.; Beier, D.R. A forward genetic screen in mice identifies mutants with abnormal cortical patterning. *Cereb. Cortex* **2013**, doi:10.1093/cercor/bht209.
28. Kantarci, S.; Al-Gazali, L.; Hill, R.S.; Donnai, D.; Black, G.C.; Bieth, E.; Chassaing, N.; Lacombe, D.; Devriendt, K.; Teebi, A.; *et al.* Mutations in lrp2, which encodes the multiligand receptor megalin, cause donnai-barrow and facio-oculo-acoustico-renal syndromes. *Nat. Genet.* **2007**, *39*, 957–959.
29. Chen, H.H.; Lee, W.J.; Wang, W.; Huang, M.T.; Lee, Y.C.; Pan, W.H. Ala55val polymorphism on ucp2 gene predicts greater weight loss in morbidly obese patients undergoing gastric banding. *Obes. Surg.* **2007**, *17*, 926–933.
30. Oktavianthi, S.; Trimarsanto, H.; Febinia, C.A.; Suastika, K.; Saraswati, M.R.; Dwipayana, P.; Arindrarto, W.; Sudoyo, H.; Malik, S.G. Uncoupling protein 2 gene polymorphisms are associated with obesity. *Cardiovasc. Diabetol.* **2012**, *11*, 41.
31. Buemann, B.; Schierning, B.; Toubro, S.; Bibby, B.M.; Sorensen, T.; Dalgaard, L.; Pedersen, O.; Astrup, A. The association between the val/ala-55 polymorphism of the uncoupling protein 2 gene and exercise efficiency. *Int. J. Obes. Relat. Metab. Disord.* **2001**, *25*, 467–471.
32. Knowles, M.R.; Daniels, L.A.; Davis, S.D.; Zariwala, M.A.; Leigh, M.W. Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease. *Am. J. Respir. Crit. Care Med.* **2013**, *188*, 913–922.
33. Bell, C.G.; Meyre, D.; Samson, C.; Boyle, C.; Lecoeur, C.; Tauber, M.; Jouret, B.; Jaquet, D.; Levy-Marchal, C.; Charles, M.A.; *et al.* Association of melanin-concentrating hormone receptor 1 5' polymorphism with early-onset extreme obesity. *Diabetes* **2005**, *54*, 3049–3055.
34. Cota, D.; Proulx, K.; Smith, K.A.; Kozma, S.C.; Thomas, G.; Woods, S.C.; Seeley, R.J. Hypothalamic mtor signaling regulates food intake. *Science* **2006**, *312*, 927–930.
35. Davenport, J.R.; Watts, A.J.; Roper, V.C.; Croyle, M.J.; van Groen, T.; Wyss, J.M.; Nagy, T.R.; Kesterson, R.A.; Yoder, B.K. Disruption of intraflagellar transport in adult mice leads to obesity and slow-onset cystic kidney disease. *Curr. Biol. CB* **2007**, *17*, 1586–1594.
36. Marion, V.; Stoetzel, C.; Schlicht, D.; Messaddeq, N.; Koch, M.; Flori, E.; Danse, J.M.; Mandel, J.L.; Dollfus, H. Transient ciliogenesis involving bardet-biedl syndrome proteins is a fundamental characteristic of adipogenic differentiation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1820–1825.

37. Sen Gupta, P.; Prodromou, N.V.; Chapple, J.P. Can faulty antennae increase adiposity? The link between cilia proteins and obesity. *J. Endocrinol.* **2009**, *203*, 327–336.
38. Szabo, N.E.; Zhao, T.; Cankaya, M.; Theil, T.; Zhou, X.; Alvarez-Bolado, G. Role of neuroepithelial sonic hedgehog in hypothalamic patterning. *J. Neurosci.* **2009**, *29*, 6989–7002.
39. British 1958 birth cohort DNA Collection. Available online: http://www.b58cgene.sgul.ac.uk/ (accessed on 23 July 2014).
40. Wang, K.; Zhang, H.; Bloss, C.S.; Duvvuri, V.; Kaye, W.; Schork, N.J.; Berrettini, W.; Hakonarson, H.; Price Foundation Collaborative Group. A genome-wide association study on common snps and rare cnvs in anorexia nervosa. *Mol. Psychiatry* **2011**, *16*, 949–959.
41. Brondani, L.A.; Assmann, T.S.; de Souza, B.M.; Boucas, A.P.; Canani, L.H.; Crispim, D. Meta-analysis reveals the association of common variants in the uncoupling protein (ucp) 1–3 genes with body mass index variability. *PLoS One* **2014**, *9*, e96411.
42. Tan, L.J.; Zhu, H.; He, H.; Wu, K.H.; Li, J.; Chen, X.D.; Zhang, J.G.; Shen, H.; Tian, Q.; Krousel-Wood, M.; *et al.* Replication of 6 obesity genes in a meta-analysis of genome-wide association studies from diverse ancestries. *PLoS One* **2014**, *9*, e96149.
43. MacArthur, D.G.; Manolio, T.A.; Dimmock, D.P.; Rehm, H.L.; Shendure, J.; Abecasis, G.R.; Adams, D.R.; Altman, R.B.; Antonarakis, S.E.; Ashley, E.A.; *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **2014**, *508*, 469–476.
44. Beaulieu, C.L.; Majewski, J.; Schwartzentruber, J.; Samuels, M.E.; Fernandez, B.A.; Bernier, F.P.; Brudno, M.; Knoppers, B.; Marcadier, J.; Dyment, D.; *et al.* Forge canada consortium: Outcomes of a 2-year national rare-disease gene-discovery project. *Am. J. Hum. Genet.* **2014**, *94*, 809–817.
45. Rabbani, B.; Tekin, M.; Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* **2014**, *59*, 5–15.

Open Access Data Sharing in Genomic Research

Stacey Pereira, Richard A. Gibbs and Amy L. McGuire

Abstract: The current emphasis on broad sharing of human genomic data generated in research in order to maximize utility and public benefit is a significant legacy of the Human Genome Project. Concerns about privacy and discrimination have led to policy responses that restrict access to genomic data as the means for protecting research participants. Our research and experience show, however, that a considerable number of research participants agree to open access sharing of their genomic data when given the choice. General policies that limit access to all genomic data fail to respect the autonomy of these participants and, at the same time, unnecessarily limit the utility of the data. We advocate instead a more balanced approach that allows for individual choice and encourages informed decision making, while protecting against the misuse of genomic data through enhanced legislation.

Reprinted from *Genes*. Cite as: Pereira, S.; Gibbs, R.A.; McGuire, A.L. Open Access Data Sharing in Genomic Research. *Genes* **2014**, *5*, 739-747.

1. Introduction

Last year marked the 10th anniversary of the completion of the Human Genome Project (HGP) [1]. One of the many accomplishments of the HGP was the broad sharing of data generated by genomic research studies in order to maximize the utility of the data and the public benefit of such projects [2]. This helped to create a culture of openness in genomic research that was codified in a joint policy from the National Human Genome Research Institute (NHGRI) and the Department of Energy (DOE) in 1991 [3] that called for the rapid public release of data generated by the HGP and subsequent projects. Additional policies in the following years, both domestic and international, reaffirmed and expanded these recommendations for publicly sharing large-scale DNA sequence data [4–7].

Initially, the means for protecting participants' privacy when these data were shared in open access (publicly accessible) databases rested upon the "de-identification" of the data by stripping them of all recognizable annotation before sharing. DNA has a very high information content, however, and in 2004, Lin *et al.* showed that it is possible to identify single individuals with as few as 30–80 single nucleotide polymorphisms (SNPs) [8,9], prompting new privacy concerns. In 2006, the U.S. National Institutes of Health (NIH) established the Database of Genotypes and Phenotypes (dbGaP) [10], which is a controlled access database, meaning that individual level genetic data are accessible only with approval from a Data Access Committee. The current NIH data sharing policy requires researchers to obtain approval from their institution before sharing genomic data in dbGaP, and provides guidance to institutions on how to review studies to ensure compliance with the policy, particularly with regard to the adequacy of informed consent documents.

In 2008, Homer *et al.* revealed further complications by showing that it was possible to uniquely identify individuals in aggregated data sets [11]. This led to the implementation of additional protections by restricting access to some aggregated data elements in dbGaP and other databases internationally [12]. Further, in early 2013, Gymrek *et al.* demonstrated that it was possible to identify individuals in the open access database of the 1000 Genomes Project by analysis of Y-chromosome short tandem repeats. They compared these data to genetic information available on a recreational genealogy website, and then used that information to link to additional publicly accessible data, such as obituaries and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository, which banks samples from one of the same populations that took part in the 1000 Genomes Project [13]. This paper was the first to show unequivocally that individuals could be uniquely identified without first obtaining a reference sample. In response, the NIH worked with the NIGMS to move age information, which was previously publicly accessible, into the controlled-access part of the database [14].

Each successive policy decision to further restrict access to genomic data has received some pushback, with critics arguing that each was an overreaction and would unnecessarily impede science [12]. Nonetheless, limiting access to increasing amounts of data continues to be the primary policy response to mounting privacy concerns. Arguments against restricted access and for more open data sharing policies must balance the social and scientific benefits of unrestricted access to and use of data, with adequate protection of the rights and interests of individuals who contribute biological specimens and information to research. The almost exclusive focus on restricting access to genomic data as a matter of policy, however, impedes research and fails to respect the autonomy of those who choose to share their information openly. It has been observed that data in controlled access databases are used less frequently than data in open access databases, and as Rodriguez *et al.* [14] remind us, researchers and other custodians have an ethical responsibility not only to minimize the risk of harm to participants, but also to maximize the utility of generated data. These considerations have led some groups to advocate for a more balanced approach that expands options for open access genomic data release [15,16]. Providing research participants the opportunity to allow their data to be shared more broadly is consistent with the principle of respect for autonomy [9], and as we show below, at least among certain populations, there are a considerable number of "information altruists" [17] who would agree, if given the choice.

2. Participant Perspective

Although studies suggest that there is significant public concern about genetic privacy [18,19], in at least one study, the majority (60%) of more than 4600 U.S. adults surveyed reported willingness to participate in genomic research [20]. Likewise, we have found that a substantial number of research participants are even willing to consent to open access release of their genomic data. In a randomized trial of consent with 323 genomic research participants, the majority (84%) agreed to open access data release. Even after being debriefed, educated about all of the consent options, including the option to consent only to the release of data into controlled access databases like dbGaP, or not at all, and surveyed about their perspectives and concerns, the majority (53%) chose to allow their data to be shared in open access databases [21].

We found a similar response from participants in the Texas Cancer Research Biobank (TCRB), which aimed to establish a fully functional open access database incorporating cancer genomes and other participant data. Controlled access data release was a condition of participation in the TCRB, but the informed consent process allowed participants to opt in to broader sharing of their genomic information via open access data release. Of the 194 participants who were offered this choice, 122 (63%) agreed to open access data release.

These studies present an encouraging picture of research participants' altruistic motivations and lend support to the argument that restrictive data sharing policies fail to respect autonomy of participants who would choose to make their data more broadly available. However, they also raise two major challenges that deserve careful consideration: (1) genomic data sharing is a complex concept that can be difficult for participants to understand; and (2) there is a diversity of perspectives about open access data sharing and certain groups may be less willing to share their data publicly.

2.1. Participant Understanding

Autonomous decision-making requires adequate understanding of the options presented. Yet, ensuring adequate understanding is a challenge in all research involving human subjects. Studies suggest that research participants generally have difficulty understanding and remembering basic information described in research-informed consent documents (e.g., the purpose and risks of the research, as well as general concepts related to study design, like randomization) [22,23]. Genomic research and data sharing are complex concepts, so it is not surprising that participants also have difficulty understanding the differences between data sharing options. For example, in the randomized trial of consent mentioned above, a majority (54%) of participants who were surveyed either could not initially recall with whom they had agreed to share their data or did not understand that by agreeing to open access data sharing it meant that their data could be accessed and used by anyone on the internet without restriction [24]. One possible solution is to try to improve understanding with targeted educational interventions, such as brochures or videos. However, studies have shown that efforts to improve understanding have had only limited success, with the most effective intervention being on-on-one education [25,26].

Another approach to ensure participant awareness is to only release data into open access databases when participants can directly exhibit adequate understanding. For example, the Personal Genome Project, which aims to create a publicly available database of genomic and health information with no expectations of privacy, requires participants to correctly answer all questions in an enrollment examination prior to being allowed to participate, although they may retake the examination multiple times [27]. Similarly, in the TCRB, mentioned above, a subset (*n* = 37) of participants who had agreed to open access data sharing took part in an education session that described the difference between controlled and open access data release and the risks and benefits of each in a question and answer format with visual aids as appropriate. After completing the education session, participants were asked to take a survey, one aim of which was to assess understanding. We found that 73% of survey participants demonstrated adequate understanding, which we defined as (1) knowing that they agreed to open access data sharing; (2) knowing who could access data in an open access database; and (3) understanding the risk of discrimination associated with open access data sharing. Only data from those who had demonstrated adequate understanding were eligible for open access data release. We also assessed participants' risk tolerance and decisional conflict. Fifty-four percent of participants reported high risk tolerance, meaning that they (1) were comfortable sharing their genetic and health information with the general public; and (2) would still participate even if they knew someone would identify their genetic data. Using an adapted version of the decisional conflict scale [28], we found that 68% demonstrated low decisional conflict, meaning that they answered all six questions in a manner indicating that they had low decision uncertainty, no pressure from others, and high perceived effective decision making. Fourteen participants (38%) changed their consent and refused open access data sharing at the completion of the education session. Of the 23 participants (62%) who still agreed to open access data release, 19 had adequate understanding and were therefore eligible for participation. Data from those with high risk tolerance and low decisional conflict were prioritized for public release.

There is considerable debate elsewhere concerning the definition of adequate understanding in research and how best to measure it [29]. Some have focused on developing educational interventions, such as those mentioned above, while others have proposed simplified consent documents as a way to improve understanding [30]. In the PGP and the TCRB, extensive measures were taken to assess understanding and to release data only from those who met a predefined threshold of comprehension. This is time consuming and resource intensive and may not be feasible in all genomic studies. Additional research is needed to identify methods of measuring and improving understanding that are not only effective, but are also efficient, especially in the context of genomic research involving open access data sharing. This is particularly important because participants' right to withdraw from the research is necessarily limited by the inability to retrieve data that has been shared publicly. As these studies suggest, however, there is a subset of participants who understand the implications of open access data release and voluntarily agree to it.

2.2. Diverse Viewpoints

It is important to note that the participants in both the randomized trial of consent and the TCRB were primarily quite ill (sometimes with terminal disease), very trusting of their physicians, and highly motivated to participate in research. Even among this group, however, there was diversity of perspectives about open access data sharing. In the randomized trial of consent, for example, Hispanic, unmarried, and more educated participants were all less likely to choose public data release, as were parents who were making decisions about the release of their child's data [21].

Other populations may exhibit even more variation in their perspectives on data sharing. For example, Lemke *et al.* [31] explored public and biobank participants' attitudes toward genomic research and data sharing via focus groups. While different levels of data sharing (*i.e.*, open *versus* controlled access) were not specifically examined in those studies, the investigators found more generally that there was wide variation in views on genomic data sharing, with some study participants more comfortable than others. Similarly, Trinidad *et al.* [32] conducted focus groups with research participants, surrogate decision-makers, and members of a health maintenance organization to investigate perspectives toward data sharing. They also found that perspectives varied, although they report that study participants were generally supportive of genomic data sharing for scientific benefit. In a commentary on conducting research with tribal communities in the U.S., Harding *et al.* [33] argue that special considerations that take into account the populations' perspectives are important when developing data sharing agreements with Native American tribes.

Our focus in this paper is data sharing in the context of the United States. Research participants in other parts of the world may feel differently about their genomic data and whether or not it should be shared for research purposes [34]. Thus, although generally reported as positive, participant perspectives on data sharing vary between populations, as well as among individuals, based on context, clinical circumstances, and personal values and beliefs.

3. Toward a More Balanced Approach

The variation in individual and group preferences for and understanding about genomic data sharing suggests that both mandatory public data release, as well as blanket restriction of access to genomic research data as a matter of policy, are misguided. Regulatory bodies in general tend to address this "heterogeneity problem" by taking the most restrictive and risk averse approach [35], which, in this case, inhibits choice by prohibiting the broader release of data from those who understand and are comfortable with open access sharing. It also reportedly impedes research [14], although studies quantifying the added benefit of open access *versus* controlled access data sharing are required. We advocate instead for a more balanced approach that allows for individual choice, but provides protection to participants by supporting adequate understanding as part of the informed consent process, and by strengthening accountability and protections against the misuse of available data.

Recent accounts demonstrate that some sophisticated patients are exercising their autonomy by sharing data themselves using existing platforms, such as social media, in order to facilitate discovery for rare and serious diseases [36]. If people are to share their own data, it is important that they are aware of the risk of identifiability and understand the challenge of obscuring segments of data in the context of public release [37]. For those whose data are shared within the research context, novel approaches have been suggested to give participants more control over decisions about who can access their data, as well as the ability to continue to manage such choices. For example, a relatively new platform called Reg4All [38] facilitates the sharing of health information in order to find relevant clinical trials, but also gives its users the ability to make finely-tuned choices about who can access their information or contact them. Others have introduced new approaches to consent that allow participants to be more nuanced in their choices, as well as change those choices over time [39], though, arguably, once data are released in an open access manner, there is no way to guarantee their removal from the public domain.

Increased participant engagement and open access data sharing could both be accomplished with modifications to the existing dbGaP model. As it is currently designed, all individual level genomic data in dbGaP is accessible only via controlled access [10]. The NIH could support more broad sharing by creating a publicly accessible segment of dbGaP that includes data from those who agree to open access data release. Participants could also be provided the option for open access data sharing in the informed consent document when agreeing to participate in NIH-funded genomic research. If a participant changes her consent over time, a request could be made to dbGaP to move the relevant data from the open access portion of the database to controlled access.

Regardless of mechanism, if genomic data are made publicly available, then the individuals from whom those data originate ought to be protected against the misuse of that information. One way of providing some protection for these participants could be the use of “click-through” data use agreements. In this model, the person accessing the data would have to read and agree to a list of conditions of use of the data, including agreeing to not attempt to identify the individuals from whom the data came. However, while this may require those accessing the data to recognize that attempting identification would be a violation of the use of the data, such click-through data use agreements are not enforceable, and as such, may not provide adequate protection.

There are existing laws in the United States that provide protection against misuse of genetic information. The vast majority of states have laws that govern the use of genetic information in health insurance and employment [40]. Likewise, the Genetic Information Nondiscrimination Act (GINA) [41], in effect as of 2009, makes it illegal for health insurers and employers with 15 or more employees to discriminate against people based on their genetic information. GINA has both corrective and monetary penalties that vary based on the intention and severity of the infraction. However, it does not protect against genetic discrimination in other types of insurance, such as long-term, disability, and life insurance, or any other realm outside of health insurance and employment. Additionally, some report not feeling fully protected by GINA, leading some to decline acceptance of DNA sequencing in both clinical and research-related contexts for fear of discrimination [42]. In contrast, the Human Tissue Act of the Parliament of the United Kingdom [43], which regulates activities with human bodies and tissues and also provides protection against the use of DNA without consent, is not limited to such contexts, and carries criminal penalties for violations that range from a fine to up to three years in prison. Though criminal law may not be the best approach to discourage the misuse of genetic data in the U.S., stricter penalties and broader protections against misuse of data by any third party may be needed to protect individuals who agree to share their data broadly for the public’s benefit.

4. Conclusions

In the context of research, investigators have a professional obligation to be good stewards of the data with which research participants have entrusted them. In order to fulfill this obligation, we need policies that respect participant autonomy and maximize the utility of the data, alongside strengthened legislation that protects those participants from the misuse of their genomic information. The field has made great progress in the 10 years since the completion of the Human Genome Project. We must find ways to protect participants, yet avoid unneeded hindrances of researchers’ access to genomic information.

Acknowledgments

This study was supported by the Cancer Prevention Research Institute of Texas (CPRIT) RP101353, and the National Human Genome Research Institute (NHGRI) HG004333 and HG003273.

Author Contributions

Stacey Pereira, Richard Gibbs, and Amy McGuire all contributed to writing, editing, and approving the final version of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; *et al.* Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
2. Wellcome trust sanger institude. The human genome project. Available online: https://www.sanger.ac.uk/about/history/hgp/ (accessed on 15 July 2014).
3. National Institutes of Health, Department of Energy. NIH-DOE guidelines for access to mapping and sequencing data and material resources. Available online: http://www.genome.gov/10000925 (accessed on 15 July 2014).
4. National human genome research institute. Reaffirmation and extension of NHGRI rapid data release policies: Large-scale sequencing and other community resource projects. Available online: http://www.genome.gov/10506537 (accessed on 3 July 2014).
5. National human genome research institute. NHGRI policy on release of human genomic sequence data. Available online: http://www.genome.gov/10000910 (accessed on 3 July 2014).
6. Wellcome trust sanger institude. Summary of principles agreed upon at the international strategy meeting on human genome sequencing, bermuda. Available online: http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1 (accessed on July 2014).
7. Wellcome Trust. Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility, Fort Lauderdale. Available online: http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf (accessed on 3 July 2014).
8. Lin, Z.; Owen, A.B.; Altman, R.B. Genomic research and human subject privacy. *Science* **2004**, *305*, 183.
9. McGuire, A.L.; Gibbs, R.A. No longer de-identified. *Science* **2006**, *312*, 370–371.
10. Mailman, M.D.; Feolo, M.; Jin, Y.; Kimura, M.; Tryka, K.; Bagoutdinov, R.; Hao, L.; Kiang, A.; Paschall, J.; Phan, L.; *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **2007**, *39*, 1181–1186.
11. Homer, N.; Szelinger, S.; Redman, M.; Duggan, D.; Tembe, W.; Muehling, J.; Pearson, J.V.; Stephan, D.A.; Nelson, S.F.; Craig, D.W.; *et al.* Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* **2008,** *4*, e1000167.
12. Gilbert, N. Researchers criticize genetic data restrictions. Available online: http://www.nature.com/news/2008/080904/full/news.2008.1083.html (accessed on 3 July 2014).

13. Gymrek, M.; McGuire, A.L.; Golan, D.; Halperin, E.; Erlich, Y. Identifying personal genomes by surname inference. *Science* **2013**, *339*, 321–324.
14. Rodriguez, L.L.; Brooks, L.D.; Greenberg, J.H.; Green, E.D. The Complexities of genomic identifiability. *Science* **2013**, *339*, 275–276.
15. National human genome research institute. Establishing a central resource of data from genome sequencing projects. Available online: http://www.genome.gov/27549169 (accessed on 15 July 2014).
16. Presidential commission for the study of bioethical issues. Privacy and progress in whole genome sequencing. Available online: http://bioethics.gov/node/764 (accessed on 15 July 2014).
17. Kohane, I.S.; Altman, R.B. Health-information altruists—A potentially critical resource. *N. Engl. J. Med.* **2005**, *353*, 2074–2077.
18. Oliver, J.M.; Slashinski, M.J.; Wang, T.; Kelly, P.A.; Hilsenbeck, S.G.; McGuire, A.L. Balancing the risks and benefits of genomic data sharing: Genome research participants' perspectives. *Publ. Health Genet.* **2012**, *15*, 106–114.
19. Williams, S.; Scott, J.; Murphy, J.; Kaufman, D.; Borchelt, R.; Hudson, K. The genetic town hall: Public opinion about research on genes, environment, and health. Available online: http://www.dnapolicy.org/pub.reports.php?action=detail&report_id=27 (accessed on 1 July 2014).
20. Kaufman, D.; Murphy, J.; Scott, J.; Hudson, K. Subjects matter: A survey of public opinions about a large genetic cohort study. *Genet. Med.* **2008**, *10*, 831–839.
21. McGuire, A.L.; Oliver, J.M.; Slashinski, M.J.; Graves, J.L.; Wang, T.; Kelly, P.A.; Fisher, W.; Lau, C.C.; Goss, J.; Okcu, M.; *et al.* To share or not to share: A randomized trial of consent for data sharing in genome research. *Genet. Med.* **2011**, *13*, 948–955.
22. Bergler, J.H.; Pennington, A.C.; Metcalfe, M.; Freis, E.D. Informed consent: How much does the patient understand? *Clin. Pharmacol. Ther.* **1980**, *27*, 435–440.
23. Joffe, S.; Cook, E.F.; Cleary, P.D.; Clark, J.W.; Weeks, J.C. Quality of informed consent in cancer clinical trials: A cross-sectional survey. *Lancet* **2001**, *358*, 1772–1777.
24. Robinson, J.O.; Slashinski, M.J.; Wang, T.; Hilsenbeck, S.G.; McGuire, A.L. Participants' recall and understanding of genomic research and large-scale data sharing. *J. Empir. Res. Hum. Res. Ethics* **2013**, *8*, 42–52.
25. Flory, J.; Emanuel, E. Interventions to improve research participants' understanding in informed consent for research: A systematic review. *JAMA* **2004**, *292*, 1593–1601.
26. Tamariz, L.; Palacio, A.; Robert, M.; Marcus, E.N. Improving the informed consent process for research subjects with low literacy: A systematic review. *J. Gen. Intern. Med.* **2013**, *28*, 121–126.
27. Ball, M.P.; Bobe, J.R.; Chou, M.F.; Clegg, T.; Estep, P.W.; Lunshof, J.E.; Vandewege, W.; Zaranek, A.; Church, G.M. Harvard personal genome project: Lessons from participatory public research. *Genome Med.* **2014**, *6*, 10.
28. O'Connor, A.M. Validation of a decisional conflict scale. *Med. Decis. Mak.* **1995**, *15*, 25–30.
29. Joffe, S.; Cook, E.F.; Cleary, P.D.; Clark, J.W.; Weeks, J.C. Quality of informed consent: A new measure of understanding among research subjects. *J. Natl. Cancer Inst.* **2001**, *93*, 139–147.

30. Beskow, L.M.; Friedman, J.Y.; Hardy, N.C.; Lin, L.; Weinfurt, K.P. Developing a simplified consent form for biobanking. *PLoS One* **2010**, *5*, e13302.
31. Lemke, A.A.; Wolf, W.A.; Hebert-Beirne, J.; Smith, M.E. Public and biobank participant attitudes toward genetic research participation and data sharing. *Publ. Health Genet.* **2010**, *13*, 368–377.
32. Trinidad, S.B.; Fullerton, S.M.; Bares, J.M.; Jarvik, G.P.; Larson, E.B.; Burke, W. Genomic research and wide data sharing: Views of prospective participants. *Genet. Med.* **2010**, *12*, 486–495.
33. Harding, A.; Harper, B.; Stone, D.; O'Neill, C.; Berger, P.; Harris, S.; Donatuto, J. Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environ. Health Perspect.* **2012**, *120*, 6–10.
34. Hart, S.; Muenke, M. Genetics and genomic medicine around the world. *Mol. Genet. Genomic Med.* **2014**, *2*, 1–2.
35. Meyer, M.N. Regulating the production of knowledge: Research risk-benefit analysis and the heterogeneity problem. *Adm. Law Rev.* **2013**, *65*, 237–298.
36. Mnookin, S. One of a kind. Available online: http://www.newyorker.com/magazine/2014/07/21/one-of-a-kind-2 (accessed on 29 July 2014).
37. Nyholt, D.R.; Yu, C.E.; Visscher, P.M. On Jim Watson's APOE status: Genetic information is hard to hide. *Eur. J. Hum. Genet.* **2009**, *17*, 147–149.
38. Reg4ALL$^{\text{Beta}}$. Learn more. Available online: https://www.reg4all.org/more.php (accessed on 29 July 2014).
39. Kaye, J.; Whitley, E.A.; Lund, D.; Morrison, M.; Teare, H.; Melham, K. Dynamic consent: A patient interface for twenty-first century research networks. *Eur. J. Hum. Genet.* **2014**, doi:10.1038/ejhg.2014.71.
40. National conference of state legislatures. Genetics and health insurance state anti-discrimination laws. Available online: http://www.ncsl.org/research/health/genetic-nondiscrimination-in-health-insurance-laws.aspx (accessed on 15 July 2014).
41. Genetic information nondiscrimination act. Available online: https://www.govtrack.us/congress/bills/110/hr493/text# (accessed on 15 July 2014).
42. Peikoff, K. Fearing punishment for bad genes. Available online: http://www.nytimes.com/2014/04/08/science/fearing-punishment-for-bad-genes.html?hp?smid=fbnytimes&WT.z_sma=SC_FPF_20140408&bicmp=AD&bicmlukp=WT.mc_id&bicmst=1388552400000&bicmet=1420088400000&_r=4 (accessed on 8 July 2014).
43. Human Tissue Act 2004. Available online: http://www.legislation.gov.uk/ukpga/2004/30/contents (accessed on 15 July 2014).

A Balanced Look at the Implications of Genomic (and Other "Omics") Testing for Disease Diagnosis and Clinical Care

Scott D. Boyd, Stephen J. Galli, Iris Schrijver, James L. Zehnder, Euan A. Ashley and Jason D. Merker

Abstract: The tremendous increase in DNA sequencing capacity arising from the commercialization of "next generation" instruments has opened the door to innumerable routes of investigation in basic and translational medical science. It enables very large data sets to be gathered, whose interpretation and conversion into useful knowledge is only beginning. A challenge for modern healthcare systems and academic medical centers is to apply these new methods for the diagnosis of disease and the management of patient care without unnecessary delay, but also with appropriate evaluation of the quality of data and interpretation, as well as the clinical value of the insights gained. Most critically, the standards applied for evaluating these new laboratory data and ensuring that the results and their significance are clearly communicated to patients and their caregivers should be at least as rigorous as those applied to other kinds of medical tests. Here, we present an overview of conceptual and practical issues to be considered in planning for the integration of genomic methods or, in principle, any other type of "omics" testing into clinical care.

Reprinted from *Genes*. Cite as: Boyd, S.D.; Galli, S.J.; Schrijver, I.; Zehnder, J.L.; Ashley, E.A.; Merker, J.D. A Balanced Look at the Implications of Genomic (and Other "Omics") Testing for Disease Diagnosis and Clinical Care. *Genes* **2014**, *5*, 748-766.

1. Introduction

Improvements in DNA sequencing technology in the past decade represent one of the most significant technological achievements in recent history, with far-reaching implications for medicine and society. Most human diseases have at least some genetic factors that contribute to their incidence, or course, either related to the germline genome inherited from an individual's parents or the somatic genetic changes that can lead to the development of malignancies. The ability to read the DNA from an individual's cells with next generation sequencing (NGS) should therefore offer insights relevant for medical care. However, there is a significant gap between our current ability to acquire sequence data and the ultimate goal of extracting all of the useful medical genetic knowledge from the sequences. In particular, societal expectations and ethical considerations require that any correlations between DNA sequences and predictions of disease risk, prognosis or optimal treatment choice should be held to higher standards of evidence than those that are typically applied in the peer review process for publication of a research article. In this overview, we initially highlight areas of recent progress and promise in clinical genomic testing (including whole genome sequence analysis, as well as analysis of selected fractions of the genome, such as the protein coding exome, or large panels of genes of clinical interest) and discuss these new approaches in the context of medical laboratory testing and the current regulatory framework governing such tests in the United States. The potential benefits of clinical genomic testing are tremendous, but devising appropriate systems for quality assurance, data sharing and validation, incorporation into clinical trials and cost-benefit analysis of this new diagnostic area will be an ongoing effort in the coming years.

2. The Promise of Genomic Methods

In the past decade, the quantity of DNA sequencing that can be performed per dollar spent has increased by several orders of magnitude, as a result of technological innovations enabling highly parallelized simultaneous sequencing of millions of spatially-separated template molecules, with optical or electronic readout of the sequencing reaction as it occurs [1,2]. With the latest generation of sequencing instruments, the cost for a whole human genome sequence with 30× coverage may approach $1,000 [3]. Although such estimates notably exclude the costs of interpreting the data, it is clear that genome sequencing is now within the range of costs for many other diagnostic methods, such as radiologic imaging studies or full evaluation of tissue biopsies by anatomic pathologists. As a result of painstaking earlier studies of the genetics of inherited diseases and the genetic changes that are found in cancer cells, there are already many known gene mutations whose significance in relation to particular diseases is described. For many of the most well-established variants, there are single-gene diagnostic sequencing tests already available, either from private companies or specialized, typically academic, diagnostic laboratories. The CDC estimates that genetic tests for use in the clinical setting have been developed for approximately 2,000 diseases [4]. The critical difference between current genome sequencing capabilities and these earlier test methods is that a significantly greater amount of data (whether genome, exome or gene panels) now can be gathered as readily and easily as the sequence from a single gene. Many medical centers and companies are hurrying to stake claims as the preferred destination for testing and interpretation of genomic sequence data for clinical purposes, as this methodology has begun to be adopted and standardized.

Already, there are multiple published medical success stories using these methods. Sequencing of genomes or exomes (which includes the protein coding portions of the genome) for the diagnosis of patients with heritable syndromic disorders has resulted in a number of exciting case reports and studies of patients in whom likely causative mutations have been discovered, and in some cases, such findings have guided successful clinical treatment decisions [5–7]. Systematic efforts in the NIH Undiagnosed Diseases Program to apply genomic methods to arrive at diagnoses for patients with unusual or mysterious clinical presentations, particularly for cases where family history suggests a possible genetic cause, have also yielded new causative mutations and discoveries in human biology [8]. Likewise, clinical laboratories have used exome sequencing since 2011 to evaluate patients with suspected genetic disorders and have identified a molecular diagnostic yield of approximately 25% [9].

Sequencing of cancer genomes has revealed many new recurrent mutations that may contribute to the development of particular cancers, and has revealed new candidates for targeted therapies [10]. These new molecular insights into cancer are already beginning to influence the ways that cancers are classified and treated [11]. Similarly, sequencing of fetal DNA from the plasma of a pregnant woman to screen for trisomies 18 and 21 demonstrates improved performance relative to prior aneuploidy screening methods [12]. Studies of viruses, bacteria and other microbes via sequencing of their small genomes is revolutionizing epidemiological tracking of infectious diseases and the rapid detection of new and emerging pathogens [13]. The most challenging area of all, predicting the risks of diseases in healthy human beings, has also show promise in particular disease categories, such as prediction of cancer risk in women having germline mutations in the *BRCA1* or *BRCA2* genes. More broadly, it is likely that only a fraction of potential disease-associated variants have been identified at present, and the cooperative or competitive relationships between the effects of different sequence variants are only beginning to be described [14]. Many common diseases (e.g., diabetes, schizophrenia and autism) do not appear to be the result of simple sequence alterations that could be easily diagnosed by DNA sequencing. In these diseases, there are many genes associated with increased risk, but the conditions may be caused by combinations of these genes acting together in association with environmental factors. In addition, while it is now possible to identify numerous variants in cancer genomes, the biological significance of many variants is unknown, and their annotation is not standardized. The number of clinically actionable variants, at present, is small.

However, even given the above limitations, these advances herald the increasing importance that genome sequencing and related methods are likely to play in the diagnosis and management of diseases across all medical specialties. We are currently in a transition period in which methods initially applied in research settings and limited small clinical studies need to be adapted for application to large numbers of patients. With that transition comes a requirement for increased standardization, reliability and monitoring of experimental steps, as well as agreed-upon standards for data analysis, storage, clinical interpretation and communication with patients and/or their guardians.

3. Research Experiments, Clinical Testing and Genomic Testing

3.1. Research Assays and Methods

The experimental methods used in medical testing typically originate in research laboratories in universities, government or private research institutes or corporations. However, there is a substantial difference between the assay performance characteristics needed for use in the published scientific literature, compared to those used for clinical diagnosis and guiding the treatment of patients. Authors using a new experimental method as part of a published peer-reviewed research study must convince the scientific reviewers selected by the journal that the assay is a valid method of measurement and has been appropriately applied to the research topic in question, but often, reviewers are not experts in all aspects of the experimental methodology and data analysis approaches used. The expectation in a research setting is that efforts by other researchers to replicate the results in question and to use the methods described will eventually test the reported results and reveal any limitations or errors. This approach has been the basis for essentially all scientific advances, despite the fact that very few published papers are entirely free from errors or only partially correct conclusions, and some are entirely false [15].

Recent studies have highlighted common problems with experimental design and statistical analysis that contribute disproportionately to preventable errors in the scientific literature, especially in attempting to identify genetic contributions to diseases [16,17]. These include choices in experimental design that can introduce biases favoring the discovery of apparently significant effects, even in randomized trials, where patient selection, problems with randomization, lack of blinded data analysis and changes in the plan for data analysis once a trial is underway can all have an impact. In case-control or retrospective studies, the potential for mistaken conclusions is even greater. Studies using large data sets are particularly vulnerable to errors from over-fitting a model to the data or insufficiently accounting for multiple hypothesis testing, especially if independent validation data sets are not used to test the robustness of initial conclusions. Other well-known factors that can distort the scientific literature are publication biases in favor of positive results and the competitive social and economic factors that disproportionately reward scientists publishing papers reporting apparently highly novel findings in prominent journals, while imposing minimal penalties for prior publications later found to be partly or completely in error.

Several suggested improvements to the research methodology have been proposed, including advance registration of a wider range of clinical studies, better documentation of experimental protocols, results and data analysis approaches, more consistent involvement of statisticians and other experts in study design and analysis, full transparency and availability of experimental data and computer scripts used in data analysis and greater attention paid to research reproducibility in the professional evaluation of scientists [16]. In practice, an excessively regulated research environment would probably serve to stifle and limit some of the creative and perhaps poorly planned, but ultimately serendipitous efforts in science that can lead to unexpected insights, so a compromise between higher standards for clinical trial research and continued freedom of inquiry (subject to ethical and safety review by institutional review boards) in more exploratory areas of medical science would probably best serve the public and the ultimate goals of funding providers.

3.2. Clinical Tests

Clinical laboratory testing in the United States is regulated and subjected to greater methodological scrutiny than basic or clinical research. All clinical laboratory testing done for purposes of patient care (as opposed to research) must be performed in a CLIA (Clinical Laboratory Improvement Amendments of 1988)-certified laboratory. CLIA established key laboratory quality standards to ensure that test performance consistently meets patient care needs. CLIA certification may be achieved through the Centers for Medicare and Medicaid Services (CMS), which administers CLIA laboratory certification, or through CMS-approved accrediting organizations (e.g., the College of American Pathologists (CAP), or The Joint Commission). CLIA requirements are stratified according to the complexity of testing performed, with genomic testing generally falling into the highest level of complexity. Such high-complexity testing laboratories must meet specified quality standards, including those related to personnel qualifications and responsibilities, proficiency testing, facilities, general laboratory systems and quality management, as well as preanalytic, analytic and postanalytic systems. (CLIA Brochure, ICN #006270, May 2013). Laboratory compliance with CLIA regulations is evaluated by biennial on-site inspections.

Depending on the nature of the tests performed, additional requirements may need to be satisfied for such laboratories, including those of the AABB (formerly, the American Association of Blood Banks) (Bethesda, MD, USA), FDA (U.S. Food and Drug Administration, Silver Spring, MD, USA), ASHI (the American Society for Histocompatibility and Immunogenetics, Mt. Laurel, NJ, USA), FAA (Federal Aviation Administration, Washington, DC, USA) and state agencies. High-complexity testing must be done by or subject to the oversight of laboratory professionals with an advanced degree and with appropriate credentials in laboratory medicine.

New types of laboratory tests, including most genomic tests, are not available in the form of FDA approved/cleared test kits, but rather are typically created within academic or commercial clinical laboratories as laboratory-developed tests (LDTs). Such development follows a strictly prescribed process of test validation prior to clinical implementation, which is when the test becomes orderable [18,19]. Recently, one high-throughput sequencing instrument and reagent kit system has received FDA clearance for use in clinical testing. In a recent position statement, the Association for Molecular Pathology (AMP) has introduced the term, laboratory-developed procedure, which is defined as "a professional service that encompasses and integrates the design, development, validation, verification, and quality systems used in laboratory testing and interpretative reporting in the context of clinical care" and which much more accurately reflects the highly complex nature of molecular laboratory testing, as well as the central contribution of highly trained and qualified laboratory professionals to the patient care process. AMP also concluded that CMS can ensure the effective oversight and validation of most molecular genetic laboratory tests [20].

3.3. Genomic Testing

The nature of genomic testing, in which large data sets of DNA sequences can be conveniently gathered, but where only a fraction of the overall data can be interpreted at present, places these data sets at the interface between research and conventional clinical testing. Interpretation of genome sequence data is not unique in requiring sophisticated understanding, both of the methods used to gather the data, and the medical literature and body of prior investigation, to arrive at accurate conclusions. For example, histologic diagnosis of cancers in modern pathology practice depends on years of training in recognizing the visual characteristics of aberrant cell populations in tissue sections and selecting appropriate confirmatory tests. The history of revised and improved tumor classification systems in modern oncology reflects the increased understanding gained over decades by studying tumors with new experimental methods, revisiting prior data and correlating the features of each cancer with its response to treatment. Genomic sequence data are vast and complex in other ways, in that the data are generated as lists of nucleotide identities that require computational tools and comparison to sequences in reference databases for analysis, before the clinical significance of sequence variants can begin to be assessed.

Despite amazing progress in the past decade, the technologies and analytical approaches for sequencing and interpreting genomes still have significant blind spots, such as the greater difficulty of detecting structural variants including insertions, deletions, inversions, and trinucleotide repeat expansions, compared to single nucleotide variants in sequences [21]. Reference databases of sequence variants also contain many artifactual annotations, such as 'variants' that are actually sequences derived from pseudogenes similar to the gene in question. It is likely to take many years to resolve such ambiguities or errors in prior gene sequencing work and to ensure that all sequence variants can be correctly annotated for medical applications. Judging from earlier work in human genetics, the interpretation of each patient's genome sequence will require a personalized approach that takes into account the population group from which that person is derived. Even for the most extensively studied genes, such as the cystic fibrosis transmembrane conductance regulator (*CFTR*), the databases of mutations and their significance are quite limited, because functional studies have not been performed for most mutations and because mutations in population groups that have not been studied as extensively as European-derived groups are not well characterized [22].

In the near future, we should not necessarily expect that any given patient's genome sequence, particularly if they are currently healthy, will reveal sequence features requiring any sort of response or clinical guidance beyond those that would already be provided by a physician in a routine checkup, such as advice about healthy diet, exercise, vaccinations and safety topics. A recent study of the potential and current limitations of whole genome sequence interpretation for clinical use in 12 healthy individuals highlighted the relatively small effects associated with most known disease-associated sequence variants, but also revealed that each individual had at least one gene variant from the Clinical Pharmacogenomics Implementation Consortium list of variants that can affect responses to drug therapies [14]. In addition, two of the 12 subjects studied received actionable information that could affect their future health, including one subject who underwent prophylactic surgery. Other studies have underlined the relatively small effect sizes of genetic variants for the most common and serious diseases, such as cardiovascular disease, but rare and highly deleterious variants can also be identified, such as those causing familial hypercholesterolemia, which would warrant immediate medical intervention, such as statin therapy [23,24]. Recently, standards of evidence for concluding that gene variants are associated with diseases have been proposed, taking into account the large amount of data gathered in genome or exome studies and the potential for false discoveries associated with such large numbers of observations of a sample; for example, in exome studies, $p = 5 \times 10^{-7}$ has been suggested as one threshold for claiming significance [25,26].

4. Quality Assurance in the Genome Sequencing Era

Molecular diagnostic laboratories have provided innovative testing since the emergence of diagnostic methods applied to DNA or RNA molecules several decades ago. Clinical molecular laboratories are experienced in validating new methods for single gene testing and data analysis. In one sense, genomic testing is "just another" such innovation. However, it could be argued that the scale of data now obtainable represents a challenge for analysis that is not merely incremental, but is qualitatively different, as human beings cannot manually go through the sequences and interpret them visually in a practical amount of time. Instead, computational methods and bioinformatics tools are required to help carry out the analysis. This represents a significant break with prior traditions of medical training in all specialties, where, with rare exceptions, computer science and bioinformatics were not learned by the generations of physicians who are currently in practice and in positions of authority. Physicians currently in training, particularly in laboratory medicine, now have the opportunity to learn and help to develop these new methods before they enter practice, and currently practicing physicians must at least learn how the results of genomic testing should be applied in their patient care decisions [27,28]. In spite of these challenges, the fundamental features required for reliable clinical laboratory tests based on genome sequencing are the same as those for other kinds of tests and are, in our view, compatible with current regulatory frameworks for ensuring the quality of medical laboratory testing. Some of the most critical initiatives underway in this area are as follows:

(a) Guidelines for Clinical NGS Implementation

Initial laboratory guidelines for clinical diagnostic NGS testing have been established and published by several laboratory professional organizations. These include initiatives by professional organizations, such as the Association for Molecular Pathology (AMP) [29] and the American College of Medical Genetics and Genomics (ACMG) [30,31], as well as those by entities, such as the Clinical Laboratory Standards Institute (CLSI) and the Division of Laboratory Science and Standards at the Centers for Disease Control (CDC) [32]. These efforts are expected eventually to result in consistent recommendations for the clinical validation process of NGS testing, as well as for performance metrics and genomic reference materials for clinical use.

(b) Checks and Balances: The College of American Pathologists Checklist for NGS Testing

The College of American Pathologists (CAP), a CMS-approved accrediting organization, has recently developed a new set of checklist requirements that are specific to NGS, which advances greater standardization in clinical NGS testing. CAP checklists are available to subscribing laboratories and cover key aspects of laboratory function: policies, procedures and pre-analytical, analytical and post-analytical aspects of clinical testing. There is a customized checklist for every section of a clinical laboratory, as well as a general checklist that applies to all sections. During a laboratory inspection, CAP inspectors use these checklists in their evaluation process, to assess whether laboratories follow regulations and practice guidelines and operate at a quality level that is worthy of CAP accreditation and CLIA certification. The NGS section of the molecular checklist contains a set of requirements for both the analytical wet bench processes, as well as for the various bio-informatics steps required for data analysis and annotation. Even though these requirements are not rigidly prescriptive, they highlight key points that must be considered when documenting the reliability and usefulness of clinical NGS testing methods. Many medical centers are within the second year or third of carrying out such testing and therefore have undergone inspection of their genomic or next-generation DNA sequencing assays by CAP or other groups that carry out inspections of CLIA-certified laboratories. Feedback from inspectors and participant laboratories will be very useful for identifying the areas in which checklists need to be revised or made more detailed, explicit or prescriptive, as well as for highlighting the more difficult or uncertain areas in genome sequence data gathering, interpretation and reporting. Some evaluation of the thoroughness of inspections in this new area and the knowledge and qualifications of the inspectors selected for this task will also be warranted as part of the laboratory medicine profession's due diligence in incorporating these new testing methods into mainstream clinical testing.

(c) Assay Validation Requirements

The assay validation conducted before a test is offered clinically documents that a test is consistently and accurately detecting what it claims to be able to identify. CLIA regulations (Code of Federal Regulations § 493.1253 (b) (2)) stipulate that certain core analytical characteristics must be assessed and documented. These include accuracy, precision, analytical sensitivity, analytical specificity, reportable range, reference intervals (normal values) and any other performance characteristic required for test performance (e.g., carryover, dilutions and calculations). The same parameters should be applied to NGS testing, which ranges in scope from single genes or mutation panels to genome sequencing. Limitations of sequence library generation and interpretation, including poorer quality analysis of repetitive sequence regions, less reliable detection or inability to detect certain categories of variation (e.g., insertions, deletions, and other structural variants), inadequately covered regions and similar problems should be reflected and noted in descriptions of the testing method. During the validation process, every single step of NGS must be evaluated, including sample library preparation, clonal fragment amplification, sequencing and all steps of the analysis.

A key need for NGS assay development and validation is the availability of well-characterized "gold-standard" reference materials. Fortunately, there are several public efforts and commercial products that are beginning to meet this need. As an example, the National Institute of Standards and Technology (NIST), with the Genome in a Bottle Consortium, has developed well-characterized single genome reference material for SNVs and small insertions and deletions [33]. Continued support of these and related efforts are needed to generate additional reference genomes and other reference materials for additional applications (e.g., somatic variants).

(d) Interpretation and Reporting of NGS Results

The ACMG has previously published recommendations for the interpretation and reporting of sequence variations for heritable disease, and updates to these recommendations that include interpretation and reporting of NGS-derived sequence changes are expected to be released soon [34]. Additional recommendations will likely be required for other applications (e.g., somatic mutation testing in cancer, pharmacogenetic variation). Despite the advances in sequencing technology, many of the key principles of interpretation still apply. SNP databases and disease-related collections of sequence variants are immensely helpful in variant annotation and interpretation, but there are significant issues that prevent them from being reliably used for clinical diagnostic purposes. Many population databases contain individuals that have developed or will develop disease, and many of the disease-specific databases include benign variants. This underscores the importance of centralized efforts to generate clinical-grade variant databases, such as ClinVar [35].

The final formal interpretation of NGS results, their official posting into the patient's medical record and their translation into clinical care by the physicians responsible for doing that requires interdisciplinary collaboration, whereby pathologists, geneticists and other laboratory professionals become even more directly involved with others in the healthcare team, in order to ensure accurate diagnostic information for individual patients in the context of their disease phenotype.

(e) Proficiency Testing for NGS Assays

Apart from the creation of NGS-specific checklist items, the CAP is in the process of developing NGS proficiency testing products, which are expected to become available in the near future. No other NGS proficiency testing is available in the U.S. from CMS or a CMS-approved accrediting organization. However, laboratories are required to participate in proficiency testing at least twice per year, and this requirement is currently met by alternative assessment. The purpose of such proficiency testing is to be a central quality assessment tool that is an integral component of laboratory inspections and regulatory requirements. To this end, laboratories commonly perform a blinded proficiency testing exchange with other laboratories.

Assessing the Utility of Genomic Information in Clinical Patient Care

We anticipate that NGS technologies will continuously improve in their ability to detect sequence changes and will increase their overall accuracy and ease of use. The increasing capabilities and enhancements to these instruments will facilitate the clinical use of genomic data. The information that is returned to the patient, however, reaches further than the technical and interpretational aspects alone and includes the perceived and, therefore, subjective value of the information. For individual patients, therefore, there is the aspect of personal usefulness (value from the patient's perspective), as well as clinical utility, which constitutes the net health benefits or the balance of benefit *versus* harm. Clinical utility is a complex metric that includes a variety of aspects, such as the patient population tested, the clinical manifestations of the finding and the rationale for testing. Currently, some clinical questions that are explored with NGS can be addressed with considerable confidence, whereas others reach beyond established knowledge. This includes an assessment of the pathogenicity of some variants detected by NGS. Full disclosure of the level of confidence in the clinical meaning of confirmed results, careful patient selection and informed consenting, with a clear understanding of the context in which NGS testing is sought, and genetic counseling before and after testing are important quality measures for the clinical use of NGS testing.

5. Physician Education and Training

It seems certain that the incorporation of genomic methods into clinical patient care will significantly change some aspects of the practice of medicine, affecting not only those who are performing and interpreting genomic testing, but virtually all healthcare professionals. Efforts to integrate genomic approaches, or approaches to analyze transcriptomes, proteomes, metabolomes, microbiomes, *etc*., into clinical care need to be paralleled by the education of our medical students, residents, clinical fellows and faculty to provide a fund of knowledge and an understanding of the possibilities, strengths and limitations of these approaches when they are translated from research to the bedside. This can be accomplished through core efforts in medical school curriculum design, residency and fellowship training programs, as well as in the form of informal and formal continuing medical education (CME) for practicing physicians. Virtually every medical specialty will need to incorporate genomics aspects, as they pertain to that specialty, into their education. A concerted educational effort in medical schools will be critical to ensure the appropriate application of genomic testing, and resources are beginning to become available from professional and medical specialty organizations (for example, the Training Residents in Genomics Working Group, [36]).

Recently, several pathology residency programs have introduced curriculum changes to include more genomic medicine teaching [27,28,37]. Admittedly, in the face of all the other knowledge that residents must acquire in the course of their training, these initial efforts will not produce pathologists who are equally expert in bioinformatics, histology and the wide scope of other laboratory testing methods, but they do represent a first step toward systematic training in the interpretation of large DNA sequence data sets.

6. Ethical and Privacy Considerations

Quality of Patient Information and Informed Consent

Informed consent is the keystone of the ethical treatment of patients in clinical care settings. Patients must have the opportunity to learn about the benefits and limitations that may be associated with genomic testing and consider whether they wish to proceed with such tests. Especially because genomic methods (and other types of "omics" testing) can, in principle, provide magnitudes more information than those traditionally derived, the informed consent process can be more challenging. The level of that challenge depends on the scope of testing and is very different for single gene tests compared to the large net that can be cast with methods that determine the entire exome or genome of the patient. In the latter scenario, incidental findings may disclose medical or personal issues that the patient was not aware of and that were not part of the reason for the current care episode, but that may impact overall health and medical care. The process of returning incidental findings, especially when they may have direct medical ramifications, is an area of active discussion in the medical genetics community [38,39].

The ACMG has issued a policy statement emphasizing the need for informed consent and the content of such a process prior to exome or genome sequencing for germline conditions [40]. Additional recommendations by the ACMG addressing incidental findings have been released and updated [39]. The initial release of these recommendations gave rise to controversy about the informed consent process, and these issues are being considered by multiple additional medical specialties. Until community consensus is reached, individual institutions need to consider how to evaluate and communicate incidental findings of known significance, as well as those genome changes that are of, as yet, unknown clinical significance. What is reported, therefore, needs to be established upfront and clearly communicated to and meaningfully discussed with the patient or his or her guardian, so that expectations are accurate. An informed consent process includes patient education and counseling prior to test ordering, with information addressing the results that will be included in the patient’s report, as well as a discussion of the limits of genomic testing and the interpretation of sequence variants. In addition, as with any other testing, patients need to be able to rely on the privacy and confidentiality of their data.

7. Guiding Principles for Clinical Genomic and Other “Omic” Testing

Any medical center or healthcare organization seeking to incorporate genomic or other “omic” testing into its system of patient care would be well served by: (1) ensuring that the testing is performed in a manner that is fully in accord with relevant legal and regulatory requirements and by personnel with the appropriate training and credentials to perform and interpret such testing; and (2) involving representatives of medical specialties, researchers in genetics, statisticians, computer scientists, bioethicists and hospital administrators in the planning, implementation and integration of this new kind of testing into clinical decision-making. Some key principles we recommend for consideration are listed in Box 1 and are described below:

Box 1. Guiding principles for clinical genomic and other “omic” testing:

(a) Clinical laboratory testing is an integral component of patient care and is held to different standards than research testing not used to guide clinical care.

(b) Clinical genomic testing requires extra effort to be dedicated to designing the informed consent and patient education processes.

(c) Education of physicians and other care-givers about genomic testing methods will be critical for appropriate use and maximal patient benefit.

(d) The use of less-extensively validated genomic testing approaches for clinical care ordinarily should progress in a graded manner from use in “innovative care” settings, followed by use in clinical research settings, before being added to “standard” clinical laboratory testing.

(e) Individual and institutional conflicts of interest in clinical genomic testing must be identified and managed.

(f) These guiding principles also apply to efforts to introduce other clinical “omics” testing into clinical care (such as transcriptomes, proteomes, metabolomes and microbiomes).

(g) Clinical genomic and other “omic” data and methodologies should, to the greatest extent possible, be shared openly with the wider medical and research communities, to accelerate the pace of medical discovery and to increase the quality and reproducibility of clinical genomic data analysis.

(a) Clinical Laboratory Testing Is an Integral Component of Patient Care and Is Held to Different Standards than Research Testing Not Used to Guide Clinical Care

Clinical laboratory testing, regardless of the assay methodology or test complexity, is done to guide patient care decisions, including making diagnoses, counseling patients regarding their prognosis or their future risk of developing disease, guiding management of the patient's condition and making recommendations about reproductive or life style choices. Multidisciplinary committees of clinician specialists and clinical laboratory geneticists, guided by recommendations from medical specialty organizations, as well as other sources of information, may be best able to decide which new genomic tests or applications are sufficiently well-supported by evidence in the scientific literature to be adapted for clinical use.

Any implementation of clinical genomic testing must, of course, comply fully with all relevant laws and regulations governing laboratory tests and, in the United States of America, meet the standards of the professional bodies, such as the College of American Pathologists (CAP) and/or ASHI (the American Society for Histocompatibility and Immunogenetics) that, together with The Joint Commission, have been deemed the status to inspect clinical laboratories on behalf of the Centers for Medicare and Medicaid Services (CMS) to ensure that requirements for CLIA certification are met and all medical tests, including genomic tests, are being carried out responsibly.

(b) Clinical Genomic Testing Requires Extra Effort to be Dedicated to Designing the Informed Consent and Patient Education Processes

Patients must be able to obtain sufficient information about the potential value, future implications and limitations of genomic testing so as to be able to give informed consent if they choose to "opt-in" to the use of such tests for their care. Patient education about genomic test results will help to ensure that any subsequent clinical decision-making is carried out as an informed collaborative process between the patient and their physician. In many cases, this may require additional time to be spent by hospital personnel with the patient to ensure that they understand what is measured and what is interpreted from genomic tests. It is likely that the development of additional educational resources for patients will be necessary for this process. As with almost any other clinical interaction with patients, the use of genomic testing should be done only on the basis of a patient decision to "opt-in", rather than as a default pathway from which patients need to "opt-out".

(c) Education of Physicians and Other Care-Givers about Genomic Testing Methods will be Critical for Appropriate Use and Maximal Patient Benefit

Medical centers and healthcare organizations should consider establishing a set of resources, including a service staffed by individuals with training in molecular genetic pathology, medical genetics and genetic counseling, to educate and advise medical personnel about the proper selection of genetic tests and the appropriate interpretation of their results. The need for such a resource has been highlighted by a recent study performed at a large reference laboratory, which documented a strikingly high rate of inappropriate selection of genetic tests (e.g., ordering the incorrect test, ordering tests that were not needed or ordering suboptimal tests given the clinical question being asked). Approximately 26% of all requests for complex genetic testing for heritable disease were changed following review [41]. These misorders result in unnecessary costs to the healthcare system and, more importantly, may result in significant clinical consequences (e.g., failure or delays in receiving necessary testing, receiving incidental or secondary findings that were not requested or desired).

If clinical findings indicate that genomic tests could be helpful for the care of a particular patient, integrated "tumor board"-style meetings are particularly important in evaluating whether genomic sequencing methods should be applied for the care of that individual patient and for discussing the results and implications of the new genomic data for that patient's care, particularly in challenging cases. It is still an open question as to how the cost of physician and other professional time and effort will be compensated for such diagnostic conferences, but the trend toward health system payments based on patient outcomes rather than the sheer volume of clinical work performed in a patient's care may be compatible with such new efforts, if genomic testing contributes significantly to optimal diagnostic and management decisions and the cost-effectiveness of caring for individuals and populations in coming years.

(d) The Use of Less-Extensively Validated Genomic Testing Approaches for Clinical Care Ordinarily Should Progress in a Graded Manner from Use in "Innovative Care" Settings, Followed by Use in Clinical Research Settings, before being Added to "Standard" Clinical Laboratory Testing

It is likely that a range of different approaches or applications of "genomic" testing will continue to be proposed by physician scientists and other medical investigators, spanning a wide range of different kinds of measurements and interpretations, with widely-varying levels of evidence for their actual clinical utility in different clinical contexts. There are preexisting good models for incorporating innovative clinical methods into practice, and these can be applied to the evaluation of genomic tests supported by various levels of prior evidence. For applications of "genomic" technology that measure already well-established genetic variants with clear clinical significance, typically by replacing older Sanger DNA sequencing assays, rapid incorporation into clinically and analytically validated molecular genetic pathology testing in the CLIA-certified clinical laboratories is advisable. The results of such sequence interpretation are applied for clinical decision-making in the same manner as equivalent results obtained using prior testing methods.

When the clinical value of genomic testing is not well established, but few or no other adequate diagnostic testing options exist, then, as with other types of "innovative clinical care" adopted by medical centers, the application of these tools on very limited numbers of patients for specific purposes at the discretion of the clinician can be considered. If genomic testing will be applied systematically on multiple subjects, without established evidence of clinical utility, it should be carried out in the context of a clinical research study with the accompanying human subject protections and regulations associated with this activity. This would, of course, include obtaining informed consent from the patient for the research, following an explicit and detailed discussion of the limitations of the novel test as a basis for making clinical decisions or healthcare recommendations, the kinds of unexpected (and in some cases, unwanted by the patient) test results that could be reported to the patient and their physician and additional confirmatory testing that would be required before making clinical decisions based on the results of the novel test.

As genomic testing and interpretation methods are validated by ongoing clinical research studies and evidence accumulates for the clinical utility of particular approaches in a given clinical context, some testing and interpretation methods would become sufficiently mature to join the list of "standard" laboratory tests that can be ordered for individual patients by clinicians without the additional safeguards and consultations described above. To the extent that patient informed consent permits, the data and interpreted results of genomic tests should be stored in databases that will permit additional research and discovery to proceed and derive additional clinical insights and knowledge from the testing process.

(e) Individual and Institutional Conflicts of Interest in Clinical Genomic Testing must be Identified and Managed

Conflicts of interest are a serious concern for physicians and others responsible for patient care, and the current period of great discovery and commercial interest in clinical genomics has presented opportunities for physician-scientists and others to become involved in the commercialization of new genomic testing or interpretation methods. All physicians and healthcare institutions must be vigilant to ensure that such potential conflicts of interest do not lead to inappropriate decisions about the kinds of testing approaches to pursue or not to pursue. Individual conflicts of interest, where a faculty member or physician has a financial stake in a private company and/or intellectual property related to genomic testing and analysis methods, are similar to those that apply to the use of other medical technologies, pharmaceuticals and devices. Institutional conflicts of interest are those where the healthcare organization or those directing it could influence decisions about which kinds of diagnostic testing would be used for the care of patients, either to encourage the use of particular test methods, instruments, analytical systems or outsourcing to particular genomic testing to particular companies, or else the avoidance of particular tests or companies. Information about any such potential conflicts should be publicly available, as well as scrutinized and managed within the organization, to ensure and document the propriety and ethical behavior of all participants.

(f) These Guiding Principles also Apply to Efforts to Introduce Other Clinical "Omics" Testing into Clinical Care (Such as Transcriptomes, Proteomes, Metabolomes and Microbiomes).

The improvement in NGS-based sequencing methods over the past five years has resulted in dramatic decreases in the cost per base of DNA sequence. Initially, these technologies revolutionized research endeavors, but clinical laboratories rapidly adopted these technologies. Other related complex testing using NGS or other technologies in the research setting shows significant potential clinical utility. These methods and technologies include RNA sequencing (complementary DNA sequencing), proteomics, metabolomics and metagenomics, as well as functional genomic studies. We suggest that these guiding principles for genomic testing can also be applied to the incorporation of other complex clinical laboratory testing into patient care when sufficient evidence is available to support clinical utility.

(g) Clinical Genomic and Other "Omic" Data and Methodologies Should, to the Greatest Extent Possible, be Shared Openly with the Wider Medical and Research Communities, to Accelerate the Pace of Medical Discovery and to Increase the Quality and Reproducibility of Clinical Genomic Data Analysis

With many technological advances, there are opportunities for private enrichment, as well as the creation of new public resources. The balance between these two components can shape the pace of adoption and the ultimate impact of the technology; the history of the development of the Internet and the role of "open source" contributions to it show the key impact that communities with some element of altruism or public spiritedness can have on the success of a technology. The recent U.S. Supreme Court ruling in *AMP v. Myriad Genetics*, *Inc*. [42], which determined that human genes themselves are not able to be patented, and the preceding ruling in *Mayo Collaborative Services v. Prometheus Laboratories, Inc.* [43], which found that correlations between measured analytes and medical interpretations of such data do not qualify as patentable inventions, may have somewhat decreased the likelihood that private companies will attempt to use litigation to deter testing for particular human gene mutations [44]. These rulings may increase the likelihood that individual companies may try to amass, and keep in private hands, human genetic information and clinical interpretations as trade secrets. In spite of this possibility, there is now an opportunity for medical centers and other healthcare institutions to cooperate in sharing data, interpretations and analysis methods, to speed the identification of correlations between genomic sequences and disease risks, prognosis and treatment outcomes. Currently, the initial frameworks for such data sharing and coordination efforts are promising, but medical organizations and, particularly, their patients will benefit from further joint activity in the public domain that advances clinical genomics [45] and that can serve as a counter-balance to siloed, competitive, inward-looking efforts (whether in academic or commercial settings).

8. Conclusions

We have entered a new era of clinical testing, in which genetic data and other types of "omics" data are much more easily obtained, but the challenges of their interpretation will likely continue for many years. A balance between efficient adoption of new genomic tests and careful consideration of the reliability and clinical value of the results derived from genomic sequence data is needed as these methods become more widely disseminated and utilized within healthcare systems. There are key differences between the quality standards for assays used in research experimentation compared to those that must be maintained for clinical testing, and these standards are under active development and refinement by laboratory professional organizations, as well as associations focusing on particular clinical conditions or specialties. We have outlined seven key principles for consideration in implementing clinical genomic testing, encompassing the technical, as well as the human elements that must be engaged and coordinated to enable optimal utilization of this new form of clinical care. Above all, we feel that this period of intense exploration and discovery in human genetics represents a major opportunity for cooperative and transparent work between different areas of laboratory and clinical medicine, for the ultimate benefit of the patients.

Author Contributions

All authors contributed to planning, writing and editing the article.

Conflicts of Interest

Scott D. Boyd has consulted for Immumetrix Inc., regarding DNA sequence analysis of immunological genes. Stephen J. Galli is a member of the Board of Directors of Atossa Genetics Inc. (Seattle, WA, USA), a publicly-traded company that develops breast health products and services, including molecular diagnostic tests. Jason D. Merker: None. Iris Schrijver: None. Euan A. Ashley is co-founder of Personalis Inc. (Menlo Park, CA, USA), a genetic diagnostics company. James L. Zehnder has received research funding from GlaxoSmithKline for genetic analysis of patient response to thrombopoietin agonists.

References

1. Boyd, S.D. Diagnostic applications of high-throughput DNA sequencing. *Annu. Rev. Pathol.* **2013**, *8*, 381–410.
2. Mardis, E.R. A decade’s perspective on DNA sequencing technology. *Nature* **2011**, *470*, 198–203.
3. Sheridan, C. Illumina claims $1,000 genome win. *Nat. Biotechnol.* **2014**, *32*, 115.
4. Centers for disease control and prevention. Aavailable online: http://www.cdc.gov/genomics/gtesting/ (accessed on 16 July 2014).
5. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; *et al*. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35.
6. Worthey, E.A.; Mayer, A.N.; Syverson, G.D.; Helbling, D.; Bonacci, B.B.; Decker, B.; Serpe, J.M.; Dasu, T.; Tschannen, M.R.; Veith, R.L.; Basehore, M.J.; *et al*. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **2011**, *13*, 255–262.
7. Lupski, J.R.; Reid, J.G.; Gonzaga-Jauregui, C.; Rio Deiros, D.; Chen, D.C.; Nazareth, L.; Bainbridge, M.; Dinh, H.; Jing, C.; Wheeler, D.A.; *et al*. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **2010**, *362*, 1181–1191.

8. Gahl, W.A.; Tifft, C.J. The NIH undiagnosed diseases program: Lessons learned. *JAMA* **2011**, *305*, 1904–1905.
9. Yang, Y.; Muzny, D.M.; Reid, J.G.; Bainbridge, M.N.; Willis, A.; Ward, P.A.; Braxton, A.; Beuten, J.; Xia, F.; Niu, Z.; *et al*. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **2013**, *369*, 1502–1511.
10. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120.
11. Verhaak, R.G.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; *et al*. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110.
12. Bianchi, D.W.; Parker, R.L.; Wentworth, J.; Madankumar, R.; Saffer, C.; Das, A.F.; Craig, J.A.; Chudova, D.I.; Devers, P.L.; Jones, K.W.; *et al*. DNA sequencing versus standard prenatal aneuploidy screening. *N. Engl. J. Med.* **2014**, *370*, 799–808.
13. Cruz-Rivera, M.; Forbi, J.C.; Yamasaki, L.H.; Vazquez-Chacon, C.A.; Martinez-Guarneros, A.; Carpio-Pedroza, J.C.; Escobar-Gutierrez, A.; Ruiz-Tovar, K.; Fonseca-Coronado, S.; Vaughan, G. Molecular epidemiology of viral diseases in the era of next generation sequencing. *J. Clin. Virol.* **2013**, *57*, 378–380.
14. Dewey, F.E.; Grove, M.E.; Pan, C.; Goldstein, B.A.; Bernstein, J.A.; Chaib, H.; Merker, J.D.; Goldfeder, R.L.; Enns, G.M.; David, S.P.; *et al*. Clinical interpretation and implications of whole-genome sequencing. *JAMA* **2014**, *311*, 1035–1045.
15. Ioannidis, J.P. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124.
16. Ioannidis, J.P.; Greenland, S.; Hlatky, M.A.; Khoury, M.J.; Macleod, M.R.; Moher, D.; Schulz, K.F.; Tibshirani, R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* **2014**, *383*, 166–175.
17. Ioannidis, J.P.; Tarone, R.; McLaughlin, J.K. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* **2011**, *22*, 450–456.
18. Halling, K.C.; Schrijver, I.; Persons, D.L. Test verification and validation for molecular diagnostic assays. *Arch. Pathol. Lab. Med.* **2012**, *136*, 11–13.
19. Jennings, L.; van Deerlin, V.M.; Gulley, M.L. Recommended principles and practices for validating clinical molecular pathology tests. *Arch. Pathol. Lab. Med.* **2009**, *133*, 743–755.
20. Ferreira-Gonzalez, A.; Emmadi, R.; Day, S.P.; Klees, R.F.; Leib, J.R.; Lyon, E.; Nowak, J.A.; Pratt, V.M.; Williams, M.S.; Klein, R.D. Revisiting oversight and regulation of molecular-based laboratory-developed tests: A position statement of the association for molecular pathology. *J. Mol. Diagn.* **2014**, *16*, 3–6.
21. Vandeweyer, G.; Kooy, R.F. Detection and interpretation of genomic structural variation in health and disease. *Expert Rev. Mol. Diagn.* **2013**, *13*, 61–82.
22. Rohlfs, E.M.; Zhou, Z.; Heim, R.A.; Nagan, N.; Rosenblum, L.S.; Flynn, K.; Scholl, T.; Akmaev, V.R.; Sirko-Osadsa, D.A.; Allitto, B.A.; *et al*. Cystic fibrosis carrier testing in an ethnically diverse US population. *Clin. Chem.* **2011**, *57*, 841–848.

23. Paynter, N.P.; Chasman, D.I.; Pare, G.; Buring, J.E.; Cook, N.R.; Miletich, J.P.; Ridker, P.M. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* **2010**, *303*, 631–637.
24. Palomaki, G.E.; Melillo, S.; Neveux, L.; Douglas, M.P.; Dotson, W.D.; Janssens, A.C.; Balkite, E.A.; Bradley, L.A. Use of genomic profiling to assess risk for cardiovascular disease and identify individualized prevention strategies—A targeted evidence-based review. *Genet. Med.* **2010**, *12*, 772–784.
25. Do, R.; Kathiresan, S.; Abecasis, G.R. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum. Mol. Genet.* **2012**, *21*, R1–R9.
26. MacArthur, D.G.; Manolio, T.A.; Dimmock, D.P.; Rehm, H.L.; Shendure, J.; Abecasis, G.R.; Adams, D.R.; Altman, R.B.; Antonarakis, S.E.; Ashley, E.A.; *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **2014**, *508*, 469–476.
27. Schrijver, I.; Natkunam, Y.; Galli, S.; Boyd, S.D. Integration of genomic medicine into pathology residency training: The stanford open curriculum. *J. Mol. Diagn.* **2013**, *15*, 141–148.
28. Haspel, R.L.; Olsen, R.J.; Berry, A.; Hill, C.E.; Pfeifer, J.D.; Schrijver, I.; Kaul, K.L. Progress and potential: Training in genomic pathology. *Arch. Pathol. Lab. Med.* **2014**, *138*, 498–504.
29. Schrijver, I.; Aziz, N.; Farkas, D.H.; Furtado, M.; Gonzalez, A.F.; Greiner, T.C.; Grody, W.W.; Hambuch, T.; Kalman, L.; Kant, J.A.; *et al.* Opportunities and challenges associated with clinical diagnostic genome sequencing: A report of the association for molecular pathology. *J. Mol. Diagn.* **2012**, *14*, 525–540.
30. Alford, R.L.; Arnos, K.S.; Fox, M.; Lin, J.W.; Palmer, C.G.; Pandya, A.; Rehm, H.L.; Robin, N.H.; Scott, D.A.; Yoshinaga-Itano, C. American college of medical genetics and genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet. Med.* **2014**, *16*, 347–355.
31. Rehm, H.L.; Bale, S.J.; Bayrak-Toydemir, P.; Berg, J.S.; Brown, K.K.; Deignan, J.L.; Friez, M.J.; Funke, B.H.; Hegde, M.R.; Lyon, E. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **2013**, *15*, 733–747.
32. Gargis, A.S.; Kalman, L.; Berry, M.W.; Bick, D.P.; Dimmock, D.P.; Hambuch, T.; Lu, F.; Lyon, E.; Voelkerding, K.V.; Zehnbauer, B.A.; *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* **2012**, *30*, 1033–1036.
33. Zook, J.M.; Chapman, B.; Wang, J.; Mittelman, D.; Hofmann, O.; Hide, W.; Salit, M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **2014**, *32*, 246–251.
34. Richards, C.S.; Bale, S.; Bellissimo, D.B.; Das, S.; Grody, W.W.; Hegde, M.R.; Lyon, E.; Ward, B.E. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet. Med.* **2008**, *10*, 294–300.
35. Landrum, M.J.; Lee, J.M.; Riley, G.R.; Jang, W.; Rubinstein, W.S.; Church, D.M.; Maglott, D.R. Clinvar: Public archive of relationships among sequence variation and human phenotype. *Nucleic. Acids. Res.* **2014**, *42*, D980–985.
36. Training Residents in Genomics Working Group. Avaliable online: http://www.ascp.org/trig (accessed on 14 July 2014).

37. Haspel, R.L.; Arnaout, R.; Briere, L.; Kantarci, S.; Marchand, K.; Tonellato, P.; Connolly, J.; Boguski, M.S.; Saffitz, J.E. A call to action: Training pathology residents in genomics and personalized medicine. *Am. J. Clin. Pathol.* **2010**, *133*, 832–834.
38. Bennette, C.S.; Trinidad, S.B.; Fullerton, S.M.; Patrick, D.; Amendola, L.; Burke, W.; Hisama, F.M.; Jarvik, G.P.; Regier, D.A.; Veenstra, D.L. Return of incidental findings in genomic medicine: Measuring what patients value—Development of an instrument to measure preferences for information from next-generation testing (IMPRINT). *Genet. Med.* **2013**, *15*, 873–881.
39. Green, R.C.; Berg, J.S.; Grody, W.W.; Kalia, S.S.; Korf, B.R.; Martin, C.L.; McGuire, A.L.; Nussbaum, R.L.; O'Daniel, J.M.; Ormond, K.E.; *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **2013**, *15*, 565–574.
40. ACMG Board of Directors. Points to consider for informed consent for genome/exome sequencing. *Genet. Med.* **2013**, *15*, 748–749.
41. Miller, C.E.; Krautscheid, P.; Baldwin, E.E.; Tvrdik, T.; Openshaw, A.S.; Hart, K.; Lagrave, D. Genetic counselor review of genetic test orders in a reference laboratory reduces unnecessary testing. *Am. J. Med. Genet. A* **2014**, *164*, 1094–1101.
42. Association for Molecular Pathology V. Myriad Genetics. 569 US, Supreme Court, 12–398, 2013.
43. Mayo Collaborative Services V. Prometheus Laboratories. Inc., 566 US, Supreme Court, 10–1150, 2012.
44. Klein, R.D. AMP V. Myriad: The supreme court gives a win to personalized medicine. *J. Mol. Diagn.* **2013**, *15*, 731–732.
45. National Research Council (US) Committee on a Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*; National Academies Press: Washington, DC, USA, 2011.

The Revolution in Human Monogenic Disease Mapping

Emma Duncan, Matthew Brown and Eileen M. Shore

Abstract: The successful completion of the Human Genome Project (HGP) was an unprecedented scientific advance that has become an invaluable resource in the search for genes that cause monogenic and common (polygenic) diseases. Prior to the HGP, linkage analysis had successfully mapped many disease genes for monogenic disorders; however, the limitations of this approach were particularly evident for identifying causative genes in rare genetic disorders affecting lifespan and/or reproductive fitness, such as skeletal dysplasias. In this review, we illustrate the challenges of mapping disease genes in such conditions through the ultra-rare disorder fibrodysplasia ossificans progressiva (FOP) and we discuss the advances that are being made through current massively parallel ("next generation") sequencing (MPS) technologies.

Reprinted from *Genes*. Cite as: Duncan, E.; Brown, M.; Shore, E.M. The Revolution in Human Monogenic Disease Mapping. *Genes* **2014**, *5*, 792-803.

1. Introduction

Until very recently, mapping the causative gene for monogenic diseases depended on finding families with demonstrable Mendelian inheritance of the disease, preferably in multiple generations. Linkage approaches in such families were successful in identifying mutations responsible for many of the more frequent monogenic diseases and traits [1]. However, identifying the underlying mutations in rare genetic diseases, especially those associated with low reproductive fitness, late onset diseases, or diseases with early lethality, proved much more challenging.

The Human Genome Project (HGP) set the stage for success in meeting these challenges. The ultra-rare disorder fibrodysplasia ossificans progressiva (FOP), with a frequency of one in two million, is an example of such a success. Linkage in FOP families identified chromosomal regions of interest; Human Genome Project databases then facilitated the identification of candidate genes within the linkage region and permitted the efficient identification of altered DNA sequences. FOP was found to be caused by a recurrent single nucleotide substitution occurring in >95% of patients. Despite the ultimate success in mapping the FOP gene, this discovery took decades of effort to identify families with inheritance of the disease (linkage analysis was eventually accomplished with just five two-generation families) followed by the time-consuming tasks of conducting genome-wide linkage analysis, and subsequent identification and re-sequencing of candidate genes within the linkage intervals. Knowing the genetic mutation in FOP has led quickly to better understanding of the underlying pathology and directed strategies for treatments, neither of which were possible before molecular aetiology was determined.

Building on the technology, computation, and scientific information generated through the HGP, the continued advances in mapping disease genes have been extraordinarily rapid. Faced today with the challenge of identifying a rare gene mutation in a disease like FOP, high-throughput exome and/or whole genome sequencing approaches would identify the genetic mutation rapidly. Moreover, gene identification may require sequencing of only a small number of unrelated cases, small families, or, in some cases, even a single proband. These breakthroughs are leading to an upsurge in disease gene discoveries with their associated benefits [2].

After mapping a disease-causing gene, many challenges remain in understanding additional genetic contributions to disease onset and progression. In FOP, for example, although the main disease characteristics are unique and readily recognized there is significant variability in the age of disease onset, the rate of disease progression, and the severity of disease, likely to arise from as yet un-identified genetic causes. Similar issues exist with many common heritable diseases such as osteogenesis imperfecta and Marfan's syndrome. With further development of genome technologies, the ability to understand phenotypic variability and the participation of genetic modifiers is becoming a reality.

2. Linkage Mapping

2.1. History of Linkage Mapping

At the turn of the twentieth century, Archibald Garrod coined the term "inborn errors of metabolism" to explain the increased incidence of alkaptonuria (and, subsequently, also cystinuria, pentosuria, and albinism) in consanguineous families compared with the general population, and suggested that these conditions were caused by transmissible elements within families [3]. Since this time, the challenge has been to identify these errors.

For most of the twentieth century, linkage mapping was the standard means of identifying the gene/s underlying an inherited disorder. Linkage is the co-segregation of a genetic region with disease phenotype within a family. Markers, such as restriction fragment length polymorphisms, microsatellites, or single nucleotide polymorphisms (SNPs), are genotyped in family members. Markers close to a disease-causing mutation will be co-inherited with the disease-causing mutation, unless separated by a meiotic event—the closer the marker to the disease-causing gene, the less likely it will be separated at meiosis. An area of linkage within a family may thus extend a considerable genetic distance. Whole genome linkage scans, in which approximately 300–400 microsatellite markers are genotyped in family members, usually result in identification of regions of linkage spanning 10–20 cM (on average ~10–20 million DNA bases). Such a region may harbor many hundreds of genes, and fine mapping (by further marker genotyping and/or sequencing of candidate genes) will usually be necessary to identify the exact causative gene.

2.2. Weaknesses of the Linkage Approach

Traditional "parametric" linkage analysis compares the likelihood of the observed transmission of genetic markers in relation to the trait or disease, in the context of a specified model of inheritance. Non-parametric methods not requiring knowledge of mode of inheritance can be used, though are less powerful when the correct mode of inheritance is known. Diseases with late onset of clinical features or with incomplete penetrance are harder to map by linkage due to possible incorrect attribution of disease status among family members. Diseases with significant gene/environment interaction present similar issues, unless all family members are exposed to the same environmental stressors. The ability to map a gene also depends on the number of informative meioses within a pedigree. Thus, large families with many affected individuals—especially distantly-related individuals who will have a high number of meioses and recombination events between them—are the most useful for linkage mapping. Diseases that affect reproductive fitness (such as skeletal dysplasias) are less likely to have such large informative pedigrees. One solution to this problem is to use many families affected by the same condition in order to identify a common linkage region shared by affected persons within each family. This strategy requires that the disease be caused by mutations in the same single common locus in all families, although the causative mutation within this common locus may be unique in each individual family. The approach will, therefore, not be successful for diseases with "phenocopies", in which a clinical phenotype may arise from mutations in more than one gene. (For example, phenocopies might result if mutations in various genes along a biological pathway resulted in a common phenotypic endpoint.) "Pooling" linkage information from families with disparate underlying causes would not be a successful strategy. Extremely rare diseases are, by definition, extremely rare; obtaining sufficient number of families to pool their linkage information will usually require international collaboration. Lastly, novel/spontaneous mutations (those newly occurring within an individual) cannot be mapped by linkage.

2.3. Successes in Linkage Mapping: Monogenic vs. Complex Diseases

Despite these limitations, mapping monogenic diseases by linkage has been quite successful, even for rare diseases, with well over 1000 monogenic (Mendelian) disorders mapped by the turn of the century [1]. Approximately two-thirds of the 400 or so recognized skeletal dysplasias were mapped by linkage or similar approaches by 2010 [4]. In contrast, mapping complex diseases by linkage was much less successful. Complex genetic diseases are typically polygenic in nature: affected individuals have different variants in multiple, but overlapping, sets of genes, with each variant contributing only a small part to the final overall phenotype. Before the era of high throughput microarray genotyping and the advent of genome-wide association studies (GWAS), only a handful of genes had been identified for complex diseases using linkage [5].

3. Mapping the FOP Gene Highlights the Challenges

Fibrodysplasia ossificans progressiva (FOP; MIM 135100) is a severe disorder of progressive and extensive extra-skeletal ossification. Heterotopic ossification in FOP begins in childhood within connective tissues, such as skeletal muscle, tendon, and ligament. Onset of bone formation can be induced by trauma, or may occur spontaneously. Bone formation is episodic, leading to cumulative disability and shortened lifespan [6].

Reproductive fitness is low in FOP, resulting in infrequent inheritance and a population frequency (about one per two million) that reflects the rate of new mutations. When the search for the FOP gene began, only very few cases of familial inheritance of FOP had been reported, with most known cases occurring *de novo* in families (reviewed in [7,8]). Although these few family pedigrees suggested an autosomal dominant mode of inheritance, the paucity of families with transmission of FOP made genome-wide linkage analysis, the state-of-the-art approach at the time, an impractical strategy for gene identification.

Eventually, four small pedigrees with autosomal dominant inheritance of FOP were identified, although some had ambiguous clinical features. An initial genome-wide linkage analysis using 240 microsatellite markers (spaced ≤ 25 cM) was conducted, although it was recognized that there was incomplete/uneven marker coverage across the genome and many markers lacked sufficient informativity [9]. Whilst the initial linkage analysis focused attention on a chromosome 4 locus [9], further analysis revealed additional linked loci on chromosomes 2 and 6. However, subsequent sequencing analysis of numerous candidate genes in the linkage regions revealed no identifiable mutations. The limited information available at this time about the human genome, including gene locations and sequence, made this process much more challenging.

As genome analysis technologies continued to develop and additional families with autosomal dominant transmission of FOP were identified, further genome-wide linkage studies were performed. These used both a higher density SNP marker panel as well as more dense and informative microsatellite panels, which were combined in a single analysis. The international team involved focused their studies on five two-generation families with stringent and unambiguous phenotypic features of FOP in all affected family members (congenital malformation of the great toes and progressive heterotopic ossification in characteristic anatomic patterns). Consistent linkage was then demonstrated with the chromosome 2q23–24 interval in all five families (LOD score 2.3) [8], with no other locus segregating with the disease in all pedigrees. Using better characterized and annotated human DNA sequences generated through the Human Genome Project (HGP), we selected candidate genes within the 16.5 Mb linkage interval for sequencing and mutation identification. The interval contained more than 40 known genes, however, concurrent investigations had identified the BMP signaling pathway as altered in FOP [10–13], and genes in this pathway were given high priority. One such gene was the *Activin A type I receptor* gene (*ACVR1*; OMIM 102576; also known as *Alk2* or *ActRI*), encoding a receptor for bone morphogenetic protein (BMP).

DNA sequence analysis of all *ACVR1* protein-coding exons and splice junctions identified a heterozygous c.617G>A (Arg206His; CGC ≥ CAC) mutation present in all affected members in these FOP families, with the same mutation present in multiple sporadic cases of FOP [8].

Cumulative data over the past eight years shows that FOP is caused by this recurrent single nucleotide substitution in >95% of patients ([14–16]). In exceptional cases of FOP, mainly those whose phenotype varies slightly from the description above, affected individuals have mutations at other amino acid positions in *ACVR1* in the glycine-serine (GS) or protein kinase domains [14,17]. Thus far, all patients clinically diagnosed with the “classical” FOP phenotype have *ACVR1* mutations, and these mutations are fully penetrant.

4. Massive Parallel Sequencing: A New Era

The mapping of rare monogenic disorders has been completely transformed by the advent of massive parallel sequencing (MPS), also known as “next-generation” sequencing. In the last few years, the causative genes for dozens of monogenic disorders have been identified using MPS, and the rate of discovery has been exponential. To illustrate this latter point, we recently published a review of MPS in skeletal dysplasias; at the time of paper submission (April 2013) 26 papers had been published using MPS to identify the causative gene for 22 skeletal dysplasias; by the time of paper acceptance (July 2013) a further ten papers had added another six novel skeletal dysplasia genes to the list [2]. Since then, further causative genes for skeletal dysplasias, as well as a host of other Mendelian disorders, have been identified, and it is likely that many of the remaining unmapped monogenic diseases will prove tractable to mapping by MPS.

4.1. MPS Technologies

Disease gene identification by MPS became possible because of three main developments: the technology of sequencing multiple genetic regions simultaneously; the success of the Human Genome Project in providing a complete and reliable reference genome for comparison with the test sequence data; and the availability of large databases of genomic information from healthy individuals, and increasingly in patients with disease, to assist in assessment of variants observed.

The pivotal technological breakthrough for MPS was the development of technologies and platforms for simultaneous sequencing of multiple regions of fragmented DNA or RNA in a single experiment. It is beyond the scope of this paper to provide a comprehensive discussion of these technologies. However, briefly: DNA is fragmented and common adapters are ligated to the fragment ends. The fragments are subsequently amplified by PCR, followed by sequencing-by-synthesis, the common adapters providing uniform starting templates for both the amplification and sequencing reactions for all fragments (more recent technological developments, so-called “third-generation” technologies, involve sequencing without this step of any preceding amplification, improving both accuracy and speed of MPS). Sequencing-by-synthesis involves addition of bases to a growing strand: as each base is added, a signal is generated and “read” by the software, thus generating the sequence of each fragment. The sequence of each fragment is then mapped against the human genome, allowing identification of genetic variants present in the sequenced individual(s).

Large databases of genetic variation (such as The HapMap project, UK10K, 1000genomes, Human Variome Project, and NCBI dbSNP) are used in interpreting sequence data obtained through MPS: identified variants can be characterized as part of the “normal” variability seen in the population, or as novel or rare variants and thus more likely to be pathogenic. Of note, population genetic variability differs among populations of different ethnicities; the sequence data of an individual should be compared against an ethnically-matched reference genome sequence. Once the sequence data have been generated and compared with the appropriate reference human genome, the data can be analyzed empirically based on the observed inheritance and population prevalence of the condition under examination.

Although MPS was developed for whole-genome sequencing (WGS), targeted sequencing proved more cost-effective and efficient initially. Thus, prior to amplification, a library of fragments containing regions of particular interest may be selected (for example, by using probes that anneal to specific regions of interest, allowing their subsequent identification and isolation for PCR amplification and sequencing). Whole exome sequencing (WES) with capture and sequencing limited to gene exons may be particularly suited for rare Mendelian disorders since prior to the advent of MPS 85% of monogenic diseases were predicted to arise from protein-coding mutations [18]—whether this will hold true as more causative mutations are identified is as of yet unknown.

The power of MPS methodology is illustrated by the many causative genetic mutations identified since its advent, especially since they are frequently mapped through sequencing of remarkably few individuals. Indeed, some disease genes have been identified through sequencing of a single proband [2], although confirmation of pathogenicity requires subsequent validation, such as genetic evidence in other affected individuals and/or functional studies of the candidate gene.

4.2. Mapping Strategies for Monogenic Diseases Using MPS

The experimental design for mapping a monogenic disease by WES does not necessarily require any prior linkage or association data. Rather it depends on the population frequency of disease, the mode of inheritance (including penetrance), and the presence or absence of consanguinity of the affected individuals. These parameters then determine an appropriately parsimonious experimental design—how many and which individuals should be sequenced and what empirical approach should be adopted for analysis of the sequence data. For example, a rare autosomal recessive condition in a non-consanguineous family is likely to arise from compound heterozygosity; identification of two novel (or very rare) damaging variants in a single gene provides strong evidence of likely causality even if only a single affected individual from the family is sequenced. In contrast, mapping an autosomal recessive condition in a consanguineous family is more difficult. In this circumstance, the disease is more likely to arise from homozygous carriage of a novel (or rare) variant carried by both parents. However, a high number of homozygous rare variants would be expected due to consanguinity anyway; determining which of these is most likely pathogenic can be difficult. For an autosomal dominant condition, the most parsimonious experimental design is to sequence distantly-related affected individuals with a large number of meioses (and, by implication, recombination events) separating the affected cases—with n meioses between the individuals, the chance of a dominant variant segregating with affection status by chance is $1/2^n$. It is also possible to map *de novo* dominant conditions by sequencing unrelated individuals and analyzing the data either for a single variant carried by all affected individuals or for unique variants carried in a common gene by each individual [19]. These last analysis strategies depend crucially upon correct clinical phenotyping of the unrelated affected individuals. The inclusion of phenocopies in the analysis would decrease the success of mapping the causative gene—unless the stringent parameters of analysis are relaxed to allow for their possible presence. For example, one can search for a common shared gene amongst only a proportion of affected individuals rather than requiring a variant to be present in the same gene in every sequenced case. An alternative approach includes pathway analysis (searching for variants in a common pathway amongst individuals with a common phenotype). From a clinical viewpoint, large online databases cataloguing observed variants/mutations in patients with common conditions are also useful in identifying likely disease-causing mutation(s) (as examples, the Leiden Open Variation Database and the Genome Medicine Database of Japan).

4.3. MPS Limitations

Like all technologies, MPS has limitations. Good sequencing data depend upon sufficient capture of the causative gene by coverage of sufficient depth of sequencing to call homozygous or heterozygous variants (typically, 10-fold coverage is required for calling a homozygous variant and 15-fold for a heterozygous variant). WES off-the-shelf target platforms vary in their coverage of the "whole exome" [20,21], which may result in a gene of interest failing to be sequenced. For example, we (and others) failed to identify the disease-causing mutation in OI type V despite sequencing several families with the condition; the causative mutation was identified in the 5' UTR of *IFITM5* [22,23], a region not captured with the whole exome capture platform we had employed. Less-than-complete whole exome capture can arise for several additional reasons, including new gene annotation subsequent to platform development and production [24]; a manufacturing decision to target only the main transcript for a gene rather than all known transcripts of a gene; and the technical challenges of capturing GC-rich sequences (which are common in the first exons of many genes [25]). Conversely, there are some regions that amplify excessively: if duplicates are not removed, strand-specific PCR-introduced errors may result, skewing variant allele frequencies with consequent effects on variant detection sensitivity and specificity [26]. Regions of genomic sequence similarity may result in non-specificity of target fragment selection—for example, a probe may anneal not only to the desired target exon but also to an unwanted region of high homology which, when incidentally captured and sequenced, results in apparently novel heterozygous variants at those points of difference between the two selected regions (a phenomenon known as multi-mapping [27]).

However, despite these limitations, faced today with the challenge of identifying a rare gene mutation in a disease like FOP, WES of a small number of FOP patients would likely rapidly reveal the recurrent c.617G>A (R206H) ACVR1 mutation, leading to quick recognition of *ACVR1* as the disease-causing gene. If MPS technologies had been available when the search for the FOP gene began, the answer could have been found in 15 weeks, not 15 years. Certainly, this has proved to be the case in many other skeletal dysplasias where researchers are faced with similar issues of small families afflicted with diseases having a detrimental effect upon reproduction and lifespan [2,4].

4.4. Rare Variants as a Cause of Complex Diseases?

Although we focused in this paper on the use of new mapping approaches for Mendelian disorders, MPS has also been used for mapping rare variants that contribute to the heritability of complex diseases. The contributions of rare variation in loci that also harbor common susceptibility alleles, or in genomic regions without common susceptibility alleles, are still the subject of active research. Whilst many examples exist of rare variants contributing to loci harboring common variant associations, these are few by comparison with the total number of common variant associations identified to date. Indeed, targeted sequencing of 25 loci associated with autoimmune disorders in nearly 25,000 individuals with six autoimmune phenotypes and just over 17,000 controls failed to identify any rare variants contributing significantly to immune-mediated disease susceptibility [28]. Styrkarsdottir *et al.* used whole genome scanning in the Icelandic population to identify a rare variant in *LGR4* associated with both bone mineral density and fracture risk [29]. WES identified mutations in *WNT1* as the cause of autosomal dominant early-onset osteoporosis in some families; however, as mutations in *WNT1* were also identified in consanguineous families with autosomal recessive osteogenesis imperfecta it would perhaps be more correct to regard the families diagnosed with AD osteoporosis as having a subtle form of OI and/or a monogenic skeletal dysplasia rather than the common polygenic disease of osteoporosis [30,31]. Studies that conducted dense rare-variant genotyping, such as Immunochip-based analyses of immune-mediated diseases, have had little success in identifying novel rare variant associations, despite large sample sizes [29,32,33].

5. Conclusions and Challenges

After mapping a disease-causing gene, many challenges remain, and many of these challenges are likely to be met through the resources and technologies that continue to build on the foundation of the Human Genome Project.

One important consideration is in understanding the multiple genetic contributions to disease onset and progression in addition to the primary causative gene in monogenic disorders. In FOP, for example, although the main disease characteristics are unique and readily recognized, variability in the age of disease onset, the rate of disease progression, and the severity of disease can be high, even in the context of the same ACVR1 c.617G>A mutation. Such phenotypic variability is likely influenced by underlying genetic causes. Identification of genetic modifiers that "protect" an individual with an FOP mutation, for example by directing a late onset or less aggressive disease progression, would provide new therapeutic targets and strategies for treating the disease. Similar issues exist for many common heritable diseases, such as osteogenesis imperfecta and Marfan's syndrome. With further development of genome technologies, the ability to understand phenotypic variability and the participation of genetic modifiers is becoming a reality.

The ultimate challenge is to elucidate the functions of the target gene and the consequences of its mutant forms, and, most importantly, the translation of this knowledge to treatments. Identification of the specific mutation in *ACVR1* has clear and important diagnostic value, providing a means to confirm suspicion of FOP based on toe malformations or, in cases of potential inheritance (and when sequencing in early life becomes more commonplace), a means to diagnose the condition before irreversible clinical manifestations occur allowing for early intervention. Identification of the target pathway and the specific mutation mechanism in FOP has opened up therapeutic strategies for this disease. Additionally, although the roles of BMP signaling in a wide range of tissue development and homeostasis functions had been well established and the signaling pathway and its components were well defined prior to identifying the FOP mutation, the roles of ACVR1/ALK2 in these processes were unrecognized and poorly understood. FOP identified ACVR1 as a key regulator of skeletal development and bone formation, providing an important new focus for skeletal biology and regenerative medicine. This has been the case for many skeletal dysplasias mapped to date, in which the causative gene was often not known to be involved in bone prior to its identification [34]. There are many examples of such findings providing important insights into musculoskeletal development and pathology, and many current treatments have been developed based on genetic discoveries—for example the development of anti-sclerostin antibodies based on the discovery that the high bone mass conditions of sclerosteosis and van Buchem's disease arise from mutations in the gene for sclerostin [35,36]. The power of MPS to map disease-associated mutations will thus benefit not only affected individuals and families, but also lead to a dramatic expansion in our understanding of human diseases. This will inform development of new therapies not only for rare monogenic disorders but also for diseases common in the general population.

Acknowledgments

The authors acknowledge the help of Murray Hargrave in editing the manuscript. This work was supported in part by the International Fibrodysplasia Ossificans Progressiva Association, the Center for Research in FOP and Related Disorders, the Cali-Weldon Professorship in FOP (to EMS), and the National Institutes of Health (NIH R01-AR41916).

Author Contributions

Emma Duncan, Matthew Brown and Eileen M. Shore all wrote the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Peltonen, L.; McKusick, V.A. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* **2001**, *291*, 1224–1229.
2. Lazarus, S.; Zankl, A.; Duncan, E.L. Next-generation sequencing: A frameshift in skeletal dysplasia gene discovery. *Osteoporos. Int.* **2014**, *25*, 407–422.
3. Garrod, A.E. The Croonian Lectures on inborn errors of metabolism. *Lancet* **1908**, *172*, 214–220.
4. Warman, M.L.; Cormier-Daire, V.; Hall, C.; Krakow, D.; Lachman, R.; LeMerrer, M.; Mortier, G.; Mundlos, S.; Nishimura, G.; Rimoin, D.L.; *et al.* Nosology and classification of genetic skeletal disorders: 2010 revision. *Am. J. Med. Genet. A* **2011**, *155A*, 943–968.
5. Glazier, A.M.; Nadeau, J.H.; Aitman, T.J. Finding genes that underlie complex traits. *Science* **2002**, *298*, 2345–2349.
6. Shore, E.M.; Kaplan, F.S. Inherited human diseases of heterotopic bone formation. *Nat. Rev. Rheumatol.* **2010**, *6*, 518–527.
7. Shore, E.M.; Feldman, G.J.; Xu, M.; Kaplan, F.S. The genetics of fibrodysplasia ossificans progressiva. *Clin. Rev. Bone Miner. Metab.* **2005**, *3*, 201–204.

8. Shore, E.M.; Xu, M.; Feldman, G.J.; Fenstermacher, D.A.; Cho, T.-J.; Choi, I.H.; Connor, J.M.; Delai, P.; Glaser, D.L.; LeMerrer, M.; *et al.* A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat. Genet.* **2006**, *38*, 525–527.
9. Feldman, G.; Li, M.; Martin, S.; Urbanek, M.; Urtizberea, J.A.; Fardeau, M.; LeMerrer, M.; Connor, J.M.; Triffitt, J.; Smith, R.; *et al.* Fibrodysplasia ossificans progressiva, a heritable disorder of severe heterotopic ossification, maps to human chromosome 4q27–31. *Am. J. Hum. Genet.* **2000**, *66*, 128–135.
10. Gannon, F.H.; Kaplan, F.S.; Olmsted, E.; Finkel, G.C.; Zasloff, M.A.; Shore, E. Bone morphogenetic protein 2/4 in early fibromatous lesions of fibrodysplasia ossificans progressiva. *Hum. Pathol.* **1997**, *28*, 339–343.
11. Virdi, A.S.; Shore, E.M.; Oreffo, R.O.; Li, M.; Connor, J.M.; Smith, R.; Kaplan, F.S.; Triffitt, J.T. Phenotypic and molecular heterogeneity in fibrodysplasia ossificans progressiva. *Calcif. Tissue Int.* **1999**, *65*, 250–255.
12. Fiori, J.L.; Billings, P.C.; Serrano de la Pena, L.S.; Kaplan, F.S.; Shore, E.M. Dysregulation of the BMP-p38 MAPK signaling pathway in cells from patients with fibrodysplasia ossificans progressiva (FOP). *J. Bone Miner. Res.* **2006**, *21*, 902–909.
13. Serrano de la Pena, L.S.; Billings, P.C.; Fiori, J.L.; Ahn, J.; Kaplan, F.S.; Shore, E.M. Fibrodysplasia ossificans progressiva (FOP), a disorder of ectopic osteogenesis, misregulates cell surface expression and trafficking of BMPRIA. *J. Bone Miner. Res.* **2005**, *20*, 1168–1176.
14. Kaplan, F.S.; Xu, M.; Seemann, P.; Connor, J.M.; Glaser, D.L.; Carroll, L.; Delai, P.; Fastnacht-Urban, E.; Forman, S.J.; Gillessen-Kaesbach, G.; *et al.* Classic and atypical fibrodysplasia ossificans progressiva (FOP) phenotypes are caused by mutations in the bone morphogenetic protein (BMP) type I receptor ACVR1. *Hum. Mutat.* **2009**, *30*, 379–390.
15. Zhang, W.; Zhang, K.; Song, L.; Pang, J.; Ma, H.; Shore, E.M.; Kaplan, F.S.; Wang, P. The phenotype and genotype of fibrodysplasia ossificans progressiva in China: A report of 72 cases. *Bone* **2013**, *57*, 386–391.
16. Shore, E.M. Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. Unpublished work, 2014.
17. Culbert, A.L.; Chakkalakal, S.A.; Convente, M.R.; Lounev, V.Y.; Kaplan, F.S.; Shore, E.M. Fibrodysplasia (myositis) ossificans progressiva. In *Genetics of Bone Biology and Skeletal Disease*; Thakker, R.V., Whyte, M.P., Eisman, J.A., Igarashi, T., Eds.; Academic Press: New York, NY, USA, 2013; pp. 375–393.
18. Cooper, D.N.; Krawczak, M.; Antonorakis, S.E. The nature and mechanisms of human gene mutation. In *The Metabolic and Molecular Bases of Inherited Disease*; Scriver, C., Beaudet, A.L., Sly, W.S., Valle, D., Eds.; McGraw-Hill: New York, NY, USA, 1995; pp. 259–291.
19. Zankl, A.; Duncan, E.L.; Leo, P.J.; Clark, G.R.; Glazov, E.A.; Addor, M.C.; Herlin, T.; Kim, C.A.; Leheup, B.P.; McGill, J.; *et al.* Multicentric carpotarsal osteolysis is caused by mutations clustering in the amino-terminal transcriptional activation domain of MAFB. *Am. J. Hum. Genet.* **2012**, *90*, 494–501.

20. McInerney-Leo, A.M.; Marshall, M.S.; Gardiner, B.; Benn, D.E.; McFarlane, J.; Robinson, B.G.; Brown, M.A.; Leo, P.J.; Clifton-Bligh, R.J.; Duncan, E.L. Whole exome sequencing is an efficient and sensitive method for detection of germline mutations in patients with phaeochromcytomas and paragangliomas. *Clin. Endocrinol.* **2014**, *80*, 25–33.
21. McInerney-Leo, A.M.; Marshall, M.S.; Gardiner, B.; Coucke, P.J.; van Laer, L.; Loeys, B.L.; Summers, K.M.; Symoens, S.; West, J.A.; West, M.J.; *et al.* Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. *Bonekey Rep.* **2013**, *2*, 456.
22. Cho, T.J.; Lee, K.E.; Lee, S.K.; Song, S.J.; Kim, K.J.; Jeon, D.; Lee, G.; Kim, H.N.; Lee, H.R.; Eom, H.H.; *et al.* A single recurrent mutation in the 5'-UTR of IFITM5 causes osteogenesis imperfecta type V. *Am. J. Hum. Genet.* **2012**, *91*, 343–348.
23. Semler, O.; Garbes, L.; Keupp, K.; Swan, D.; Zimmermann, K.; Becker, J.; Iden, S.; Wirth, B.; Eysel, P.; Koerber, F.; *et al.* A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. *Am. J. Hum. Genet.* **2012**, *91*, 349–357.
24. Brunham, L.R.; Hayden, M.R. Hunting human disease genes: Lessons from the past, challenges for the future. *Hum. Genet.* **2013**, *132*, 603–617.
25. Hoischen, A.; Gilissen, C.; Arts, P.; Wieskamp, N.; van der Vliet, W.; Vermeer, S.; Steehouwer, M.; de Vries, P.; Meijer, R.; Seiqueros, J.; *et al.* Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum. Mutat.* **2010**, *31*, 494–499.
26. Koboldt, D.C.; Ding, L.; Mardis, E.R.; Wilson, R.K. Challenges of sequencing human genomes. *Brief Bioinform.* **2010**, *11*, 484–498.
27. Hashimoto, T.; de Hoon, M.J.; Grimmond, S.M.; Daub, C.O.; Hayashizaki, Y.; Faulkner, G.J. Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRescueLite. *Bioinformatics* **2009**, *25*, 2613–2614.
28. Hunt, K.A.; Mistry, V.; Bockett, N.A.; Ahmad, T.; Ban, M.; Barker, J.N.; Barrett, J.C.; Blackburn, H.; Brand, O.; Burren, O.; *et al.* Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* **2013**, *498*, 232–235.
29. Styrkarsdottir, U.; Thorleifsson, G.; Sulem, P.; Gudbjartsson, D.F.; Sigurdsson, A.; Jonasdottir, A.; Jonasdottir, A.; Oddsson, A.; Helgason, A.; Magnusson, O.Y.; *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **2013**, *497*, 517–520.
30. Laine, C.M.; Joeng, K.S.; Campeau, P.M.; Kiviranta, R.; Tarkkonen, K.; Grover, M.; Lu, J.T.; Pekkinen, M.; Wessman, M.; Heino, T.J.; *et al.* WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. *N. Engl. J. Med.* **2013**, *368*, 1809–1816.
31. Pyott, S.M.; Tran, T.T.; Leistritz, D.F.; Pepin, M.G.; Mendelsohn, N.J.; Temme, R.T.; Fernandez, B.A.; Elsayed, S.M.; Elsobky, E.; Verma, I.; *et al.* WNT1 mutations in families affected by moderately severe and progressive recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* **2013**, *92*, 590–597.

32. Parkes, M.; Cortes, A.; van Heel, D.A.; Brown, M.A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **201**3, *14*, 661–673.
33. Keupp, K.; Beleggia, F.; Kayserili, H.; Barnes, A.M.; Steiner, M.; Semler, O.; Fischer, B.; Yigit, G.; Janda, C.Y.; Becker, J.; *et al.* Mutations in WNT1 cause different forms of bone fragility. *Am. J. Hum. Genet.* **2013**, *92*, 565–574.
34. Glazov, E.A.; Beleggia, F.; Kayserili, H.; Barnes, A.M.; Steiner, M.; Semler, O.; Fischer, B.; Yigit, G.; Janda, C.Y.; Becker, J.; *et al.* Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet.* **2011**, *7*, e1002027.
35. Balemans, W.; Patel, N.; Ebeling, M.; van Hul, E.; Wuyts, W.; Lacza, C.; Dioszegi, M.; Dikkers, F.G.; Hildering, P.; Willems, P.J.; *et al.* Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.* **2002**, *39*, 91–97.
36. Van Hul, W.; Balemans, W.; van Hul, E.; Dikkers, F.G.; Obee, H.; Stokroos, R.J.; Hildering, P.; Vanhoenacker, F.; van Camp, G.; Willems, P.J. Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12–q21. *Am. J. Hum. Genet.* **1998**, *62*, 391–399.

DNA Methylation Biomarkers: Cancer and Beyond

Thomas Mikeska and Jeffrey M. Craig

Abstract: Biomarkers are naturally-occurring characteristics by which a particular pathological process or disease can be identified or monitored. They can reflect past environmental exposures, predict disease onset or course, or determine a patient's response to therapy. Epigenetic changes are such characteristics, with most epigenetic biomarkers discovered to date based on the epigenetic mark of DNA methylation. Many tissue types are suitable for the discovery of DNA methylation biomarkers including cell-based samples such as blood and tumor material and cell-free DNA samples such as plasma. DNA methylation biomarkers with diagnostic, prognostic and predictive power are already in clinical trials or in a clinical setting for cancer. Outside cancer, strong evidence that complex disease originates in early life is opening up exciting new avenues for the detection of DNA methylation biomarkers for adverse early life environment and for estimation of future disease risk. However, there are a number of limitations to overcome before such biomarkers reach the clinic. Nevertheless, DNA methylation biomarkers have great potential to contribute to personalized medicine throughout life. We review the current state of play for DNA methylation biomarkers, discuss the barriers that must be crossed on the way to implementation in a clinical setting, and predict their future use for human disease.

Reprinted from *Genes*. Cite as: Mikeska, T.; Craig, J.M. DNA Methylation Biomarkers: Cancer and Beyond. *Genes* **2014**, *5*, 821-864.

1. Introduction

A biomarker is any biological characteristic that can be objectively measured and evaluated as an indicator of normal biological process, pathogenic process, or pharmacological response to a therapeutic intervention [1]. Biomarkers can be used at any stage of a disease and can be associated with its cause or latency (risk biomarkers), onset (diagnostic biomarkers), clinical course (prognostic biomarkers), or response to treatment (predictive biomarkers) ([2–4] and references therein). Biomarkers can also be associated with specific environments (exposure biomarkers). As almost all complex human diseases are caused by a mixture of genetic and environmental variation, biomarkers, especially those antecedent to disease, can be influenced by either of these factors. Biomarkers can also reflect the mechanisms by which exposure and disease are related. They can stratify individuals according to risk or prognosis and they can be used as targets or surrogate endpoints in clinical trials. An ideal biomarker must be able to provide clinically-relevant information, be accurately measurable in multiple individuals, ideally across multiple populations [2,4]. In this review we focus on DNA methylation biomarkers, review the current state of the field, and discuss limitations and our expectations for the future.

2. Epigenetics and Disease Latency

Epigenetics refers to the molecular marks that influence gene function in a mitotically-heritable manner [5]. Epigenetic marks are themselves influenced by a mix of genetic and environmental variation [6]. A typical gene will be regulated by epigenetic marks present at one or more gene promoters, which are usually but not exclusively close to its transcriptional start site, and by one or more enhancers, which can be within the gene or a large distance away from the gene [7]. Such regions of transcriptional control exhibit molecular characteristics in the form of multiple, synergistic epigenetic marks.

Epigenetic marks include methylation of DNA at the cytosine residue of cytosine-phosphate-guanine (CpG) dinucleotides and covalent modifications of amino acid residues within histone proteins that are responsible for the primary packaging of DNA. Other cellular components, such as those involved in writing, reading, and erasing epigenetic marks, determine the local chromatin structure, which at two extremes can be open and active or closed and inactive [8].

In the human genome, DNA methylation occurs almost exclusively at CpG dinucleotides. The cytosine residue of a CpG dinucleotide can be covalently modified by adding a methyl group to its carbon-5 atom resulting in 5-methylcytosine. The methyl group is transferred from *S*-adenosyl-L-methionine to a cytosine residue via DNA methyltransferases (reviewed in [9,10]). CpG dinucleotides are unevenly distributed throughout the genome and are generally methylated [11]. Some CpG dinucleotides are clustered in regions known as CpG-islands, which can span hundreds to thousands of base pairs and are generally unmethylated [11].

The definition of a CpG island has been quite arbitrary and two algorithms have found widespread use throughout the scientific community to identify CpG-islands in genomic DNA sequences [12,13]. However, genome-wide studies have vastly increased our understanding of the human genome over the last few years, and more sophisticated algorithms for the identification of CpG-islands have been developed [14–16].

CpG islands are often, but not exclusively, located at gene promoters, where the methylation status is generally correlated with transcriptional gene activity [11]. DNA methylation can have other (regulatory) functions outside promoter regions, for example in intragenic regions [17,18], intergenic regions [19] and in regions of low CpG density [20]. DNA methylation performs a regulatory role at local and global levels. Global methylation is mainly determined by methylated CpG dinucleotides in highly repeated DNA sequences, such as satellite DNAs, which play an important function in maintaining genome stability [21]. DNA methylation level changes, namely local hypermethylation (gain of DNA methylation) and global hypomethylation (loss of DNA methylation), are often associated with a diseased state.

Most studies of the role of epigenetics in human disease have focused on investigating disease-associated DNA methylation changes and on determining the environmental influence on DNA methylation variations. Most of these have focused on cancer. It is now widely accepted that cancer results from a combination of genetic and epigenetic disruption or dysfunction (reviewed in [22]). Whereas the underlying causes of cancer remains largely elusive, it has also become clear that certain environmental factors such as the exposure to certain chemicals, toxins or heavy metals are capable of altering the epigenome and ultimately increase the risk of developing cancer [23–25].

Outside cancer, environmental influences on DNA methylation are the centre of the developmental origins of health and disease (DOHaD) phenomenon [26,27]. In this phenomenon, which grew out of the "fetal origins" hypothesis [28], adverse environment, *in utero* or in early postnatal life, programs the body for complex, non-communicable diseases including diabetes, cardiovascular disease (CVD) and neurodevelopmental disorders. Central to this phenomenon is the hypothesis that disease predisposition results when postnatal environment is mismatched to prenatal environment [29].

The DOHaD phenomenon involves a period of disease latency between the early origins and the later clinical manifestation. This latency may be in the form of a few years, for example with obesity and autism, or many decades, in the case of CVD. Non-epigenetic biomarkers of latent conditions such as CVD are already being developed and these include plasma high sensitivity C-reactive peptide, blood pressure, body mass index and artery wall thickness [30,31]. We discuss below how epigenetic biomarkers, in particular DNA methylation biomarkers, are being identified within the context of cancer and DOHaD.

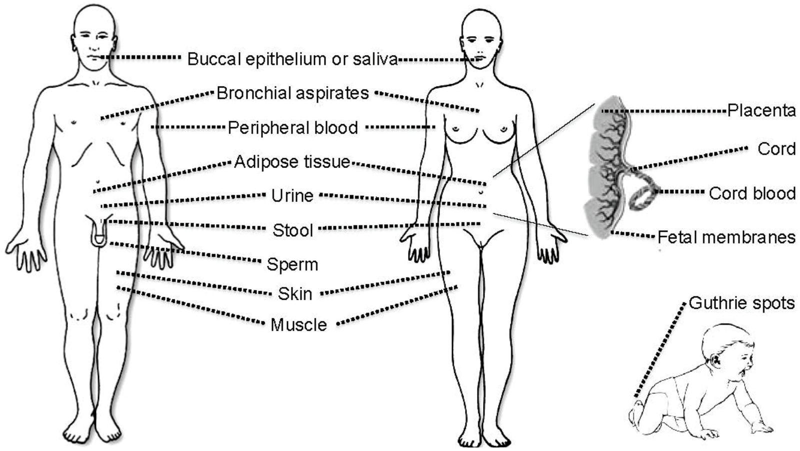
3. Tissues and Bodily Fluids Suitable for Analysis of DNA Methylation Biomarkers

Almost any biological tissue sample or bodily fluid can be used for DNA methylation analysis. DNA methylation is the most robust epigenetic mark and will survive most sample storage conditions including, in the case of Guthrie neonatal blood spots, long-term drying [32]. DNA methylation can also be studied in histological specimens such as formalin-fixed paraffin-embedded (FFPE) samples [33] and microscopic preparations [34]. The robustness of DNA methylation marks makes DNA methylation analysis very attractive in a clinical environment as the analysis of gene expression pattern and histone modifications require more careful storage conditions, either with an RNA-preserving agent, by snap-freezing, or by cryopreservation of viable cells. In most cancers, (primary) tumor biopsies can be sampled but for the early detection of cancer and most other non-communicable diseases, only peripheral, easy-to-access tissues or bodily fluids can be collected. Such samples include venous peripheral blood, buccal epithelium or saliva, urine, stools, bronchial aspirates, and, even in some cases, muscle or adipose tissue [35–39] (Figure 1). At birth, placenta, umbilical cords and fetal membranes are also suitable tissues for analysis of DNA methylation [40–42].

It is important to note that even though it is desirable to measure disease-associated methylation biomarkers in a disease-relevant tissue, this condition does not always need to be met if a methylation biomarker is tightly associated with disease state. This is especially the case for tissues such as the brain and heart that can only be sampled *post mortem*.

Cellular homogeneity within a tissue is also a desirable characteristic for a DNA methylation biomarker [43]. Tissues such as blood or even blood fractions such as mononuclear cells, exhibit cellular heterogeneity [44–46]. However, methods have been developed to control for such heterogeneity, using either differential cell counts [47] or *post hoc* in regression models [48–50].

Figure 1. Illustration of the variety of tissues that can be used to investigate DNA methylation biomarkers. Note that tumor tissue is not listed individually as a cancer can affect any part of the body.



4. Parameters for Developing DNA Methylation Biomarkers

Before we go into more detail about specific DNA methylation biomarkers, we will review the measures of particular importance for assay performance and the barriers that must be breached in developing DNA methylation biomarkers. The nomenclature we use in this review is generally already in use, although it has not been previously summarized in such a way. It is as follows: single studies provide *potential biomarkers*, which could be *validated* using an independent technique and *replicated* in an independent cohort, also known as external validation. Following the systematic review and/or meta-analysis of a large number of independent studies, they become *candidate clinical biomarkers* that can enter clinical trials. Once approved, they become *proven clinical biomarkers* (Table 1).

Table 1. Nomenclature used in this review for the stages of DNA methylation biomarker development.

Nomenclature	Description
Potential biomarker	Results of a single study
Validated biomarker	Same finding using an independent method
Replicated biomarker	Same finding in independent cohort(s)
Candidate clinical biomarker	Replicated in multiple cohorts and subjected to systematic review and meta-analysis; most likely undergoing clinical trials
Proven clinical biomarker	Used in clinical practice

4.1. Methods for DNA Methylation Biomarker Discovery

Genome-wide profiling of DNA methylation patterns of healthy and diseased individuals has enabled the identification of potential methylation biomarkers for many diseases, most prominently in cancer but also other diseases such as metabolic or neurodevelopmental disorders. Following initial studies using pre-selected candidate gene approaches [51–53], many different genome-wide methods have been developed and used in the scientific community for DNA methylation biomarker discovery and good overviews are provided elsewhere [54–58]. Other scientific publications review such methods in the context of methylome-wide association studies (MWAS), which utilize a variety of platforms [59,60]. Typically, MWAS, as a subset of epigenome-wide association studies (EWAS), involves regression of DNA methylation at thousands to millions of CpG dinucleotides or CpG-rich regions on disease phenotype, outcomes or interventions. Such analyses usually adjust for multiple testing to produce potential methylation biomarkers in the form of differentially-methylated probes (DMPs) or regions (DMRs). Often, DMPs or DMRs are validated using locus-specific methods. The next stages of discovery following replication involve longitudinal analysis to resolve the issue of cause *vs.* effect in MWAS, and importantly to show whether replicated biomarkers can be used to predict a disease before its clinical onset or predict clinical outcomes after onset or after therapeutic intervention. Following discovery of such replicated biomarkers, further replication followed by meta-analysis and/or systematic review are required, at which stage these candidate clinical methylation biomarkers are ready for clinical trials leading to clinical proven methylation biomarkers. In this review we will focus on single locus DNA methylation biomarkers at all stages of discovery.

4.2. DNA Methylation Assay Sensitivity and Specificity

Assay sensitivity describes the proportion of patients with disease who have a positive test result (true positive rate), whereas the assay specificity describes the proportion of patients without disease who have a negative test result (true negative result) [61]. The ideal assay would show 100% sensitivity and 100% specificity. In other words, the test is never positive for a disease-free patient and never negative for a patient with disease. However, this ideal scenario is rarely achieved. It is also important to note that an assay with a sensitivity of 50% and a specificity of 50% is no better than tossing a coin to decide if the patient is harboring the disease or is disease-free [61].

The receiver operating characteristic (ROC) curve is a fundamental tool for diagnostic test or biomarker evaluation and visually displays the interdependency of specificity and sensitivity [62,63]. In a ROC curve the true positive rate (sensitivity; y-axis) is plotted in function of the false positive rate (1-specificity; x-axis). The area under the curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. In other words, for a well performing diagnostic test or biomarker the curve is located towards the upper left corner. On the other hand a less well-performing diagnostic test or biomarker is characterized by a curve close to a diagonal line, representing a state in which sensitivity and specificity are similar.

It is desirable to achieve values for sensitivity and specificity as high as possible. However, for some tests it might be acceptable to achieve a higher sensitivity by sacrificing assay specificity or *vice versa*. This could be the case in particular for diseases for which a misclassification would result in severe consequences for the patient [61]. Acceptable values for sensitivity and specificity of a testing procedure can be determined by comparing to existing values of a test currently considered as gold standard. It is also important to consider that a diagnostic test is providing information independent of the experience of a clinician, which sometimes varies dramatically among hospitals and countries. However, it remains to be determined how easily the new testing procedure can be implemented in a clinical environment.

4.3. Barriers to Developing, Testing and Using DNA Methylation Biomarkers

Despite the promise of epigenetic biomarkers, so far only a few DNA methylation-based candidate biomarkers have reached the potential for use in a clinical setting, and all these are mainly related to the field of cancer. As with disease phenotypes, each clinical DNA methylation biomarker would need to be measured accurately and reproducibly. Differences in DNA methylation between cases and controls may be large (e.g., more than 50%) in cancer but in other non-communicable diseases may often be less than 5%. Methods used to measure methylation must be accurate to well below this level of resolution. The analytical sensitivity of specific methods is discussed below. Next, variability within the population needs to be small to maximize assay sensitivity and specificity. Predictive power also needs to be high. Positive predictive power is the percentage of people with a positive test who actually get the disease. These hurdles are all similar to those for any clinical trial.

5. Methods Suitable for the Analysis of Locus-Specific DNA Methylation Biomarkers

Many different methods have been described for the investigation of locus-specific DNA methylation (reviewed in [58,64–67]). Whereas some methods use genomic DNA for methylation analysis, the majority of methods require bisulfite-treated DNA as starting material [68,69]. Bisulfite treatment converts unmethylated cytosines to uracil, whereas 5-methylcytosines are relatively inert under reaction conditions. Subsequent use of bisulfite-treated DNA in PCR replaces the uracils with thymines and 5-methylcytosines with cytosines. Therefore, the methylation status of a particular CpG dinucleotide is detected indirectly [70].

The use of bisulfite-treated DNA has three important consequences for downstream applications for DNA methylation detection. Firstly, a considerable loss of initial input DNA can occur, due to extensive DNA degradation during the preparation and purification of bisulfite-treated DNA [71–74]. Loss of amplifiable DNA can be critical in particular for those samples where only a limited amount of genomic DNA is available, such as those from very small biopsies. Secondly, a poor bisulfite conversion rate can lead to false-positive results. This is of particular importance for very sensitive DNA methylation detection methods, such as those based on methylation-specific PCR (MSP) [75]. However, the use of a commercially-available bisulfite conversion kit can help to improve DNA recovery and to control for a proper bisulfite conversion rate [67,72]. Thirdly, PCR amplification may sometimes be biased towards unmethylated or methylated templates due to differences in CG content [76]. However, different approaches have been described in the literature to overcome or at least to minimize a potential PCR amplification bias [77–80].

Another problem for most downstream applications is the presence of heterogeneous DNA methylation patterns at many gene loci [81–83]. Heterogeneous methylation patterns are characterized by the presence of multiple epialleles (alleles which differ in the pattern of methylated and unmethylated CpG dinucleotides across the analyzed region). As every sample has its own set of epialleles, it can complicate quantification of methylation (reviewed in [84]) and cut-off value settings for when to call a sample unmethylated or methylated. The need for cut-off values also demands the use of quantitative DNA methylation detection methods, in particular for those gene loci, which are hypomethylated (loss of DNA methylation), or where already variable background methylation is present in healthy individuals [81,85].

Despite the many methodologies available for DNA methylation analysis the methodological considerations and requirements of a molecular diagnostics laboratory renders only a fraction of these methods suitable for DNA methylation analysis in a clinical setting. Such methods would need to use small quantities of DNA of varying quality. The latter is of particular importance for formalin-fixed paraffin-embedded (FFPE) specimens where the DNA is often degraded and chemically modified [86]. Ideally, DNA methylation detection methods for clinical settings should be low cost, easy to use, automatable, and capable of processing many samples in parallel in order to minimize costs of future tests. In the following sections we will discuss methods for DNA methylation detection suitable for use in clinical settings or in a molecular diagnostic laboratory.

Bisulfite pyrosequencing (Qiagen, Hilden, Germany) is based on sequencing-by-synthesis methodology and uses bisulfite-treated DNA as starting material [87–89]. This method is relatively cost- and time-effective, and is suitable for DNA methylation analysis of single gene loci. DNA methylation can be determined at single CpG dinucleotide resolution but methylation levels are provided in a quantitative manner for each CpG site as an average across all epialleles amplified during PCR. The analytical sensitivity is about 5%–10% for individual CpG dinucleotides [90,91]. This approach has a high-throughput capacity and is well suited for the analysis of small PCR amplicons, such as those typically generated from FFPE specimens. Importantly, this approach allows to quality control for a sufficient bisulfite conversion rate. However, the downside of this approach is that the instrument required to perform DNA methylation analysis is rather costly.

The MassARRAY EpiTYPER (Sequenom Inc., San Diego, CA, USA) also requires bisulfite-treated DNA as starting material and uses matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry to extract (semi-) quantitative DNA methylation information from shifts and intensities of fragment signals after base-specific cleavage of PCR amplified epialleles present at single gene loci [92]. DNA methylation levels are determined as an average for a single CpG dinucleotide, or for multiple CpG dinucleotides, clusters of CpGs on the same fragment or for multiple CpGs across all fragments of amplicons generated during PCR [93]. Nevertheless, this approach is suitable for providing an almost complete methylation profile across the region-of-interest [92]. The analytical sensitivity is similar to bisulfite pyrosequencing [93] and DNA methylation data obtained by both methods for the same set of CpG dinucleotides has been shown to be highly concordant [81]. Like bisulfite pyrosequencing, MassARRAY EpiTYPER is suitable for high sample throughput and also requires the purchase of an expensive instrument.

Methylation-sensitive high-resolution melting (MS-HRM) is an inexpensive, fast and medium to high throughput screening methodology for DNA methylation analysis at single gene loci [94,95]. This approach requires bisulfite-treated DNA as starting material and exploits the differential melting behavior of PCR products generated from unmethylated and methylated epialleles. The melting profile of an unknown sample is compared to melting profiles of a DNA methylation standard series. This allows the reliably detection of homogeneous methylation levels down to 1%–5%, and can detect the presence of heterogeneous methylation patterns. However, the presence of heterogeneous DNA methylation allows the estimation of methylation levels in a semi-quantitative or qualitative manner. This is because the presence of heterogeneous DNA methylation results in a complex melting profile that does not allow the ready estimation of the amount of methylated epialleles; the result is largely qualitative [84]. MS-HRM is quite attractive in a clinical environment as PCR amplification and subsequent DNA methylation analysis is performed in one tube, which minimizes the risk of sample mix-up and sample cross contamination [96]. However, MS-HRM is not suitable on its own for use in a clinical setting as this method is not capable to deliver quantitative methylation information. Nevertheless, MS-HRM PCR products can be further quantified for DNA methylation using bisulfite pyrosequencing [97].

Another group of important approaches for DNA methylation detection is based on methylation-specific PCR (MSP) [98]. The strength of MSP-based approaches comes from the high analytical sensitivity, which allows them to detect only few methylated epialleles in a large background of unmethylated epialleles. The high analytical sensitivity originates from PCR primers containing CpG dinucleotides that selectively amplify only methylated epialleles. However, conventional MSP is not suitable for use in clinical settings as this approach detects DNA methylation only in a qualitative manner [98]. This can result in an overestimation of methylation in particular for those samples where background methylation is already present in normal tissues [81]. Moreover, conventional MSP is difficult to standardize between different laboratories and is also well known to generate false-positive, as well false-negative results, especially when DNA of low quality is used as starting material, for example FFPE-derived DNA [99,100]. Nevertheless, quantitative offshoots of conventional MSP, such as MethyLight [101], ConLight-MSP [102], MS-FLAG [103], SMART-MSP [104] and HeavyMethyl [105] are potentially suitable approaches for use in a clinical environment. The latter approach has been already successfully applied for DNA methylation detection of *SHOX2* and *PITX2* (see below).

Methylation-sensitive multiplex ligation-dependent probe amplification (MS-MLPA) plays a key role in the diagnosis of genomic imprinting disorders (see below) [106]. Different to the methods described above, this method uses genomic DNA as starting material to produce semi-quantitative DNA methylation information for single CpG dinucleotides. MS-MLPA relies on CpG dinucleotide-specific probes and a digestion step using the methylation-sensitive restriction endonuclease *HhaI* prior to PCR amplification to distinguish unmethylated from methylated epialleles. DNA methylation levels are determined by comparing peak sizes of patient samples with control samples and the analytical sensitivity is approximately 5%–20% [107–110]. MS-MLPA is suitable for high-throughput screening, is relatively cost-effective and does not require non-standard laboratory instruments as the PCR amplification products are separated by capillary electrophoresis on a DNA analyzer instrument.

The use of genomic DNA for methylation analysis is quite attractive as it avoids problems associated with bisulfite treatment. However, as MS-MLPA is based on a digestion step with a methylation-sensitive restriction endonuclease, false-positive results can occur as a result of incomplete digestion, in particular with DNA of poor quality. The use of the restriction endonuclease *HhaI* also limits the investigation of DNA methylation to *HhaI* recognition sites and therefore provides only a limited view of the DNA methylation landscape of any region of interest. However, MS-MLPA is capable of analyzing up to 50 CpG dinucleotides at any one time and allows the determination of DNA methylation levels at different gene loci simultaneously. Moreover, MS-MLPA can be combined with gene copy number and point mutation detection, which makes it a quite flexible methodology [110].

Lessons Learned from the DNA Methylation Biomarker MGMT

The DNA repair gene O^6-methylguanine-DNA methyltransferase (*MGMT*) was first characterized in the early 1990s [111,112] and its key role in the resistance of malignant glioma to alkylating drugs was proposed repeatedly [113–115]. Approximately ten years later, a first link was established between *MGMT* methylation and improved patient outcome in response of malignant gliomas to the alkylating drug carmustine [116]. However, the relatively small number of patients investigated as well as some flaws in study design raised concerns of the validity of the results and warranted confirmation of the potential predictive biomarker *MGMT* (see comments in [116,117]).

Subsequently, *MGMT* methylation as a predictive biomarker of a patient’s response to alkylating drug regimens was replicated on different sample cohorts with mixed success. Methylation of *MGMT* was shown to serves as a predictive biomarker for determining response of glioma and glioblastoma patients treated with the alkylating agent Temozolomide [117,118]. Nevertheless, another study was not able to replicate *MGMT* methylation as a predictive biomarker in glioblastoma patients treated with alkylating drug regimens [119].

However, the seminal findings of a clinical trial reported in 2005, conducted by Hegi and colleagues, clearly showed that glioblastoma patients treated with Temozolomide showed a survival benefit if the promoter-associated CpG-island of the *MGMT* gene was methylated [120]. Since then, several clinical trials have confirmed *MGMT* methylation as a candidate clinical biomarker for determining patient response to Temozolomide treatment and it is now a proven clinical biomarker (reviewed in [121]).

Since 2005, many research groups and commercial companies (Table 2) have spent much effort developing assays to investigate the methylation status of *MGMT* by using various methods and platforms [122–126]. However, these methods varied in analytical sensitivity and provided methylation information ranging from purely qualitative to quantitative. As consequence, the general lack of consensus for an agreed methodology and the widespread use of inappropriate methodologies slowed down the implementation of *MGMT* methylation analysis in molecular diagnostics [127].

Table 2. Commercially-available DNA methylation test kits for cancer. References are either systematic reviews/meta-analyses [1] or a set of corroborating references [2]. This table is an updated version of that shown in [127].

Gene(s)	Type of Biomarker	Type of Cancer	Diagnostic Test Kit: Brand Name (Manufacturer)	References
VIM	diagnostic	Colorectal	Cologuard (Exact Sciences)	[128] [1]
SEPT9	diagnostic	Colorectal	Epi proColon (Epigenomics) ColoVantage (Quest Diagnostics) RealTime mS9 (Abbott)	[129] [1]
SHOX2	diagnostic	Lung	Epi prolong (Epigenomics)	[130–135] [2]
GSTP1/*APC*/*RASSF1A*	diagnostic	Prostate	ConfirmMDx (MDx Health)	[136–138] [1]
MGMT	predictive	Glioblastoma	PredictMDx Glioblastoma (MDx Health) SALSA MS-MLPA probemix ME011 Mismatch Repair genes (MRC-Holland) PyroMark MGMT Kit (Qiagen)	[121,139,140] [1]

Several recent studies assessing the clinical utility of different methodologies for *MGMT* methylation detection favor quantitative approaches such as bisulfite pyrosequencing [141,142]. Quantitative approaches are necessary to determine cut-off values for methylation ranges related to clinical information such as prognosis [143,144]. However, methylation cut-off values are not universal for a particular gene and strongly depend on the method used for DNA methylation analysis. Even by using the same methodology for methylation analysis requires determination of cut-off values for each assay as these values also depend on the region of the gene investigated, PCR primers and PCR conditions used as well as minimal tumor content required [143–145]. It has also been recognized that careful studies of the entire *MGMT* promoter-associated CpG-island are required to determine those CpG dinucleotides or CpG clusters suitable as a surrogate biomarker for biological or clinical relevant information [83,146].

Keeping in mind that *MGMT* methylation was one of the first DNA methylation biomarkers to be identified, it is not surprising that it took a considerable amount of time until it found its way into the clinic. Advancements in study and clinical trial design will certainly help to speed up replication and clinical implementation of new DNA methylation biomarker. However, the current lack of an agreed methodology as the gold standard for DNA methylation analysis is still a roadblock to overcome. For a more detailed view on which milestones need to be achieved in bringing a DNA methylation biomarker into clinical practice we refer the interested reader elsewhere [127].

6. DNA Methylation Biomarkers

To date, the vast majority of replicated and candidate clinical DNA methylation biomarkers come from cancer research. Clinically-relevant DNA methylation biomarkers outside cancer exist for diseases originating from genomic imprinting disorders, such as Prader-Willi and Angelman syndrome (see below), and are currently being developed for a wide range of environmental agents and the chronic diseases to which they predispose. The following sections will give an overview of promising DNA methylation biomarkers for potential clinical use.

6.1. Candidate Clinical DNA Methylation Biomarkers for Cancer

A selection of candidate clinical DNA methylation biomarkers for cancer will be discussed below; many others have been described in greater detail elsewhere (e.g., [127,147–161]) or have been subject of systematic reviews and meta-analysis (e.g., [129,136–140,162]). Not surprisingly, much effort has been spent in identifying diagnostic DNA methylation biomarkers suitable for early detection and diagnosis of cancer. Early detection allows treatment of the cancer at a stage that is generally considered beneficial for disease outcome. Such tests could be blood-based or use other bodily fluids collected less invasively, which makes it very convenient to the patient. Prognostic biomarkers would provide information on a patient's overall survival if the disease is left untreated, whereas predictive biomarkers would be suitable for determining a patient's response to a certain drug regimen. The latter are of particular importance as they may help to minimize the health burden of patients, as well as to minimize costs for healthcare providers for unnecessary drug treatment.

DNA methylation-based candidate clinical biomarker genes for the early detection include vimentin (*VIM*) [128,163], septin 9 (*SEPT9*) [129,164], and syndecan 2 (*SDC2*) [165,166] for colorectal cancer, glutathione S-transferase pi 1 (*GSTP1*) for prostate cancer [136,167,168], and cyclin-dependent kinase inhibitor 2A (*CDKN2A*) [169,170] and short stature homeobox 2 (*SHOX2*) (see below) for lung cancer. These have already reached clinical potential and for some diagnostic test kits are commercially-available (Table 2). In the next sections we will provide an overview of *SHOX2*, *PITX2* and *MGMT* as good examples of diagnostic, prognostic and predictive biomarkers in cancer.

6.1.1. *SHOX2*

DNA methylation of the short stature homeobox 2 (*SHOX2*) gene was found to be a diagnostic clinical biomarker candidate for the detection of malignant lung disease even in patients where histology and cytology results are equivocal [135]. *SHOX2* methylation allowed the specific detection of malignant lung disease with a sensitivity of 60% and a specificity of 90% in blood plasma using HeavyMethyl, a quantitative methylation-specific PCR-based approach [134]. The highest assay sensitivity was achieved for small cell lung cancer (SCLC) cases with 80% and squamous cell carcinoma (SCC) with 63%, respectively, when compared to adenocarcinomas (AC) cases with a sensitivity of only 39%. However, the poor sensitivity for detecting adenocarcinomas could be improved by the addition of a second (or more) adenocarcinoma-specific biomarker. Not surprisingly, the sensitivity values obtained of the blood-plasma-based assay were lower compared to sensitivities seen from bronchial aspirates (SCLC: 97% (80%), SCC: 82% (63%), and AC: 47% (39%)); overall sensitivity and specificity were 68% (60%) and 95% (90%), respectively) as the tumor-derived amount of DNA is expected to be lower in blood than a lung-specific analyte [134,135]. However, a blood-based assay has the advantage of using specimens which have been collected with a far less invasive procedure compared to those obtained from bronchoscopy. Furthermore, a blood-based assay enables screening of asymptomatic patients whereas availability of bronchoscopy is limited to patients with suspected lung cancer. Noteworthy, elevated *SHOX2* methylation levels in pleural effusions do not only allow the detection of lung cancer but also the detection of other malignancies, such as breast cancer and gastrointestinal cancers [132,133]. However, assay sensitivity and specificity for these was not as good as for bronchial aspirates or blood. *SHOX2* methylation level in lymph node tissue obtained by endobrochial ultrasound with transbronchial needle aspiration (EBUS-TBNA) improved endoscopic lung cancer staging with an assay sensitivity and specificity of 94% and 99%, respectively [130].

DNA methylation of *SHOX2* not only provides diagnostic but also provides prognostic information for cancer patients [131,132]. Pleural effusion samples obtained from patients with different malignancies (including lung cancer) showed a shorter overall survival if elevated levels of *SHOX2* methylation were detected [132]. Contrarily, gain of *SHOX2* methylation in tumor tissues has been shown to be associated with good prognosis in lung cancer patients. The prognostic power of *SHOX2* methylation was further improved when combined with DNA methylation analysis of *PITX2* [131].

6.1.2. *PITX2*

The paired-like homeodomain 2 (*PITX2*) gene encodes the PITX2 transcription factor. DNA methylation status of the *PITX2* promoter has been identified as a candidate clinical biomarker in tumor tissues. This has provided prognostic information for breast cancer, prostate cancer, and lung cancer. *PITX2* methylation in estrogen receptor alpha positive breast cancer patients without lymph node metastasis has been associated with poor prognosis when treated without any systemic adjuvant therapy [171] as well as a higher risk of disease recurrence after surgery when treated with the antiestrogen Tamoxifen only [172,173]. Furthermore, methylation of the *PITX2* promoter was also associated with poor patient outcome in estrogen receptor alpha positive, HER-2/*neu*-negative breast cancer patients positive for lymph node metastasis when treated with an anthracycline-based adjuvant chemotherapy [174]. Methylation of *PITX2* in prostate cancer patients has also been shown to be a prognostic biomarker for an increased risk of biochemical recurrence after radical prostatectomy [175–177]. Importantly, the prognostic value of *PITX2* methylation was particularly high in tumor-enriched samples of patients at intermediate risk for whom further risk stratification is quite often challenging [176]. Interestingly, and different to breast and prostate cancer, increased DNA methylation levels of *PITX2* were associated with prolonged survival in lung cancer patients and requires further investigation [131].

6.1.3. *MGMT*

O^6-methylguanine DNA methyltransferase is a DNA repair protein that is encoded by the *MGMT* gene and is capable of removing alkyl residues directly from the *O*^{6}-position of guanines. However, if the DNA repair capacity of *MGMT* is impaired or inactivated, for example by DNA methylation, affected cells are less protected against mutagenic DNA adducts [178,179]. Therefore, tumor *MGMT* promoter methylation renders cancer cells susceptible to the cell damaging effects of drug regimens utilizing alkylating agents [116,180] (see also *Lessons learned from the DNA methylation biomarker MGMT*). *MGMT* was one of the first predictive DNA methylation biomarkers to determine a patient's response to alkylating chemotherapeutics and it was shown that glioblastoma patients with tumor *MGMT* promoter methylation have a survival benefit from Temozolomide chemotherapy [120,121].

The more frequent use of quantitative approaches such as bisulfite pyrosequencing to detect and measure *MGMT* methylation have revealed that the DNA methylation biomarker *MGMT* does not only have a predictive but also a prognostic clinical component (reviewed in [121,126]). Glioblastoma patients with more than 29% *MGMT* promoter methylation showed a longer progression-free and overall survival when treated with radiotherapy and Temozolomide [143]. A methylation cut-off value of 25% separated elderly glioblastoma patients into two groups with those having more than 25% of methylation had a better survival rate when treated with alkylating agents alone [144]. Tumor *MGMT* methylation status was also shown to have a prognostic value for progression-free survival of anaplastic glioma patients treated with radiotherapy alone [181,182].

7. DNA Methylation Biomarkers for Genomic Imprinting Disorders

Whereas most genes are expressed from both the maternal and paternal allele, imprinted genes are monoallelically expressed in a parent-of-origin-specific manner either from the maternal or the paternal allele. Only a small proportion of all human genes are imprinted and are often found clustered in imprinted domains and mono-allelic gene expression is controlled by differentially methylated regions (DMRs) (reviewed in [183]). Disrupted or altered imprinting patterns have been linked to pathological conditions termed genomic imprinting disorders (reviewed in [184]). Examples of imprinting disorders include Prader-Willi syndrome (PWS), Angelman syndrome (AS), Beckwith-Wiedemann syndrome (BWS) and Silver-Russell syndrome (SRS), which will be discussed briefly below.

PWS and AS are clinically distinct neurodevelopmental imprinting disorders, which have been linked to a region on the long arm of human chromosome 15 (15q11–q13; reviewed in [185]). This region consists of several imprinted genes and the absence of paternally expressed genes in this imprinting domain results in PWS, whereas the loss of maternally-expressed genes causes AS. Additionally, point mutations in the E6-AP ubiquitin-protein ligase (*UBE3A*) gene, which is also part of the imprinting domain account for approximately 10% of AS patients. In cases where PWS or AS is suspected, DNA methylation analysis of the PWS/AS critical region allows the reliable identification of more than 99% of PWS patients and about 80% of AS patients [186].

Two approaches are commonly used in molecular diagnostics for DNA methylation analysis of the PWS/AS critical region [186,187]. The first approach determines the methylation status at a single gene locus, the small nuclear ribonucleoprotein polypeptide N (*SNRPN*) gene, whereas the second approach determines the methylation status and gene copy number changes at several sites across the region [186]. DNA methylation analysis of the *SNRPN* gene is frequently determined by MSP [188,189] whereas the simultaneous investigation of methylation levels and gene copy numbers is determined by methylation-sensitive multiplex ligation-dependent probe amplification

(MS-MLPA) [190]. Molecular diagnostics of PWS and AS is quite complex and challenging, and guidelines for molecular genetic testing and reporting PWS and AS have been developed [186]. Furthermore, a WHO international genetic reference panel for PWS and AS has been established and was successfully validated in an international multicentre study [187].

BWS and SRS are imprinting disorders, which have been associated with imprinted genes on chromosome region 11p15.5 [191–193]. This region is functionally divided into two domains: the first domain consists of the imprinted insulin-like growth factor gene 2 (*IGF2*) and the non-coding RNA *H19* and is controlled by DMR1 whereas the second region contains several imprinted genes, including cyclin-dependent kinase inhibitor 1C (*CDKN1C*), potassium voltage-gated channel, KQT-like subfamily, member 1 (*KCNQ1*) and KCNQ1 opposite strand/antisense transcript 1 (*KCNQ1OT1*), is controlled by DMR2. Loss of methylation at DMR2 (*KCNQ1OT1* hypomethylation), is the most frequent alteration, in around 50% of BWS patients [194] whereas loss of methylation at DMR1 (*H19* hypomethylation) is typically observed in SRS is found in around 40% of SRS patients [192,195]. As mentioned before, MS-MLPA allows the simultaneous investigation of methylation levels and gene copy numbers and has thus been considered well suited for detecting the majority of (epi-) genetic alterations associated with BWS and SRS in region 11p15.5 [196–198].

Most approaches for routine clinical DNA methylation analysis at single-gene loci in genomic imprinting disorders rely, most probably for historical reasons, on qualitative methylation detection methods. However, the diagnostic advantages of quantitative DNA methylation detection methodologies, such as bisulfite pyrosequencing [191,199,200], are being increasingly recognized and will be probably the preferred methods of choice for analyzing single gene loci in the near future.

8. DNA Methylation Biomarkers of Outcome in Chronic Diseases Other than Cancer

Given the likely early life origins for non-communicable disease, there are plenty of opportunities in which DNA methylation biomarkers could be used. Biomarkers for intrauterine environmental exposures such as maternal alcohol consumption or smoking could provide a way to measure exposures without the need for time-consuming, hard-to-administer questionnaires and where access to mothers is not possible. DNA methylation risk biomarkers could be used to stratify risk for latent non-communicable disease before the onset of disease. They could also be used to monitor progression from first symptoms to full disease. After disease onset, they could be used for predicting survival and response to therapy as they are beginning to do with cancer. Below, we review data from the most promising studies of environmental, risk, diagnostic, prognostic, and predictive DNA methylation biomarkers.

8.1. DNA Methylation Biomarkers for Adverse Environments

There have been a large number of environmental agents linked to epigenetic change, including toxins, stress and nutrition, and these have been reviewed elsewhere [201–203]. Below, we focus on two that have yielded replicated DNA methylation biomarkers: smoking and stress.

8.1.1. *AHRR* Methylation and Smoking

Exposure to adverse environments at all stages of life have been shown to influence the epigenome (reviewed in [39,42,204,205]). However, a replicated association has been found for only one: the effect of DNA methylation on the aryl hydrocarbon receptor repressor (*AHRR*) gene involved in the detoxification of chemicals found in tobacco smoke. As of June 2014, ten independent methylome-wide studies using Illumina Infinium HM450 arrays (containing probes for about 480,000 CpG dinucleotides located in functionally-relevant regions of the genome [206]) had all identified the same smoking-associated probe, cg05575921, located in a region of intermediate CpG density (CpG-island shore) 450 bp upstream of a CpG island in the third intron of the *AHRR* gene [207–216] (Table 3). Two studies focused on the effect of maternal smoking in umbilical cord blood [209,215], which they and others [217] replicated in independent sample cohorts. Others found an association of adult smoking with *AHRR* methylation in blood [207,208,210–212,214], lung tissue [211] or blood lymphoblastoid cell lines [213]. No effects were seen at birth in placenta or buccal epithelium [217] and effects were seen elsewhere in the *AHRR* gene in lung alveolar macrophage DNA but not at the cg05575921 probe [213]. Three studies performed within-cohort validation using locus-specific DNA methylation analysis [207,211,212] and six studies replicated their findings in adults in independent cohorts [208,209,211,212,214,215]. Two studies showed evidence of a role for the region surrounding probe cg05575921 in regulation of *AHRR* expression [211,213]. All found an inverse relationship between smoking and DNA methylation with an effect size ranging from −4% in neonates of mothers who smoked throughout pregnancy [215] to −24.4% in adult current smokers [212].

Similar effects were seen in Europeans, African Americans [207], and South Asians [208]. The latter study found that current smokers were identified with 100% sensitivity and 97% specificity in Europeans and with 80% sensitivity and 95% specificity in South Asians. Timing-specific effects were also found; prenatal smoking only exerted an effect when mothers smoked during a significant part of gestation [217,218]. Furthermore, associations found at birth were also present at 18 months of age [217] but in adulthood, DNA methylation levels were similar in never smokers and in former smokers [212]. Clearly, loss of methylation at and around the *AHRR* cg05575921 probe is strongly associated with first or second hand exposure to smoking. Importantly, one study found an association in adults with smoking, but not tobacco snuff consumption, implicating that a product(s) of tobacco combustion is responsible for the loss of DNA methylation rather than tobacco itself. Further work is needed to link this loss to the timing of prenatal smoking, and postnatal passive and active smoking, and its relationship with downstream health outcomes previously associated with *AHRR* polymorphisms such as cancers [219–221] and endometriosis [222].

Table 3. Summary of findings for the relationship between smoking and DNA methylation within the *AHRR* gene. Data refer to *AHRR* HM450 probe cg05575921 unless otherwise stated. Summary includes details of assay platform, age of subjects, details of exposure, tissues examined, number of subjects, whether *AHRR* expression was also measured, whether findings were validated or replicated and effect size (methylation levels in smokers compared to non-smokers).

Reference	Platform	Age, Median	Exposure	Tissue	N	Effects Elsewhere in AHRR	*AHRR* Expression	Vali-dation	Repli-cation	Effect Size	Notes
[213]	HM450	Adults, 45	Current smoking	LCLs & alveolar MP	119/19 [1]	yes	Yes [2]	No	No	−15%/NS	
[209]	HM450	Birth	Maternal smoking [3]	Whole CB	1062/36 [4]	Yes	No	No	Yes [4]	−7.5%/−7.7% [4]	Multiple hits in the aryl hydrocarbon signaling pathway. Authors have since shown that effects are specific for maternal smoking through at least gestational week 18 [218]
[207]	HM450	Adults, 49	Current smoking	PBMC	111	Yes	No	Yes	No	−15%	African Americans
[208]	HM450	Adults, 48	Current smoking	Whole PB	81/84 [5]	No	No	No	Yes [6]	−22%	Former smokers same as never smokers; changed only slightly after adjusting for cell composition
[210]	HM450	Adults, 22	Current serum cotinine	PBMC	107	yes	No	No	No	−20% [7]	
[211]	HM450	Adults, 51/55/49/? [8]	Current smoking	Whole PB, lung tissue	184/190/180/27	yes	Yes [9]	Yes	Yes	−17%/−14%/NS/NS [10]	Replicated in a mouse model of smoking exposure

Table 3. *Cont.*

Reference	Platform	Age, Median	Exposure	Tissue	N	Effects Elsewhere in AHRR	*AHRR* Expression	Vali-dation	Repli-cation	Effect Size	Notes
[212]	HM450	Adults, 60/53 [11]	Current smoking	Whole PB	749/232 [11]	yes	No	Yes	Yes [11]	−24/−23% [11]	methylation-specific protein binding patterns were observed for cg05575921; levels in former smokers revert to levels similar to never smokers over time
[215]	HM450	Birth	Maternal smoking	Whole CB	889	yes	No	No	Yes	−4%	Replicated a previous study [209]
[214]	HM450	Adults, 43	Current smoking	Whole PB	432	yes	No [13]	No	Yes	−7.4%	Replicated a previous study [212]; no effect with tobacco snuff
[216]	HM450	Female adults, 57	Current smoking	Whole PB	200	No	No	No	Yes	−8%	Former and never smokers had similar methylation levels
[217]	Sequenom EpiTyper	Birth & 18 months	Maternal smoking	CBMC, buccal epithelium, placenta	46/15/24 [12]	yes	Yes [14]	n/a	Y	−10%/NS/NS [12]	No effect if mother smoked early pregnancy only; effects of smoking stable to 18 months of age

[1] refers to the two different cell types tested; [2] *AHRR* expression in alveolar macrophages was inversely correlated with methylation of probe cg05575921; [3] measured using plasma cotinine at gestational week 18; [4] replicated using data from maternal smoking in pregnancy in an independent cohort; [5] data on Europeans replicated in South East Asians; [6] replicated across two ethnic groups; [7] effect size calculated from the regression line, highest to lowest plasma cotinine; [8] discovery, replication and validation groups are subsets of the same cohort and were analyzed along with lung tissue samples from a separate cohort; [9] *AHRR* expression in lung tissue was inversely correlated with methylation of probe cg05575921; [10] no difference with probe cg05575921, differences found for *AHRR* probes cg21161138 and cg23576855 (magnitudes similar to those seen in blood); [11] discovery and replication subsets of the same cohort; [12] significant associations between methylation and expression seen at six genes other than *AHRR*; [13] CBMC/buccal epithelium/placenta; [14] *AHRR* expression non-significantly higher in CBMCs in newborns exposed to smoking in pregnancy than those not exposed. Abbreviations: LCL, lymphoblastoid cell lines; MP, macrophages; PB, peripheral blood; CB, cord blood; MC, mononuclear cells; NS, not significant.

In addition to probe cg05575921, a number of CpG dinucleotides have been significantly associated with prior smoking. Table 4 lists these probes, using a cut-off of those that have been identified by four or more studies. These include two further CpG dinucleotides from *AHRR* [207,209–213], one from the thrombin receptor-like 3 (*F2RL3*) gene [208,211,214,216], one from the growth factor independent 1 transcription repressor (*GFI1*) gene and two from the myosin 1G (*MYO1G*) gene. In addition, two intergenic smoking-associated CpG sites have been replicated across multiple studies [207–209,211–214], all coinciding with regions of DNAse hypersensitivity, suggesting functional significance. Potentially, one or more of these CpG dinucleotides could be used in combination with probe cg05575921 as DNA methylation biomarkers for smoking.

Table 4. Other HM450 probes with significant correlations with smoking in at least four studies. Probes are included if found to be significantly associated with smoking in at least four independent studies. DHS, DNAse hypersensitive site, indicative of regulatory potential.

Probe	**Gene**	**References**
cg03991871	*AHRR*	[209,212,213,215]
cg21161138	*AHRR*	[207,209–212,215]
cg03636183	*F2RL3*	[208,211,214,216]
cg09935388	*GFI1*	[208,209,212,214,215]
cg22132788	*MYO1G*	[208–210,214]
cg12803068	*MYO1G*	[210,212,215,218]
cg21566642	Intergenic (CpG island, DHS)	[207,208,211,212]
cg06126421	Intergenic (enhancer, DHS)	[207,208,211,212,214]

8.1.2. *NR3C1* Methylation and Stress

Stress triggers the activation of the hypothalamus-pituitary-adrenal axis, resulting in the production of glucocorticoids by the adrenal glands. By binding to receptors in the brain, glucocorticoids induce changes in gene expression and in turn, health and behavior [223]. Landmark studies with rats have shown that lack of maternal licking and grooming at birth resulted in an increased level of DNA methylation within the exon 1_7 promoter of the glucocorticoid receptor gene *Nr3c1* in rat hippocampus, in particular at a region that binds nerve growth factor-inducible protein-A (NGFI-A) [224,225]. Since then, studies of the equivalent region in humans (exon 1F of the *NR3C1* gene) have found decreased DNA methylation in cord blood [226,227] and placenta [228] associated with maternal anxiety during pregnancy. Others have shown that violence towards women during pregnancy can have a similar effect [229,230]. Even extremes of stress experienced prior to conception, in the form of the holocaust, were also found to correlate with *NR3C1* exon 1F methylation, albeit in opposite directions depending on the sex of the parent [231]. Methylation analysis of various tissues from adults, either alive or *post mortem*, have found long-lasting effects of abuse [232–235] or death of a parent [235,236] during childhood on *NR3C1* exon 1F. In addition, adults with post-traumatic stress disorder had decreased DNA methylation at the same [237] or alternate [238] *NR3C1* promoters. Of further interest, three studies have shown that methylation of *NR3C1* exon 1F can predict health outcomes, whether predicting quality of movement and attention at birth [239], response to psychotherapy in adults with posttraumatic stress disorder [240] or response to threat-associated stress in adult females [241]. In the latter study, DNA methylation levels at *NR3C1*, the estrogen receptor alpha gene *ESR1* and the serotonin transporter gene *SL6A4* each had independent predictive power. Furthermore, a model containing data from all those genes accounted for half of the variance in total cortisol output. Rat studies showing that the adverse effects and DNA methylation changes associated with early neglect could be reversed in adulthood by methyl-donor rich diet [242] or the histone deacetylase inhibitor Trichostatin A [243], suggesting that *NR3C1* methylation could be use to monitor response to future interventions in humans.

Clearly, methylation at *NR3C1* promoters has the potential to be developed into a variety of candidate biomarkers. In addition, despite yielding no replicated stress biomarkers to date, the small (typically 1%–2%) effect sizes for *NR3C1* methylation would suggest that there may be better DNA methylation-based stress biomarkers out there, discoverable using epigenome-wide approaches [244–248].

8.2. DNA Methylation Risk Biomarkers at Birth

Measuring DNA methylation in five candidate genes in DNA from umbilical cords, Godfrey and colleagues found that methylation of two genes correlated with childhood adiposity as measured by fat mass and trunk/limb fat ratio in 78 nine-year-olds [42]. Methylation of the retinoic acid X receptor alpha (*RXRA*) gene and the endothelial nitric oxide synthase (*ENOS*) gene, together with sex, explained 25% of the variance in adiposity at age nine. Data for *RXRA* were replicated in a second cohort of 239 six-year-olds [42]. Other studies have identified associations between *RXRA* methylation in cord blood at birth and bone mineral density at age four [249] and between methylation of the alkaline phosphatase *ALPL* and body mass index at nine years of age [250]. However, the first association could not be replicated in another sample cohort whereas for the second association no replication study was performed.

8.3. DNA Methylation Biomarkers during Childhood

Rakyan and colleagues identified 132 CpG dinucleotides whose methylation levels differed significantly in twin pairs discordant for type 1 diabetes and which were subsequently validated with an independent method and replicated in a further set of twin pairs [251]. Two-thirds of these CpG dinucleotides were also present in singletons prior to the onset of overt symptoms of type 1 diabetes but positive for diabetes-associated autoantibodies. If those findings can be further replicated, this could provide single or panels of DNA methylation candidate clinical biomarkers predicting the onset of type 1 diabetes. A potential biomarker study found that DNA methylation within the promoter of the peroxisomal proliferator activated receptor gamma (*PPARG*) gene in blood at age five to seven years predicted obesity risk from nine to 14 years [252]. However, these results have yet to be replicated.

Autism spectrum disorder (ASD) describes a related set of neurodevelopmental disorders of childhood characterized by social deficits and communication difficulties, stereotyped or repetitive behaviors and interests, and in some cases, cognitive delays. To date, a small number of ASD MWAS have been performed, using a variety of platforms, on lymphoblastoid cell lines [253], peripheral blood [254,255], buccal epithelium [256], *post mortem* occipital cortex and cerebellum [257], and dorsolateral prefrontal cortex, temporal cortex and cerebellum [258]. ASD-specific DNA methylation was found in all but one study [257] and in the rest, although ASD-specific methylation was often validated within the study, only one study attempted to replicate across cohorts and tissues [258]. In this study, three significant ASD-associated array probes discovered in temporal cortex were replicated in such a manner. ASD-specific DNA methylation found within the proline-rich transmembrane protein 1 (*PRRT1*) gene was replicated in prefrontal *post mortem* cortex and cerebellum, methylation of *c11orf21* was replicated in prefrontal cortex and methylation at an intergenic site near the zinc finger gene *ZFP37* was replicated in a sex-specific manner in cerebellum. The only differentially methylated gene replicated in two separate studies is the olfactory receptor gene *OR2L13* found in buccal epithelium [256] and peripheral blood [254]. Further replication will be required to develop this potential biomarker for ASD.

8.4. DNA Methylation Biomarkers in Adults

Cardiovascular disease (CVD) and its precursors are receiving arguably the greatest attention in MWAS outside cancer [259–261]. DNA methylation biomarkers could help ascertain risk early in life, help with diagnosis and predict response to interventions. Below, we report some of the more advanced such studies.

Levels of fasting glucose and insulin and measures of insulin resistance are used to test for early signs of diabetes and they have been subject to a recent MWAS [262]. This study divided up a cohort of 837 non-diabetic individuals at a median age of 48 years into discovery and replication subsets. Using HM450 arrays on DNA from CD4+ T cells, the investigators found significant associations between methylation of two CpG sites with the ATP-binding cassette gene *ABCG1*, involved in macrophage cholesterol and phospholipids transport, with insulin resistance, with one associated with insulin "of borderline significance". The CpG site with the strongest association with insulin and insulin resistance was also strongly associated with nearby single-nucleotide polymorphisms, implying that differences in genetic sequence can alter the epigenetic functionality of a genomic region. Another recent study replicated across two cohorts a DNA methylation biomarker for triglyceride levels at the carnitine palmitoyltransferase gene *CPT1A* in the same cell type [263]. In this study, *CPT1A* methylation explained 11.6% and 5.5% of the variation in triglyceride levels in the discovery and replication cohorts, respectively.

Although several studies have discovered associations between DNA methylation and obesity [264], few studies have searched for risk or predictive DNA methylation biomarkers in adulthood. In one study that did, males with a history of CVD had higher global DNA methylation than those without [57]. Those who went on to develop symptoms of CVD six years later had intermediate levels of global DNA methylation. In other study, a type 2 diabetes-specific CpG dinucleotide in the first intron of the fat mass and obesity-associated gene *FTO* predicted the onset of symptoms between ages 30 and 43 in a cohort of initially asymptomatic adults [265]. Replication is required for both studies.

Two unreplicated studies resulted in potential predictive DNA methylation biomarkers for response to weight loss programs in adults. In the first, obese women with better response to dietary intervention showed significantly lower levels of DNA methylation at promoters of the leptin (*LEP*) and TNF-alpha (*TNF*) genes than the non-responder group [266]. Although no differences were found between responder and non-responder groups in *LEP* and *TNF* gene expression, if replicated, the potential predictive methylation biomarker would still be valid on its own. In a similar study of obese men, DNA methylation levels in several CpG dinucleotides located in the ATPase *ATP10A* and the CD44 antigen (*CD44*) genes showed statistical baseline differences depending on the weight-loss outcome [266]. Again, these finding have not yet been replicated.

In a search for potential DNA methylation biomarkers of postpartum depression using MWAS and a parallel study in mice, Guintivano and colleagues found that DNA methylation at the heterochromatin protein 1 binding protein 3 (*HP1BP3*), and tetratricopeptide repeat domain 9B (*TTC9B*) genes predicted postpartum depression in the original and replication cohorts [267]. Adjustment for blood cell heterogeneity resulted in a higher specificity (96%) in both cohorts compared to unadjusted values.

Schizophrenia is a psychotic disorder, and bipolar disorder is a mood disorder but both have similar symptoms and they are often studied together. Many potential DNA methylation biomarker studies and MWAS have been conducted for these disorders ([268–274] and references therein). Despite the heterogeneity of platforms and tissues used in these studies, a small number of potential diagnostic schizophrenia- and/or bipolar disorder-associated biomarkers have been identified. The serotonin receptor 2A (*HTR2A*) gene was differentially methylated in both disorders in two brain regions (frontal cortex and the anterior cingulate) [270], replicating the findings of a previous study [275,276]. Similar results were also found in saliva of patients with these disorders [277]. Another gene differentially methylated in two brain regions in both disorders was the dystrobrevin binding protein gene *DTNBP1*, also found in an MWAS of frontal cortex of females with both disorders [278] and in all individuals with schizophrenia [268]. The reelin (*RLN*) gene was differentially methylated in individuals with schizophrenia using an MWAS [271], as it was for schizophrenia and bipolar disorder in a MWAS of brain regions [270], replicating previous findings [279,280]. Other potential DNA methylation biomarkers for psychoses include the human leukocyte antigen (HLA) gene *HCG9* and the serotonin transported gene *SCL6A4* (*5HTT*). *HCG9* was identified in patients with schizophrenia or bipolar disorder in an MWAS of frontal cortex [278] and in brain, blood and sperm in an MWAS for bipolar disorder [281]. *SLC6A4* was differentially methylated in an MWAS of saliva and frontal cortex in individuals with schizophrenia [272], similar to previous findings in lymphoblastoid cell lines and brain tissue of individuals with bipolar disorder in a study that included cross-cohort replication [282]. No studies have investigated the possibility of using above associative biomarkers as potential risk biomarkers in early life. However, a subset of studies has found associations between DNA methylation and medication for schizophrenia or bipolar disorder [273,276]. Clearly, there is much promise for future potential biomarkers of risk, diagnosis and prognosis in schizophrenia and bipolar disorder.

More longitudinal studies at stages of life are required to generate DNA methylation biomarkers for exposure and outcome in chronic diseases other than cancer. Birth cohorts and the retrospective utility of birth dried blood spot Guthrie cards [283] will be essential for this search.

8.5. DNA Methylation Biomarkers of Aging

A number of individual MWAS have looked at the relationship of DNA methylation and aging, with the intention of developing age-specific biomarkers for forensic applications and for investigating premature cellular aging. Three independent meta-analyses have been performed on such datasets [284–286]. The first [284] reviewed six MWAS datasets from Infinium HM27 arrays containing probes for about 27,000 CpG sites [287] on a variety of cell types. None of the 1,093 age-associated probe CpG dinucleotides replicated across all six studies. However, probes at two genes, neuronal pentraxin II (*NPTX2*) and phosphodiesterase 4C (*PDE4C*), did overlap in five of the six studies. The second study [285] performed an analysis of DNA methylation from whole blood from 575 individuals ranging from newborns to age 78 from published HM27 datasets and replicated with a further group of four similar datasets. This yielded 99 significantly age-associated probes including the same *PDE4C* CpG probe cg17861230 as the first study. An even more extensive study of 39 "training" and 32 "test" HM27 and HM450 datasets of more than 7,000 samples from multiple tissues yielded 353 "age predictor" CpG dinucleotides, which included one (cg13899108) in *PDE4C* [286] just 420 bp from the CpG site identified in the first two studies. Although this locus is the most validated age-related CpG dinucleotide, these analyses show that sometimes, a combination of several CpG dinucleotides may be more accurate than a single CpG site. A recent large single analysis measured age-associated DNA methylation in whole blood DNA from 656 individuals using HM450 arrays [288]. In this tissue, investigators identified 70,387 significant age-associated CpG dinucleotides, of which, 53,670 were replicated in an independent dataset. Data was not available to identify whether the *PDE4C* locus mentioned above was among this dataset. The study went on to develop a predictive model of aging that included methylation data and clinical parameters such as gender and body mass index. We predict that this is how most DNA methylation biomarkers will be used in the future. The model selected a set of 71 age-associated methylation biomarkers that were highly predictive of age. Although *PDE4C* was not among this subset, another probe within the subset, cg09809672 associated with the EDAR-associated death domain (*EDARADD*) gene was also identified as age-associated in two of the other studies [285,286]. Importantly, this study also found evidence of an accelerated epigenetic aging in tumor tissue [288] and a further study has since identified epigenetic age acceleration as a risk factor for mortality [289]. Clearly, age-associated DNA methylation biomarkers have more applications than forensic medicine.

9. Integrating Epigenetic Data into Disease Risk Models

Although DNA methylation biomarkers can be used by themselves, the emerging field of molecular pathological epidemiology proposes that they can be integrated into models of disease risk together with other factors [4,290,291]. Such factors include transcriptomic, proteomic, metabolomic, microbiome, and neuroimaging data. The logic is that combinations of risk biomarkers will provide more accurate estimation of disease risk, particularly when dealing with individuals, due to inter- and intra-individual biological variation. Based on principles similar to systems and network biology and a variety of modeling methods, this field is in its infancy but is the next logical step for DNA methylation biomarkers and is already yielding promising results for genetic biomarkers [292].

10. Future Prospects

An increasing tendency to harmonize appropriate methods for DNA methylation detection and reference standards will accelerate the development of DNA methylation biomarkers for cancer and for other diseases. This tendency will be synergistically enhanced by next generation sequencing methodology, which has unlocked a new area of possibilities. This relatively new methodology opens the avenue for routine testing of DNA methylation biomarker panels rather than the selective choice of individual biomarkers. The use of appropriate DNA methylation biomarker panels will prove beneficial where the disease phenotype is quite heterogeneous. It is also expected that the genetic component of disease will be further revealed, which will subsequently allows the strengthening of biomarker panels by combining genetic and DNA methylation biomarker panels [293].

It is not only important to have appropriate epi(genetic) biomarker panels available for certain diseases or risk stratification but also to translate them into clinical actionable information. If no clinical action is available there is a risk of adverse psychological impacts among patients and a risk of those patients being disadvantaged by healthcare providers. However, there is also an enormous potential that affected patients can use the knowledge to their benefit allowing them to actively prevent or delay the early onset of certain diseases.

11. Conclusions

DNA methylation biomarkers are promising and valuable biomarkers which are heading for the molecular diagnostic laboratory. This is particular true for methylation biomarkers in cancer where the biomarkers are currently being used for early detection. However, the uptake of DNA methylation biomarkers is quite slow and will still require a considerable amount of time until the field reaches its full potential. The development of DNA methylation biomarkers for cancer and other diseases has also been slowed down by the lack of standardized methodologies and reference standards for use in DNA methylation detection. The still widespread use of inappropriate methods in combination with inappropriate controls still produces potential DNA methylation biomarkers, which may not be replicated. The need for methods of quantitative DNA methylation detection is becoming more and more obvious and is critical where only small differences in methylation values determine a diseased or disease-free state. Finally, the availability of DNA methylation biomarkers in diseases other than cancer is still in its very early steps but in time, their transition to a clinical setting will follow as it has for cancer.

Acknowledgments

Jeffrey M. Craig is supported by the Murdoch Childrens Research Institute (MCRI), the National Health and Medical Research Council, the National Institutes of Health, and the Australian Twin Registry. This work was supported by the Victorian Government's Operational Infrastructure Support Program.

Author Contributions

Jeffrey M. Craig conceived the idea; Jeffrey M. Craig and Thomas Mikeska were involved in planning, writing and editing the manuscript.

Conflicts of Interest

Thomas Mikeska is the co-inventor of intellectual property on approaches for the detection of *MGMT* promoter methylation in clinical samples. Views and opinions of, and endorsements by the authors do not reflect those of Genetic Technologies Ltd.

References

1. Naylor, S. Biomarkers: Current perspectives and future prospects. *Expert Rev. Mol. Diagn.* **2003**, *3*, 525–529.
2. Mayeux, R. Biomarkers: Potential uses and limitations. *NeuroRx* **2004**, *1*, 182–188.
3. Strimbu, K.; Tavel, J.A. What are biomarkers? *Curr. Opin. HIV AIDS* **2010**, *5*, 463–466.
4. Ogino, S.; Lochhead, P.; Chan, A.T.; Nishihara, R.; Cho, E.; Wolpin, B.M.; Meyerhardt, J.A.; Meissner, A.; Schernhammer, E.S.; Fuchs, C.S.; *et al.* Molecular pathological epidemiology of epigenetics: Emerging integrative science to analyze environment, host, and disease. *Mod. Pathol.* **2013**, *26*, 465–484.
5. Bird, A. Perceptions of epigenetics. *Nature* **2007**, *447*, 396–398.
6. Teh, A.L.; Pan, H.; Chen, L.; Ong, M.L.; Dogra, S.; Wong, J.; Macisaac, J.L.; Mah, S.M.; McEwen, L.M.; Saw, S.M.; *et al.* The effect of genotype and *in utero* environment on inter-individual variation in neonate DNA methylomes. *Genome Res.* **2014**, doi:10.1101/gr.171439.113.
7. De Laat, W.; Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **2013**, *502*, 499–506.
8. Quina, A.S.; Buschbeck, M.; di Croce, L. Chromatin structure and epigenetics. *Biochem. Pharmacol.* **2006**, *72*, 1563–1569.
9. Jurkowska, R.Z.; Jurkowski, T.P.; Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *Chembiochem* **2011**, *12*, 206–222.

10. Jeltsch, A.; Jurkowska, R.Z. New concepts in DNA methylation. *Trends Biochem. Sci.* **2014**, *39*, 310–318.
11. Deaton, A.M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010–1022.
12. Gardiner-Garden, M.; Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **1987**, *196*, 261–282.
13. Takai, D.; Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3740–3745.
14. Zhao, Z.; Han, L. CpG islands: Algorithms and applications in methylation studies. *Biochem. Biophys. Res. Commun.* **2009**, *382*, 643–645.
15. Wu, H.; Caffo, B.; Jaffee, H.A.; Irizarry, R.A.; Feinberg, A.P. Redefining CpG islands using hidden Markov models. *Biostatistics* **2010**, *11*, 499–514.
16. Chuang, L.Y.; Huang, H.C.; Lin, M.C.; Yang, C.H. Particle swarm optimization with reinforcement learning for the prediction of CpG islands in the human genome. *PLoS One* **2011**, *6*, e21036.
17. Kulis, M.; Queiros, A.C.; Beekman, R.; Martin-Subero, J.I. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta* **2013**, *1829*, 1161–1174.
18. Oberdoerffer, S. A conserved role for intragenic DNA methylation in alternative pre-mRNA splicing. *Transcription* **2012**, *3*, 106–109.
19. Illingworth, R.S.; Gruenewald-Schneider, U.; Webb, S.; Kerr, A.R.; James, K.D.; Turner, D.J.; Smith, C.; Harrison, D.J.; Andrews, R.; Bird, A.P. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **2010**, *6*, e1001134.
20. Irizarry, R.A.; Ladd-Acosta, C.; Wen, B.; Wu, Z.; Montano, C.; Onyango, P.; Cui, H.; Gabo, K.; Rongione, M.; Webster, M.; *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **2009**, *41*, 178–186.
21. Ehrlich, M. DNA methylation and cancer-associated genetic instability. *Adv. Exp. Med. Biol.* **2005**, *570*, 363–392.
22. Choi, J.D.; Lee, J.S. Interplay between Epigenetics and Genetics in Cancer. *Genomics Inform.* **2013**, *11*, 164–173.
23. Heng, H.H.; Liu, G.; Stevens, J.B.; Bremer, S.W.; Ye, K.J.; Ye, C.J. Genetic and epigenetic heterogeneity in cancer: The ultimate challenge for drug therapy. *Curr. Drug Targets* **2010**, *11*, 1304–1316.
24. Huang, S. Genetic and non-genetic instability in tumor progression: Link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* **2013**, *32*, 423–448.
25. Stecklein, S.R.; Jensen, R.A.; Pal, A. Genetic and epigenetic signatures of breast cancer subtypes. *Front. Biosci.* **2012**, *4*, 934–949.
26. Gluckman, P.D.; Hanson, M.A.; Buklijas, T. A conceptual framework for the developmental origins of health and disease. *J. Dev. Orig. Health Dis.* **2010**, *1*, 6–18.

27. Gluckman, P.D.; Hanson, M.A.; Pinal, C. The developmental origins of adult disease. *Matern. Child Nutr.* **2005**, *1*, 130–141.
28. Barker, D.J. The fetal and infant origins of adult disease. *BMJ* **1990**, *301*, 1111.
29. Godfrey, K.M.; Lillycrop, K.A.; Burdge, G.C.; Gluckman, P.D.; Hanson, M.A. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr. Res.* **2007**, *61*, R5–R10.
30. McNeal, C.J.; Wilson, D.P.; Christou, D.; Bush, R.L.; Shepherd, L.G.; Santiago, J.; Wu, G.Y. The use of surrogate vascular markers in youth at risk for premature cardiovascular disease. *J. Pediatr. Endocrinol. Metab.* **2009**, *22*, 195–211.
31. Urbina, E.M.; Williams, R.V.; Alpert, B.S.; Collins, R.T.; Daniels, S.R.; Hayman, L.; Jacobson, M.; Mahoney, L.; Mietus-Snyder, M.; Rocchini, A.; *et al.* Noninvasive assessment of subclinical atherosclerosis in children and adolescents: Recommendations for standard assessment for clinical research: A scientific statement from the American Heart Association. *Hypertension* **2009**, *54*, 919–950.
32. Wong, N.C.; Morley, R.; Saffery, R.; Craig, J.M. Archived guthrie blood spots as a novel source for quantitative DNA methylation analysis. *Biotechniques* **2008**, *45*, 423–428.
33. Thirlwell, C.; Eymard, M.; Feber, A.; Teschendorff, A.; Pearce, K.; Lechner, M.; Widschwendter, M.; Beck, S. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods* **2010**, *52*, 248–254.
34. Wong, N.C.; Ashley, D.; Chatterton, Z.; Parkinson-Bates, M.; Ng, H.K.; Halemba, M.S.; Kowalczyk, A.; Bedo, J.; Wang, Q.; Bell, K.; *et al.* A distinct DNA methylation signature defines pediatric pre-B cell acute lymphoblastic leukemia. *Epigenetics* **2012**, *7*, 535–541.
35. Ribel-Madsen, R.; Fraga, M.F.; Jacobsen, S.; Bork-Jensen, J.; Lara, E.; Calvanese, V.; Fernandez, A.F.; Friedrichsen, M.; Vind, B.F.; Hojlund, K.; *et al.* Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. *PLoS One* **2012**, *7*, e51302.
36. Talens, R.P.; Boomsma, D.I.; Tobi, E.W.; Kremer, D.; Jukema, J.W.; Willemsen, G.; Putter, H.; Slagboom, P.E.; Heijmans, B.T. Variation, patterns, and temporal stability of DNA methylation: Considerations for epigenetic epidemiology. *FASEB J.* **2010**, *24*, 3135–3144.
37. Souren, N.Y.; Tierling, S.; Fryns, J.P.; Derom, C.; Walter, J.; Zeegers, M.P. DNA methylation variability at growth-related imprints does not contribute to overweight in monozygotic twins discordant for BMI. *Obesity* **2011**, *19*, 1519–1522.
38. Belinsky, S.A.; Palmisano, W.A.; Gilliland, F.D.; Crooks, L.A.; Divine, K.K.; Winters, S.A.; Grimes, M.J.; Harms, H.J.; Tellez, C.S.; Smith, T.M.; *et al.* Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res.* **2002**, *62*, 2370–2377.
39. Barnes, S.K.; Ozanne, S.E. Pathways linking the early environment to long-term health and lifespan. *Prog. Biophys. Mol. Biol.* **2011**, *106*, 323–336.

40. Ollikainen, M.; Smith, K.R.; Joo, E.J.; Ng, H.K.; Andronikos, R.; Novakovic, B.; Abdul Aziz, N.K.; Carlin, J.B.; Morley, R.; Saffery, R.; *et al.* DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum. Mol. Genet.* **2010**, *19*, 4176–4188.
41. Dittrich, B.; Buiting, K.; Gross, S.; Horsthemke, B. Characterization of a methylation imprint in the Prader-Willi syndrome chromosome region. *Hum. Mol. Genet.* **1993**, *2*, 1995–1999.
42. Godfrey, K.M.; Sheppard, A.; Gluckman, P.D.; Lillycrop, K.A.; Burdge, G.C.; McLean, C.; Rodford, J.; Slater-Jefferies, J.L.; Garratt, E.; Crozier, S.R.; *et al.* Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes* **2011**, *60*, 1528–1534.
43. Lowe, R.; Rakyan, V.K. Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Med.* **2014**, *6*, 23.
44. Adalsteinsson, B.T.; Gudnason, H.; Aspelund, T.; Harris, T.B.; Launer, L.J.; Eiriksdottir, G.; Smith, A.V.; Gudnason, V. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One* **2012**, *7*, e46705.
45. Loh, M.; Liem, N.; Lim, P.L.; Vaithilingam, A.; Cheng, C.L.; Salto-Tellez, M.; Yong, W.P.; Soong, R. Impact of sample heterogeneity on methylation analysis. *Diagn. Mol. Pathol.* **2010**, *19*, 243–247.
46. Montano, C.M.; Irizarry, R.A.; Kaufmann, W.E.; Talbot, K.; Gur, R.E.; Feinberg, A.P.; Taub, M.A. Measuring cell-type specific differential methylation in human brain tissue. *Genome Biol.* **2013**, *14*, R94.
47. Moverare-Skrtic, S.; Mellstrom, D.; Vandenput, L.; Ehrich, M.; Ohlsson, C. Peripheral blood leukocyte distribution and body mass index are associated with the methylation pattern of the androgen receptor promoter. *Endocrine* **2009**, *35*, 204–210.
48. Zou, J.; Lippert, C.; Heckerman, D.; Aryee, M.; Listgarten, J. Epigenome-wide association studies without the need for cell-type composition. *Nat. Methods* **2014**, *11*, 309–311.
49. Accomando, W.P.; Wiencke, J.K.; Houseman, E.A.; Nelson, H.H.; Kelsey, K.T. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol.* **2014**, *15*, R50.
50. Houseman, E.A.; Accomando, W.P.; Koestler, D.C.; Christensen, B.C.; Marsit, C.J.; Nelson, H.H.; Wiencke, J.K.; Kelsey, K.T. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **2012**, *13*, 86.
51. Herceg, Z.; Hainaut, P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. *Mol. Oncol.* **2007**, *1*, 26–41.
52. Iliopoulos, D.; Guler, G.; Han, S.Y.; Johnston, D.; Druck, T.; McCorkell, K.A.; Palazzo, J.; McCue, P.A.; Baffa, R.; Huebner, K. Fragile genes as biomarkers: Epigenetic control of WWOX and FHIT in lung, breast and bladder cancer. *Oncogene* **2005**, *24*, 1625–1633.
53. Verma, M.; Manne, U. Genetic and epigenetic biomarkers in cancer diagnosis and identifying high risk populations. *Crit. Rev. Oncol./Hematol.* **2006**, *60*, 9–18.
54. Zilberman, D.; Henikoff, S. Genome-wide analysis of DNA methylation patterns. *Development* **2007**, *134*, 3959–3965.
55. Zuo, T.; Tycko, B.; Liu, T.M.; Lin, J.J.; Huang, T.H. Methods in DNA methylation profiling. *Epigenomics* **2009**, *1*, 331–345.

56. Kalari, S.; Pfeifer, G.P. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.* **2010**, *70*, 277–308.
57. Kim, M.; Long, T.I.; Arakawa, K.; Wang, R.; Yu, M.C.; Laird, P.W. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One* **2010**, *5*, e9692.
58. How Kit, A.; Nielsen, H.M.; Tost, J. DNA methylation based biomarkers: Practical considerations and applications. *Biochimie* **2012**, *94*, 2314–2337.
59. Michels, K.B.; Binder, A.M.; Dedeurwaerder, S.; Epstein, C.B.; Greally, J.M.; Gut, I.; Houseman, E.A.; Izzi, B.; Kelsey, K.T.; Meissner, A.; *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat. Methods* **2013**, *10*, 949–955.
60. Rakyan, V.K.; Down, T.A.; Balding, D.J.; Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **2011**, *12*, 529–541.
61. Wians, F.H. Clinical laboratory tests: Which, why, and what do the results mean? *Lab Med.* **2009**, *40*, 105–113.
62. Linnet, K.; Bossuyt, P.M.; Moons, K.G.; Reitsma, J.B. Quantifying the accuracy of a diagnostic test or marker. *Clin. Chem.* **2012**, *58*, 1292–1301.
63. Wentzensen, N.; Wacholder, S. From differences in means between cases and controls to risk stratification: A business plan for biomarker development. *Cancer Discov.* **2013**, *3*, 148–157.
64. Fraga, M.F.; Esteller, M. DNA methylation: A profile of methods and applications. *Biotechniques* **2002**, *33*, 632, 634, 636–649.
65. Dahl, C.; Guldberg, P. DNA methylation analysis techniques. *Biogerontology* **2003**, *4*, 233–250.
66. Kristensen, L.S.; Hansen, L.L. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin. Chem.* **2009**, *55*, 1471–1483.
67. Hernandez, H.G.; Tse, M.Y.; Pang, S.C.; Arboleda, H.; Forero, D.A. Optimizing methodologies for PCR-based DNA methylation analysis. *Biotechniques* **2013**, *55*, 181–197.
68. Jorda, M.; Peinado, M.A. Methods for DNA methylation analysis and applications in colon cancer. *Mutat. Res.* **2010**, *693*, 84–93.
69. Shen, L.; Waterland, R.A. Methods of DNA methylation analysis. *Curr. Opin. Clin. Nutr. Metab. Care* **2007**, *10*, 576–581.
70. Frommer, M.; McDonald, L.E.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831.
71. Grunau, C.; Clark, S.J.; Rosenthal, A. Bisulfite genomic sequencing: Systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **2001**, *29*, e65.
72. Holmes, E.E.; Jung, M.; Meller, S.; Leisse, A.; Sailer, V.; Zech, J.; Mengdehl, M.; Garbe, L.A.; Uhl, B.; Kristiansen, G.; *et al.* Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One* **2014**, *9*, e93933.

73. Tanaka, K.; Okamoto, A. Degradation of DNA by bisulfite treatment. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912–1915.

74. Munson, K.; Clark, J.; Lamparska-Kupsik, K.; Smith, S.S. Recovery of bisulfite-converted genomic sequences in the methylation-sensitive QPCR. *Nucleic Acids Res.* **2007**, *35*, 2893–2903.

75. Brandes, J.C.; Carraway, H.; Herman, J.G. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene* **2007**, *26*, 6229–6237.

76. Warnecke, P.M.; Stirzaker, C.; Melki, J.R.; Millar, D.S.; Paul, C.L.; Clark, S.J. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.* **1997**, *25*, 4422–4426.

77. Chhibber, A.; Schroeder, B.G. Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing. *Anal. Biochem.* **2008**, *377*, 46–54.

78. Moskalev, E.A.; Zavgorodnij, M.G.; Majorova, S.P.; Vorobjev, I.A.; Jandaghi, P.; Bure, I.V.; Hoheisel, J.D. Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Res.* **2011**, *39*, e77.

79. Shen, L.; Guo, Y.; Chen, X.; Ahmed, S.; Issa, J.P. Optimizing annealing temperature overcomes bias in bisulfite pcr methylation analysis. *Biotechniques* **2007**, *42*, 48–58.

80. Wojdacz, T.K.; Hansen, L.L.; Dobrovic, A. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res. Notes* **2008**, *1*, 54.

81. Claus, R.; Wilop, S.; Hielscher, T.; Sonnet, M.; Dahl, E.; Galm, O.; Jost, E.; Plass, C. A systematic comparison of quantitative high-resolution DNA methylation analysis and methylation-specific PCR. *Epigenetics* **2012**, *7*, 772–780.

82. Raval, A.; Tanner, S.M.; Byrd, J.C.; Angerman, E.B.; Perko, J.D.; Chen, S.S.; Hackanson, B.; Grever, M.R.; Lucas, D.M.; Matkovic, J.J.; *et al.* Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell* **2007**, *129*, 879–890.

83. Shah, N.; Lin, B.; Sibenaller, Z.; Ryken, T.; Lee, H.; Yoon, J.G.; Rostad, S.; Foltz, G. Comprehensive analysis of MGMT promoter methylation: Correlation with MGMT expression and clinical response in GBM. *PLoS One* **2011**, *6*, e16146.

84. Mikeska, T.; Candiloro, I.L.; Dobrovic, A. The implications of heterogeneous DNA methylation for the accurate quantification of methylation. *Epigenomics* **2010**, *2*, 561–573.

85. Reddy, A.N.; Jiang, W.W.; Kim, M.; Benoit, N.; Taylor, R.; Clinger, J.; Sidransky, D.; Califano, J.A. Death-associated protein kinase promoter hypermethylation in normal human lymphocytes. *Cancer Res.* **2003**, *63*, 7694–7698.

86. Srinivasan, M.; Sedmak, D.; Jewell, S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am. J. Pathol.* **2002**, *161*, 1961–1971.

87. Colella, S.; Shen, L.; Baggerly, K.A.; Issa, J.P.; Krahe, R. Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques* **2003**, *35*, 146–150.

88. Tost, J.; Dunker, J.; Gut, I.G. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* **2003**, *35*, 152–156.

89. Mikeska, T.; Felsberg, J.; Hewitt, C.A.; Dobrovic, A. Analysing DNA methylation using bisulphite pyrosequencing. *Methods Mol. Biol.* **2011**, *791*, 33–53.
90. Dejeux, E.; Audard, V.; Cavard, C.; Gut, I.G.; Terris, B.; Tost, J. Rapid identification of promoter hypermethylation in hepatocellular carcinoma by pyrosequencing of etiologically homogeneous sample pools. *J. Mol. Diagn.* **2007**, *9*, 510–520.
91. Lillycrop, K.A.; Phillips, E.S.; Torrens, C.; Hanson, M.A.; Jackson, A.A.; Burdge, G.C. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. *Br. J. Nutr.* **2008**, *100*, 278–282.
92. Ehrich, M.; Nelson, M.R.; Stanssens, P.; Zabeau, M.; Liloglou, T.; Xinarianos, G.; Cantor, C.R.; Field, J.K.; van den Boom, D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15785–15790.
93. Coolen, M.W.; Statham, A.L.; Gardiner-Garden, M.; Clark, S.J. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: Critical evaluation and improvements. *Nucleic Acids Res.* **2007**, *35*, e119.
94. Wojdacz, T.K.; Dobrovic, A. Methylation-sensitive high resolution melting (MS-HRM): A new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* **2007**, *35*, e41.
95. Mikeska, T.; Dobrovic, A. Methylation-sensitive high resolution melting for the rapid analysis
of DNA methylation. In *Epigenetics: A Reference Manual*; Craig, J.M., Wong, N.C., Eds.; Caister Academic Press: Norwich, UK, 2011; pp. 325–335.
96. Candiloro, I.L.; Mikeska, T.; Dobrovic, A. Closed-tube PCR methods for locus-specific DNA methylation analysis. *Methods Mol. Biol.* **2011**, *791*, 55–71.
97. Candiloro, I.L.; Mikeska, T.; Dobrovic, A. Assessing combined methylation-sensitive high resolution melting and pyrosequencing for the analysis of heterogeneous DNA methylation. *Epigenetics* **2011**, *6*, 500–507.
98. Herman, J.G.; Graff, J.R.; Myohanen, S.; Nelkin, B.D.; Baylin, S.B. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9821–9826.
99. Preusser, M.; Elezi, L.; Hainfellner, J.A. Reliability and reproducibility of PCR-based testing of O^6-methylguanine-DNA methyltransferase gene (MGMT) promoter methylation status in formalin-fixed and paraffin-embedded neurosurgical biopsy specimens. *Clin. Neuropathol.* **2008**, *27*, 388–390.
100. Hamilton, M.G.; Roldan, G.; Magliocco, A.; McIntyre, J.B.; Parney, I.; Easaw, J.C. Determination of the methylation status of MGMT in different regions within glioblastoma multiforme. *J. Neurooncol.* **2011**, *102*, 255–260.
101. Eads, C.A.; Danenberg, K.D.; Kawakami, K.; Saltz, L.B.; Blake, C.; Shibata, D.; Danenberg, P.V.; Laird, P.W. MethyLight: A high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* **2000**, *28*, e32.

102. Rand, K.; Qu, W.; Ho, T.; Clark, S.J.; Molloy, P. Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* **2002**, *27*, 114–120.

103. Bonanno, C.; Shehi, E.; Adlerstein, D.; Makrigiorgos, G.M. MS-FLAG, a novel real-time signal generation method for methylation-specific PCR. *Clin. Chem.* **2007**, *53*, 2119–2127.

104. Kristensen, L.S.; Mikeska, T.; Krypuy, M.; Dobrovic, A. Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): High-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res.* **2008**, *36*, e42.

105. Cottrell, S.E.; Distler, J.; Goodman, N.S.; Mooney, S.H.; Kluth, A.; Olek, A.; Schwope, I.; Tetzner, R.; Ziebarth, H.; Berlin, K. A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res.* **2004**, *32*, e10.

106. Nygren, A.O.; Ameziane, N.; Duarte, H.M.; Vijzelaar, R.N.; Waisfisz, Q.; Hess, C.J.; Schouten, J.P.; Errami, A. Methylation-specific MLPA (MS-MLPA): Simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res.* **2005**, *33*, e128.

107. Serizawa, R.R.; Ralfkiaer, U.; Dahl, C.; Lam, G.W.; Hansen, A.B.; Steven, K.; Horn, T.; Guldberg, P. Custom-designed MLPA using multiple short synthetic probes: Application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. *J. Mol. Diagn.* **2010**, *12*, 402–408.

108. Gatta, V.; Gennaro, E.; Franchi, S.; Cecconi, M.; Antonucci, I.; Tommasi, M.; Palka, G.; Coviello, D.; Stuppia, L.; Grasso, M. MS-MLPA analysis for FMR1 gene: Evaluation in a routine diagnostic setting. *BMC Med. Genet.* **2013**, *14*, 79.

109. Jeuken, J.W.; Cornelissen, S.J.; Vriezen, M.; Dekkers, M.M.; Errami, A.; Sijben, A.; Boots-Sprenger, S.H.; Wesseling, P. MS-MLPA: An attractive alternative laboratory assay for robust, reliable, and semiquantitative detection of MGMT promoter hypermethylation in gliomas. *Lab. Investig.* **2007**, *87*, 1055–1065.

110. Homig-Holzel, C.; Savola, S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagn. Mol. Pathol.* **2012**, *21*, 189–206.

111. Tano, K.; Shiota, S.; Collier, J.; Foote, R.S.; Mitra, S. Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O^6-alkylguanine. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 686–690.

112. Natarajan, A.T.; Vermeulen, S.; Darroudi, F.; Valentine, M.B.; Brent, T.P.; Mitra, S.; Tano, K. Chromosomal localization of human O^6-methylguanine-DNA methyltransferase (MGMT) gene by in situ hybridization. *Mutagenesis* **1992**, *7*, 83–85.

113. Nutt, C.L.; Costello, J.F.; Bambrick, L.L.; Yarosh, D.B.; Swinnen, L.J.; Chambers, A.F.; Cairncross, J.G. O^6-methylguanine-DNA methyltransferase in tumors and cells of the oligodendrocyte lineage. *Can. J. Neurol. Sci.* **1995**, *22*, 111–115.

114. Mineura, K.; Yanagisawa, T.; Watanabe, K.; Kowada, M.; Yasui, N. Human brain tumor O(6)-methylguanine-DNA methyltransferase mRNA and its significance as an indicator of selective chloroethylnitrosourea chemotherapy. *Int. J. Cancer* **1996**, *69*, 420–425.

115. Silber, J.R.; Bobola, M.S.; Ghatan, S.; Blank, A.; Kolstoe, D.D.; Berger, M.S. O^6-methylguanine-DNA methyltransferase activity in adult gliomas: Relation to patient and tumor characteristics. *Cancer Res.* **1998**, *58*, 1068–1073.

116. Esteller, M.; Garcia-Foncillas, J.; Andion, E.; Goodman, S.N.; Hidalgo, O.F.; Vanaclocha, V.; Baylin, S.B.; Herman, J.G. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N. Engl. J. Med.* **2000**, *343*, 1350–1354.

117. Hegi, M.E.; Diserens, A.C.; Godard, S.; Dietrich, P.Y.; Regli, L.; Ostermann, S.; Otten, P.; van Melle, G.; de Tribolet, N.; Stupp, R. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. *Clin. Cancer Res.* **2004**, *10*, 1871–1874.

118. Paz, M.F.; Yaya-Tur, R.; Rojas-Marcos, I.; Reynes, G.; Pollan, M.; Aguirre-Cruz, L.; Garcia-Lopez, J.L.; Piquer, J.; Safont, M.J.; Balana, C.; *et al.* CpG island hypermethylation of the DNA repair enzyme methyltransferase predicts response to temozolomide in primary gliomas. *Clin. Cancer Res.* **2004**, *10*, 4933–4938.

119. Blanc, J.L.; Wager, M.; Guilhot, J.; Kusy, S.; Bataille, B.; Chantereau, T.; Lapierre, F.; Larsen, C.J.; Karayan-Tapon, L. Correlation of clinical features and methylation status of MGMT gene promoter in glioblastomas. *J. Neurooncol.* **2004**, *68*, 275–283.

120. Hegi, M.E.; Diserens, A.C.; Gorlia, T.; Hamou, M.F.; de Tribolet, N.; Weller, M.; Kros, J.M.; Hainfellner, J.A.; Mason, W.; Mariani, L.; *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 997–1003.

121. Wick, W.; Weller, M.; van den Bent, M.; Sanson, M.; Weiler, M.; von Deimling, A.; Plass, C.; Hegi, M.; Platten, M.; Reifenberger, G. MGMT testing-the challenges for biomarker-based glioma treatment. *Nat. Rev. Neurol.* **2014**, *10*, 372–385.

122. Christians, A.; Hartmann, C.; Benner, A.; Meyer, J.; von Deimling, A.; Weller, M.; Wick, W.; Weiler, M. Prognostic value of three different methods of MGMT promoter methylation analysis in a prospective trial on newly diagnosed glioblastoma. *PLoS One* **2012**, *7*, e33449.

123. Karayan-Tapon, L.; Quillien, V.; Guilhot, J.; Wager, M.; Fromont, G.; Saikali, S.; Etcheverry, A.; Hamlat, A.; Loussouarn, D.; Campion, L.; *et al.* Prognostic value of O^6-methylguanine-DNA methyltransferase status in glioblastoma patients, assessed by five different methods. *J. Neurooncol.* **2010**, *97*, 311–322.

124. Mikeska, T.; Bock, C.; el-Maarri, O.; Hubner, A.; Ehrentraut, D.; Schramm, J.; Felsberg, J.; Kahl, P.; Buttner, R.; Pietsch, T.; *et al.* Optimization of quantitative MGMT promoter methylation analysis using pyrosequencing and combined bisulfite restriction analysis. *J. Mol. Diagn.* **2007**, *9*, 368–381.

125. Quillien, V.; Lavenu, A.; Karayan-Tapon, L.; Carpentier, C.; Labussiere, M.; Lesimple, T.; Chinot, O.; Wager, M.; Honnorat, J.; Saikali, S.; *et al.* Comparative assessment of 5 methods (methylation-specific polymerase chain reaction, MethyLight, pyrosequencing, methylation-sensitive high-resolution melting, and immunohistochemistry) to analyze O^6-methylguanine-DNA-methyltranferase in a series of 100 glioblastoma patients. *Cancer* **2012**, *118*, 4201–4211.

126. Weller, M.; Stupp, R.; Reifenberger, G.; Brandes, A.A.; van den Bent, M.J.; Wick, W.; Hegi, M.E. MGMT promoter methylation in malignant gliomas: Ready for personalized medicine? *Nat. Rev. Neurol.* **2010**, *6*, 39–51.
127. Mikeska, T.; Bock, C.; Do, H.; Dobrovic, A. DNA methylation biomarkers in cancer: Progress towards clinical implementation. *Expert Rev. Mol. Diagn.* **2012**, *12*, 473–487.
128. Li, Y.W.; Kong, F.M.; Zhou, J.P.; Dong, M. Aberrant promoter methylation of the vimentin gene may contribute to colorectal carcinogenesis: A meta-analysis. *Tumour Biol.* **2014**, *35*, 6783–6790.
129. Payne, S.R. From discovery to the clinic: The novel DNA methylation biomarker (m)sept9 for the detection of colorectal cancer in blood. *Epigenomics* **2010**, *2*, 575–585.
130. Darwiche, K.; Zarogoulidis, P.; Baehner, K.; Welter, S.; Tetzner, R.; Wohlschlaeger, J.; Theegarten, D.; Nakajima, T.; Freitag, L. Assessment of shox2 methylation in ebus-tbna specimen improves accuracy in lung cancer staging. *Ann. Oncol.* **2013**, *24*, 2866–2870.
131. Dietrich, D.; Hasinger, O.; Liebenberg, V.; Field, J.K.; Kristiansen, G.; Soltermann, A. DNA methylation of the homeobox genes pitx2 and shox2 predicts outcome in non-small-cell lung cancer patients. *Diagn. Mol. Pathol.* **2012**, *21*, 93–104.
132. Dietrich, D.; Jung, M.; Puetzer, S.; Leisse, A.; Holmes, E.E.; Meller, S.; Uhl, B.; Schatz, P.; Ivascu, C.; Kristiansen, G. Diagnostic and prognostic value of shox2 and sept9 DNA methylation and cytology in benign, paramalignant and malignant pleural effusions. *PloS One* **2013**, *8*, e84225.
133. Ilse, P.; Biesterfeld, S.; Pomjanski, N.; Fink, C.; Schramm, M. Shox2 DNA methylation is a tumour marker in pleural effusions. *Cancer Genomics Proteomics* **2013**, *10*, 217–223.
134. Kneip, C.; Schmidt, B.; Seegebarth, A.; Weickmann, S.; Fleischhacker, M.; Liebenberg, V.; Field, J.K.; Dietrich, D. Shox2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J. Thorac. Oncol.* **2011**, *6*, 1632–1638.
135. Schmidt, B.; Liebenberg, V.; Dietrich, D.; Schlegel, T.; Kneip, C.; Seegebarth, A.; Flemming, N.; Seemann, S.; Distler, J.; Lewin, J.; *et al.* Shox2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer* **2010**, *10*, 600.
136. Wu, T.; Giovannucci, E.; Welge, J.; Mallick, P.; Tang, W.Y.; Ho, S.M. Measurement of gstp1 promoter methylation in body fluids may complement psa screening: A meta-analysis. *Br. J. Cancer* **2011**, *105*, 65–73.
137. Chen, Y.; Li, J.; Yu, X.; Li, S.; Zhang, X.; Mo, Z.; Hu, Y. Apc gene hypermethylation and prostate cancer: A systematic review and meta-analysis. *Eur. J. Hum. Genet.* **2013**, *21*, 929–935.
138. Pan, J.; Chen, J.; Zhang, B.; Chen, X.; Huang, B.; Zhuang, J.; Mo, C.; Qiu, S. Association between rassf1a promoter methylation and prostate cancer: A systematic review and meta-analysis. *PloS One* **2013**, *8*, e75283.
139. Yin, A.A.; Zhang, L.H.; Cheng, J.X.; Dong, Y.; Liu, B.L.; Han, N.; Zhang, X. The predictive but not prognostic value of mgmt promoter methylation status in elderly glioblastoma patients: A meta-analysis. *PloS One* **2014**, *9*, e85102.

140. Zhang, K.; Wang, X.Q.; Zhou, B.; Zhang, L. The prognostic value of mgmt promoter methylation in glioblastoma multiforme: A meta-analysis. *Fam. cancer* **2013**, *12*, 449–458.
141. Preusser, M.; Berghoff, A.S.; Manzl, C.; Filipits, M.; Weinhausel, A.; Pulverer, W.; Dieckmann, K.; Widhalm, G.; Wohrer, A.; Knosp, E.; *et al.* Clinical neuropathology practice news 1–2014: Pyrosequencing meets clinical and analytical performance criteria for routine testing of mgmt promoter methylation status in glioblastoma. *Clin. Neuropathol.* **2014**, *33*, 6–14.
142. Quillien, V.; Lavenu, A.; Sanson, M.; Legrain, M.; Dubus, P.; Karayan-Tapon, L.; Mosser, J.; Ichimura, K.; Figarella-Branger, D. Outcome-based determination of optimal pyrosequencing assay for mgmt methylation detection in glioblastoma patients. *J. Neurooncol.* **2014**, *116*, 487–496.
143. Dunn, J.; Baborie, A.; Alam, F.; Joyce, K.; Moxham, M.; Sibson, R.; Crooks, D.; Husband, D.; Shenoy, A.; Brodbelt, A.; *et al.* Extent of mgmt promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. *Br. J. Cancer* **2009**, *101*, 124–131.
144. Reifenberger, G.; Hentschel, B.; Felsberg, J.; Schackert, G.; Simon, M.; Schnell, O.; Westphal, M.; Wick, W.; Pietsch, T.; Loeffler, M.; *et al.* Predictive impact of mgmt promoter methylation in glioblastoma of the elderly. *Int. J. Cancer.* **2012**, *131*, 1342–1350.
145. Oberstadt, M.C.; Bien-Moller, S.; Weitmann, K.; Herzog, S.; Hentschel, K.; Rimmbach, C.; Vogelgesang, S.; Balz, E.; Fink, M.; Michael, H.; *et al.* Epigenetic modulation of the drug resistance genes mgmt, abcb1 and abcg2 in glioblastoma multiforme. *BMC Cancer* **2013**, *13*, 617.
146. Everhard, S.; Tost, J.; el Abdalaoui, H.; Criniere, E.; Busato, F.; Marie, Y.; Gut, I.G.; Sanson, M.; Mokhtari, K.; Laigle-Donadey, F.; *et al.* Identification of regions correlating mgmt promoter methylation and gene expression in glioblastomas. *Neuro-oncology* **2009**, *11*, 348–356.
147. Heyn, H.; Esteller, M. DNA methylation profiling in the clinic: Applications and challenges. *Nat. Rev. Genet.* **2012**, *13*, 679–692.
148. Delpu, Y.; Cordelier, P.; Cho, W.C.; Torrisani, J. DNA methylation and cancer diagnosis. *Int. J. Mol. Sci.* **2013**, *14*, 15029–15058.
149. Heichman, K.A.; Warren, J.D. DNA methylation biomarkers and their utility for solid cancer diagnostics. *Clin. Chem. Lab. Med.* **2012**, *50*, 1707–1721.
150. Bardhan, K.; Liu, K. Epigenetics and colorectal cancer pathogenesis. *Cancers* **2013**, *5*, 676–713.
151. Colussi, D.; Brandi, G.; Bazzoli, F.; Ricciardiello, L. Molecular pathways involved in colorectal cancer: Implications for disease behavior and prevention. *Int. J. Mol. Sci.* **2013**, *14*, 16365–16385.
152. Gyparaki, M.T.; Basdra, E.K.; Papavassiliou, A.G. DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. *J. Mol. Med. (Berlin, Germany)* **2013**, *91*, 1249–1256.
153. Balgkouranidou, I.; Liloglou, T.; Lianidou, E.S. Lung cancer epigenetics: Emerging biomarkers. *Biomark. Med.* **2013**, *7*, 49–58.

154. Jones, A.; Lechner, M.; Fourkala, E.O.; Kristeleit, R.; Widschwendter, M. Emerging promise of epigenetics and DNA methylation for the diagnosis and management of women's cancers. *Epigenomics* **2010**, *2*, 9–38.

155. Day, T.K.; Bianco-Miotto, T. Common gene pathways and families altered by DNA methylation in breast and prostate cancers. *Endocr. Relat. Cancer* **2013**, *20*, R215–R232.

156. Kandimalla, R.; van Tilborg, A.A.; Zwarthoff, E.C. DNA methylation-based biomarkers in bladder cancer. *Nat. Rev. Urol.* **2013**, *10*, 327–335.

157. Fukushige, S.; Horii, A. Road to early detection of pancreatic cancer: Attempts to utilize epigenetic biomarkers. *Cancer Lett.* **2014**, *342*, 231–237.

158. Greenberg, E.S.; Chong, K.K.; Huynh, K.T.; Tanaka, R.; Hoon, D.S. Epigenetic biomarkers in skin cancer. *Cancer Lett.* **2014**, *342*, 170–177.

159. Zoratto, F.; Rossi, L.; Verrico, M.; Papa, A.; Basso, E.; Zullo, A.; Tomao, L.; Romiti, A.; Lo Russo, G.; Tomao, S. Focus on genetic and epigenetic events of colorectal cancer pathogenesis: Implications for molecular diagnosis. *Tumour Biol.* **2014**, *35*, 6195–6206.

160. McDevitt, M.A. Clinical applications of epigenetic markers and epigenetic profiling in myeloid malignancies. *Semin. Oncol.* **2012**, *39*, 109–122.

161. Rodriguez-Vicente, A.E.; Diaz, M.G.; Hernandez-Rivas, J.M. Chronic lymphocytic leukemia: A clinical and molecular heterogenous disease. *Cancer Genet.* **2013**, *206*, 49–62.

162. Li, Y.S.; Xie, Q.; Yang, D.Y.; Zheng, Y. Role of rassf1a promoter methylation in the pathogenesis of hepatocellular carcinoma: A meta-analysis of 21 cohort studies. *Mol. Biol. Rep.* **2014**, *41*, 3925–3933.

163. Chen, W.D.; Han, Z.J.; Skoletsky, J.; Olson, J.; Sah, J.; Myeroff, L.; Platzer, P.; Lu, S.; Dawson, D.; Willis, J.; *et al.* Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J. Natl. Cancer Inst.* **2005**, *97*, 1124–1132.

164. deVos, T.; Tetzner, R.; Model, F.; Weiss, G.; Schuster, M.; Distler, J.; Steiger, K.V.; Grutzmann, R.; Pilarsky, C.; Habermann, J.K.; *et al.* Circulating methylated sept9 DNA in plasma is a biomarker for colorectal cancer. *Clin. Chem.* **2009**, *55*, 1337–1346.

165. Oh, T.; Kim, N.; Moon, Y.; Kim, M.S.; Hoehn, B.D.; Park, C.H.; Kim, T.S.; Kim, N.K.; Chung, H.C.; An, S. Genome-wide identification and validation of a novel methylation biomarker, sdc2, for blood-based detection of colorectal cancer. *J. Mol. Diagn.* **2013**, *15*, 498–507.

166. Mitchell, S.M.; Ross, J.P.; Drew, H.R.; Ho, T.; Brown, G.S.; Saunders, N.F.; Duesing, K.R.; Buckley, M.J.; Dunne, R.; Beetson, I.; *et al.* A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer* **2014**, *14*, 54.

167. Goessl, C.; Krause, H.; Muller, M.; Heicappell, R.; Schrader, M.; Sachsinger, J.; Miller, K. Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. *Cancer Res.* **2000**, *60*, 5941–5945.

168. Goessl, C.; Muller, M.; Heicappell, R.; Krause, H.; Straub, B.; Schrader, M.; Miller, K. DNA-based detection of prostate cancer in urine after prostatic massage. *Urology* **2001**, *58*, 335–338.

169. Merlo, A.; Herman, J.G.; Mao, L.; Lee, D.J.; Gabrielson, E.; Burger, P.C.; Baylin, S.B.; Sidransky, D. 5' cpg island methylation is associated with transcriptional silencing of the tumour suppressor p16/cdkn2/mts1 in human cancers. *Nat. Med.* **1995**, *1*, 686–692.
170. Sterlacci, W.; Tzankov, A.; Veits, L.; Zelger, B.; Bihl, M.P.; Foerster, A.; Augustin, F.; Fiegl, M.; Savic, S. A comprehensive analysis of p16 expression, gene status, and promoter hypermethylation in surgically resected non-small cell lung carcinomas. *J. Thorac. Oncol.* **2011**, *6*, 1649–1657.
171. Nimmrich, I.; Sieuwerts, A.M.; Meijer-van Gelder, M.E.; Schwope, I.; Bolt-de Vries, J.; Harbeck, N.; Koenig, T.; Hartmann, O.; Kluth, A.; Dietrich, D.; *et al.* DNA hypermethylation of pitx2 is a marker of poor prognosis in untreated lymph node-negative hormone receptor-positive breast cancer patients. *Breast Cancer Res. Treat.* **2008**, *111*, 429–437.
172. Harbeck, N.; Nimmrich, I.; Hartmann, A.; Ross, J.S.; Cufer, T.; Grutzmann, R.; Kristiansen, G.; Paradiso, A.; Hartmann, O.; Margossian, A.; *et al.* Multicenter study using paraffin-embedded tumor tissue testing pitx2 DNA methylation as a marker for outcome prediction in tamoxifen-treated, node-negative breast cancer patients. *J. Clin. Oncol.* **2008**, *26*, 5036–5042.
173. Maier, S.; Nimmrich, I.; Koenig, T.; Eppenberger-Castori, S.; Bohlmann, I.; Paradiso, A.; Spyratos, F.; Thomssen, C.; Mueller, V.; Nahrig, J.; *et al.* DNA-methylation of the homeodomain transcription factor PITX2 reliably predicts risk of distant disease recurrence in tamoxifen-treated, node-negative breast cancer patients—Technical and clinical validation in a multi-centre setting in collaboration with the European Organisation for Research and Treatment of Cancer (EORTC) PathoBiology group. *Eur. J. Cancer* **2007**, *43*, 1679–1686.
174. Hartmann, O.; Spyratos, F.; Harbeck, N.; Dietrich, D.; Fassbender, A.; Schmitt, M.; Eppenberger-Castori, S.; Vuaroqueaux, V.; Lerebours, F.; Welzel, K.; *et al.* DNA methylation markers predict outcome in node-positive, estrogen receptor-positive breast cancer with adjuvant anthracycline-based chemotherapy. *Clin. Cancer Res.* **2009**, *15*, 315–323.
175. Banez, L.L.; Sun, L.; van Leenders, G.J.; Wheeler, T.M.; Bangma, C.H.; Freedland, S.J.; Ittmann, M.M.; Lark, A.L.; Madden, J.F.; Hartman, A.; *et al.* Multicenter clinical validation of PITX2 methylation as a prostate specific antigen recurrence predictor in patients with post-radical prostatectomy prostate cancer. *J. Urol.* **2010**, *184*, 149–156.
176. Dietrich, D.; Hasinger, O.; Banez, L.L.; Sun, L.; van Leenders, G.J.; Wheeler, T.M.; Bangma, C.H.; Wernert, N.; Perner, S.; Freedland, S.J.; *et al.* Development and clinical validation of a real-time PCR assay for PITX2 DNA methylation to predict prostate-specific antigen recurrence in prostate cancer patients following radical prostatectomy. *J. Mol. Diagn.* **2013**, *15*, 270–279.
177. Weiss, G.; Cottrell, S.; Distler, J.; Schatz, P.; Kristiansen, G.; Ittmann, M.; Haefliger, C.; Lesche, R.; Hartmann, A.; Corman, J.; *et al.* DNA methylation of the PITX2 gene promoter region is a strong independent prognostic marker of biochemical recurrence in patients with prostate cancer after radical prostatectomy. *J. Urol.* **2009**, *181*, 1678–1685.

178. Kaina, B.; Christmann, M.; Naumann, S.; Roos, W.P. MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair* **2007**, *6*, 1079–1099.

179. Sharma, S.; Salehi, F.; Scheithauer, B.W.; Rotondo, F.; Syro, L.V.; Kovacs, K. Role of MGMT in tumor development, progression, diagnosis, treatment and prognosis. *Anticancer Res.* **2009**, *29*, 3759–3768.

180. Esteller, M.; Gaidano, G.; Goodman, S.N.; Zagonel, V.; Capello, D.; Botto, B.; Rossi, D.; Gloghini, A.; Vitolo, U.; Carbone, A.; *et al.* Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. *J. Natl. Cancer Inst.* **2002**, *94*, 26–32.

181. Van den Bent, M.J.; Dubbink, H.J.; Sanson, M.; van der Lee-Haarloo, C.R.; Hegi, M.; Jeuken, J.W.; Ibdaih, A.; Brandes, A.A.; Taphoorn, M.J.; Frenay, M.; *et al.* MGMT promoter methylation is prognostic but not predictive for outcome to adjuvant PCV chemotherapy in anaplastic oligodendroglial tumors: A report from EORTC Brain Tumor Group Study 26951. *J. Clin. Oncol.* **2009**, *27*, 5881–5886.

182. Wick, W.; Hartmann, C.; Engel, C.; Stoffels, M.; Felsberg, J.; Stockhammer, F.; Sabel, M.C.; Koeppen, S.; Ketter, R.; Meyermann, R.; *et al.* NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. *J. Clin. Oncol.* **2009**, *27*, 5874–5880.

183. Skaar, D.A.; Li, Y.; Bernal, A.J.; Hoyo, C.; Murphy, S.K.; Jirtle, R.L. The human imprintome: Regulatory mechanisms, methods of ascertainment, and roles in disease susceptibility. *ILAR J.* **2012**, *53*, 341–358.

184. Peters, J. The role of genomic imprinting in biology and disease: An expanding view. *Nat. Rev. Genet.* **2014**, *15*, 517–530.

185. Butler, M.G. Genomic imprinting disorders in humans: A mini-review. *J. Assist. Reprod. Genet.* **2009**, *26*, 477–486.

186. Ramsden, S.C.; Clayton-Smith, J.; Birch, R.; Buiting, K. Practice guidelines for the molecular analysis of Prader-Willi and Angelman syndromes. *BMC Med. Genet.* **2010**, *11*, 70.

187. Boyle, J.; Hawkins, M.; Barton, D.E.; Meaney, K.; Guitart, M.; O'Grady, A.; Tobi, S.; Ramsden, S.C.; Elles, R.; Gray, E.; *et al.* Establishment of the first WHO international genetic reference panel for Prader Willi and Angelman syndromes. *Eur. J. Hum. Genet.* **2011**, *19*, 857–864.

188. Kubota, T.; Das, S.; Christian, S.L.; Baylin, S.B.; Herman, J.G.; Ledbetter, D.H. Methylation-specific PCR simplifies imprinting analysis. *Nat. Genet.* **1997**, *16*, 16–17.

189. Zeschnigk, M.; Lich, C.; Buiting, K.; Doerfler, W.; Horsthemke, B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur. J. Hum. Genet.* **1997**, *5*, 94–98.

190. Henkhaus, R.S.; Kim, S.J.; Kimonis, V.E.; Gold, J.A.; Dykens, E.M.; Driscoll, D.J.; Butler, M.G. Methylation-specific multiplex ligation-dependent probe amplification and identification of deletion genetic subtypes in Prader-Willi syndrome. *Genet. Test. Mol. Biomark.* **2012**, *16*, 178–186.

191. Bourque, D.K.; Penaherrera, M.S.; Yuen, R.K.; van Allen, M.I.; McFadden, D.E.; Robinson, W.P. The utility of quantitative methylation assays at imprinted genes for the diagnosis of fetal and placental disorders. *Clin. Genet.* **2011**, *79*, 169–175.

192. Eggermann, T.; Begemann, M.; Binder, G.; Spengler, S. Silver-Russell syndrome: Genetic basis and molecular genetic testing. *Orphanet J. Rare Dis.* **2010**, *5*, 19.

193. Weksberg, R.; Shuman, C.; Beckwith, J.B. Beckwith-Wiedemann syndrome. *Eur. J. Hum. Genet.* **2010**, *18*, 8–14.

194. Smilinich, N.J.; Day, C.D.; Fitzpatrick, G.V.; Caldwell, G.M.; Lossie, A.C.; Cooper, P.R.; Smallwood, A.C.; Joyce, J.A.; Schofield, P.N.; Reik, W.; *et al.* A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8064–8069.

195. Horike, S.; Ferreira, J.C.; Meguro-Horike, M.; Choufani, S.; Smith, A.C.; Shuman, C.; Meschino, W.; Chitayat, D.; Zackai, E.; Scherer, S.W.; *et al.* Screening of DNA methylation at the H19 promoter or the distal region of its ICR1 ensures efficient detection of chromosome 11p15 epimutations in Russell-Silver syndrome. *Am. J. Med. Genet. A* **2009**, *149A*, 2415–2423.

196. Lukova, M.; Todorova, A.; Todorov, T.; Mitev, V. Different methylation patterns in BWS/SRS cases clarified by MS-MLPA. *Mol. Biol. Rep.* **2013**, *40*, 263–268.

197. Priolo, M.; Sparago, A.; Mammi, C.; Cerrato, F.; Lagana, C.; Riccio, A. MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. *Eur. J. Hum. Genet.* **2008**, *16*, 565–571.

198. Scott, R.H.; Douglas, J.; Baskcomb, L.; Nygren, A.O.; Birch, J.M.; Cole, T.R.; Cormier-Daire, V.; Eastwood, D.M.; Garcia-Minaur, S.; Lupunzina, P.; *et al.* Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. *J. Med. Genet.* **2008**, *45*, 106–113.

199. Calvello, M.; Tabano, S.; Colapietro, P.; Maitz, S.; Pansa, A.; Augello, C.; Lalatta, F.; Gentilin, B.; Spreafico, F.; Calzari, L.; *et al.* Quantitative DNA methylation analysis improves epigenotype-phenotype correlations in Beckwith-Wiedemann syndrome. *Epigenetics* **2013**, *8*, 1053–1060.

200. White, H.E.; Durston, V.J.; Harvey, J.F.; Cross, N.C. Quantitative analysis of SNRPN(correction of SRNPN) gene methylation by pyrosequencing as a diagnostic test for Prader-Willi syndrome and Angelman syndrome. *Clin. Chem.* **2006**, *52*, 1005–1013.

201. Feil, R.; Fraga, M.F. Epigenetics and the environment: Emerging patterns and implications. *Nat. Rev. Genet.* **2011**, *13*, 97–109.

202. Mazzio, E.A.; Soliman, K.F. Basic concepts of epigenetics: Impact of environmental signals on gene expression. *Epigenetics* **2012**, *7*, 119–130.

203. Cortessis, V.K.; Thomas, D.C.; Levine, A.J.; Breton, C.V.; Mack, T.M.; Siegmund, K.D.; Haile, R.W.; Laird, P.W. Environmental epigenetics: Prospects for studying epigenetic mediation of exposure-response relationships. *Hum. Genet.* **2012**, *131*, 1565–1589.

204. Hogg, K.; Price, E.M.; Hanna, C.W.; Robinson, W.P. Prenatal and perinatal environmental influences on the human fetal and placental epigenome. *Clin. Pharmacol. Ther.* **2012**, *92*, 716–726.

205. Mirbahai, L.; Chipman, J.K. Epigenetic memory of environmental organisms: A reflection of lifetime stressor exposures. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2014**, *764–765*, 10–17.

206. Bibikova, M.; Barnes, B.; Tsan, C.; Ho, V.; Klotzle, B.; Le, J.M.; Delano, D.; Zhang, L.; Schroth, G.P.; Gunderson, K.L.; *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **2011**, *98*, 288–295.

207. Dogan, M.V.; Shields, B.; Cutrona, C.; Gao, L.; Gibbons, F.X.; Simons, R.; Monick, M.; Brody, G.H.; Tan, K.; Beach, S.R.; *et al.* The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics* **2014**, *15*, 151.

208. Elliott, H.R.; Tillin, T.; McArdle, W.L.; Ho, K.; Duggirala, A.; Frayling, T.M.; Davey Smith, G.; Hughes, A.D.; Chaturvedi, N.; Relton, C.L. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin. Epigenetics* **2014**, *6*, 4.

209. Joubert, B.R.; Haberg, S.E.; Nilsen, R.M.; Wang, X.; Vollset, S.E.; Murphy, S.K.; Huang, Z.; Hoyo, C.; Midttun, O.; Cupul-Uicab, L.A.; *et al.* 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ. Health Perspect.* **2012**, *120*, 1425–1431.

210. Philibert, R.A.; Beach, S.R.; Lei, M.K.; Brody, G.H. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin. Epigenetics* **2013**, *5*, 19.

211. Shenker, N.S.; Polidoro, S.; van Veldhoven, K.; Sacerdote, C.; Ricceri, F.; Birrell, M.A.; Belvisi, M.G.; Brown, R.; Vineis, P.; Flanagan, J.M. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum. Mol. Genet.* **2013**, *22*, 843–851.

212. Zeilinger, S.; Kuhnel, B.; Klopp, N.; Baurecht, H.; Kleinschmidt, A.; Gieger, C.; Weidinger, S.; Lattka, E.; Adamski, J.; Peters, A.; *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* **2013**, *8*, e63812.

213. Monick, M.M.; Beach, S.R.; Plume, J.; Sears, R.; Gerrard, M.; Brody, G.H.; Philibert, R.A. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2012**, *159B*, 141–151.

214. Besingi, W.; Johansson, A. Smoke-related DNA methylation changes in the etiology of human disease. *Hum. Mol. Genet.* **2014**, *23*, 2290–2297.

215. Markunas, C.A.; Xu, Z.; Harlid, S.; Wade, P.A.; Lie, R.T.; Taylor, J.A.; Wilcox, A.J. Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy. *Environ. Health Perspect.* **2014**, doi:10.1289/ehp.1307892.

216. Harlid, S.; Xu, Z.; Panduri, V.; Sandler, D.P.; Taylor, J.A. CpG Sites Associated with Cigarette Smoking: Analysis of Epigenome-Wide Data from the Sister Study. *Environ. Health Perspect.* **2014**, *122*, 673–678.

217. Novakovic, B.; Ryan, J.; Pereira, N.; Boughton, B.; Craig, J.M.; Saffery, R. Postnatal stability, tissue, and time specific effects of AHRR methylation change in response to maternal smoking in pregnancy. *Epigenetics* **2014**, *9*, 377–386.

218. Joubert, B.R.; Haberg, S.E.; Bell, D.A.; Nilsen, R.M.; Vollset, S.E.; Midttun, O.; Ueland, P.M.; Wu, M.C.; Nystad, W.; Peddada, S.D.; *et al.* Maternal Smoking and DNA Methylation in Newborns: *In Utero* Effect or Epigenetic Inheritance? *Cancer Epidemiol. Biomark. Prev.* **2014**, doi:10.1158/1055-9965.EPI-13-1256.

219. Brokken, L.J.; Lundberg-Giwercman, Y.; Meyts, E.R.; Eberhard, J.; Stahl, O.; Cohn-Cedermark, G.; Daugaard, G.; Arver, S.; Giwercman, A. Association between polymorphisms in the aryl hydrocarbon receptor repressor gene and disseminated testicular germ cell cancer. *Front. Endocrinol.* **2013**, *4*, 4.

220. Cauchi, S.; Stucker, I.; Cenee, S.; Kremers, P.; Beaune, P.; Massaad-Massade, L. Structure and polymorphisms of human aryl hydrocarbon receptor repressor (AhRR) gene in a French population: Relationship with CYP1A1 inducibility and lung cancer. *Pharmacogenetics* **2003**, *13*, 339–347.

221. Liang, Y.; Li, W.W.; Yang, B.W.; Tao, Z.H.; Sun, H.C.; Wang, L.; Xia, J.L.; Qin, L.X.; Tang, Z.Y.; Fan, J.; *et al.* Aryl hydrocarbon receptor nuclear translocator is associated with tumor growth and progression of hepatocellular carcinoma. *Int. J. Cancer* **2012**, *130*, 1745–1754.

222. Tsuchiya, M.; Katoh, T.; Motoyama, H.; Sasaki, H.; Tsugane, S.; Ikenoue, T. Analysis of the AhR, ARNT, and AhRR gene polymorphisms: Genetic contribution to endometriosis susceptibility and severity. *Fertil. Steril.* **2005**, *84*, 454–458.

223. Lupien, S.J.; McEwen, B.S.; Gunnar, M.R.; Heim, C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* **2009**, *10*, 434–445.

224. Meaney, M.J.; Szyf, M. Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues Clin. Neurosci.* **2005**, *7*, 103–123.

225. Weaver, I.C.; Cervoni, N.; Champagne, F.A.; D'Alessio, A.C.; Sharma, S.; Seckl, J.R.; Dymov, S.; Szyf, M.; Meaney, M.J. Epigenetic programming by maternal behavior. *Nat. Neurosci.* **2004**, *7*, 847–854.

226. Oberlander, T.F.; Weinberg, J.; Papsdorf, M.; Grunau, R.; Misri, S.; Devlin, A.M. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics* **2008**, *3*, 97–106.

227. Hompes, T.; Izzi, B.; Gellens, E.; Morreels, M.; Fieuws, S.; Pexsters, A.; Schops, G.; Dom, M.; van Bree, R.; Freson, K.; *et al.* Investigating the influence of maternal cortisol and emotional state during pregnancy on the DNA methylation status of the glucocorticoid receptor gene (NR3C1) promoter region in cord blood. *J. Psychiatr. Res.* **2013**, *47*, 880–891.

228. Conradt, E.; Lester, B.M.; Appleton, A.A.; Armstrong, D.A.; Marsit, C.J. The roles of DNA methylation of NR3C1 and 11beta-HSD2 and exposure to maternal mood disorder *in utero* on newborn neurobehavior. *Epigenetics* **2013**, *8*, 1321–1329.
229. Radtke, K.M.; Ruf, M.; Gunter, H.M.; Dohrmann, K.; Schauer, M.; Meyer, A.; Elbert, T. Transgenerational impact of intimate partner violence on methylation in the promoter of the glucocorticoid receptor. *Transl. Psychiatry* **2011**, *1*, e21.
230. Mulligan, C.J.; D'Errico, N.C.; Stees, J.; Hughes, D.A. Methylation changes at NR3C1 in newborns associate with maternal prenatal stress exposure and newborn birth weight. *Epigenetics* **2012**, *7*, 853–857.
231. Yehuda, R.; Daskalakis, N.P.; Lehrner, A.; Desarnaud, F.; Bader, H.N.; Makotkine, I.; Flory, J.D.; Bierer, L.M.; Meaney, M.J. Influences of Maternal and Paternal PTSD on Epigenetic Regulation of the Glucocorticoid Receptor Gene in Holocaust Survivor Offspring. *Am. J. Psychiatry* **2014**, *171*, 872–880.
232. McGowan, P.O.; Sasaki, A.; D'Alessio, A.C.; Dymov, S.; Labonte, B.; Szyf, M.; Turecki, G.; Meaney, M.J. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat. Neurosci.* **2009**, *12*, 342–348.
233. Perroud, N.; Paoloni-Giacobino, A.; Prada, P.; Olie, E.; Salzmann, A.; Nicastro, R.; Guillaume, S.; Mouthon, D.; Stouder, C.; Dieben, K.; *et al.* Increased methylation of glucocorticoid receptor gene (NR3C1) in adults with a history of childhood maltreatment: A link with the severity and type of trauma. *Transl. Psychiatry* **2011**, *1*, e59.
234. Labonte, B.; Yerko, V.; Gross, J.; Mechawar, N.; Meaney, M.J.; Szyf, M.; Turecki, G. Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol. Psychiatry* **2012**, *72*, 41–48.
235. Tyrka, A.R.; Price, L.H.; Marsit, C.; Walters, O.C.; Carpenter, L.L. Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: Preliminary findings in healthy adults. *PLoS One* **2012**, *7*, e30148.
236. Melas, P.A.; Wei, Y.; Wong, C.C.; Sjoholm, L.K.; Aberg, E.; Mill, J.; Schalling, M.; Forsell, Y.; Lavebratt, C. Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities. *Int. J. Neuropsychopharmacol.* **2013**, *16*, 1513–1528.
237. Yehuda, R.; Flory, J.D.; Bierer, L.M.; Henn-Haase, C.; Lehrner, A.; Desarnaud, F.; Makotkine, I.; Daskalakis, N.P.; Marmar, C.R.; Meaney, M.J. Lower Methylation of Glucocorticoid Receptor Gene Promoter 1 in Peripheral Blood of Veterans with Posttraumatic Stress Disorder. *Biol. Psychiatry* **2014**, doi:10.1016/j.biopsych.2014.02.006.
238. Labonte, B.; Azoulay, N.; Yerko, V.; Turecki, G.; Brunet, A. Epigenetic modulation of glucocorticoid receptors in posttraumatic stress disorder. *Transl. Psychiatry* **2014**, *4*, e368.
239. Bromer, C.; Marsit, C.J.; Armstrong, D.A.; Padbury, J.F.; Lester, B. Genetic and epigenetic variation of the glucocorticoid receptor (NR3C1) in placenta and infant neurobehavior. *Dev. Psychobiol.* **2013**, *55*, 673–683.

240. Yehuda, R.; Daskalakis, N.P.; Desarnaud, F.; Makotkine, I.; Lehrner, A.L.; Koch, E.; Flory, J.D.; Buxbaum, J.D.; Meaney, M.J.; Bierer, L.M. Epigenetic Biomarkers as Predictors and Correlates of Symptom Improvement Following Psychotherapy in Combat Veterans with PTSD. *Front. Psychiatry* **2013**, *4*, 118.

241. Edelman, S.; Shalev, I.; Uzefovsky, F.; Israel, S.; Knafo, A.; Kremer, I.; Mankuta, D.; Kaitz, M.; Ebstein, R.P. Epigenetic and genetic factors predict women's salivary cortisol following a threat to the social self. *PLoS One* **2012**, *7*, e48597.

242. Weaver, I.C.; Champagne, F.A.; Brown, S.E.; Dymov, S.; Sharma, S.; Meaney, M.J.; Szyf, M. Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: Altering epigenetic marking later in life. *J. Neurosci.* **2005**, *25*, 11045–11054.

243. Weaver, I.C.; Meaney, M.J.; Szyf, M. Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3480–3485.

244. Essex, M.J.; Thomas Boyce, W.; Hertzman, C.; Lam, L.L.; Armstrong, J.M.; Neumann, S.M.; Kobor, M.S. Epigenetic Vestiges of Early Developmental Adversity: Childhood Stress Exposure and DNA Methylation in Adolescence. *Child Dev.* **2011**, doi:10.1111/j.1467-8624.2011.01641.x.

245. Uddin, M.; Aiello, A.E.; Wildman, D.E.; Koenen, K.C.; Pawelec, G.; de Los Santos, R.; Goldmann, E.; Galea, S. Epigenetic and immune function profiles associated with posttraumatic stress disorder. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9470–9475.

246. Borghol, N.; Suderman, M.; McArdle, W.; Racine, A.; Hallett, M.; Pembrey, M.; Hertzman, C.; Power, C.; Szyf, M. Associations with early-life socio-economic position in adult DNA methylation. *Int. J. Epidemiol.* **2012**, *41*, 62–74.

247. Mehta, D.; Klengel, T.; Conneely, K.N.; Smith, A.K.; Altmann, A.; Pace, T.W.; Rex-Haffner, M.; Loeschner, A.; Gonik, M.; Mercer, K.B.; *et al.* Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8302–8307.

248. Naumova, O.Y.; Lee, M.; Koposov, R.; Szyf, M.; Dozier, M.; Grigorenko, E.L. Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. *Dev. Psychopathol.* **2012**, *24*, 143–155.

249. Harvey, N.C.; Sheppard, A.; Godfrey, K.M.; McLean, C.; Garratt, E.; Ntani, G.; Davies, L.; Murray, R.; Inskip, H.M.; Gluckman, P.D.; *et al.* Childhood Bone Mineral Content Is Associated With Methylation Status of the RXRA Promoter at Birth. *J. Bone Miner. Res.* **2014**, *29*, 600–607.

250. Relton, C.L.; Groom, A.; St Pourcain, B.; Sayers, A.E.; Swan, D.C.; Embleton, N.D.; Pearce, M.S.; Ring, S.M.; Northstone, K.; Tobias, J.H.; *et al.* DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS One* **2012**, *7*, e31821.

251. Rakyan, V.K.; Beyan, H.; Down, T.A.; Hawa, M.I.; Maslau, S.; Aden, D.; Daunay, A.; Busato, F.; Mein, C.A.; Manfras, B.; *et al.* Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* **2011**, *7*, e1002300.
252. Clarke-Harris, R.; Wilkin, T.J.; Hosking, J.; Pinkney, J.; Jeffery, A.N.; Metcalf, B.S.; Godfrey, K.M.; Voss, L.D.; Lillycrop, K.A.; Burdge, G.C. Peroxisomal proliferator activated receptor-gamma-co-activator-1alpha promoter methylation in blood at 5–7 years predicts adiposity from 9 to 14 years (earlybird 50). *Diabetes* **2014**, *63*, 2528–2537.
253. Nguyen, A.; Rauch, T.A.; Pfeifer, G.P.; Hu, V.W. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J.* **2010**, *24*, 3036–3051.
254. Wong, C.C.; Meaburn, E.L.; Ronald, A.; Price, T.S.; Jeffries, A.R.; Schalkwyk, L.C.; Plomin, R.; Mill, J. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol. Psychiatry* **2014**, *19*, 495–503.
255. Wang, Y.; Fang, Y.; Zhang, F.; Xu, M.; Zhang, J.; Yan, J.; Ju, W.; Brown, W.T.; Zhong, N. Hypermethylation of the enolase gene (ENO2) in autism. *Eur. J. Pediatr.* **2014**, *173*, 1233–1244.
256. Berko, E.R.; Suzuki, M.; Beren, F.; Lemetre, C.; Alaimo, C.M.; Calder, R.B.; Ballaban-Gil, K.; Gounder, B.; Kampf, K.; Kirschen, J.; *et al.* Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genet.* **2014**, *10*, e1004402.
257. Ginsberg, M.R.; Rubin, R.A.; Falcone, T.; Ting, A.H.; Natowicz, M.R. Brain transcriptional and epigenetic associations with autism. *PLoS One* **2012**, *7*, e44736.
258. Ladd-Acosta, C.; Hansen, K.D.; Briem, E.; Fallin, M.D.; Kaufmann, W.E.; Feinberg, A.P. Common DNA methylation alterations in multiple brain regions in autism. *Mol. Psychiatry* **2013**, doi:10.1038/mp.2013.114.
259. Sun, C.; Burgner, D.P.; Ponsonby, A.L.; Saffery, R.; Huang, R.C.; Vuillermin, P.J.; Cheung, M.; Craig, J.M. Effects of early-life environment and epigenetics on cardiovascular disease risk in children: Highlighting the role of twin studies. *Pediatr. Res.* **2013**, *73*, 523–530.
260. Chang, C.P.; Bruneau, B.G. Epigenetics and cardiovascular development. *Annu. Rev. Physiol.* **2012**, *74*, 41–68.
261. Shirodkar, A.V.; Marsden, P.A. Epigenetics in cardiovascular disease. *Curr. Opin. Cardiol.* **2011**, *26*, 209–215.
262. Hidalgo, B.; Irvin, M.R.; Sha, J.; Zhi, D.; Aslibekyan, S.; Absher, D.; Tiwari, H.K.; Kabagambe, E.K.; Ordovas, J.M.; Arnett, D.K. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. *Diabetes* **2014**, *63*, 801–807.
263. Irvin, M.R.; Zhi, D.; Joehanes, R.; Mendelson, M.; Aslibekyan, S.; Claas, S.A.; Thibeault, K.S.; Patel, N.; Day, K.; Waite Jones, L.; *et al.* Epigenome-Wide Association Study of Fasting Blood Lipids in the Genetics of Lipid Lowering Drugs and Diet Network Study. *Circulation* **2014**, *130*, 565–572.

264. Campion, J.; Milagro, F.I.; Martinez, J.A. Individuality and epigenetics in obesity. *Obes. Rev.* **2009**, *10*, 383–392.

265. Toperoff, G.; Aran, D.; Kark, J.D.; Rosenberg, M.; Dubnikov, T.; Nissan, B.; Wainstein, J.; Friedlander, Y.; Levy-Lahad, E.; Glaser, B.; *et al.* Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum. Mol. Genet.* **2012**, *21*, 371–383.

266. Cordero, P.; Campion, J.; Milagro, F.I.; Goyenechea, E.; Steemburgo, T.; Javierre, B.M.; Martinez, J.A. Leptin and TNF-alpha promoter methylation levels measured by MSP could predict the response to a low-calorie diet. *J. Physiol. Biochem.* **2011**, *67*, 463–470.

267. Guintivano, J.; Arad, M.; Gould, T.D.; Payne, J.L.; Kaminsky, Z.A. Antenatal prediction of postpartum depression with blood DNA methylation biomarkers. *Mol. Psychiatry* **2014**, *19*, 633.

268. Wockner, L.F.; Noble, E.P.; Lawford, B.R.; Young, R.M.; Morris, C.P.; Whitehall, V.L.; Voisey, J. Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl. Psychiatry* **2014**, *4*, e339.

269. Nishioka, M.; Bundo, M.; Koike, S.; Takizawa, R.; Kakiuchi, C.; Araki, T.; Kasai, K.; Iwamoto, K. Comprehensive DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia. *J. Hum. Genet.* **2013**, *58*, 91–97.

270. Xiao, Y.; Camarillo, C.; Ping, Y.; Arana, T.B.; Zhao, H.; Thompson, P.M.; Xu, C.; Su, B.B.; Fan, H.; Ordonez, J.; *et al.* The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. *PLoS One* **2014**, *9*, e95875.

271. Aberg, K.A.; McClay, J.L.; Nerella, S.; Clark, S.; Kumar, G.; Chen, W.; Khachane, A.N.; Xie, L.; Hudson, A.; Gao, G.; *et al.* Methylome-wide association study of schizophrenia: Identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* **2014**, *71*, 255–264.

272. Abdolmaleky, H.M.; Nohesara, S.; Ghadirivasfi, M.; Lambert, A.W.; Ahmadkhaniha, H.; Ozturk, S.; Wong, C.K.; Shafa, R.; Mostafavi, A.; Thiagalingam, S. DNA hypermethylation of serotonin transporter gene promoter in drug naive patients with schizophrenia. *Schizophr. Res.* **2014**, *152*, 373–380.

273. Nishioka, M.; Bundo, M.; Kasai, K.; Iwamoto, K. DNA methylation in schizophrenia: Progress and challenges of epigenetic studies. *Genome Med.* **2012**, *4*, 96.

274. Kato, T.; Iwamoto, K. Comprehensive DNA methylation and hydroxymethylation analysis in the human brain and its implication in mental disorders. *Neuropharmacology* **2014**, *80*, 133–139.

275. Abdolmaleky, H.M.; Smith, C.L.; Faraone, S.V.; Shafa, R.; Stone, W.; Glatt, S.J.; Tsuang, M.T. Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2004**, *127B*, 51–59.

276. Abdolmaleky, H.M.; Yaqubi, S.; Papageorgis, P.; Lambert, A.W.; Ozturk, S.; Sivaraman, V.; Thiagalingam, S. Epigenetic dysregulation of HTR2A in the brain of patients with schizophrenia and bipolar disorder. *Schizophr. Res.* **2011**, *129*, 183–190.

277. Ghadirivasfi, M.; Nohesara, S.; Ahmadkhaniha, H.R.; Eskandari, M.R.; Mostafavi, S.; Thiagalingam, S.; Abdolmaleky, H.M. Hypomethylation of the serotonin receptor type-2A gene (HTR2A) at T102C polymorphic site in DNA derived from the saliva of patients with schizophrenia and bipolar disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2011**, *156*, 536–545.

278. Mill, J.; Tang, T.; Kaminsky, Z.; Khare, T.; Yazdanpanah, S.; Bouchard, L.; Jia, P.; Assadzadeh, A.; Flanagan, J.; Schumacher, A.; *et al.* Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* **2008**, *82*, 696–711.

279. Abdolmaleky, H.M.; Cheng, K.H.; Russo, A.; Smith, C.L.; Faraone, S.V.; Wilcox, M.; Shafa, R.; Glatt, S.J.; Nguyen, G.; Ponte, J.F.; *et al.* Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: A preliminary report. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2005**, *134B*, 60–66.

280. Grayson, D.R.; Jia, X.; Chen, Y.; Sharma, R.P.; Mitchell, C.P.; Guidotti, A.; Costa, E. Reelin promoter hypermethylation in schizophrenia. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9341–9346.

281. Kaminsky, Z.; Tochigi, M.; Jia, P.; Pal, M.; Mill, J.; Kwan, A.; Ioshikhes, I.; Vincent, J.B.; Kennedy, J.L.; Strauss, J.; *et al.* A multi-tissue analysis identifies HLA complex group 9 gene methylation differences in bipolar disorder. *Mol. Psychiatry* **2012**, *17*, 728–740.

282. Sugawara, H.; Iwamoto, K.; Bundo, M.; Ueda, J.; Miyauchi, T.; Komori, A.; Kazuno, A.; Adati, N.; Kusumi, I.; Okazaki, Y.; *et al.* Hypermethylation of serotonin transporter gene in bipolar disorder detected by epigenome analysis of discordant monozygotic twins. *Transl. Psychiatry* **2011**, *1*, e24.

283. Cruickshank, M.N.; Pitt, J.; Craig, J.M. Going back to the future with Guthrie-powered epigenome-wide association studies. *Genome Med.* **2012**, *4*, 83.

284. Tsai, P.C.; Spector, T.D.; Bell, J.T. Using epigenome-wide association scans of DNA methylation in age-related complex human traits. *Epigenomics* **2012**, *4*, 511–526.

285. Weidner, C.I.; Lin, Q.; Koch, C.M.; Eisele, L.; Beier, F.; Ziegler, P.; Bauerschlag, D.O.; Jockel, K.H.; Erbel, R.; Muhleisen, T.W.; *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **2014**, *15*, R24.

286. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**, *14*, R115.

287. Bibikova, M.; Le, J.; Barnes, B.; Saedinia-Melnyk, S.; Zhou, L.; Shen, R.; Gunderson, K.L. Geneom-wide methylation profiling using Infinium assay. *Epigenomics* **2009**, *1*, 177–200.

288. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sadda, S.; Klotzle, B.; Bibikova, M.; Fan, J.B.; Gao, Y.; *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **2013**, *49*, 359–367.

289. Gibbs, W.W. Biomarkers and ageing: The clock-watcher. *Nature* **2014**, *508*, 168–170.

290. Ogino, S.; Stampfer, M. Lifestyle factors and microsatellite instability in colorectal cancer: The evolving field of molecular pathological epidemiology. *J. Natl. Cancer Inst.* **2010**, *102*, 365–367.

291. Bishehsari, F.; Mahdavinia, M.; Vacca, M.; Malekzadeh, R.; Mariani-Costantini, R. Epidemiological transition of colorectal cancer in developing countries: Environmental factors, molecular pathways, and opportunities for prevention. *World J. Gastroenterol.* **2014**, *20*, 6055–6072.

292. Whelan, R.; Watts, R.; Orr, C.A.; Althoff, R.R.; Artiges, E.; Banaschewski, T.; Barker, G.J.; Bokde, A.L.; Buchel, C.; Carvalho, F.M.; *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **2014**, *512*, 185–189.

293. Kim, D.S.; Lee, J.Y.; Lee, S.M.; Choi, J.E.; Cho, S.; Park, J.Y. Promoter methylation of the RGC32 gene in nonsmall cell lung cancer. *Cancer* **2011**, *117*, 590–596.

Mouse ENU Mutagenesis to Understand Immunity to Infection: Methods, Selected Examples, and Perspectives

Grégory Caignard, Megan M. Eva, Rebekah van Bruggen, Robert Eveleigh, Guillaume Bourque, Danielle Malo, Philippe Gros and Silvia M. Vidal

Abstract: Infectious diseases are responsible for over 25% of deaths globally, but many more individuals are exposed to deadly pathogens. The outcome of infection results from a set of diverse factors including pathogen virulence factors, the environment, and the genetic make-up of the host. The completion of the human reference genome sequence in 2004 along with technological advances have tremendously accelerated and renovated the tools to study the genetic etiology of infectious diseases in humans and its best characterized mammalian model, the mouse. Advancements in mouse genomic resources have accelerated genome-wide functional approaches, such as gene-driven and phenotype-driven mutagenesis, bringing to the fore the use of mouse models that reproduce accurately many aspects of the pathogenesis of human infectious diseases. Treatment with the mutagen *N*-ethyl-*N*-nitrosourea (ENU) has become the most popular phenotype-driven approach. Our team and others have employed mouse ENU mutagenesis to identify host genes that directly impact susceptibility to pathogens of global significance. In this review, we first describe the strategies and tools used in mouse genetics to understand immunity to infection with special emphasis on chemical mutagenesis of the mouse germ-line together with current strategies to efficiently identify functional mutations using next generation sequencing. Then, we highlight illustrative examples of genes, proteins, and cellular signatures that have been revealed by ENU screens and have been shown to be involved in susceptibility or resistance to infectious diseases caused by parasites, bacteria, and viruses.

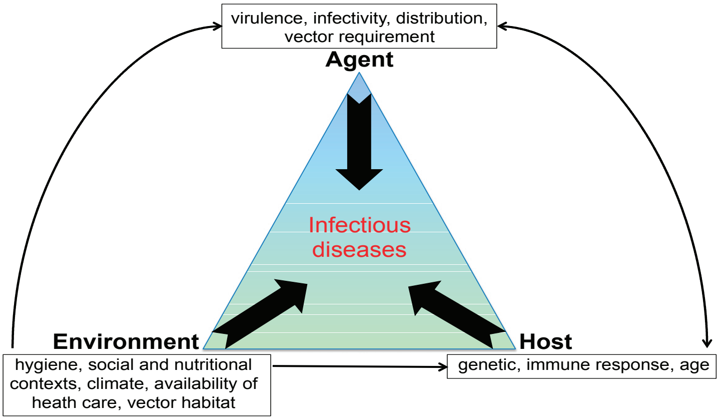
Reprinted from *Genes*. Cite as: Caignard, G.; Eva, M.M.; van Bruggen, R.; Eveleigh, R.; Bourque, G.; Malo, D.; Gros, P.; Vidal, S.M. Mouse ENU Mutagenesis to Understand Immunity to Infection: Methods, Selected Examples, and Perspectives. *Genes* **2014**, *5*, 887-925.

1. Introduction

The Neolithic Era, which began around 10,000 years B.C., constituted a turning point in human civilization. Its importance stems not only from the establishment of the first human settlements, but also from the development of farming activities involving the domestication of wild plants and animals. These changes in societal organization brought humans into close contact with animals and soil, exposing them to potential new pathogens, and with each other, allowing the spread of any new infection. It therefore comes as no surprise that the Neolithic Era saw the emergence of several human infectious diseases [1]. Indeed, given this close proximity, trans-species infections became more likely and ultimately resulted in the appearance of diseases such as measles and smallpox [2]. As a result, from the Neolithic Era until the Industrial Revolution, human life expectancy did not exceed 25 years of age [3]. Fortunately, life expectancy has been steadily increasing over the last 150 years for two main reasons. First, public hygiene measures implemented in the mid-19th century reduced the transmission of infection. Additionally, the advent of vaccination and antimicrobial drugs in the late 19th and early 20th century meant that many deadly infections were now curable or preventable. On a larger scale, diseases such as polio and measles were drastically reduced, while the dreaded smallpox was completely eradicated.

Nevertheless, infectious diseases remain directly responsible for close to 25% of all deaths globally and constitute a perpetual burden for humankind [4]. Numerous circumstances favor the emergence or reemergence of pathogens, or their spread to new ecological niches; these include pathogen virulence factors, as well as changing environmental conditions and host factors (e.g., aging populations, a heavier chronic disease burden, and therapeutic suppression of host defenses) (Figure 1). Of course, the eradication of most infectious diseases is highly unlikely. Instead, we are often involved in an unremitting struggle to control infection, for which a constant influx of novel countermeasure strategies is needed.

Figure 1. Factors involved in susceptibility to infectious diseases.


virulence, infectivity, distribution, vector requirement
Agent
Infectious diseases
Environment
Host
hygiene, social and nutritional contexts, climate, availability of heath care, vector habitat
genetic, immune response, age

The development of these novel countermeasure strategies largely relies on a better understanding of the molecular mechanisms of disease pathogenesis. This requires not only basic research on the pathogen side but also on its interaction with the host. A possible approach is to exploit the observed variability in the outcome of infection, since at any given time, even during epidemics, clinical disease only develops in a subset of exposed persons. A large body of evidence indicates that the human genome is a major determinant of the variability in the onset, progression, and severity of infectious diseases [5–8]. In light of this evidence, research efforts aiming to better understand the pathogenesis of infectious diseases have shifted their focus from the pathogen to the host. Investigators are thus now attempting to identify host genes that are essential for successful pathogen infection, instead of focusing solely on pathogen genes. Candidate gene analysis studies have revealed a handful of single gene variants associated with increased susceptibility or resistance to specific infectious diseases (reviewed in [5]). Some remarkable examples identified in human populations include the malaria-protective effect of heterozygosity in the case of otherwise disease-causing hemoglobinopathies, such as sickle cell anemia and thalassemia [9], the protective effects of *CCR5* mutations against HIV [10], and resistance to norovirus infection conferred by loss-of-function alleles of the *FUT2* gene [11]. Further, the study of children with rare monogenic defects has revealed a considerable number of rare human genetic variations in innate immune pathways that underlie susceptibility to certain infectious diseases. For example, *IRAK* and *MYD88* deficiencies predispose to life-threatening infection by some bacterial species [12]. Another example is Mendelian Susceptibility to Mycobacterial Disease (MSMD), a primary immunodeficiency characterized by genetic defects in the IFNγ pathway, leading to susceptibility to *Mycobacterium bovis* (BCG) or other environmental mycobacteria species innocuous to the general population and to non-typhoidal, extra-intestinal salmonellosis (for review, see [5]). Thus, the fact that individuals exposed to life-threatening pathogens display differential susceptibility to infection and varying disease outcome not only reflects the genetic variability within the human population, but also the functional genetic diversity of the immune response itself.

The growing awareness of the importance of host genetic makeup in infectious disease outcome has motivated large-scale investigations of the human genome, made possible by recent technological advances. Namely, sequencing of the human genome [13], the International HapMap project [14], and microarray-based high-throughput genotyping technology have paved the way to Genome Wide Association Studies (GWAS) of major infectious diseases. In these GWAS, millions of single nucleotide polymorphisms (SNPs) can be tested for association with major infectious diseases, and this can be done simultaneously in thousands of individuals (for review, see [5]). Results emanating from these large datasets are certainly improving our understanding of infectious disease pathogenesis. However, full interpretation of the genes and pathways identified by GWAS studies is complicated by several factors including the modest effect size of most signals and the fact that even together these signals can explain only a fraction of the genetic predisposition to disease. Furthermore, the SNPs showing the strongest association are usually found near gene-coding regions rather than within obvious structural or regulatory regions making it difficult to pinpoint the gene directly involved in the disease phenotype. Such results are not entirely surprising given the inherent genetic heterogeneity of the human population, the variable exposure to the microbe during natural infection, the inherent variation in the microbe itself, and the difficulty associated with assembling the large cohorts required for GWAS. Yet, another key roadblock of GWAS studies is the lack of functional annotation for the majority of genes and encoded proteins, which is often limited to general ontology terms but lacks experimental validation for a possible role in an infectious disease phenotype.

2. Mice to the Rescue

An alternative and successful approach to identifying and characterizing the genetic component of the host response to infection in human studies has been the use of the mouse model. Owing to their striking physiological and genetic similarity with humans, mice have become a prime model for the study of human diseases. Numerous inbred strains exist that display natural resistance or susceptibility to a similar range of fungal, viral, parasitic, and bacterial pathogens, as well as the disease phenotypes associated with these infections [15–18]. These inbred strains represent homogeneous populations that serve to test different routes of inoculation, and various pathogen doses, all in a controlled environment, thus lessening many of the confounding effects encountered in human genetic studies. Due to its prominent role in biomedical research, the mouse was selected as the first non-human mammal to have its genome sequenced [19], revealing an astonishing genetic homology between the two species. The mouse and human genomes are approximately the same size, contain the same number of genes and show extensive conservation in gene order. Namely, 80% of human genes had 1:1 orthologous relationships with mouse genes, likely maintaining ancestral function in both species [20]. Mutations that cause diseases in humans often cause similar diseases in mice, including defects in the genes of the immune system [21]. Yet another advantage of the mouse is the string of unique technological advantages to manipulate the mouse genome.

Using the mouse model, two major genetic approaches have been employed to dissect the genetic architecture of the host defense against pathogens. The first is the so-called reverse genetic or gene-driven approach. In this approach, the sequence or expression of a gene of interest is altered, the effects of which are then investigated. Genetic modification of the mouse genome can be undertaken in various ways: (1) transgenesis or the introduction of gene DNA sequences into oocytes; (2) targeted mutation using embryonic stem cells (ES) which are modified to create knock-out alleles, whereby the function of the gene is abolished and equivalent to a null allele, or knock-in alleles resulting from the introduction of putative mutations in a given gene. In addition, recently developed genomic resources have further facilitated the use of genetically modified mice by the scientific community. These include large libraries of knock-out and conditional knock-out mice produced by international consortia aiming to target every gene in the mouse genome [22] and their accompanying large-scale phenotyping initiatives [23]; (3) targeted mutation in zygotes using the Clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated (Cas9) system [24]. With this approach it is possible to efficiently produce mice with mutations in both copies of multiple genes in a matter of weeks [25]. The phenotypes of these genetically modified mice can then be thoroughly scrutinized to determine the function of a gene in the context of the whole organism. These tools are dramatically improving our understanding of the genetic etiology of infectious diseases in both mice and humans. However, in many instances, these reverse genetics experiments can prove to be inconclusive. This is the case, for example, when the inactivation of a gene results in embryonic lethality or, conversely, when the resulting phenotypes are only slightly different from the wild-type or even undistinguishable because of gene redundancy. The reverse genetics approach also requires a preliminary hypothesis for gene function. Yet, as of 2014, less than 50% of about 34,000 known mouse genes (coding or not) have some form of functional annotation based on experimental evidence [26–28], which shows how our understanding of gene function still lags behind our knowledge of gene sequence.

The second approach is known as forward genetics, sometimes called phenotype-driven. The forward genetics approach begins with an inherited phenotype, with the aim of identifying the genomic regions and variant(s) underlying it. This involves the production of segregating crosses of inbred mouse strains or panels of specialized strains that display varying responses to infection, followed by linkage or association analyses. This approach is unbiased and requires no prior knowledge of gene function, allowing the discovery of unsuspected mechanisms. Numerous laboratory mouse resources are readily accessible for use in these studies: homozygous inbred strains, panels of selectively bred strains, consomic strains [29], recombinant congenic strains [30–32] or recombinant inbred strains from the collaborative cross [33]. A growing number of wild-derived inbred strains [34] or outbred crosses [35] can also be obtained, increasing the pool of genomic variation available for these studies. Whole genome sequencing has been performed on 18 of the most commonly used inbred mouse strains; the results are now public [36,37], facilitating the identification of candidate genes underlying a given phenotypic variation. Moreover, forward genetics studies in mice have already been shown to work; some elegant examples have allowed the identification of a number of genes and proteins that are essential for the early detection of and response to many invading pathogens (for review see [38]). In some cases, the human orthologues of these mouse genes (e.g., *NRAMP1*, *TLR4*, *IRF8*) have also been associated with predisposition to infection in humans, providing evidence of evolutionary conservation of immune defense mechanisms. However, there are limitations to this forward genetics strategy. Namely, a given genetic effect may be complex, making it difficult for investigators to determine the contribution of individual genes, as this requires subsequent breeding of congenic mice over several generations followed by positional cloning. Identifying the precise nature of a genetic lesion in a given candidate gene can also be complicated for other reasons, such as the presence of multigenic families or unrelated genes within the candidate interval bearing various coding polymorphisms, or predictive regulatory mutations or splicing variants rendering it difficult to identify the causative variant. Many of these drawbacks, however, can be overcome by the use of mutagens that introduce random mutations in the germ line. As presented later, in these models the causative mutation can be more easily identified by comparison with the parental non-mutagenized strain. This functional genomic strategy has successfully advanced our understanding of the intricate cellular and molecular cascades involved in immunodeficiency, autoimmunity, or behavioral disorders, which have already been well documented by others (see [39–42]). In the remainder of this review, we present the advantages and how-to of experiments using chemical mutagenesis of the mouse germ-line to dissect the genetic architecture of immunity to infection in mice. We also detail the procedures required to identify causal mutations underlying altered phenotypes using next generation sequencing. Finally, we highlight some of the most important findings from *in vivo* screens in the area of infectious disease research and discuss perspectives for mouse ENU approaches.

3. Chemical Mutagenesis and Generation of Mice Carrying Homozygous ENU-Induced Mutations

To better understand the link between genotypes and phenotypes, and ultimately gene function, mouse geneticists have elaborated upon several methods capable of introducing random mutations in the mouse germ-line, with the aim of expanding the phenotypic diversity in inbred mice and thus providing a wider range of research objects. These methods include the use of whole mouse radiation [43], infection of pre-implantation embryos with retroviruses [44], and injection with chemicals, such as procarbazine, methyl ethane sulfonate (MES), and *N*-ethyl-*N*-nitrosourea (ENU) [45]. ENU mutagenesis, however, has become the most popular technique to induce germ-line mutations due to its advantageous attributes: potency, preferential activity in spermatogonial stem cells, and a propensity to introduce point mutations.

As early as 1979, W. L. Russell demonstrated that a single dose of ENU was significantly more active than X-ray or procarbazine treatment, the most commonly used mouse mutagens at the time [46]. Later, studies showed that the mutation frequency could be increased if the ENU dose was fractionated and injected on a weekly schedule instead of being administered in one large dose, as this allowed a higher total dose to be tolerated [47]. In these conditions, the activity of ENU was 12 times that of X-rays and 36 times that of procarbazine, as well as being over 200 times the rate of spontaneous mutation [48]. The rate of ENU mutation appears variable for each gene, ranging from 1.5 to 10^{-3} per locus, which is equivalent to obtaining a mutation in a gene of choice at a rate of one in every 200–700 gametes screened. Additionally, it was noted that compared to X-ray-generated deletions, ENU rarely induced mutations in closely linked loci, suggesting that mutations introduced by ENU are subtler. Finally, compared to procarbazine, which is more active in transient post-meiotic cells, ENU preferentially affects spermatogonial stem cells, which are multiplied and replenished during the mouse lifetime, allowing the genetic lesions to be recovered indefinitely.

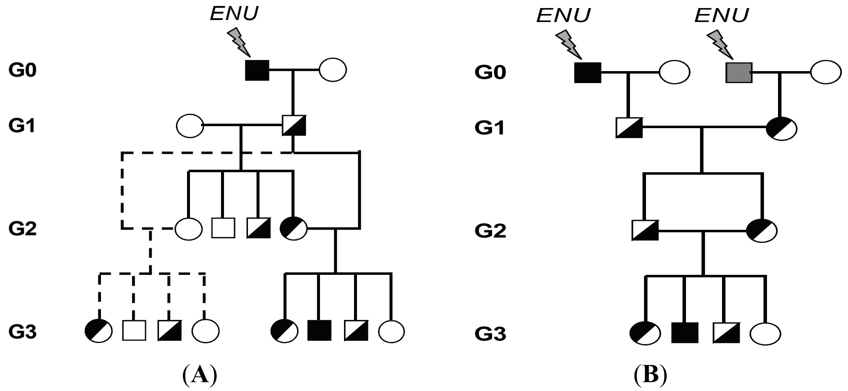
ENU is an alkylating agent that acts by preferential transfer of its ethyl group to O and N radicals in genomic DNA within mammalian cells [49,50]. Binding of the ethyl to the nucleoradicals creates DNA adducts that provoke mispairing, resulting mainly in base-pair substitutions if not restored by enzymatic DNA repair mechanisms during replication [51,52]. Systematic analysis of the type and frequency of ENU mutations was recently done using whole-exome and whole-genome sequencing [53–55]. Genome-wide, ENU has an average point mutation rate of 1.5 per Mb of genomic DNA [55], with a bias for AT to GC transitions (45%) compared to AT to TA transversions (28%). The size of a given target gene and its AT density can therefore explain, at least in part, the variable sensitivity to the mutagenic effects of ENU. With a mouse genome size of about 2.7 Mb including 1.5% of protein coding sequence, one can expect about 1,900 new sequence variants per genome of which about 30 are coding.

With a few exceptions (microRNA and cis-elements) [56,57], to date most ENU-induced phenotypes, whose corresponding genotype has been identified, result from nucleotide changes that alter the coding sequencing. A current survey of the Mutagenix database (http://mutagenetix.utsouthwestern.edu/home.cfm) which contains the largest collection of ENU-induced phenotypic mutations (N = 185) [58], revealed that 61% were missense mutations, 19% nonsense alleles, 18% splicing defects, and 2% were frame-shift mutations. Therefore, while targeted mutations producing null alleles are necessary for genetic dissection of phenotypic traits, ENU-induced point mutations can also be used in parallel, revealing the multiple functions of a gene by altering individual protein domains and splicing products. Further, point mutations can produce various types of allelic series: (1) hypermorphic or hypomorphic alleles (increased or reduced activity of the gene product, respectively); (2) antimorphic alleles (the gene product is antagonistic to the wild-type allele); or (3) neomorphic alleles (new molecular function) [59] which can display a broad range of possible phenotypes.

The phenotypes that arise following ENU mutagenesis segregate with different inheritance patterns. Autosomal recessive (68%) is the most commonly observed, followed by dominant or co-dominant segregation (23%); X-linked recessive (4%) or X-linked dominant (1%) are rare, though 4% remain uncharacterized [58]. Once male mice have been treated with ENU, they are crossed to female mice. The resulting large cohorts of offspring are then tested to identify the phenotypically distinct mice most likely to bear a large-effect mutation; this is usually done with dominant or recessive screens. The above data illustrates how recessive screens, which require a three-generation breeding scheme (see below), constitute a more efficient and inclusive design than dominant screens, although the latter are logistically simpler and quicker to conduct since only the first generation offspring are analyzed. Using different breeding schemes, these recessive screens have successfully advanced our understanding of the intricate cellular and molecular cascades involved in immunodeficiency and autoimmunity, as well as in neurological or behavioral disorders, as already reviewed by others (see [39–42]).

Methods for mutagenizing male mice and breeding protocols to recover homozygous mutations have been described previously [60–62]. In our laboratory, we use a recessive screen involving genetically related mouse strains to generate the collections of mutant mice (Figure 2). By using genetically related inbred strains, the number of animals used and the timeline of the experiment can be reduced, as the mice that are screened also serve for mapping of the ENU-induced phenotypes. Moreover, using closely related strains alleviates any possible second-site modifier gene effects that could be present in the mapping strain. Briefly, we use well-validated protocols to induce single nucleotide mutations in 129S1/SvImJ (129S1) and C57BL/6 (B6) mice. This is done using a single intraperitoneal (i.p.) injection of 150 mg/kg of ENU (129S1) or three weekly i.p. injections of 90 mg/kg (B6) [63]. Following treatment, spermatogenesis ceases transiently and fertility is then regained after 11–13 weeks. In a general breeding strategy (Figure 2A), generation 0 (G0) males are then out-crossed with wild-type female mice to produce G1 offspring. These G1 hybrids carry one full set of mutagenized chromosomes and one full set of wild-type chromosomes. Individual G1 males are bred as founders of separate pedigrees, with the aim of bringing B6 or 129S1 sequence variants to homozygosity. To achieve it, G1 males are first crossed with genetically related wild-type females (129X1/SvJ (129X1) females for 129S1 males and C57BL/10 (B10) females for B6 males) to distinguish mutation-bearing chromosomes while preventing the introduction of additional genetic modifiers. The mutations present in the G1 founders are thus propagated in the G2 progeny. Since each G2 offspring should inherit only 50% of sequence variants present in the G1 males, two G2 daughters are backcrossed to their G1 father. This produces G3 progeny, where 12.5% of the G1 sequence variants should come to homozygosity in any given G3 offspring. On average, each G3 offspring is thus expected to be homozygous for about four to five loss-of-function sequence variants of the 30 present in the G1. Therefore, if there is a recessive Mendelian immune variant segregating within a pedigree, researchers can expect to identify 25% of individuals with the same trait or a cluster of two to four deviants by initially screening about 16 G3 offspring in that pedigree. The clustering of heritable variants within a pedigree filters out unavoidable false positives, which occur at a low rate (~5%) in screens for host susceptibility to infection; typically only one individual constitutes a false positive in a given pedigree. Variations of this breeding strategy have been used (Figure 2B) and will be described in the corresponding sections.

Figure 2. Breeding strategies used in our program to produce mice carrying homozygous *N*-ethyl-*N*-nitrosourea (ENU) mutations. (**A**) Treatment with ENU introduces mutations (indicated by a black or gray square) in the germ-line of males of generation 0 (G0). The mutagenized G0 male is out-crossed to a wild-type female to produce first generation (G1) animals. First generation G1 mice are carriers of ENU-induced mutations (indicated by half-filled black squares). G1 males are mated to wild-type females, to produce second generation (G2) animals, which carry about half of the mutation load present in the parental G1. Two G2 daughters are backcrossed to their G1 father to yield third-generation (G3) mice, where the original mutations have been brought to homozygosity (filled black squares). About 25% G3 progeny are expected to present a deviant phenotype in pedigrees that bear a given relevant recessive mutations; (**B**) In this strategy, the G1 progeny of two independent G0 males are intercrossed to produce G2 animals, which in turn are intercrossed to produce G3 mice.



This pedigree structure allows early mapping of heritable variants. At this point, breeding and screening of additional G3 siblings confirm the inheritance of Mendelian recessive infectious traits in one quarter of the offspring. If eight to ten G3 animals displaying a new recessive immune trait are obtained out of 40–50 G3 mice in the pedigree, a genome-wide scan can be performed to establish linkage of the variant to a large initial segment. Before the advent of next generation sequencing (NGS), a time consuming and labor intensive positional cloning approach had to be undertaken to identify candidate genes bearing new genetic variants. The use of NGS techniques has dramatically increased the pace of mutation identification.

4. Gene Identification

The materials and methods underlying phenotype-driven or forward genetics approaches have become considerably more powerful over the years. Traditionally, these approaches required laborious genetic and fine mapping procedures in order to refine regions of interest to large megabase (Mb) chromosomal loci for subsequent PCR amplification and direct sequencing. Nonetheless, they were the methods of choice for the discovery of novel genes and/or novel gene functions in both humans and mice. The introduction of NGS has revolutionized forward genetics approaches, as it allowed the elaboration of robust methods of systematic mutation discovery, thus further closing the gap between phenotype and genotype. However, the sequencing and analysis of whole mammalian genomes remain a substantial bottleneck for many laboratories, both financially and computationally. Instead, inexpensive alternatives have been favored in order to sequence mouse mutations, namely targeting approaches using minimal mapping data. Moreover, targeted sequencing of coding regions of the genomes, or exomes, are particularly relevant for large mutational collections and have become the standard in cases where high-throughput gene mutation discovery methods are needed [64–66]. We describe below some of the standard techniques for sequencing and analysis of *de novo* mutations generated within ENU mouse models in a rapid and unbiased fashion.

Currently, the most widely used commercial mouse exome capture panels (Agilent and NimbleGen) target approximately 37 Mb of the sequences contained within the consensus coding sequence (CCDS) database of the genome, as well as other genomic features (e.g., microRNAs) [53,67] (see Table 1). The protocols contained in each of these kits are very similar. First, labeled DNA (or RNA) baits ranging from 55 to 120 bases are hybridized to fragmented genomic DNA. The baits are pulled down using magnetic beads, and the "captured" genomic fragments are then sequenced using NGS instruments such as SOLiD, Illumina or Roche 454.

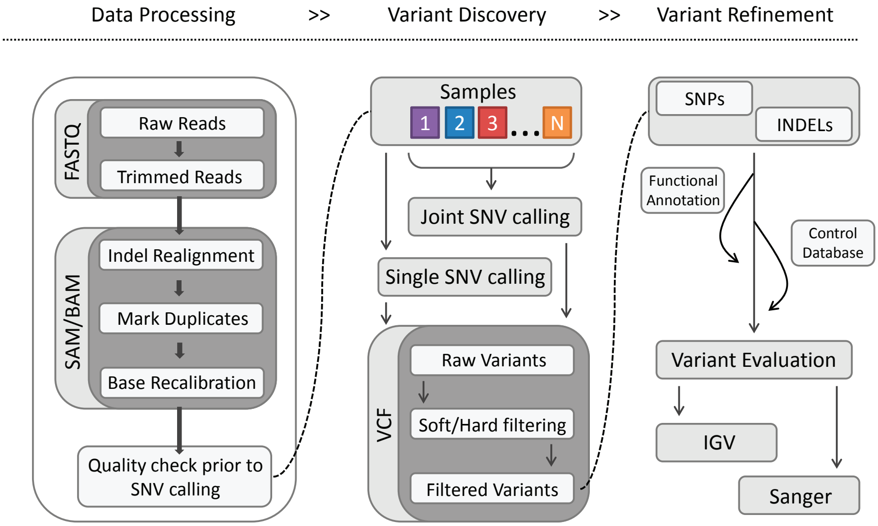
Table 1. Comparison between two standard whole exome mouse capture kits.

	Agilent Sureselect Mouse All Exon	Nimblegen SeqCap Ez
Probe size	120 bases	55–105 bases
Target Region size	49.6 Mb	54.3 Mb
Probe Type	RNA	DNA
Number of Targeted Exons	221,784	203,225

Mutation identification and ultimately gene discovery in the context of ENU-designed projects require significant computational analyses, where sequenced DNA fragments are mapped to a mouse reference sequence (C57BL/6J) [68] or to that of a specific mouse strain when available [36], followed by post alignment and variant calling procedures. For a given mouse sample, these procedures typically produce a large amount of single nucleotide variants (SNVs) and insertion/deletions (INDELs), which, depending on the sample's genetic background and coverage, can range from a few thousand to hundreds of thousands in more divergent strains. Further steps are required to filter the strain specific variants if the reference sequence of the mouse background is not used. This can be accomplished, for example, by adding more controls.

Numerous workflows (e.g., Genome Analysis Toolkit (GATK) best practices [69] and McGill University and Genome Quebec Innovation Centre (MUQGIC) [70]) have been designed for mutation discovery. Although each design may vary with regards to the steps and computational programs utilized, the underlying principle of these workflows remains the same. Each one divides the processing and analysis of sequencing data into three key steps: (1) data processing for quality control and filtering of sequenced reads; (2) variant discovery through alignment of filtered reads to known reference genomes; and (3) variant refinement leading to variant calling to identify mutations of interest. A flow diagram similar to GATK best practices [71] but with subdivided steps in file format is shown (Figure 3).

Figure 3. A typical workflow to identify causative mutations in genomic data. The procedures are separated into three general processes: (1) data processing, where raw sequencing data (fastq format) is aligned (sam/bam file format) to a known genome reference followed by alignment improvement steps (*i.e.*, indel realignment, mark duplicates and base recalibration); (2) a variant discovery step in which single nucleotide variants (SNVs) are called from aligned data followed by subsequent filtering (using variant quality thresholds; hard filtering, or Genome Analysis Toolkit (GATK) variant recalibration; and soft filtering); (3) and a variant refinement step to reduce the number of candidate mutations to a manageable number for further validation using Integrative Genomics Viewer (IGV) and/or Sanger sequencing [71].



The sequenced reads (in fastq file format) are usually derived from the instrument specific base-calling algorithm (or subsequent steps therein) and contain an identifier for each raw DNA fragment, as well as a phred quality score for each base in the fragment. The raw reads are aligned to a reference genome following a quality control step or “trimmed” to obtain a high quality set of reads for sequence alignment file (sam/bam) generation. The trimming step removes adaptor sequences from the raw reads and optionally removes bases at the 3' end using a specified phred quality threshold, and/or performs a size selection filtering step (e.g., trimmomatic [72]; Figure 3). The trimmed reads are aligned by using either a “hashing” or an effective data compression algorithm called the “Burrows-Wheeler transform” (BWT). Fast, memory-efficient BWT-based aligners, such as BWA [73], are often used in NGS studies. However, these aligners tend to be less sensitive than recent hash-based aligners, such as Novoalign [74], which conversely tend to require more computational resources [75].

Numerous software packages such as GATK [69], samtools [76], and Picard [77] have been developed to attempt to correct for biases incorporated at the sequencing and alignment phases, thus improving variant detection (Figure 3). During library construction and sequencing, duplicated DNA fragments produced by polymerase chain reaction (PCR) amplification and optical duplicates can occur. Software package such as Picard markDup and Samtools rmdup remove or flag potential PCR duplicates if both mates (in the case of paired-end reads) contain the same 5' alignment positions.

At the alignment phase, due in part to the heuristics of the alignment algorithm and the alignment scoring procedure, refinement of mapped reads near indels (GATK indel realigner [69]) and quality scores (GATK base recalibration [69]) are typically required to help reduce false positive variants in downstream analysis. Utilizing these two post-alignment programs, GATK indel realigner transforms regions with misalignments generally introduced by indels into clean reads containing fewer mismatches, whereas base recalibration improves the quality score to better reflect the true base-calling error rates by correcting for variation in quality with respect to machine cycle, sequence context, and other attributes.

To identify the protein-encoding mutations induced by ENU, numerous variant-calling procedures can be employed to convert base calls and quality scores into a set of genotypes on a per sample basis. The most recent variant callers, such as GATK [69], Samtools [75], and FreeBayes [78], use sophisticated statistical models that can be extended to incorporate additional information regarding allele frequencies and/or linkage disequilibrium (LD) patterns. Furthermore, joint analysis of multiple individuals can further improve genotype calling for single samples by taking into account allele frequencies or genotype frequencies [79].

Variant detection programs convert the refined base-calls and quality scores resulting from the post-alignment process and generate variant data containing information regarding the genomic position, SNV quality, *etc.*, of each variant. Generally, thousands of SNVs are generated by the detection protocol. Further annotations and filtering procedures are thus required to identify the expected 50–100 ENU-induced mutations [80]. The use of functional annotation programs such as snpEff [81] and VEP [82], coupled with the exclusion of known variants (for example, on the basis of SNP data from the dbSNP database [83]) and of variants falling below acceptable quality metrics (QUAL, genotype quality (GQ), strand bias, *etc.*), can help to preferentially identify protein coding mutations. However, despite rigorous post-alignment refinement and variant exclusion criteria, recurrent false positive SNVs remain. By comparing a set of ENU samples to unrelated genome or exome sequencing data sets, as well as to mouse genomes data from the Sanger Institute [68] generated using the same analysis workflow, variants commonly shared between related strains or systematic false positives arising from mapping issues related to genome structure (e.g., repetitive or paralogous sequences) or errors (e.g., miss-annotated reference allele) can be flagged for removal. In numerous studies this procedure has proven successful in prioritizing candidate mutations and decreasing their numbers [54,80], and has helped reduce the time requirements and cost of visual inspection (e.g., Integrative Genomics Viewer (IGV) [84]), of Sanger sequencing [85], of validation, and ultimately of novel mutation/gene discovery.

ENU experiments have successfully identified candidate causative mutations residing in protein coding sequences, splice sites or UTRs. However, these causative mutations are not always successfully identified due to either the fact that they may reside in uncaptured regions (*i.e.*, non-coding regions, regulatory regions or un-annotated coding sequences that are not captured by the capture design) or to biases in standard mapping and variant calling procedures. Therefore, further improvements are required in the development of software tools in order to better deal with regions of the genome that are difficult to map (e.g., paralogous sequences and GC-rich regions). The design of exome capture kits must also be improved to extend the set of captured regions. Alternatively, whole genome sequencing may also be a way to identify mutations in regions not captured by whole exome sequencing.

5. Infectious Screens

Establishing an ENU mutagenesis program with the aim of identifying genes involved in the host response to pathogens presents particular challenges. The first is the choice of a pathogen relevant to human health. Mouse models of infection with this pathogen must be available and representative of the corresponding human pathology. Also, the contribution of genetic factors in human and/or mouse response to this pathogen must be proven to support the feasibility of a genetic screen. The second challenge is the choice of the inbred mouse strain to be used for mutagenesis. There is ample evidence that the ENU sensitivity of inbred mice is genetically controlled and thus widely variable across strains [63]. This must be balanced with the varying susceptibility or resistance of inbred strains to infection with specific pathogens. The third challenge is the choice of the screening phenotype. Cell-based phenotypes have been used successfully to identify fundamental mechanisms of innate and acquired immunity [36,37]. The findings, however, require further validation in mouse models to determine a possible role in the infectious process. A clinically relevant, robust, and unequivocal *in vivo* phenotype is also attractive, as it will lead to the identification of the most important molecular determinants for a given infection; it will also minimize the appearance of false positives. Such phenotypes include severe disease (in terms of clinical evaluation or pathogen load) or death, when the mutagenized strain is resistant, or survival, when the mutagenized strain is innately susceptible, following infection. As presented below both screening approaches have led to the identification of key molecules involved in susceptibility or resistance to infectious diseases caused by parasites, bacteria, and viruses.

6. Malaria Parasites

Infecting hundreds of thousands of people every year, malaria is a significant cause of morbidity and mortality in developing countries (www.who.org). Having co-existed with humans for centuries, malaria has exerted a significant selective pressure on the human genome [16,86]. Likely the best-known selection has been the retention of deleterious hemoglobinopathies, such as sickle cell anemia, in malaria endemic regions [87,88]. Other variants associated with reduced susceptibility to malaria infections include those affecting erythrocyte proteins [89–94], the scavanger receptor CD36 [95,96], and elements of the host immune response, including human leukocyte antigen

(HLA) [97] and tomor necrosis factor-alpha (TNF-α) [98], among others [99,100]. Despite these clear examples, the genetic component influencing the human response to malarial parasites is complex, multigenic, and influenced by various environmental factors, including parasite virulence [101–103].

Cerebral malaria (CM) is the most severe and lethal complication of *Plasmodium falciparum* infection in humans [104,105]. Prevalent in immunologically naïve children, CM is characterized by high fever and a rapid progression to severe cerebral symptoms including impaired consciousness, seizures, and coma [106,107], resulting in death in about 20% of all cases [16,107]. During CM, parasitized erythrocytes (pRBCs) become trapped within the brain microvasculature [103], triggering a strong pro-inflammatory response [104,105] leading to the activation of the vascular endothelium [106], as well as the recruitment of immune cells and activated platelets [108–110]. This host-directed immune response results in the disruption of blood-brain barrier integrity [111], suggesting that CM pathogenesis is at least partially caused by over-activation of the inflammatory response [16,106,107]. By gaining a more thorough understanding of this disease, including of the host genetic factors affecting differences in susceptibility, novel and more effective prophylactic and therapeutic interventions can be developed.

Mice infected with *Plasmodium berghei* ANKA (PbA) have been used as a model of CM (experimental cerebral malaria, ECM). Mice susceptible to ECM develop neurological symptoms between days five to eight post-infection, including ataxia, hind limb paralysis, coma, and death [112]. ECM-resistant mice survive the cerebral malaria phase, but subsequently succumb to hyperparasitemia and resultanting anemia within three weeks post-infection [16]. Informative crosses between mouse strains of varying degrees of susceptibility to PbA have revealed at least nine quantitative trait loci (QTL) that modulate the host response to ECM [113–118]. However, these methods have failed to identify the causative genes, due in part to the large size of the genomic region and to the high number of positional candidates under the QTL peaks [119]. By introducing random point-mutations and small deletions within a susceptible genetic background, such as B6, B10, or 129S1, ENU-mutagenesis allows for the interrogation and determination of genes that are involved in resistance to ECM.

6.1. Screening for Acquired Resistance to Cerebral Malaria

We have successfully utilized ENU-mutagenesis to identify genes responsible for controlling susceptibility to ECM [119]. Male B6 mice (G0) were mutagenized with the administration of three consecutive i.p. injections of ENU. These G0 males were then bred to wild-type B10, 129S1, or B6 females to establish heterozygous G1 offspring. G1 males were out-crossed a second time to wild-type susceptible females to form the G2 generation. One to two G2 females per pedigree were backcrossed to the paternal G1 to produce G3 offspring, fixing mutations to homozygousity in approximately 25% of all animals (Figure 2A). G3 mice were infected with 10^6 *Plasmodium berghei* ANKA-parasitized RBCs by intravenous injection. The appearance of neurological symptoms and survival time were used as phenotypic markers of ECM disease [119]. Phenodeviant pedigrees were defined as those exhibiting >17% resistant pups in at least three litters or 10 offspring, whichever came first.

Enhanced laboratory resources and technological advances have allowed us to implement three variations of the general protocol outlined above. The first screen out-crossed mutagenized G0 males to the B10 genetic background. G3 animals from this cross were phenotyped for ECM-resistance. To facilitate linkage mapping, G1 males identified as segregating an ECM-resistant phenotype were out-crossed to 129S1 wild-type females. The resulting F1s were intercrossed randomly to generate F2 offspring, which were then phenotyped. Pedigrees identified as resistant were then analyzed for linkage analysis using a genome scan. A total of 6062 G3 mice from 244 G1 males were screened, generating nine phenodeviant pedigrees, with a background survival of approximately 2.8%. From this screen, we have identified an ECM protective mutation in *Jak3* (*Jak3*^{W81R}) [119]. A cytosolic tyrosine kinase that interacts with the common γc chain of cytokine receptors (IL-2, -4, -7, -9, -15, -21), JAK3 is required for STAT family members dependent transcriptional development and activation of inflammatory pathways in NK, T, and B cells [120]. *Jak3*^{W81R} mutants exhibit reduced numbers of NK cells, CD8^{+} T cells, and B cells, as well as severely reduced production levels of IFNγ by CD4^{+} T cells. We also demonstrated that tasocitinib, a JAK3 inhibitor used clinically to treat rheumatoid arthritis (RA) and Crohn's disease (CD), can reduce neuroinflammation and increase survival of *Jak3*$^{-/+}$ heterozygotes in the ECM model [119]. Genetic variants in JAK and STAT family proteins have been identified as causing certain primary immunodeficiencies and are also associated with chronic inflammatory diseases, such as inflammatory bowel disease (IBD), multiple sclerosis (MS), and systemic lupus erythematosus (SLE) in humans [121–123].

With respect to the second screen, we have out-crossed the mutagenized G0 males to the 129S1 genetic background. The 129S1 strain produces larger litters, allowing for the generation of larger numbers of G3 animals. Additionally, out-crossing directly to the 129S1 background eliminated the requirement to complete additional out-crossing of phenodeviant animals. Twenty-eight phenodeviant pedigrees were identified following the screening of 7705 G3 animals in 220 pedigrees, with a background survival of almost 8%. An epistatic interaction between the B6 and 129S1 genetic backgrounds on Chromosomes 4 and 1 was identified in 10 of the 28 phenodeviant pedigrees, potentially masking the effect of ENU-mutagenesis [124]. However, several mutations were identified in this screen, including an abrogated splicing mutation of Exon 6 in the winged-helix transcriptional regulator *Foxn1* gene [125] *Foxn1* mouse mutants are athymic and severely immuno-compromised, while human *FOXN1* mutations cause T-cell immunodeficiency [126]. Heterozygosity for the *Foxn1* mutant allele confers partial protection against ECM, suggesting that FOXN1 transcriptional targets may be relevant to reducing neuroinflammation.

The epistatic interaction between the B6 and 129S1 genetic backgrounds highlights both the limitations and advantages of different variations of the ENU-mutagenesis screen. Both the B6 and 129S1 strains are susceptible to *Plasmodium berghei* ANKA infection, developing neurological symptoms between Days 5 and 10 post-infection. However, in over a third of the phenodeviant pedigrees identified in the mixed background screen, an enrichment of B6 alleles on distal chromosome 4 was associated with resistance to ECM. With such a high percentage of phenodeviant pedigrees mapping to the same locus, we hypothesized that the likelihood of this effect being caused by a single causative ENU-induced mutation was minimal, and that this effect was due to genetic background rearrangements. Additional analysis revealed that ECM resistance on Chromosome 4 (*Berghei* resistance locus 8, named *Berr8*,) was being modulated by a second locus on Chromosome 1 (named *Berr7*). Although we had expected to discover single point mutations due to ENU-mutagenesis, these results highlight the complex nature of cerebral malaria, as well as the difficulty inherent to finding point mutations that are solely responsible for trait modulation.

Due to improvements in technology and the resulting cost reduction, we switched from linkage analysis to exome sequencing analysis for the identification of ENU-induced mutations, removing the requirement for genetic background variations. Hence, the third and final screen was executed on a pure B6 genetic background, wherein the mutagenized G0 males were outcrossed to wild-type B6 females. Switching to the pure genetic background eliminated the likelihood of epistatic interactions between genetic backgrounds, as exhibited in the reduction of background survival rate from almost 8% in the B6x129S1 screen to less than 5% in the pure B6 screen. However, due to smaller litter sizes, almost 40% fewer G3 animals were produced from the 109 screened pedigrees. Even so, eight phenodeviant pedigrees were identified and are currently being investigated.

6.2. Screening for Acquired Resistance to Blood-Stage Malaria

ENU-mutagenesis has also been used to identify genes implicated in host resistance to blood-stage malaria. A dominant ENU-mutagenesis screen for erythrocyte production and maturation defects linked to malaria resistance identified two mutations in the *Ank1* gene: an alternative splice acceptor mutation resulting in a frameshift mutation and premature stop codon was identified in *Mpl*$^{-/-}$ mice mutagenized on a BALB/c background [127], and a single nonsense mutation was identified in mutagenized SJL/J mice [128]. Both mutations result in early truncation of the ANK1 protein, encoded by *Ank1*. Implicated in hereditary spherocytosis, an inherited form of hemolytic anemia, mouse erythrocytes harboring mutations in *Ank1* exhibit increased resistance to *P. chabaudi*, a model of blood stage malaria, potentially due to parasite maturation impairment [127,128].

6.3. Conclusion

ENU-mutagenesis has enabled the identification of individual genes involved in modulating the host response to both cerebral and blood-stage malaria. We have identified mutations in host inflammatory genes involved in T cell development and/or function (*Jak3* and *Foxn1*), thymus development, and immune cell function [119,125]. These results are consistent with the current understanding of the role of T cells in cerebral malaria pathogenesis [129–132]. Additionally, these genes have been associated with the modulation of other models of acute inflammation as well as of chronic inflammatory conditions [99]. Other labs have identified mutations in the erythrocyte protein ANK1, an important factor in the erythrocyte cytoskeleton [127,128]. Mutations in erythrocytic proteins, including the cell surface Duffy antigen [92] and structural component Band 3 [89–91], have been associated with increased resistance to malaria in humans for several years. Together, these findings advance our understanding of the host response to malaria, and may aid in the discovery of novel drug targets against this devastating disease.

7. *Salmonella* Bacteria Infections

Salmonella enterica infections in humans represent an increasingly significant economic and public health challenge that is associated with high morbidity and mortality in both developing and industrialized countries [133]. In fact, the increase in global population, the emergence of antimicrobial resistance in bacteria, and the prevalence of co-infections (e.g., *Plasmodium*, HIV) exacerbate the burden of this infectious disease [38]. *Salmonella* infection in humans can cause a range of food and waterborne illnesses, from a localized diarrheal disease to the more severe systemic disease, typhoid fever. In fact, nontyphoidal *Salmonella enterica* serovars (e.g., *S. typhimurium*, *S. enteritidis*) are the second leading cause of bacterial food poisoning in the United States. Importantly, about 1%–4% of these *Salmonella*-infected individuals are at an increased risk of developing sepsis, chronic infection or clinical sequelae (ex. chronic arthritis) [134–136]. *Salmonella enterica* Typhi is the etiologic agent of typhoid fever, which is endemic primarily in areas with poor sanitation and a lack of clean drinking water. *Salmonella typhi* causes twenty-one million infections annually, with 220,000 deaths [133]. The outcome of infection depends on the activation of early innate functions, neutrophilic infiltration, phagocytosis by tissue macrophages, and inflammatory cytokine/chemokine secretion (e.g., IFNγ, IL-12, IL-18, TNFα, and IL-6). However, ultimately, the resolution of systemic infection is dependent on both humoral and cell-mediated immune responses [137,138].

In humans, the contribution of host genetics to *Salmonella* infection has been proven by the candidate gene approach and by exome sequencing in patients. Individuals with defects in the IL-12/IL-23 (IL-12β, IL-12Rβ1) and IFNγ (IFNγR1, IFNγR2, STAT1) pathways are in fact predisposed to Mendelian susceptibility to mycobacterial disease (MSMD) and/or disseminated *Salmonella* infection [139–144]. Furthermore, major histocompatabilty complex (MHC) class II and III loci, as well as the TNF haplotype, were significantly associated with typhoid fever in a Vietnamese cohort [145]. Although clinical evidence supports a strong role for host genetics, susceptibility to *Salmonella*-related infections is complex and also influenced by environmental factors and bacterial serotype.

Salmonella typhimurium infection is a recognized experimental model for studying systemic typhoid-like disease in mice [146,147]. Various classical inbred strains of mice demonstrate differential susceptibility/survival following sub-lethal intravenous infection with *S. typhimurium* strain Keller [148]. In particular, the 129 substrains (129S1, 129X1) of mice are highly resistant to virulent infection, compared to DBA/2J mice, which display intermediate mortality, and to the highly susceptible B6 strain. Although the genetic and molecular basis of several mutations important in resistance to *Salmonella* infection in mice have been identified, namely *Nramp1/Slc11a1*, *Tlr4*, and *Pklr*, the low frequency of naturally occurring spontaneous mutations has prompted the use of novel genomic approaches like ENU mutagenesis to identify novel host susceptibility genes to *Salmonella* infection [148–153].

7.1. Screening for Acquired Susceptibility to Salmonella typhimurium

We used ENU mutagenesis to further decipher the host genetic component of susceptibility to *Salmonella* infection *in vivo*. In the screen, G3 ENU-mutagenized mice between 7 and 9 weeks of age were challenged intravenously through the caudal vein with an infectious dose of virulent *Salmonella typhimurium* strain Keller, varying between 1000 to 10,000 colony forming units (CFUs), depending on the background strains used for breeding. Over the course of 14 days, infected mice were monitored for clinical manifestations of illness including a body score index of less than two, muscle wasting, fur ruffling (fever), inactivity, twirling, and shaking. Susceptible mutants were defined as those presenting severe clinical signs between Days 3 to 7 post-infection (prior to background control mice). On average, a minimum of six to eight G3 mice per G2 female were infected with the expectation of identifying two to five heritable deviant pedigrees following the screening of G3 mice derived from roughly 100 G1 males.

Two prototype breeding schemes differing in the genetic contribution of background strains (B6, 129S1, 129X1, and DBA/2J) have been used in five rounds of screening for *Salmonella* susceptibility. Male 129S1 (G0) mice were mutagenized using a single i.p. injection of 150mg/kg of ENU at 8–10 weeks of age. The first breeding scheme involved the generation of G1 mice produced by two independent G0 males (Figure 2B). The G0 males were crossed to B6 females. For each G1 pedigree, four G2 brother-sister pairs were bred to produce G3 progeny. Using this breeding scheme, the *Salmonella* susceptibility allele *Slc11a1*^{Asp169} from B6 mice was segregated into the G2 population. G2 animals carrying the wild-type *Slc11a1* alleles were then selected for further breeding. As the introduction of susceptibility to the B6 background was interfering with our capacity to capture recessive alleles acting in later infection stages (past Day 4), we subsequently modified the breeding scheme as in Figure 2A. Hence in the second round of screening, G0 males were out-crossed to wild-type 129X1 females to generate G1 heterozygote offspring. G1 males were further backcrossed to 129X1 females to generate G2 mice. G2 females were then backcrossed to the G1 male to give rise to G3 progeny, which were then used for primary phenotyping of susceptibility to infection using survival analysis with 10,000 CFUs. Using the following scheme, 643 G3 mice derived from 39 G1 males were screened and two deviant pedigrees were identified: *Oxie & Celie* (*Ity14)* (*Immunity to Typhimurium* locus *14*) and *Jody & Cloe* (*Ity15).* In this particular case, we used a strain that was closely related to the mutagenized males to prevent or minimize the impact of the genetic background on the expressivity of the phenotype while allowing mapping in the G3 animals. We identified 105 SNPs between 129S1 and 129X1. However, their clustering in the genome did not allow the mapping of some pedigrees. Variations of these protocols (Figure 2) were used to facilitate mapping resolution using SNPs between 129S1 and DBA/2J directly in the G3 population. In the third round of screening, G1 males were out-crossed to DBA/2J, and the resulting G2 mice were randomly intercrossed to generate G3 progeny. G3 mice were then screened with an infectious dose of 5000 CFUs. Using this scheme, 1570 G3 mice derived from 65 G1 males were screened, and one deviant pedigree, *Ity16*, was identified, validated, and cloned [154]. In the fourth round of screening, G0 males were out-crossed directly to DBA/2J in order to introduce genetic variability as early as possible in the breeding scheme, thus facilitating mapping (Figure 2B). In this round, 3,348 G3 mice derived from 208 G1 males were screened and four deviant pedigrees were identified: *Cherrie & Walter* (*Ity17*), *Jeanine & Harman (Ity18)*, *Lexie & Leona*, and *Philippe & Desiree*. Lastly, with the onset of whole-exome sequencing as an alternative to mapping using genetic variation between parental strains, the breeding scheme shown in Figure 2B was carried out on an 129S1 background. From the following screen we have infected 580 G3 mice derived from 41 G1 males, and two deviant pedigrees, *Rakeem & Athena* and *Lessie & Virgie*, were identified.

In summary, 8,389 G3 mice derived from 491 G1 males were screened for increased susceptibility to *Salmonella typhimurium* infection as measured by survival analysis. A total of 10 deviant pedigrees have been identified (Table 2). From this screen, we have to date identified, cloned, and characterized *Salmonella* susceptible mutations in *Usp18* (*Usp18^{L361F}*), *Ank1* (*Ank1^{Gln1357Ter}*), and *Stat4* (*Stat4$^{G418\_E445}$*) [154–156]. USP18 (Ubiquitin Specific Peptidase 18) both regulates type I IFN signaling and functions as a protease to remove ISG15 adducts from substrate proteins [157,158]. We have reported that decreased survival in mice that carry the *Usp18^{L361F}* mutation results from increased bacterial loads in the spleen and liver, as well as increased inflammatory response leading to septic shock [156,159]. In more recent studies, we have shown that regulation of type I IFN signaling is the predominant mechanism affecting the susceptibility of *Usp18^{L361F}* mice to bacterial infection. Also, we have found that hyperactivation of type I IFN signaling leads to increased ISGylation and IL-10 production, as well as decreased expression of markers of autophagy [160]. Additionally, we have shown that *Usp18^{L361F}* mice are more susceptible to infection with *Mycobacterium tuberculosis* (same as above).

Table 2. Summary of the three ENU-mutagenesis screens for experimental cerebral malaria, *Salmonella*, and herpes simplex virus (HSV)-1.

	Malaria	***Salmonella***	**HSV-1**
G1 males	573	491	265
G3 mice	16,411	8,415	7,802
Deviant pedigrees (in progress)	45	16	11
Confirmed pedigrees	5	3	2

The transcription factor STAT4 (Signal Transducer and Activator of Transcription Factor 4) is a critical mediator of IL-12 signaling. It plays an important role in both innate and adaptive immunity by regulating the transcription of target genes such as *Ifng* and those mediating NK cell cytotoxicity, T helper 1 cell differentiation, and immunoglobulin isotype switching to IgG1. The *Stat4$^{G418\_E445}$* mutation results in impaired innate IFNγ secretion, primarily from splenic NK and NKT cells, contributing to increased hepatosplenic bacterial loads. These findings support the importance of the IL-12/IFNγ axis in resistance to *Salmonella* infection.

ANK1 is a structural protein of the erythrocyte membrane, which plays an important role in membrane stability by mediating the attachment of band 3 (SLC4A1) and protein 4.2 (EPB4.2) to the spectrin-based membrane cytoskeleton [161]. Mice homozygous for the *Ank1^{Gln1357Ter}* mutation develop hemolytic anemia and present clinicopathological features of human hereditary spherocytosis, the most common form of congenital chronic hemolysis in Europe and North

America [162]. On one hand, as observed with other mutations affecting red blood cell turnover [163], *Ank1* deficits protect mice against malaria [128]. On the other hand, normal ANK1 function is critical for an effective host response against infection with *Salmonella*. *Salmonella* susceptibility in *Ank1^{Gln1357Ter}* mutant mice is the result of a combination of factors, namely the concomitant deposition of iron in tissues, which favors bacterial growth, and low levels of the iron regulatory hormone hepcidin [154]. In addition, the strong induction of heme oxygenase 1 (*Hmox1*) expression observed during malaria infection and in *Ank1^{Gln1357Ter}* mutant results in impaired oxidative burst function, which favors the intracellular replication of bacteria [154,164].

7.2. Ex Vivo and in Vivo ENU Screens for Susceptibility to Bacteria Infections

Additional ENU initiatives have uncovered novel genetic determinants of resistance to bacterial infections. Different primary screens in G3 offspring were used, including: (1) measurement of TNF bioactivity after *ex vivo* challenge of thioglycolate-induced peritoneal macrophages with various pathogen-associated molecular patterns (PAMPs) (*Cd36*, *Tnf*, *Map3k8*) [165–167]; (2) measurement of type I IFN bioactivity after *ex vivo* challenge of thioglycolate-induced peritoneal macrophages with *Listeria monocytogenes* (*Tmem173/Sting*) [168]; (3) *in vivo* screen for other classes of pathogens (*Slfn2*) [169]; (4) mutations affecting hematopoetic cell development (Genista-*Gfi1*) [170]; and (5) visible phenodeviants presenting inflammatory lesions of the skin (*Scd1*) [171] or of the feet (*Ptpn6/Shp1*) [172]. For example, a TLR2 agonist screen in macrophages identified the *Oblivious* pedigree, which possesses a mutation in *Cd36* resulting in increased susceptibility to infection with Gram positive bacterium *Staphylococcus aureus* [165]. In addition, the *Sluggish* pedigree, which carries a mutation in the *Map3k8* kinase, has impaired type I IFN production downstream of TLR7 and TLR9 signaling, rendering it susceptible to Group B streptococcus infection *in vivo* [166]. Another example is the ENU-induced mutation in *Gfi1* within the *Genista* pedigree, wherein depletion of PMNs confers resistance to *Brucella abortus* infection [173,174] and increased susceptibility to oral infection with *Salmonella typhimurium sfiA^{-}* [170]. Moreover, the *ex vivo* ENU screen using *Listeria monocytogenes* identified the *Goldenticket* pedigree as carrying a mutation in *Tmem173/Sting,* further demonstrating the importance of type I IFN signaling during bacterial infection [168].

7.3. Conclusion

ENU-mutagenesis identified single gene effects (novel allele and novel function) within critical pathways involved in immunity to bacterial infection that could potentially be translatable to infection with other classes of pathogens and/or to chronic inflammatory diseases. The findings have emphasized the importance of IFN signaling (*Usp18*, *Stat4*, *Sting*, *Map3k8*) during bacterial infections [155,156,166,168], as well as erythropoeisis and iron metabolism, (*Ank1*) in the case of *Salmonella* pathogenesis [159].

8. Herpes Viruses

The *Herpesviridae* family is a large ancient family with a long history of coevolution with their hosts probably predating the origin of the primate lineage. Altogether the nine human herpesviruses infect 90% of the world population causing different types of pathologies that vary considerably according to the immune status of the infected individual. These ubiquitous viruses constitute a striking example of the intricate interplay that can be gradually established between host and pathogen, and show that important information can be gleaned from the study of host-pathogen interactions, namely the contribution of both viral immune evasion and host resistance genes to the outcome of infection.

8.1. Cytomegaloviruses

Human cytomegaly virus (HCMV) is the most frequent congenital viral infection in developing countries, potentially leading to blindness, deafness or mental retardation in affected infants. Primary infection or reactivation of the virus can result in severe morbidity and mortality, especially in immune-compromised individuals such as transplant recipients, leukemia or lymphoma patients and AIDS patients. Fortunately, HCMV is closely related to its murine homologue, mouse cytomegalovirus and both cause death in immunocompromised individuals [175,176]. Thus, infection of mice with MCMV represents an excellent model for the study of HCMV pathology and indeed it is an important tool for virologists, immunologists, and geneticists, all of whom have benefited from the well-developed state of the model. Forward genetic studies in inbred mouse strains identified major epistatic (*Klra16*/*H2*^{k}) or single gene effects (*Klra7*, *Klra8*) demonstrating the crucial role that natural killer (NK) cell specific activating (Ly49H, Ly49P) and inhibitory (Ly49G) receptors play in response to virus infections (reviewed in [177]).

8.2. Screening for Altered Immune Responses to MCMV

Beutler and colleagues were the ones to initiate the ENU screen for MCMV susceptibility (for the latest review, see [178]). With this strategy, over 20,000 G3 B6 mice carrying ENU mutations were infected i.p. with 10^5 plaque forming units (pfu) of MCMV. This viral dose was chosen because wild-type B6 mice are uniformly resistant in this infectious experimental situation. However, the pheno-deviant offspring that exhibited clinical signs of disease or/and high viral titers in the spleen were considered susceptible. Several mice with immunodeficiency phenotypes identified from other screens made by Beulter's group, such as defects in toll-like receptor (TLR) signaling or adaptive immunity, were also tested for their potential MCMV susceptibility. Here, we highlight some of the most important findings that have been made using ENU-mutagenesis to test susceptibility to MCMV infection.

Dendritic cells (DCs) are specialized cells of the hematopoietic system that alert the immune system to the presence of infection. Therefore, they generally represent the first line of defense against pathogens. In the context of MCMV infection, DCs recognize the virus through TLR3 and TLR9, which are able to respectively detect double-stranded RNA (an intermediate product of viral replication) and viral double stranded DNA. Following MCMV recognition, DCs and plasmacytoid

DCs (pDCs) in particular, produce large amounts of antiviral type I IFN cytokines (IFN-α/β), which are essential mediators of the innate and adaptive immune responses. Thus, loss-of-function mutations in genes that encode components necessary for the expression of IFN-α/β (such as *Tlr9*, *Tlr3*, *Myd88*, *Trif*, and *Unc93b1*), or that are involved in the IFN-α/β signaling pathway (ie, downstream of IFN-α/β receptor), like *Stat1*, have been shown to increase susceptibility to MCMV infection [179–182]. It should be noted that among all of these ENU mutations, only the one in *Stat1* was initially identified from the MCMV screen, the others being deduced from immune screens. The NF-κB signaling pathway is also essential for survival to MCMV infection. This is attested by the identification of a loss-of-function mutation in the *Ikbkg* gene encoding NEMO, a regulatory subunit of the IKK complex responsible for the nuclear translocation of NF-κB [183]. *Ex vivo* screens for increased susceptibility to MCMV infection have been performed on peritoneal macrophages isolated from ENU-mutagenized mice, and revealed a missense mutation in the *Eif2ak4* gene encoding GCN2 [184]. This protein is related to PKR, an effector known to inhibit viral replication via phosphorylation of the alpha subunit of eukaryotic initiation factor 2 (eIF2α). The loss-of-function mutation identified in *Eif2ak4* affects the phosphorylation of eIF2α in response to MCMV infection and was therefore associated with an increased susceptibility to MCMV. The MCMV screen, together with the immune screens, led to the identification of several phenodeviants with mutations in genes that contribute to the establishment of an efficient immune response against pathogens, as they act at different levels of IFN-α/β production (TLR9, TRIF and UNC93B1), of IFN-α/β signaling (STAT1), and of the antiviral response (GCN2).

DCs are not the only sites of MCMV recognition. Natural killer (NK) cells are also important responders to MCMV infection, playing a crucial role in containing it at early times post-infection [185,186]. This was initially demonstrated by *in vivo* depletion studies, in which specific antibodies were used to transiently eliminate NK cells before infection with the virus [186–188]. Then, the differential susceptibility of the BALB/c and B6 strains was shown to be due to the presence of the NK-activating receptor Ly49H in the latter [189,190]. This receptor engages the MCMV viral protein m157 [191,192], leading to NK cell proliferation and target cell killing [193]. ENU studies allowed the initial discovery of mutations in the *Gimap5* and *Unc13d* genes, in the context of two screens that had been designed to detect *in vivo* defective NK cells and cytotoxic T lymphocyte (CTL) responses [194] and MCMV susceptibility [195], respectively. In both cases, *Gimap5*^{G38C} and *Unc13d*$^{jinx/jinx}$ were shown to be associated with defects in NK cell activity and impaired resistance to MCMV infection, which are consistent with the crucial function of NK cells in the early control of MCMV replication. *Gimap5*^{G38C} affects NK cell development, whereas *Unc13d*$^{jinx/jinx}$ NK cells fail to degranulate, a deficit also observed in activated CD8^{+} T cells. Individuals carrying another deleterious mutation, this time in the *Itgb2* gene encoding the integrin β2 CD18, which partially affects NK cell development, are, however, fully resistant to MCMV [196]. In this case, it suggests that even if the β2 integrins are required for optimal NK cell maturation, their partial deficiency could be overcome during MCMV infection, highlighting the robustness of antiviral protective responses.

Other ENU mutations revealed from the screen for host survival against MCMV infection were independently identified in the *Flt3* [197] and *Slfn2* [169] genes. *Flt3*$^{wmfl/wmfl}$ mice have been shown to have impaired DC development, making these cells incapable of supporting the effector function of NK cells [197]. In contrast to *Flt3*$^{wmfl/wmfl}$, neither DCs, nor NK cells are impaired in *Slfn2*^{I1135N} mice [169]. However, both bacterial and viral infections trigger death by apoptosis of peripheral T cells and inflammatory monocytes in *Slfn2*^{I1135N} mice, indicating the crucial role of Slfn2 in maintaining quiescence in some immune cells. In addition to these ENU mutants recovered from the MCMV screen, four unrelated mutants, called *Mayday*, *Solitaire*, *Goodnight*, and *Slumber*, were shown to die very early post-infection (*i.e.*, D2-D3 p.i.) before high viral titers could be observed in the spleen and the liver [198]. Their abrupt death was probably not caused by the direct lytic effects of the virus, but mostly by collateral damage, such as the accompanying inflammatory reaction in response to MCMV infection, since this phenotype was also observed after lipopolysaccharides (LPS) or CpG administration. Based on the comparative sequence analysis of these four mutants, their MCMV susceptibility has been shown to be due to a genetic rearrangement of the *Kcnj8* locus that is likely to have occurred in B6 mice prior to ENU treatment. *Kcnj8* encodes the potassium channel Kir6.1, which maintains the host homeostatic state during the innate immune response. Altogether, these mutations highlight genes that are directly involved in the immune system, but also show the importance of other non-immune signaling pathways, such as homoestasis, in host survival.

8.3. Herpes Simplex Virus 1

HSV-1 is the causative agent of herpes simplex encephalitis (HSE), a lethal neurological disease. It is acknowledged that environmental factors have no effect on the pathogenesis of "HSE, and no geographical or seasonal patterns in the distribution of the disease have been observed [199,200]. Despite the high seroprevalence of HSV-1 (up to 90%) [201], HSE pathology is rare and affects only a small proportion of otherwise healthy individuals. Therefore, in addition to HSV-1 infection, the second major cause of the disease is the presence of rare host genetic factors, which play a large part in determining the susceptibility of an individual to HSE. Loss-of-function mutations in the *UNC93B1*, *TLR3*, *TRIF*, *TRAF3*, and *TBK1* genes have been associated with a human genetic predisposition to HSE [202–207], illustrating the critical role of the UNC93B-TLR3-type I IFN pathway in protection against HSV-1. However, these mutations exhibit incomplete penetrance and represent only a minority of HSE cases. This indicates the likely existence of other anti-HSE pathways and may reflect the effects of additional host genetics factors.

8.4. Screening for Acquired Susceptibility to HSE

Two breeding schemes have been used in the mutagenesis screen to identify host susceptibility genes to HSV-1 infection. We started with the B6/B10 screen, where mutagenized B6 G0 males were out-crossed to B10. This allowed linkage mapping with the use of a panel of 255 B6/B10 polymorphic markers (SNPs) distributed across the genome [208]. We then switched to a pure B6 genetic background to eliminate the likelihood of epistatic interactions between the B6 and B10 genetic backgrounds. In total (Table 2), 7,802 G3 B6 mice carrying ENU mutations derived from 265 G1 males were infected i.p. with 10^4 pfu of HSV-1 strain 17. This dose led to lethal encephalitis in susceptible A/J mice, whereas wild-type B6 mice remained unaffected. Following infection, the ENU-mutagenized mice were monitored for two weeks. The phenodeviant offsprings that exhibited clinical signs of disease or succumbed to the infection were considered susceptible. Using this strategy, we revealed eleven deviant pedigrees. One of these led to the identification of a premature stop codon (L3X) in the Ptprc gene, which encodes the leukocyte common antigen CD45. Ptprc^{L3X} mutant mice showed reduced numbers of CD3^{+} T and mature follicular B cells, suggesting defects in T and B cell development [209]. In this report, we also demonstrated that CD4^{+} Th1 cells, by producing IFNγ, help CD8^{+} T cell recruitment to prevent the dissemination of HSV-1 into the central nervous system, thus protecting mice from lethal HSV-1 infection. Altogether, our data point to CD45 as the first host component involved in the adaptive immune response that directly contributes to susceptibility to HSV-1 and HSE pathology. We are currently investigating the 10 other deviant pedigrees, which have, once again, shown the crucial role of T cells in host survival, but have also revealed that anti-inflammatory factors are critical to protection against HSV-1-induced encephalitis [210].

9. Conclusions and Perspectives

ENU-mutagenesis constitutes an inherently unbiased and powerful approach to the production of new alleles. Technological improvements in high-throughput DNA sequencing, combined with the completion of the mouse genome project [68], have greatly facilitated their identification. The recent introduction of NGS has led to a faster and more efficient identification of ENU mutations, which is particularly helpful for analyzing large mutant collections, especially when mapping data are not available to guide an analysis. New variants generated by ENU-mutagenesis mirror those existing in the human population and also represent a natural complement to null alleles being produced by gene targeting. Finding new ENU-induced alleles will also benefit from the new CRISPR/Cas9 technology. ENU variants, although easier to pinpoint by sequencing, need to be validated experimentally as in any forward genetic approach of gene identification. The CRISPR/Cas9 system appears to be an excellent complement to ENU mutagenesis, allowing candidate point mutations identified by NGS to be efficiently confirmed as causative mutations. The ENU mutagenesis approach has proven to be extremely useful in dissecting the genetic architecture of host defenses against infectious diseases. The approach promises to remain current in the field, being constantly renewed by technological advances such as NGS or genome editing.

As summarized in Table 2, over 30,000 G3 mice were screened by our group for either resistance to *Plasmodium berghei* or susceptibility to *Salmonella typhimurium* and HSV-1 infection. In total, 72 deviant pedigrees have been identified and we have to date confirmed ENU-induced mutations for 10 pedigrees. These mutations highlight gene functions that are directly involved in the immune system (*Foxn1*, *Jak3*, *Stat4*, *Usp18* and *Ptprc*), but also show the importance of other non-immune pathways, such as erythropoeisis and iron metabolism (*Ank1*), in host survival (Table 3). Beutler and colleagues also used the ENU mutagenesis approach, and over 20,000 G3 mice were screened for their susceptibility to MCMV. In parallel, they also developed several “immune” ENU screens, where some phenodeviant pedigrees, characterized by defects in the TLR signaling pathway and/or in T/NK cells functions, were then tested for their potential

MCMV susceptibility. Of these, it should be noted that among the ENU mutations identified by the group of Beutler, only few were initially revealed by the MCMV *in vivo* screen (*Stat1*, *Unc13d*, *Flt3* and *Slfn2*), the others being deduced from other screens (*Tlr9*, *Trif*, *Unc93b1*, *Ikbkg*, *Eif2ak4*, *Gimap5*) [178]. This observation can be explained by the fact that *in vivo* models are more complex than *in vitro* systems. Indeed, deficiencies in one particular immune cell or signaling pathway can be compensated by the presence of other competent immune cells, making the identification of defective alleles more difficult *in vivo*.

The ENU mutations identified in *Jak3* (*Jak3*^{W81R}) and *Ptprc* (*Ptprc*^{L3X}) highlighted the critical nature of T cell function for CM pathogenesis and protection against HSV1 infection, respectively. The robustness of these mouse models of neuroinflammation and their ability to detect genetic effects regulating common pathways critical for neuroinflammation are highlighted by the complementary observations that the *Jak3*^{W81R} mutant allele (protective in the ECM screen) confers susceptibility to HSV encephalitis (HSE), while the *Ptprc*^{L3X} (causing susceptibility to HSE screen) is protective in the ECM model [211]. This approach could be generalized to other interesting pedigrees, where the role of the ENU mutations could be assessed in these different mouse models of infectious diseases. By cross-testing these mutant pedigrees, it should be possible to reveal common and specific pathways, as well as cells and proteins, that are crucial in the protection against malaria and *Salmonella* or viral infections. Moreover, the role of ENU mutations identified in the neuroinflammatory models of ECM and HSE could also be tested in other models of inflammation, such as the model of experimental encephalitis (EAE) that mimics MS, or DSS colitis that models IBD. Preliminary experiments using the EAE model have already suggested that *Ptprc*^{L3X} mice are more resistant to EAE symptoms than wild-type and heterozygous littermate controls [210]. Thus, the cross-testing of these mutant pedigrees in different models of inflammation may provide additional information on the gene function, including its role in the pro- and anti-inflammatory balance. It can also provide novel targets for the development of new drugs that could be used in therapy for acute and chronic inflammatory diseases. As an example, a JAK3 inhibitor, currently in clinical use for the treatment of RA and CD (tasocitinib; Pfizer, New York, NY, USA), has been shown to reduce neuroinflammation and increase survival of *Jak3*$^{-/+}$ heterozygotes in our ECM model [119]. Therefore, pharmacological modulation of JAK3 mimics the effect of its genetic inactivation, indicating that the ECM screen can identify novel pharmacological targets for drug discovery.

One objective of the ENU-mutagenesis approach is to translate and validate knowledge obtained in the mouse infectious context to an improved understanding of human immunity and susceptibility to infection. As a starting point, mouse studies are fundamental for exploring host-pathogen interactions, especially when orthologous human genes exist. One striking example came from the discovery of the ENU-induced mutation in the mouse *Unc93b1* gene that causes susceptibility to MCMV [181]. Based on this finding, the group of JL Casanova identified an autosomal recessive UNC93B deficiency in two human patients with HSE [202]. Furthermore, a survey of the literature has shown that human variants identified in our ECM and HSE screens are risk factors for inflammatory diseases. For example, genetic variants in *JAK* and *STAT* family members have been associated with IBD, MS, RA, and SLE [121,122]. *PTPRC* polymorphisms are associated with autoimmune and inflammatory conditions including MS, SLE, and myasthenia gravis [212]. Thus, the ENU-mutagenesis approach should be continued in combination with GWAS studies, thus providing important insights into the pathways, cells, and proteins that directly impact susceptibility to pathogens, as it constitutes an invaluable resource for identifying novel therapeutic treatments.

Table 3. Genes and pathways identified in ENU screens described in this review.

Pathway	Gene	Screen	Phenotype	Reference
TLR signaling	*Cd36*	Immunity→*S. aureus*	Susceptible	[165]
	Map3k8	Immunity→Group B streptococcus	Susceptible	[166]
	Ptpn6	Autoimmunity→*L. monocytogenes*	Susceptible	[172]
	Tlr9	Immunity→MCMV	Susceptible	[179]
	Trif	Immunity→MCMV	Susceptible	[180]
	Unc93b1	Immunity→MCMV	Susceptible	[181]
	Ikbkg	Immunity→MCMV	Susceptible	[183]
Type I IFN signal	*Usp18*	*S.* Typhimurium	Susceptible	[156,159]
	Stat1	MCMV	Susceptible	[182]
Effector	*Eif2ak4*	MCMV	Susceptible	[184]
Cellular immunity	*Jak3*	*P. Berghei*	Resistant	[119]
	Foxn1	*P. Berghei*	Resistant	[125]
	Stat4	*S.* Typhimurium	Susceptible	[155,156]
	Tnf	Immunity→L. *monocytogenes*	Resistant	[167]
	Gfi1	Immunity→*S.* Typhimurium	Susceptible	[170]
	Gimap5	Immunity→MCMV	Susceptible	[194]
	Unc13d	MCMV	Susceptible	[195]
	Flt3	MCMV	Susceptible	[197]
	Slfn2	MCMV	Susceptible	[169]
	Ptprc	HSV-1	Susceptible	[209]
Red cell cytoskeleton	*Ank1*	*S.* Typhimurium	Susceptible	[154]
	Ank1	*P. Chabaudi*	Resistant	[127,128]
Homeostasis	*Kcnj8*	MCMV	Susceptible	[198]
Lipid metabolism	*Scd1*	Immunity→*S. Pyogenes*	Susceptible	[171]

Acknowledgments

This work was support by a Team grant (Danielle Malo, Philippe Gros, Silvia M. Vidal) from the Canadian Institutes of Health Research (CTP-87520). Grégory Caignard is supported by the Canadian Institutes of Health Research. Megan M. Eva is supported by the McGill University Health Centre. Silvia M. Vidal is supported by the Canada Research Chairs Program. We thank Eve-Marie Gendron-Pontbriand for proofreading the manuscript.

Author Contributions

Robert Eveleigh and Guillaume Bourque contributed to the section on gene identification (section 4); Rebekah van Bruggen and Philippe Gros contributed to the section about Malarial parasites infection and screen (section 6); Megan M. Eva and Danielle Malo contributed to the section about Salmonella infection and secreen (section 7), Grégory Caignard and Silvia M. Vidal contributed to the sections about herpes virus infection and screen (section 8) as well as the introductory sections (sections 1, 2, 3 and 5) and the conclusions (section 9).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Brussow, H. Europe, the bull and the Minotaur: The biological legacy of a Neolithic love story. *Environ. Microbiol.* **2009**, *11*, 2778–2788.
2. McMichael, A.J. Environmental and social influences on emerging infectious diseases: Past, present and future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359*, 1049–1058.
3. Casanova, J.L.; Abel, L. Inborn errors of immunity to infection: The rule rather than the exception. *J. Exp. Med.* **2005**, *202*, 197–201.
4. Fauci, A.S.; Morens, D.M. The perpetual challenge of infectious diseases. *N. Engl. J. Med.* **2012**, *366*, 454–461.
5. Chapman, S.J.; Hill, A.V. Human genetic susceptibility to infectious disease. *Nat. Rev. Genet.* **2012**, *13*, 175–188.
6. Cobat, A.; Orlova, M.; Barrera, L.F.; Schurr, E. Host genomics and control of tuberculosis infection. *Publ. Health Genet.* **2013**, *16*, 44–49.
7. Plantinga, T.S.; Johnson, M.D.; Scott, W.K.; Joosten, L.A.; van der Meer, J.W.; Perfect, J.R.; Kullberg, B.J.; Netea, M.G. Human genetic susceptibility to Candida infections. *Med. Mycol.* **2012**, *50*, 785–794.
8. Keynan, Y.; Malik, S.; Fowke, K.R. The role of polymorphisms in host immune genes in determining the severity of respiratory illness caused by pandemic H1N1 influenza. *Publ. Health Genet.* **2013**, *16*, 9–16.
9. Min-Oo, G.; Gros, P. Erythrocyte variants and the nature of their malaria protective effect. *Cell Microbiol.* **2005**, *7*, 753–763.
10. Lederman, M.M.; Penn-Nicholson, A.; Cho, M.; Mosier, D. Biology of CCR5 and its role in HIV infection and treatment. *JAMA* **2006**, *296*, 815–826.
11. Lindesmith, L.; Moe, C.; Marionneau, S.; Ruvoen, N.; Jiang, X.; Lindblad, L.; Stewart, P.; LePendu, J.; Baric, R. Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **2003**, *9*, 548–553.
12. Von Bernuth, H.; Picard, C.; Puel, A.; Casanova, J.L. Experimental and natural infections in MyD88- and IRAK-4-deficient mice and humans. *Eur. J. Immunol.* **2012**, *42*, 3126–3135.
13. International Human Genome Sequencing. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.
14. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **2005**, *437*, 1299–1320.

15. Hohl, T.M. Overview of vertebrate animal models of fungal infection. *J. Immunol. Methods* **2014**, doi:10.1016/j.jim.2014.03.022.
16. Longley, R.; Smith, C.; Fortin, A.; Berghout, J.; McMorran, B.; Burgio, G.; Foote, S.; Gros, P. Host resistance to malaria: Using mouse models to explore the host response. *Mamm. Genet.* **2011**, *22*, 32–42.
17. Sancho-Shimizu, V.; Zhang, S.Y.; Abel, L.; Tardieu, M.; Rozenberg, F.; Jouanguy, E.; Casanova, J.L. Genetic susceptibility to herpes simplex virus 1 encephalitis in mice and humans. *Curr. Opin. Allergy Clin. Immunol*. **2007**, *7*, 495–505.
18. Wick, M.J. Innate immune control of *Salmonella enterica* serovar Typhimurium: Mechanisms contributing to combating systemic *Salmonella* infection. *J. Innate Immun*. **2011**, *3*, 543–549.
19. Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* **2002**, *420*, 520–562.
20. Church, D.M.; Goodstadt, L.; Hillier, L.W.; Zody, M.C.; Goldstein, S.; She, X.; Bult, C.J.; Agarwala, R.; Cherry, J.L.; DiCuccio, M.; *et al*. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol*. **2009**, *7*, e1000112.
21. Guenet, J.L. Animal models of human genetic diseases: Do they need to be faithful to be useful? *Mol. Genet. Genet.* **2011**, *286*, 1–20.
22. Guan, C.; Ye, C.; Yang, X.; Gao, J. A review of current large-scale mouse knockout efforts. *Genesis* **2010**, *48*, 73–85.
23. Ayadi, A.; Birling, M.C.; Bottomley, J.; Bussell, J.; Fuchs, H.; Fray, M.; Gailus-Durner, V.; Greenaway, S.; Houghton, R.; Karp, N.; *et al*. Mouse large-scale phenotyping initiatives: Overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* **2012**, *23*, 600–610.
24. Gaj, T.; Gersbach, C.A.; Barbas, C.F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. **2013**, *31*, 397–405.
25. Wang, H.; Yang, H.; Shivalila, C.S.; Dawlaty, M.M.; Cheng, A.W.; Zhang, F.; Jaenisch, R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **2013**, *153*, 910–918.
26. Blake, J.A.; Bult, C.J.; Eppig, J.T.; Kadin, J.A.; Richardson, J.E. The Mouse Genome Database: Integration of and access to knowledge about the laboratory mouse. *Nucl. Acids Res*. **2014**, *42*, D810–D817.
27. Smith, C.M.; Finger, J.H.; Hayamizu, T.F.; McCright, I.J.; Xu, J.; Berghout, J.; Campbell, J.; Corbani, L.E.; Forthofer, K.L.; Frost, J,P. The mouse Gene Expression Database (GXD): 2014 update. *Nucl. Acids Res*. **2014**, *42*, D818–D824.
28. Begley, D.A.; Krupke, D.M.; Neuhauser, S.B.; Richardson, J.E.; Bult, C.J.; Eppig, J.T.; Sundberg, J.P. The Mouse Tumor Biology Database (MTB): A central electronic resource for locating and integrating mouse tumor pathology data. *Vet. Pathol*. **2012**, *49*, 218–223.
29. Wiltshire, S.A.; Leiva-Torres, G.A.; Vidal, S.M. Quantitative trait locus analysis, pathway analysis, and consomic mapping show genetic variants of Tnni3k, Fpgt, or H28 control susceptibility to viral myocarditis. *J. Immunol*. **2011**, *186*, 6398–6405.

30. Di Pietrantonio, T.; Hernandez, C.; Girard, M.; Verville, A.; Orlova, M.; Belley, A.; Behr, M.A.; Loredo-Osti, J.C.; Schurr, E. Strain-specific differences in the genetic control of two closely related mycobacteria. *PLoS Pathog*. **2010**, *6*, e1001169.
31. Toth, L.A.; Trammell, R.A.; Williams, R.W. Mapping complex traits using families of recombinant inbred strains: An overview and example of mapping susceptibility to Candida albicans induced illness phenotypes. *Pathog. Dis*. **2014**, *71*, 232–246.
32. Boivin, G.A.; Pothlichet, J.; Skamene, E.; Brown, E.G.; Loredo-Osti, J.C.; Sladek, R.; Vidal, S.M. Mapping of clinical and expression quantitative trait loci in a sex-dependent effect of host susceptibility to mouse-adapted influenza H3N2/HK/1/68. *J. Immunol*. **2012**, *188*, 3949–3960.
33. Ferris, M.T.; Aylor, D.L.; Bottomly, D.; Whitmore, A.C.; Aicher, L.D.; Bell, T.A.; Bradel-Tretheway, B.; Bryan, J.T.; Buus, R.J.; Gralinski, L.E.; *et al*. Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog*. **2013**, *9*, e1003196.
34. Guenet, J.L.; Bonhomme, F. Wild mice: An ever-increasing contribution to a popular mammalian model. *Trends Genet.* **2003**, *19*, 24–31.
35. Chia, R.; Achilli, F.; Festing, M.F.; Fisher, E.M. The origins and uses of mouse outbred stocks. *Nat. Genet.* **2005**, *37*, 1181–1186.
36. Keane, T.M.; Goodstadt, L.; Danecek, P.; White, M.A.; Wong, K.; Yalcin, B.; Heger, A.; Agam, A.; Slater, G.; Goodson, M.; *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **2011**, *477*, 289–294.
37. Simon, M.M.; Greenaway, S.; White, J.K.; Fuchs, H.; Gailus-Durner, V.; Wells, S.; Sorg, T.; Wong, K.; Bedu, E.; Cartwright, E.J.; *et al.* A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genet. Biol*. **2013**, *14*, R82.
38. Vidal, S.M.; Malo, D.; Marquis, J.F.; Gros, P. Forward genetic dissection of immunity to infection in the mouse. *Annu. Rev. Immunol*. **2008**, *26*, 81–132.
39. Cook, M.C.; Vinuesa, C.G.; Goodnow, C.C. ENU-mutagenesis: Insight into immune function and pathology. *Curr. Opin. Immunol*. **2006**, *18*, 627–633.
40. Hoebe, K.; Beutler, B. Forward genetic analysis of TLR-signaling pathways: An evaluation. *Adv. Drug Deliv. Rev*. **2008**, *60*, 824–829.
41. Hoyne, G.F.; Goodnow, C.C. The use of genomewide ENU mutagenesis screens to unravel complex mammalian traits: Identifying genes that regulate organ-specific and systemic autoimmunity. *Immunol. Rev*. **2006**, *210*, 27–39.
42. Oliver, P.L.; Davies, K.E. New insights into behaviour using mouse ENU mutagenesis. *Hum. Mol. Genet.* **2012**, *21*, R72–R81.
43. Russell, W.L. Radiation and chemical mutagenesis and repair in mice. *Johns Hopkins Med. J. Suppl*. **1972**, *1*, 239–247.
44. Jaenisch, R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 1260–1264.
45. Russell, L.B.; Russell, W.L. Frequency and nature of specific-locus mutations induced in female mice by radiations and chemicals: A review. *Mutat. Res*. **1992**, *296*, 107–127.

46. Russell, W.L.; Kelly, E.M.; Hunsicker, P.R.; Bangham, J.W.; Maddux, S.C.; Phipps, E.L. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 5818–5819.
47. Russell, W.L.; Hunsicker, P.R.; Carpenter, D.A.; Cornett, C.V.; Guinn, G.M. Effect of dose fractionation on the ethylnitrosourea induction of specific-locus mutations in mouse spermatogonia. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3592–3593.
48. Hitotsumachi, S.; Carpenter, D.A.; Russell, W.L. Dose-repetition increases the mutagenic effectiveness of N-ethyl-N-nitrosourea in mouse spermatogonia. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6619–6621.
49. Singer, B. All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature* **1976**, *264*, 333–339.
50. Bignami, M.; Vitelli, A.; di Muccio, A.; Terlizzese, M.; Calcagnile, A.; Zapponi, G.A.; Lohman, P.H.M.; den Engelse, L.; Dogliotti1, E. Relationship between specific alkylated bases and mutations at two gene loci induced by ethylnitrosourea and diethyl sulfate in CHO cells. *Mutat. Res*. **1988**, *193*, 43–51.
51. Van Zeeland, A.A. Molecular dosimetry of alkylating agents: Quantitative comparison of genetic effects on the basis of DNA adduct formation. *Mutagenesis* **1988**, *3*, 179–191.
52. Vogel, E.W.; Natarajan, A.T. DNA damage and repair in somatic and germ cells *in vivo*. *Mutat. Res*. **1995**, *330*, 183–208.
53. Fairfield, H.; Gilbert, G.J.; Barter, M.; Corrigan, R.R.; Curtain, M.; Ding, Y.M.; Ascenzo, M.D.; Gerhardt, D.J.; He, C.; Huang, W.H.; *et al.* Mutation discovery in mice by whole exome sequencing. *Genet. Biol*. **2011**, *12*, R86.
54. Andrews, T.D.; Whittle, B.; Field, M.A.; Balakishnan, B.; Zhang, Y.; Shao, Y.; Cho, V.; Kirk, M.; Singh, M.; Xia, Y.; *et al.* Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: An immediate source for thousands of new mouse models. *Open Biol.* **2012**, *2*, 120061.
55. Bull, K.R.; Rimmer, A.J.; Siggs, O.M.; Miosge, L.A.; Roots, C.M.; Enders, A.; Bertram, E.M.; Crockford, T.L.; Whittle, B.; Potter, P.K.; *et al.* Unlocking the bottleneck in forward genetics using whole-genome sequencing and identity by descent to isolate causative mutations. *PLoS Genet.* **2013**, *9*, e1003219.
56. Lewis, M.A.; Quint, E.; Glazier, A.M.; Fuchs, H.; de Angelis, M.H.; Langford, C.; van Dongen, S.; Abreu-Goodger, C.; Piipari, M.; Redshaw, N.; *et al.* An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat. Genet.* **2009**, *41*, 614–618.
57. Masuya, H.; Sezutsu, H.; Sakuraba, Y.; Sagai, T.; Hosoya, M.; Kanedaa, H.; Miuraa, I.; Kobayashia, K.; Sumiyamad, K.; Shimizu, A.; *et al.* A series of ENU-induced single-base substitutions in a long-range cis-element altering sonic hedgehog expression in the developing mouse limb bud. *Genomics* **2007**, *89*, 207–214.
58. Arnold, C.N.; Barnes, M.J.; Berger, M.; Blasius, A.L.; Brandl, K.; Croker, B.; Crozat, K.; Du, X.; Eidenschenk, C.; Georgel, P.; *et al.* ENU-induced phenovariance in mice: Inferences from 587 mutations. *BMC Res. Notes* **2012**, *5*, 577.

59. Puk, O.; Moller, G.; Geerlof, A.; Krowiorz, K.; Ahmad, N.; Wagner, S.; Adamski, J.; de Angelis, M.H.; Graw, J. The pathologic effect of a novel neomorphic Fgf9(Y162C) allele is restricted to decreased vision and retarded lens growth. *PLoS ONE* **2011**, *6*, e23678.
60. Caspary, T. Phenotype-driven mouse ENU mutagenesis screens. *Methods Enzymol.* **2010**, *477*, 313–327.
61. Probst, F.J.; Justice, M.J. Mouse mutagenesis with the chemical supermutagen ENU. *Methods Enzymol.* **2010**, *477*, 297–312.
62. Georgel, P.; Du, X.; Hoebe, K.; Beutler, B. ENU mutagenesis in mice. *Methods Mol. Biol.* **2008**, *415*, 1–16.
63. Justice, M.J.; Carpenter, D.A.; Favor, J.; Neuhauser-Klaus, A.; de Angelis, M.H.; Soewarto, D.; Moser, A.; Cordes, S.; Miller, D.; Chapman, V.; *et al.* Effects of ENU dosage on mouse strains. *Mamm. Genet.* **2000**, *11*, 484–488.
64. Bainbridge, M.N.; Wang, M.; Burgess, D.L.; Kovar, C.; Rodesch, M.J.; D'Ascenzo, M.; Kitzman, J.; Wu, Y.-Q.; Newsham, I.; Richmond, T.A.; *et al.* Whole exome capture in solution with 3 Gbp of data. *Genet. Biol.* **2010**, *11*, R62.
65. Choi, M.; Scholl, U.I.; Ji, W.; Liu, T.; Tikhonova, I.R.; Zumbob, P.; Nayirc, A.; Bakkaloğlud, A.; Özend, S.; Sanjad, S.; *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19096–19101.
66. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35.
67. Pruitt, K.D.; Harrow, J.; Harte, R.A.; Wallin, C.; Diekhans, M.; Maglott, D.R.; Searle, S.; Farrell, C.M.; Loveland, J.E.; Ruef, B.J.; *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genet. Res.* **2009**, *19*, 1316–1323.
68. Mouse Genome Project. Available online: http://www.sanger.ac.uk/resources/mouse/genomes/ (accessed on 28 August 2014).
69. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498.
70. Pipeline Space Home. Available online: https://biowiki.atlassian.net/wiki/display/PS/Pipeline+Space+Home (accessed on 28 August 2014).
71. GATK Best Practices. Available online: http://www.broadinstitute.org/gatk/guide/best-practices (accessed on 8 August 2014).
72. Lohse, M.; Bolger, A.M.; Nagel, A.; Fernie, A.R.; Lunn, J.E.; Stitt, M.; Usadel, B. RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucl. Acids Res.* **2012**, *40*, W622–W627.
73. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows—Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595.

74. Li, H.; Homer, N. A survey of sequence alignment algorithms for next-generation sequencing. *Br. Bioinform*. **2010**, *11*, 473–483.
75. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.
76. Skliris, G.P.; Rowan, B.G.; Al-Dhaheri, M.; Williams, C.; Troup, S.; Begic, S.; Parisien, M.; Watson, P.H.; Murphy, L.C. Immunohistochemical validation of multiple phospho-specific epitopes for estrogen receptor alpha (ERalpha) in tissue microarrays of ERalpha positive human breast carcinomas. *Breast Cancer Res. Treat* **2009**, *118*, 443–453.
77. Picard. Available online: http://picard.sourceforge.net (accessed on 28 August 2014).
78. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. Available online: http://arxiv.org/abs/1207.3907 (accessed on 28 August 2014).
79. Nielsen, R.; Paul, J.S.; Albrechtsen, A.; Song, Y.S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **2011**, *12*, 443–451.
80. Moresco, E.M.; Li, X.; Beutler, B. Going forward with genetics: Recent technological advances and forward genetics in mice. *Am. J. Pathol*. **2013**, *182*, 1462–1473.
81. Cingolani, P.; Platts, A.; Wang le, L.; Coon, M.; Nguyen, T.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **2012**, *6*, 80–92.
82. McLaren, W.; Pritchard, B.; Rios, D.; Chen, Y.; Flicek, P.; Cunningham, F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **2010**, *26*, 2069–2070.
83. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. DbSNP: The NCBI database of genetic variation. *Nucl. Acids Res*. **2001**, *29*, 308–311.
84. Robinson, J.T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol*. **2011**, *29*, 24–26.
85. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
86. Kwiatkowski, D.P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **2005**, *77*, 171–192.
87. Ayi, K.; Turrini, F.; Piga, A.; Arese, P. Enhanced phagocytosis of ring-parasitized mutant erythrocytes: A common mechanism that may explain protection against falciparum malaria in sickle trait and beta-thalassemia trait. *Blood* **2004**, *104*, 3364–3371.
88. Friedman, M.J. Erythrocytic mechanism of sickle cell resistance to malaria. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 1994–1997.
89. Allen, S.J.; O'Donnell, A.; Alexander, N.D.; Mgone, C.S.; Peto, T.E.; Clegg, J.B.; Alpers, M.P.; Weatherall, D.J. Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 1056–1060.

90. Foo, L.C.; Rekhraj, V.; Chiang, G.L.; Mak, J.W. Ovalocytosis protects against severe malaria parasitemia in the Malayan aborigines. *Am. J. Trop. Med. Hyg*. **1992**, *47*, 271–275.

91. Genton, B.; al-Yaman, F.; Mgone, C.S.; Alexander, N.; Paniu, M.M.; Alpers, M.P.; Mokela, D. Ovalocytosis and cerebral malaria. *Nature* **1995**, *378*, 564–565.

92. Miller, L.H.; Mason, S.J.; Clyde, D.F.; McGinniss, M.H. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N. Engl. J. Med.* **1976**, *295*, 302–304.

93. Ruwende, C.; Khoo, S.C.; Snow, R.W.; Yates, S.N.; Kwiatkowski, D.; Gupta, S.; Warn, P.; Allsopp, C.E.; Gilbert, S.C.; Peschu, N.; *et al*. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature* **1995**, *376*, 246–249.

94. Tishkoff, S.A.; Varkonyi, R.; Cahinhinan, N.; Abbes, S.; Argyropoulos, G.; Destro-Bisol, G.; Drousiotou, A.; Dangerfield, B.; Lefranc, G.; Loiselet, J.; *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science* **2001**, *293*, 455–462.

95. Aitman, T.J.; Cooper, L.D.; Norsworthy, P.J.; Wahid, F.N.; Gray, J.K.; Curtis, B.R.; McKeigue, P.M.; Kwiatkowski, D.; Greenwood, B.M.; Snow, R.W.; *et al.* Malaria susceptibility and CD36 mutation. *Nature* **2000**, *405*, 1015–1016.

96. Omi, K.; Ohashi, J.; Patarapotikul, J.; Hananantachai, H.; Naka, I.; Pottere, S.; Medanad, I.M.; Miua, J.; Ball, H.J. CD36 polymorphism is associated with protection from cerebral malaria. *Am. J. Hum. Genet.* **2003**, *72*, 364–374.

97. Hill, A.V.; Allsopp, C.E.; Kwiatkowski, D.; Anstey, N.M.; Twumasi, P.; Rowe, P.A.; Bennett, S.; Brewster, D.; McMichael, A.J.; Greenwood, B.M. Common west African HLA antigens are associated with protection from severe malaria. *Nature* **1991**, *352*, 595–600.

98. McGuire, W.; Hill, A.V.; Allsopp, C.E.; Greenwood, B.M.; Kwiatkowski, D. Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* **1994**, *371*, 508–510.

99. Bongfen, S.E.; Laroque, A.; Berghout, J.; Gros, P. Genetic and genomic analyses of host-pathogen interactions in malaria. *Trends Parasitol*. **2009**, *25*, 417–422.

100. Verra, F.; Mangano, V.D.; Modiano, D. Genetics of susceptibility to Plasmodium falciparum: From classical malaria resistance genes towards genome-wide association studies. *Parasit. Immunol*. **2009**, *31*, 234–253.

101. Fortin, A.; Stevenson, M.M.; Gros, P. Complex genetic control of susceptibility to malaria in mice. *Genes Immun*. **2002**, *3*, 177–186.

102. Hunt, N.H.; Golenser, J.; Chan-Ling, T.; Parekh, S.; Rae, C.; Pottere, S.; Medanad, I.M.; Miua, J.; Ball, H.J. Immunopathogenesis of cerebral malaria. *Int. J. Parasitol.* **2006**, *36*, 569–582.

103. Miller, L.H.; Baruch, D.I.; Marsh, K.; Doumbo, O.K. The pathogenic basis of malaria. *Nature* **2002**, *415*, 673–679.

104. Newton, C.R.; Hien, T.T.; White, N. Cerebral malaria. *J. Neurol. Neurosurg. Psychiatry* **2000**, *69*, 433–441.

105. Tripathi, A.K.; Sha, W.; Shulaev, V.; Stins, M.F.; Sullivan, D.J., Jr. Plasmodium falciparum-infected erythrocytes induce NF-kappaB regulated inflammatory pathways in human cerebral endothelium. *Blood* **2009**, *114*, 4243–4252.
106. Brown, H.; Hien, T.T.; Day, N.; Mai, N.T.; Chuong, L.V.; Chau, T.T.; Loc, P.P.; Phu, N.H.; Bethell, D.; Farrar, J.; *et al.* Evidence of blood-brain barrier dysfunction in human cerebral malaria. *Neuropathol. Appl. Neurobiol.* **1999**, *25*, 331–340.
107. Mishra, S.K.; Newton, C.R. Diagnosis and management of the neurological complications of falciparum malaria. *Nat. Rev. Neurol.* **2009**, *5*, 189–198.
108. Hafalla, J.C.; Claser, C.; Couper, K.N.; Grau, G.E.; Renia, L.; de Souza, J.B.; Riley, E.M. The CTLA-4 and PD-1/PD-L1 inhibitory pathways independently regulate host resistance to Plasmodium-induced acute immune pathology. *PLoS Pathog.* **2012**, *8*, e1002504.
109. Lacerda-Queiroz, N.; Rodrigues, D.H.; Vilela, M.C.; Rachid, M.A.; Soriani, F.M.; Sousa, L.P.; Campos, R.D.; Quesniaux, V.F.; Teixeira, M.M.; Teixeira, A.L. Platelet-activating factor receptor is essential for the development of experimental cerebral malaria. *Am. J. Pathol.* **2012**, *180*, 246–255.
110. Lou, J.; Lucas, R.; Grau, G.E. Pathogenesis of cerebral malaria: Recent experimental data and possible applications for humans. *Clin. Microbiol. Rev.* **2001**, *14*, 810–820.
111. De Souza, J.B.; Riley, E.M. Cerebral malaria: The contribution of studies in animal models to our understanding of immunopathogenesis. *Microbes Infect.* **2002**, *4*, 291–300.
112. Senaldi, G.; Vesin, C.; Chang, R.; Grau, G.E.; Piguet, P.F. Role of polymorphonuclear neutrophil leukocytes and their integrin CD11a (LFA-1) in the pathogenesis of severe murine malaria. *Infect. Immun.* **1994**, *62*, 1144–1149.
113. Bagot, S.; Idrissa Boubou, M.; Campinom, S.; Behrschmidt, C.; Gorgette, O.; Guénet, J.L.; Penha-Gonçalves, C.; Mazier, D.; Pied, S.; Cazenave, P.A. Susceptibility to experimental cerebral malaria induced by Plasmodium berghei ANKA in inbred mouse strains recently derived from wild stock. *Infect. Immun.* **2002**, *70*, 2049–2056.
114. Berghout, J.; Min-Oo, G.; Tam, M.; Gauthier, S.; Stevenson, M.M.; Gros, P. Identification of a novel cerebral malaria susceptibility locus (Berr5) on mouse chromosome 19. *Genes Immun.* **2010**, *11*, 310–318.
115. Bopp, S.E.; Rodrigo, E.; Gonzalez-Paez, G.E.; Frazer, M.; Barnes, S.W.; Valim, C.; Watson, J.; Walker, J.R.; Schmedt, C.; Winzeler, E.A. Identification of the Plasmodium berghei resistance locus 9 linked to survival on chromosome 9. *Malar. J.* **2013**, *12*, 316.
116. Campino, S.; Bagot, S.; Bergman, M.L.; Almeida, P.; Sepulveda, N.; Pied, S.; Penha-Gonçalves, C.; Holmberg, D.; Cazenave, P.-A. Genetic control of parasite clearance leads to resistance to Plasmodium berghei ANKA infection and confers immunity. *Genes Immun.* **2005**, *6*, 416–421.
117. Ohno, T.; Nishimura, M. Detection of a new cerebral malaria susceptibility locus, using CBA mice. *Immunogenetics* **2004**, *56*, 675–678.

118. Bopp, S.E.; Ramachandran, V.; Henson, K.; Luzader, A.; Lindstrom, M.; Spooner, M.; Steffy, B.M.; Suzuki, O.; Janse, C.; Waters, A.P.; *et al.* Genome wide analysis of inbred mouse lines identifies a locus containing Ppar-gamma as contributing to enhanced malaria survival. *PLoS ONE* **2010**, *5*, e10903.

119. Bongfen, S.E.; Rodrigue-Gervais, I.G.; Berghout, J.; Torre, S.; Cingolani, P.; Wiltshire, S.A.; Leiva-Torres, G.A.; Letourneau, L.; Sladek, R.; Blanchette, M.; *et al.* An N-ethyl-N-nitrosourea (ENU)-induced dominant negative mutation in the JAK3 kinase protects against cerebral malaria. *PLoS ONE* **2012**, *7*, e31012.

120. Nosaka, T.; van Deursen, J.M.; Tripp, R.A.; Thierfelder, W.E.; Witthuhn, B.A.; McMickle, A.P.; Doherty, P.C.; Grosveld, G.C.; Ihle, J.N. Defective lymphoid development in mice lacking Jak3. *Science* **1995**, *270*, 800–802.

121. Jostins, L.; Ripke, S.; Weersma, R.K.; Duerr, R.H.; McGovern, D.P.; Hui, K.Y.; Lee, J.C.; Schumm, L.P.; Sharma, Y.; Anderson, C.A.; *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **2012**, *491*, 119–124.

122. International Multiple Sclerosis Genetics Consortium. The genetic association of variants in CD6, TNFRSF1A and IRF8 to multiple sclerosis: A multicenter case-control study. *PLoS ONE* **2011**, *6*, e18813.

123. Jakkula, E.; Leppa, V.; Sulonen, A.M.; Varilo, T.; Kallio, S.; Kemppinen, A.; Purcell, S.; Koivisto, K.; Tienari, P.; Sumelahti, M.; *et al.* Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am. J. Hum. Genet.* **2010**, *86*, 285–291.

124. Torre, S.; van Bruggen, R.; Kennedy, J.M.; Berghout, J.; Bongfen, S.E.; Langat, P.; Lathrop, M.; Vidal, S.M.; Gros, P. Susceptibility to lethal cerebral malaria is regulated by epistatic interaction between chromosome 4 (Berr6) and chromosome 1 (Berr7) loci in mice. *Genes Immun*. **2013**, *14*, 470.

125. Torre, S.; Gros, P. Department of Human Genetics, Department of Biochemistry, and Complex Traits Group, McGill University, Montréal, QC, Canada. Unpublished data, 2014.

126. Pignata, C.; Fusco, A.; Amorosi, S. Human clinical phenotype associated with FOXN1 mutations. *Adv. Exp. Med. Biol*. **2009**, *665*, 195–206.

127. Rank, G.; Sutton, R.; Marshall, V.; Lundie, R.J.; Caddy, J.; Romeo, T.; Fernandez, K.; Mccormack, M.P.; Cooke, B.M.; Foote, S.J.; *et al.* Novel roles for erythroid Ankyrin-1 revealed through an ENU-induced null mouse mutant. *Blood* **2009**, *113*, 3352–3362.

128. Greth, A.; Lampkin, S.; Mayura-Guru, P.; Rodda, F.; Drysdale, K.; Roberts-Thomson, M.; McMorran, B.J.; Foote, S.J.; Burgio, G. A novel ENU-mutation in ankyrin-1 disrupts malaria parasite maturation in red blood cells of mice. *PLoS One* **2012**, *7*, e38999.

129. Belnoue, E.; Kayibanda, M.; Vigario, A.M.; Deschemin, J.C.; van Rooijen, N.; Viguier, M.; Snounou, G.; Rénia, L. On the pathogenic role of brain-sequestered alphabeta CD8+ T cells in experimental cerebral malaria. *J. Immunol*. **2002**, *169*, 6369–6375.

130. Finley, R.W.; Mackey, L.J.; Lambert, P.H.; Virulent, P. berghei malaria: Prolonged survival and decreased cerebral pathology in cell-dependent nude mice. *J. Immunol.* **1982**, *129*, 2213–2218.

131. Renia, L.; Potter, S.M.; Mauduit, M.; Rosa, D.S.; Kayibanda, M.; Deschemina, J.; Snounoub, G.; Grüner, A.C. Pathogenic T cells in cerebral malaria. *Int. J. Parasitol.* **2006**, *36*, 547–554.
132. Grau, G.E.; Piguet, P.F.; Engers, H.D.; Louis, J.A.; Vassalli, P.; Lambert, P.H. L3T4+ T lymphocytes play a major role in the pathogenesis of murine cerebral malaria. *J. Immunol.* **1986**, *137*, 2348–2354.
133. Crump, J.A.; Luby, S.P.; Mintz, E.D. The global burden of typhoid fever. *Bull. World Health Organ.* **2004**, *82*, 346–353.
134. Mastroeni, P.; Maskell, D. *Salmonella Infections: Clinical*, *Immunological*, *and Molecular Aspects*; Cambridge University Press: Cambridge, UK, 2006; Volume 13, p. 381.
135. Gordon, M.A.; Banda, H.T.; Gondwe, M.; Gordon, S.B.; Boeree, M.J.; Walsh, A.L.; Corkill, J.E.; Hart, C.A.; Gilks, C.F.; Molyneux, M.E. Non-typhoidal *Salmonella* bacteraemia among HIV-infected Malawian adults: High mortality and frequent recrudescence. *AIDS* **2002**, *16*, 1633–1641.
136. Majowicz, S.E.; Musto, J.; Scallan, E.; Angulo, F.J.; Kirk, M.; O'Brien, S.J.; Jones, T.F.; Fazil, A.; Hoekstra, R.M. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin. Infect. Dis*. **2010**, *50*, 882–889.
137. Mittrucker, H.W.; Kaufmann, S.H. Immune response to infection with *Salmonella typhimurium* in mice. *J. Leukoc. Biol.* **2000**, *67*, 457–463.
138. Dougan, G.; John, V.; Palmer, S.; Mastroeni, P. Immunity to salmonellosis. *Immunol. Rev.* **2011**, *240*, 196–210.
139. Alcais, A.; Abel, L.; Casanova, J.L. Human genetics of infectious diseases: Between proof of principle and paradigm. *J. Clin. Invest*. **2009**, *119*, 2506–2514.
140. Bustamante, J.; Zhang, S.Y.; von Bernuth, H.; Abel, L.; Casanova, J.L. From infectious diseases to primary immunodeficiencies. *Immunol. Allergy Clin. North Am*. **2008**, *28*, 235–258.
141. Casanova, J.L.; Fieschi, C.; Zhang, S.Y.; Abel, L. Revisiting human primary immunodeficiencies. *J. Intern. Med*. **2008**, *264*, 115–127.
142. De Beaucoudrey, L.; Samarina, A.; Bustamante, J.; Cobat, A.; Boisson-Dupuis, S.; Feinberg, J.; Al-Muhsen, S.; Jannière, L.; Rose, Y.; Desurenaim, M.; *et al*. Revisiting human IL-12Rbeta1 deficiency: A survey of 141 patients from 30 countries. *Med. (Baltim.)* **2010**, *89*, 381–402.
143. Gordon, M. *Salmonella* infections in immunocompromised adults. *J. Infect*. **2008**, *56*, 413–422.
144. Lammas, D.A.; Casanova, J.L.; Kumararatne, D.S. Clinical consequences of defects in the IL-12-dependent interferon-gamma (IFN-gamma) pathway. *Clin. Exp. Immunol*. **2000**, *121*, 417–425.
145. Dunstan, S.J.; Stephens, H.A.; Blackwell, J.M.; Duc, C.M.; Lanh, M.N.; Dudbridge, F.; Phuong, C.X.; Luxemburger, C.; Wain, J.; Ho, V.A.; *et al.* Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam. *J. Infect. Dis.* **2001**, *183*, 261–268.

146. House, D.; Bishop, A.; Parry, C.; Dougan, G.; Wain, J. Typhoid fever: Pathogenesis and disease. *Curr. Opin. Infect. Dis. 2001*, *14*, 573–578.

147. Santos, R.L.; Zhang, S.; Tsolis, R.M.; Kingsley, R.A.; Adams, L.G.; Bäumler, A.J. Animal models of *Salmonella* infections: Enteritis *versus* typhoid fever. *Microbes Infect.* **2001**, *3*, 1335–1344.

148. Roy, M.F.; Malo, D. Genetic regulation of host responses to *Salmonella* infection in mice. *Genes Immun*. **2002**, *3*, 381–393.

149. Malo, D.; Vogan, K.; Vidal, S.; Hu, J.; Cellier, M.; Schurr, E.; Fuks, A.; Bumstead, N.; Morgan, K.; Gros, P. Haplotype mapping and sequence analysis of the mouse Nramp gene predict susceptibility to infection with intracellular parasites. *Genomics* **1994**, *23*, 51–61.

150. Qureshi, S.T.; Lariviere, L.; Leveque, G.; Clermont, S.; Moore, K.J.; Gros, P.; Malo, D. Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J. Exp. Med.* **1999**, *189*, 615–625.

151. Poltorak, A.; He, X.; Smirnova, I.; Liu, M.Y.; van Huffel, C.; Du, X.; Birdwell, D.; Alejos, E.; Silva, M.; Galanos, C.; *et al*. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: Mutations in Tlr4 gene. *Science* **1998**, *282*, 2085–2088.

152. Vidal, S.M.; Malo, D.; Vogan, K.; Skamene, E.; Gros, P. Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg. *Cell* **1993**, *73*, 469–485.

153. Roy, M.F.; Riendeau, N.; Bedard, C.; Helie, P.; Min-Oo, G.; Turcotte, K.; Gros, P.; Canonne-Hergaux, F.; Malo, D. Pyruvate kinase deficiency confers susceptibility to *Salmonella typhimurium* infection in mice. *J. Exp. Med.* **2007**, *204*, 2949–2961.

154. Yuki, K.E.; Eva, M.M.; Richer, E.; Chung, D.; Paquet, M.; Cellier, M.; Canonne-Hergaux, F.; Vaulont, S.; Vidal, S.M.; Malo, D.; *et al*. Suppression of hepcidin expression and iron overload mediate *Salmonella* susceptibility in ankyrin 1 ENU-induced mutant. *PLoS One* **2013**, *8*, e55331.

155. Eva, M.M.; Yuki, K.E.; Dauphinee, S.M.; Schwartzentruber, J.A.; Pyzik, M.; Paquet, M.; Lathrop, M.; Majewski, J.; Vidal, S.M.; Malo, D.; *et al.* Altered IFN-gamma-mediated immunity and transcriptional expression patterns in N-Ethyl-N-nitrosourea-induced STAT4 mutants confer susceptibility to acute typhoid-like disease. *J. Immunol*. **2014**, *192*, 259–270.

156. Richer, E.; Prendergast, C.; Zhang, D.E.; Qureshi, S.T.; Vidal, S.M.; Malo, D. N-ethyl-N-nitrosourea-induced mutation in ubiquitin-specific peptidase 18 causes hyperactivation of IFN-alphass signaling and suppresses STAT4-induced IFN-gamma production, resulting in increased susceptibility to *Salmonella typhimurium*. *J. Immunol.* **2010**, *185*, 3593–3601.

157. Malakhova, O.; Malakhov, M.; Hetherington, C.; Zhang, D.E. Lipopolysaccharide activates the expression of ISG15-specific protease UBP43 via interferon regulatory factor 3. *J. Biol. Chem*. **2002**, *277*, 14703–14711.

158. Kim, K.I.; Yan, M.; Malakhova, O.; Luo, J.K.; Shen, M.F.; Zou, W.; de la Torre, J.C.; Zhang, D. Ube1L and protein ISGylation are not essential for alpha/beta interferon signaling. *Mol. Cell. Biol.* **2006**, *26*, 472–479.

159. Richer, E.; Yuki, K.E.; Dauphinee, S.M.; Lariviere, L.; Paquet, M.; Malo, D. Impact of Usp18 and IFN signaling in *Salmonella*-induced typhlitis. *Genes Immun*. **2011**, *12*, 531–543.
160. Dauphinee, S.M.; Richer, E.; Eva, M.M.; McIntosh, F.; Paquet, M.; Dangoor, D.; Burkart, C.; Zhang, D.E.; Gruenheid, S.; Gros, P.; *et al*. Contribution of increased ISG15, ISGylation and deregulated type I IFN signaling in Usp18 mutant mice during the course of bacterial infections. *Genes Immun*. **2014**, *15*, 282–292.
161. Perrotta, S.; Gallagher, P.G.; Mohandas, N. Hereditary spherocytosis. *Lancet* **2008**, *372*, 1411–1426.
162. Eber, S.W.; Gonzalez, J.M.; Lux, M.L.; Scarpa, A.L.; Tse, W.T.; Dornwell, M.; Herbers, J.; Kugler, W.; Ozcan, R.; Pekrun, A.; *et al*. Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis. *Nat. Genet.* **1996**, *13*, 214–218.
163. Min-Oo, G.; Fortin, A.; Tam, M.F.; Nantel, A.; Stevenson, M.M.; Gros, P. Pyruvate kinase deficiency in mice protects against malaria. *Nat. Genet.* **2003**, *35*, 357–362.
164. Cunnington, A.J.; de Souza, J.B.; Walther, M.; Riley, E.M. Malaria impairs resistance to *Salmonella* through heme- and heme oxygenase-dependent dysfunctional granulocyte mobilization. *Nat. Med*. **2012**, *18*, 120–127.
165. Hoebe, K.; Georgel, P.; Rutschmann, S.; Du, X.; Mudd, S.; Crozat, K.; Sovath, S.; Shame, L.; Hartung, T.; Zähringer, U.; *et al.* CD36 is a sensor of diacylglycerides. *Nature* **2005**, *433*, 523–527.
166. Xiao, N.; Eidenschenk, C.; Krebs, P.; Brandl, K.; Blasius, A.L.; Xia, Y.; Khovananth, K.; Smart, N.G.; Beutler, B. The Tpl2 mutation Sluggish impairs type I IFN production and increases susceptibility to group B streptococcal disease. *J. Immunol*. **2009**, *183*, 7975–7983.
167. Rutschmann, S.; Hoebe, K.; Zalevsky, J.; Du, X.; Mann, N.; Dahiyat, B.I.; Steed, P.; Beutler, B. PanR1, a dominant negative missense allele of the gene encoding TNF-alpha (Tnf), does not impair lymphoid development. *J. Immunol*. **2006**, *176*, 7525–7532.
168. Sauer, J.D.; Sotelo-Troha, K.; von Moltke, J.; Monroe, K.M.; Rae, C.S.; Brubaker, S.W.; Hyodo, M.; Hayakawa, Y.; Woodward, J.J.; Portnoy, D.A.; *et al.* The N-ethyl-N-nitrosourea-induced Goldenticket mouse mutant reveals an essential function of Sting in the *in vivo* interferon response to Listeria monocytogenes and cyclic dinucleotides. *Infect. Immun*. **2011**, *79*, 688–694.
169. Berger, M.; Krebs, P.; Crozat, K.; Li, X.; Croker, B.A.; Siggs, O.M.; Popkin, D.; Du, X.; Lawson, B.R.; Theofilopoulos, A.N.; *et al*. An Slfn2 mutation causes lymphoid and myeloid immunodeficiency due to loss of immune cell quiescence. *Nat. Immunol*. **2010**, *11*, 335–343.
170. Ordonez-Rueda, D.; Jonsson, F.; Mancardi, D.A.; Zhao, W.; Malzac, A.; Liang, Y.; Bertosio, E.; Grenot, P.; Blanquet, V.; Sabrautzki, S.; *et al.* A hypomorphic mutation in the Gfi1 transcriptional repressor results in a novel form of neutropenia. *Eur. J. Immunol*. **2012**, *42*, 2395–2408.
171. Georgel, P.; Crozat, K.; Lauth, X.; Makrantonaki, E.; Seltmann, H.; Sovath, S.; Hoebe, K.; Du, X.; Rutschmann, S.; Jiang, Z.; *et al.* A toll-like receptor 2-responsive lipid effector pathway protects mammals against skin infections with gram-positive bacteria. *Infect. Immun*. **2005**, *73*, 4512–4521.

172. Croker, B.A.; Lawson, B.R.; Rutschmann, S.; Berger, M.; Eidenschenk, C.; Blasius, A.L.; Moresco, E.M.Y.; Sovath, S.; Cengia, L.; Shultz, L.D.; *et al*. Inflammation and autoimmunity caused by a SHP1 mutation depend on IL-1, MyD88, and a microbial trigger. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15028–15033.

173. Jaeger, B.N.; Donadieu, J.; Cognet, C.; Bernat, C.; Ordonez-Rueda, D.; Barlogis, V.; Mahlaoui, N.; Fenis, A.; Narni-Mancinelli, E.; Beaupain, B.; *et al*. Neutrophil depletion impairs natural killer cell maturation, function, and homeostasis. *J. Exp. Med.* **2012**, *209*, 565–580.

174. Barquero-Calvo, E.; Martirosyan, A.; Ordonez-Rueda, D.; Arce-Gorvel, V.; Alfaro-Alarcon, A.; Lepidi, H.; Malissen, B.; Malissen, M.; Gorvel, J.; Moreno, E. Neutrophils exert a suppressive effect on Th1 responses to intracellular pathogen Brucella abortus. *PLoS Pathog*. **2013**, *9*, e1003167.

175. Krmpotic, A.; Bubic, I.; Polic, B.; Lucin, P.; Jonjic, S. Pathogenesis of murine cytomegalovirus infection. *Microbes Infect*. **2003**, *5*, 1263–1277.

176. Rawlinson, W.D.; Farrell, H.E.; Barrell, B.G. Analysis of the complete DNA sequence of murine cytomegalovirus. *J. Virol*. **1996**, *70*, 8833–8849.

177. Pyzik, M.; Gendron-Pontbriand, E.M.; Vidal, S.M. The impact of Ly49-NK cell-dependent recognition of MCMV infection on innate and adaptive immune responses. *J. Biomed. Biotechnol.* **2011**, *2011*, 641702.

178. Moresco, E.M.; Beutler, B. Resisting viral infection: The gene by gene approach. *Curr. Opin. Virol*. **2011**, *1*, 513–518.

179. Tabeta, K.; Georgel, P.; Janssen, E.; Du, X.; Hoebe, K.; Crozat, K.; Mudd, S.; Shamel, L.; Sovath, S.; Goode, J.; *et al.* Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3516–3521.

180. Hoebe, K.; Du, X.; Georgel, P.; Janssen, E.; Tabeta, K.; Kim, S.O.; Goode, J.; Lin, P.; Mann, N.; Mudd, S.; *et al*. Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **2003**, *424*, 743–748.

181. Tabeta, K.; Hoebe, K.; Janssen, E.M.; Du, X.; George, P.; Crozat, K.; Mudd, S.; Mann, N.; Sovath, S.; Goode, J.; *et al*. The Unc93b1 mutation 3d disrupts exogenous antigen presentation and signaling via Toll-like receptors 3, 7 and 9. *Nat. Immunol*. **2006**, *7*, 156–164.

182. Crozat, K.; Georgel, P.; Rutschmann, S.; Mann, N.; Du, X.; Hoebe, K.; Beutlerm, B. Analysis of the MCMV resistome by ENU mutagenesis. *Mamm. Genome* **2006**, *17*, 398–406.

183. Siggs, O.M.; Berger, M.; Krebs, P.; Arnold, C.N.; Eidenschenk, C.; Huberb, C.; Piriea, E.; Smarta, N.G.; Khovanantha, K.; Xia, Y.; *et al.* A mutation of Ikbkg causes immune deficiency without impairing degradation of IkappaB alpha. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3046–3051.

184. Won, S.; Eidenschenk, C.; Arnold, C.N.; Siggs, O.M.; Sun, L.; Brandla, K.; Mullenb, T.; Nemerowb, G.R.; Morescoa, E.M.Y.; Beutler, B. Increased susceptibility to DNA virus infection in mice with a GCN2 mutation. *J. Virol.* **2012**, *86*, 1802–1808.

185. Biron, C.A. Initial and innate responses to viral infections—Pattern setting in immunity or disease. *Curr. Opin. Microbiol.* **1999**, *2*, 374–381.

186. Bukowski, J.F.; Woda, B.A.; Habu, S.; Okumura, K.; Welsh, R.M. Natural killer cell depletion enhances virus synthesis and virus-induced hepatitis *in vivo*. *J. Immunol.* **1983**, *131*, 1531–1538.

187. Bukowski, J.F.; Woda, B.A.; Welsh, R.M. Pathogenesis of murine cytomegalovirus infection in natural killer cell-depleted mice. *J. Virol.* **1984**, *52*, 119–128.

188. Welsh, R.M.; Dundon, P.L.; Eynon, E.E.; Brubaker, J.O.; Koo, G.C.; O'Donnell, C.L. Demonstration of the antiviral role of natural killer cells *in vivo* with a natural killer cell-specific monoclonal antibody (NK 1.1). *Nat. Immun. Cell Growth Regul.* **1990**, *9*, 112–120.

189. Brown, M.G.; Dokun, A.O.; Heusel, J.W.; Smith, H.R.; Beckman, D.L.; Blattenberger, E.A.; Dubbelde, C.E.; Stone, L.R.; Scalzo, A.A.; Yokoyama, W.M. Vital involvement of a natural killer cell activation receptor in resistance to viral infection. *Science* **2001**, *292*, 934–937.

190. Lee, S.H.; Girard, S.; Macina, D.; Busa, M.; Zafer, A.; Belouchi, A.; Gros, P.; Vidal, S.M. Susceptibility to mouse cytomegalovirus is associated with deletion of an activating natural killer cell receptor of the C-type lectin superfamily. *Nat. Genet.* **2001**, *28*, 42–45.

191. Arase, H.; Mocarski, E.S.; Campbell, A.E.; Hill, A.B.; Lanier, L.L. Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. *Science* **2002**, *296*, 1323–1326.

192. Smith, H.R.; Heusel, J.W.; Mehta, I.K.; Kim, S.; Dorner, B.G.; Naidenko, O.V.; Iizuka, K.; Furukawa, H.; Beckman, D.L.; Pingel, J.T.; *et al.* Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8826–8831.

193. Dokun, A.O.; Kim, S.; Smith, H.R.; Kang, H.S.; Chu, D.T.; Yokoyama, W.M. Specific and nonspecific NK cell activation during virus infection. *Nat. Immunol.* **2001**, *2*, 951–956.

194. Barnes, M.J.; Aksoylar, H.; Krebs, P.; Bourdeau, T.; Arnold, C.N.; Xia, Y.; Khovananth, K.; Engel, I.; Sovath, S.; Lampe, K.; *et al.* Loss of T cell and B cell quiescence precedes the onset of microbial flora-dependent wasting disease and intestinal inflammation in Gimap5-deficient mice. *J. Immunol.* **2010**, *184*, 3743–3754.

195. Crozat, K.; Hoebe, K.; Ugolini, S.; Hong, N.A.; Janssen, E.; Rutschmann, S.; Mudd, S.; Sovath, S.; Vivier, E.; Beutler, B. Jinx, an MCMV susceptibility phenotype caused by disruption of Unc13d: A mouse model of type 3 familial hemophagocytic lymphohistiocytosis. *J. Exp. Med.* **2007**, *204*, 853–863.

196. Crozat, K.; Eidenschenk, C.; Jaeger, B.N.; Krebs, P.; Guia, S.; Beutler, B.; Vivier, E.; Ugolini, S. Impact of beta2 integrin deficiency on mouse natural killer cell development and function. *Blood* **2011**, *117*, 2874–2882.

197. Eidenschenk, C.; Crozat, K.; Krebs, P.; Arens, R.; Popkin, D.; Arnolda, C.N.; Blasiusa, A.L.; Benedictb, C.A.; Morescoa, E.M.Y.; Xiaa, Y.; *et al.* Flt3 permits survival during infection by rendering dendritic cells competent to activate NK cells. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9759–9764.

198. Croker, B.; Crozat, K.; Berger, M.; Xia, Y.; Sovath, S.; Schaffer, L.; Eleftherianos, I.; Imler, J.; Beutler, B. ATP-sensitive potassium channels mediate survival during infection in mammals and insects. *Nat. Genet.* **2007**, *39*, 1453–1460.
199. Abel, L.; Plancoulaine, S.; Jouanguy, E.; Zhang, S.Y.; Mahfoufi, N.; Nicolas, N.; Sancho-Shimizu, V.; Alcaïs, A.; Guo, Y.; Cardon, A.; *et al.* Age-dependent Mendelian predisposition to herpes simplex virus type 1 encephalitis in childhood. *J. Pediatr.* **2010**, *157*, 623–629.
200. Yao, H.W.; Ling, P.; Chen, S.H.; Tung, Y.Y.; Chen, S.H. Factors affecting herpes simplex virus reactivation from the explanted mouse brain. *Virology* **2012**, *433*, 116–123.
201. Kennedy, P.G.; Chaudhuri, A. Herpes simplex encephalitis. *J. Neurol. Neurosurg. Psychiatry* **2002**, *73*, 237–238.
202. Casrouge, A.; Zhang, S.Y.; Eidenschenk, C.; Jouanguy, E.; Puel, A.; Yang, K.; Alcais, A.; Picard, C.; Mahfoufi, N.; Nicolas, N.; *et al.* Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* **2006**, *314*, 308–312.
203. Zhang, S.Y.; Jouanguy, E.; Ugolini, S.; Smahi, A.; Elain, G.; Romero, P.; Segal, D.; Sancho-Shimizu, V.; Lorenzo, L.; Puel, A.; *et al.* TLR3 deficiency in patients with herpes simplex encephalitis. *Science* **2007**, *317*, 1522–1527.
204. Sancho-Shimizu, V.; Perez de Diego, R.; Lorenzo, L.; Halwani, R.; Alangari, A.; Israelsson, E.; Fabrega, S.; Cardon, A.; Maluenda, J.; Tatematsu, M.; *et al.* Herpes simplex encephalitis in children with autosomal recessive and dominant TRIF deficiency. *J. Clin. Invest.* **2011**, *121*, 4889–4902.
205. Perez de Diego, R.; Sancho-Shimizu, V.; Lorenzo, L.; Puel, A.; Plancoulaine, S.; Picard, C.; Herman, M.; Cardon, A.; Durandy, A.; Bustamante, J.; *et al.* Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity* **2010**, *33*, 400–411.
206. Guo, Y.; Audry, M.; Ciancanelli, M.; Alsina, L.; Azevedo, J.; Herman, M.; Anguiano, E.; Sancho-Shimizu, V.; Lorenzo, L.; Pauwels, E.; *et al.* Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *J. Exp. Med.* **2011**, *208*, 2083–2098.
207. Herman, M.; Ciancanelli, M.; Ou, Y.H.; Lorenzo, L.; Klaudel-Dreszler, M.; Pauwels, E.; Sancho-Shimizu, V.; de Diego, R.P.; Abhyankar, A.; Israelsson, E.; *et al.* Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes simplex encephalitis of childhood. *J. Exp. Med.* **2012**, *209*, 1567–1582.
208. Xia, Y.; Won, S.; Du, X.; Lin, P.; Ross, C.; la Vine, D.; Wiltshire, S.; Leiva, G.; Vidal, S.M.; Whittle, B.; *et al.* Bulk segregation mapping of mutations in closely related strains of mice. *Genetics* **2010**, *186*, 1139–1146.
209. Caignard, G.; Leiva-Torres, G.A.; Leney-Greene, M.; Charbonneau, B.; Dumaine, A.; Fodil-Cornu, N.; Pyzik, M.; Cingolani, P.; Schwartzentruber, J.; Dupaul-Chicoine, J.; *et al.* Genome-wide mouse mutagenesis reveals CD45-mediated T cell function as critical in protective immunity to HSV-1. *PLoS Pathog.* **2013**, *9*, e1003637.

210. Caignard, G.; Vidal, S.M. Department of Human Genetics and Complex Traits Group, McGill University, Montréal, QC, Canada. Unpublished data, 2014.

211. Caignard, G.; Gros, P.; Vidal, S.M. Department of Human Genetics, Department of Biochemistry, and Complex Traits Group, McGill University, Montréal, QC, Canada. Unpublished data, 2014.

212. Tchilian, E.Z.; Beverley, P.C. Altered CD45 expression and disease. *Trends Immunol.* **2006**, *27*, 146–153.

Delivery of a Clinical Genomics Service

William G. Newman and Graeme C. Black

Abstract: Over the past five years, next generation sequencing has revolutionised the discovery of genes responsible for rare inherited diseases previously resistant to traditional discovery techniques. This review considers how this new technology is being introduced into clinical practice to aid diagnosis and improve the clinical management of individuals and families affected by rare diseases where access to genetic testing was previously limited. We compare and contrast the different approaches that have been adopted including panel based tests, exome and genome sequencing. We provide insights from our own clinical practice demonstrating the challenges and benefits of this new technology.

Reprinted from *Genes*. Cite as: Newman, W.G.; Black, G.C. Delivery of a Clinical Genomics Service. *Genes* **2014**, *5*, 1001-1017.

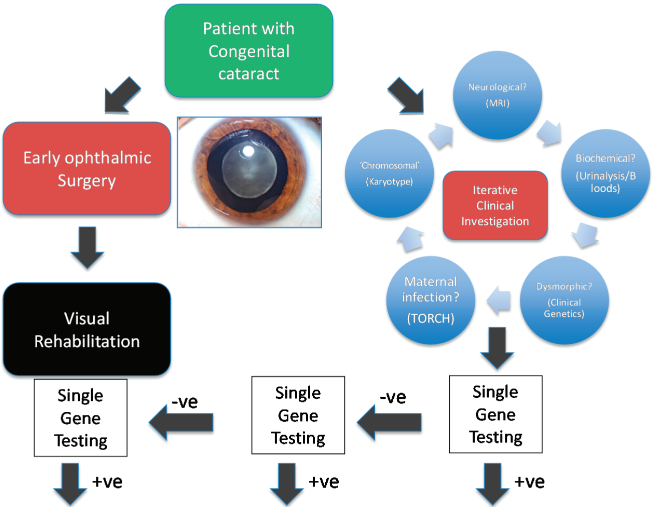
1. Clinical Genetic Testing for Inherited Disorders

Rare diseases are individually rare but affect a large number of individuals—for example, it is estimated that collectively they affect around 1 in 17 individuals in Western populations [1]. The identification of a specific genetic variant, in a patient DNA sample, that is responsible for a rare inherited disease can establish or confirm a clinical diagnosis, inform screening programmes and the implementation of personalised approaches to medical management. The information also facilitates risk assessment for affected families and enables reproductive decision-making.

Molecular genetic testing for rare diseases has been managed by a small number of expert clinicians, Clinical geneticists, over the past three to four decades, but is now becoming relevant to more patients seen across all clinical specialties. However, for clinicians within such so-called "mainstream" specialties (*i.e.*, outside of clinical genetics) it is often difficult to know how to access genetic testing for their patients. Over the past thirty years, Medical Genetics laboratories have been providing mutation testing for a relatively small number of inherited disorders due to variants in single genes. Of the approximately 7000 rare inherited disorders that have been defined, 3500 have so far been characterized at a molecular level [2]. The Genetic Testing Registry has collated the details on 16,000 tests for 4200 conditions analysing 2800 different genes [3]. The majority of these tests are still undertaken on a research basis. Clinically accredited testing provided by diagnostic laboratories is often limited.

The traditional testing model has been driven by clinical hypotheses (Figure 1, using congenital cataract, a genetically heterogeneous condition, as an exemplar). A clinician usually defines, through detailed clinical investigation, a specific phenotype and subsequently develops a testable clinical hypothesis. The resulting clinical question leads to the request of a specific (usually single gene) test or at most the testing of a very small number of potentially relevant genes. This aims to confirm or refute the clinician's suspicions and historically has been limited in great part by the technological limitations of nucleic acid sequencing. The pick up rate of such a testing approach varies considerably from approximately 0.6% for Fragile X syndrome [4] to over 40% for CHARGE syndrome [5]. In general, this has been a highly targeted approach, that is expensive, iterative and inefficient because of the limited number of target genes that can be tested and by the tendency to institute a large number of simultaneous investigations. By its very nature, it has also been limited to patients, and their relatives, with clinical features indicative of a specific genetic disease. Even where genetic testing is well established in familial breast cancer, genetic testing for *BRCA1* and *BRCA2* mutations has been limited to those with a very strong family history of the condition. Genetic molecular analysis has been especially challenging for genetically heterogeneous conditions, that is those conditions of identical phenotype cause by mutations in a wide range of genes, including intellectual and developmental delay, deafness, retinal dystrophies, congenital cataract, neuropathies and cerebellar ataxias.

Figure 1. Classical clinical hypothesis-driven diagnostic approach. Traditional investigation of genetically and clinically heterogeneous conditions, such as congenital cataract, require an inefficient and iterative process based upon the development and testing of multiple clinical hypotheses, which leads to testing of many genes in series.


Patient with Congenital cataract
Early ophthalmic Surgery
Visual Rehabilitation
Neurological? (MRI)
'Chromosomal' (Karyotype)
Biochemical? (Urinalysis/Bloods)
Iterative Clinical Investigation
Maternal infection? (TORCH)
Dysmorphic? (Clinical Genetics)
Single Gene Testing
-ve
Single Gene Testing
-ve
Single Gene Testing
+ve
+ve
+ve

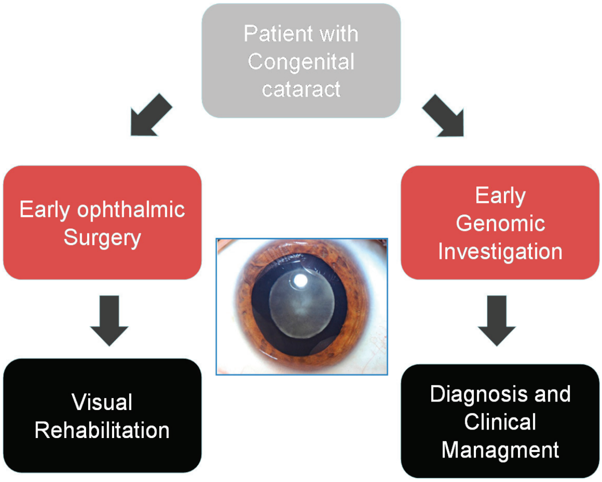
2. Next Generation Sequencing as a Diagnostic Tool

In 2009, the first proof of principle studies were published exploring the application of massively paralleled or so called "next generation" sequencing (NGS) to identify the novel causes of rare inherited diseases [6,7]. These conditions had previously not been amenable to standard gene discovery approaches, e.g., *de novo* autosomal dominant disorders could not be refined by linkage analysis and/or candidate gene approaches had proved unsuccessful. The technology and bioinformatic approach demonstrated an extremely powerful ability to identify disease-causing genes from large genomic regions using small patient cohorts. This technology has led to the molecular characterization of numerous rare disorders and has been hailed as a revolution in medical research and practice [8]. NGS has already been applied in many disciplines across medicine, including in microbiology, virology, transplantation medicine and in the identification of acquired (somatic) mutations in tumours. However, this paper considers the use of clinical application of NGS in the molecular diagnosis of rare diseases.

NGS, when first applied to Mendelian disorders focussed on gene discovery where the majority of studies used either approaches focussing of targeted sequencing of genomic regions or most commonly on whole exome sequencing (WES). The WES approach is focused on approximately 1% of the genome, which includes coding and non-coding exons, some intronic and untranslated regions and promoters. It is important to note that the terms whole exome and whole genome sequencing are misnomers as the entire sequence of the exome or genome is not covered using the currently available techniques [9]. Focussing on the protein-coding DNA sequence such an approach generates manageable datasets; although large when compared to conventional sequencing, these are comparatively small when compared to the data from complete genomes. These present challenging, but surmountable, computing challenges [8].

In the clinical setting, the commonest initial approach—that has been introduced by many clinical laboratories—is the targeted sequencing of a panel of genes relevant to a specific disease indication (Figure 2). Here, NGS has already had a major impact. Our own experience with testing of a panel of 105 inherited retinal dystrophy (IRD) genes has seen an increase in detection of the causal variant from 14% to 60% over the past two years of providing this service, allowing earlier implementation of genetic diagnosis and a reduction in the use of other diagnostic options [10]. More recently an "exome" approach to clinical diagnostic NGS sequencing has been adopted due to considerable practical advantages from the ability to develop a single diagnostic pathway for a huge range of clinical indications [11]. Please provide the original file (in ppt or other format) or a copy in tiff format of Figure 2 with high resolution.

Figure 2. Genomic diagnostic approach Genomic technologies allow early genetic investigation of heterogeneous disorders, allowing much improved diagnostic pick-up, early diagnosis and reduced cost of investigation compared to a classical approach.



Patient with Congenital cataract

Early ophthalmic Surgery

Early Genomic Investigation

Visual Rehabilitation

Diagnosis and Clinical Managment

2.1. Targeted Next Generation Sequencing for Diagnostic Molecular Testing

To sequence relatively small numbers of genes many approaches have been introduced to harness the power of NGS to improve throughput, reduce costs and improve turnaround times. For example, long range PCR has been used in our laboratory for *BRCA1* and *BRCA2* mutation analysis to generate large overlapping amplicons, which can then be sequenced. For panels of genes many technologies have been used to target the specific sequences, e.g., amplicon generation, Haloplex and hybridisation capture. Each of these methods has advantages and disadvantages in terms of labor intensity, cost and the specificity of the sequence generated. Further, each method has some limitations when identifying small insertion/deletion mutations has meant that, when using it *as a replacement for Sanger-based diagnosis,* care needs to be taken in using it as an equally effective diagnostic mechanism of excluding mutations in given genes [12].

Testing of panels of genes sequenced by NGS has been introduced with considerable success. The capture of selected sequences has been employed on both a research and diagnostic basis to study groups of genes—focused around biochemical pathways or those known to cause specific phenotypes, usually to analyse 20–200 genes in genetically heterogeneous disorders, such as IRD [10].

Such panels have many advantages over exome-based approaches: since they sequence fewer targets than genome-wide approaches they currently remain cheaper in absolute cost terms, although not when cost per base is used as the basis for evaluation. Panel-based approaches can now achieve even and very high levels of coverage of the targets selected, ≥99% coverage, allowing considerable clarity in diagnostic reporting. Importantly, many of the methodologies for capture-based exome sequencing, in particular at lower levels of overall coverage, have resulted in patchy coverage with a significant dropout of many, in particular GC-rich exons [13]. From a diagnostic viewpoint this dropout has been seen as challenging as it hinders the ability of the clinical scientist to deliver a report that provides a confident definition that a comprehensive screening of the selected genes has been undertaken and that, for exonic and flanking sequences, it is unlikely that a given individual carries a pathogenic variant. Such clarity is important as there are increasing numbers of clinicians, who are unfamiliar with complex genetic terminology and mechanisms, requesting genetic testing to inform their clinical practice.

Gene panel testing has other attractions when applied in the diagnostic sphere. The ability to define the genes that are screened lowers the likelihood that unexpected and potentially actionable findings may be encountered. However, it should be recognised that even amongst panel testing unexpected findings will nonetheless be found: for example, for two panels designed for ophthalmic disorders by our group [10,14], a wide range of conditions are covered by approximately 100–200 genes, many of broad pleiotropic effect [15]. Taking the example of retinal disease, it should be remembered that the ability to diagnose, in those with apparently isolated retinal dystrophy, syndromic conditions such as Senior-Loken and Bardet-Biedl syndromes can be—*for the patient*—unexpected and can result in altered management. When compared to single gene testing this then requires a more detailed approach to consent and counselling when implementing NGS testing.

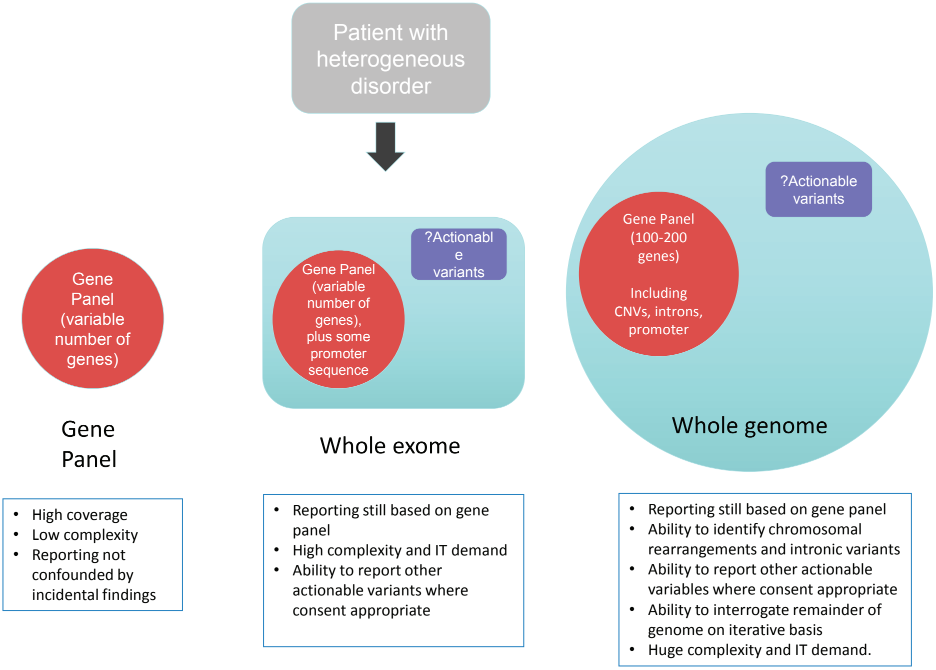
When compared to exome (WES) or whole genome (WGS) sequencing, gene panel testing thus offers apparent simplicity and has consequently been employed to improve diagnosis of genetically heterogeneous monogenic diseases (e.g., retinitis pigmentosa, congenital deafness, cardiomyopathy). The relatively small numbers of potentially novel or pathogenic variants identified enable a detailed and focussed approach to variant interpretation that is more manageable for clinical scientists analysing and interpreting variants discovered through WES. However, even at this level of complexity, there are challenges presented in patient reporting. For example, in a series of 700 patients with IRD that has been evaluated for variation in 105 genes, in 40 (approximately 12%) of those for whom a molecular cause for their condition was found were shown to be heterozygous carriers of a pathogenic variant in another gene known to cause autosomal recessive IRD (Black; personal communication [16]). Here, the issue of disclosure is not straightforward and requires clear policy decisions by the diagnostic team delivering NGS [17]. Furthermore, it is essential that these policies are complementary to, and understood by, the clinicians consenting to testing. Since families with higher levels of consanguinity may not be identified to clinical scientists, it may be necessary to report all incidental carriers in such circumstances. Such an approach may differ from WES, where the numbers of heterozygous recessive variants present in each individual is high and the identification of carrier status relating to conditions not similar to the primary indication for testing is potentially more complex.

As gene panels are adopted for clinically and phenotypically heterogeneous disorders, it becomes possible for gene testing to be employed earlier in the diagnostic pathway (Figure 2). The breadth of variants that are identified—even in small gene panels—means that interpretation is highly context dependent and in our experience this has led to the development of dialogue between the clinical reporting scientist and the diagnostic clinician. The development of multidisciplinary reporting processes allows sharing of complex phenotypic, family, clinical and genomic data. For example, amongst the multi-systemic diseases that cause congenital cataract, such as cerebroteninous xanthomatosis, Stickler or Cockayne syndromes, genomic discoveries may uncover unexpected or overlooked clinical features that require re-evaluation in the clinic. In addition, genomic discoveries in conditions such as inborn errors of metabolism may define a range of secondary clinical investigations that support genomic findings and facilitate precise diagnosis [18]. In delivering NGS multi-gene panels, the identification of variants of uncertain significance is common. While, from a clinician's viewpoint these cannot necessarily be acted upon, they represent a considerable workload for the team reporting genomic sequencing. At the current time the laboratory methodologies for NGS and the informatics tools to process the data have been honed substantially and have reached a point where this can be relatively easily automated. However, variant interpretation is both gene and phenotype specific. While there may well be certain guiding principles that are generally applicable, nonetheless this remains a labour intensive and complex aspect of NGS panel testing that must be factored into the costs of delivering testing in a healthcare setting. The diagnostic power of gene panel testing via NGS is remarkable and, alongside research developments, has allowed NGS very rapidly to contribute to clinical care. However, in planning the adoption of such processes, the hidden costs of testing are easily overlooked including the need for segregation studies, increased uptake of cascade testing and the need to evaluate the increasing demand for testing. These may be offset, potentially, by reduced adoption of clinical investigations that are superseded by NGS testing, but in many circumstances the costs of such tests are held in separate budgets to other aspects of clinical care. Finally, when considering multigene panels it is important to realise that, while compared to genome-wide approaches there is less data generated and analysed, there is nonetheless a considerable need for IT support, including sufficient computing hardware for data analysis and storage. Ensuring that data governance—in the diagnostic setting—fulfils those required in a healthcare setting immediately places a significant extra financial burden.

2.2. The Use of Genome-Wide NGS Approaches as a Diagnostic Tool

The speed of technological advance in NGS is remarkable, and has led to the technology being described as "disruptive" [19]. Recapitalisation and standardisation of approaches that are key to secure delivery of accredited diagnostics remain challenging in an environment that is yet to fully mature. The panel-based approaches, discussed above, are inherently prone to redundancy as new genes relevant to a particular condition are discovered. As wet lab sequencing and bioinformatic processing and analysis become standardized and provided by increasing numbers of diagnostic laboratories, a single test and pipeline that leads to rapid diagnosis is appealing, with economies of scale and resultant rapid turnaround. Consequently genome-wide approaches, which facilitate sequencing of all known genes, are increasingly seen to be an important step in the delivery of genomic medicine—and we will now consider both exome-based and genome based approaches (Figure 3).

Figure 3. Diagnostic approaches using next generation sequencing.


Patient with heterogeneous disorder
Gene Panel (variable number of genes)
Gene Panel
Gene Panel (variable number of genes), plus some promoter sequence
?Actionabl e variants
Whole exome
Gene Panel (100-200 genes)
Including CNVs, introns, promoter
?Actionable variants
Whole genome

- High coverage
- Low complexity
- Reporting not confounded by incidental findings

- Reporting still based on gene panel
- High complexity and IT demand
- Ability to report other actionable variants where consent appropriate

- Reporting still based on gene panel
- Ability to identify chromosomal rearrangements and intronic variants
- Ability to report other actionable variables where consent appropriate
- Ability to interrogate remainder of genome on iterative basis
- Huge complexity and IT demand.

2.2.1. Clinical Exome Sequencing

The ability of NGS to sequencing the entire exome—that is all of the coding exons of the expressed component of the genome, has fuelled gene discovery and accelerated the understanding of the pathogenesis of many monogenic diseases. As a result, *clinical exome sequencing* has been launched at a number of Centres in the United States [5], Australia and Europe and is being actively developed by clinical laboratories across the world [20]. Interestingly, in order to develop workable pipelines and a cost effective manner, at present in most clinical centres clinical reports are generated providing genetic sequencing data that is directly related to the specific phenotype of the tested individual—that is such an approach is based upon an *in silico* panel of genes that are analysed bioinformatically and reported (Figure 3). Such a targeted approach to analysis reduces substantially the cost of analysis, validation and variant interpretation. However, as discussed above, it is important in the consent process that patients and their families understand such focused analytical approaches.

In addition to a focussed approach, extended clinical reports may also be delivered that can provide information about:

(i) Carrier status for a range of recessive disorders to inform future reproductive risks.

(ii) Inherited disorders that are not predicted on the basis of family history or clinical presentation and for which treatment or preventive screening may be appropriate—so-called actionable variants. An example would be the detection of a variant in the low density lipoprotein receptor (*LDLR*) that would consistent with a diagnosis of familial hypercholesterolemia for which dietary intervention and statin treatment can reduce the risk of cardiovascular disease.

(iii) Pharmacogenetic data that may reduce the risk of adverse drug reactions, e.g., detection of variants in thiopurine methyl transferase that predict adverse response to thiopurines, e.g., azathiopurine.

There has been, and remains, extensive debate about the optimal approach to clinical exome sequencing, including uncertainty in defining the optimal population who should be tested and what information should be reported back to health care professionals and tested individuals. In one recent study of 250 cases referred for clinical exome sequencing, 80% of referrals were of children with neurological problems. In this group molecular diagnoses were confirmed in 62 (25%) with analysis confined to genes known to cause inherited disorders [5]. Of note, demonstrating the power of this approach, a significant number of the causative genes defined in this cohort had been discovered in the previous twelve months. The utility of exome testing has been explored in a number of other clinical settings, including improving diagnosis of children on intensive care units [21] or in children affected by likely recessive disorders when born to consanguineous parents [22].

Overall, exome sequencing lends itself to a high diagnostic yield in a range of clinical scenarios, including the molecular diagnosis of heterogeneous disorders, including primary immunodeficiencies and metabolic disorders. This precise diagnosis will result in reduced expenditure on alternative diagnostic tests and importantly provide patients and parents of affected children with diagnostic certainty. In addition to providing diagnostic information, reports are emerging of exome sequencing that has led to successful changes in clinical management—for example in the diagnosis and treatment of early onset inflammatory bowel disease [23] and in sepiapterin reductase deficiency in twins leading to supplementation of L-dopa therapy with 5-hydroxytryptophan [24].

2.2.2. Whole Genome Sequencing

Whole genome sequencing (WGS) is considered to be the most comprehensive form of genetic test currently available [25]. In contrast to exome sequencing relatively few studies have used (WGS) in rare disease gene discovery. Initially successes have mainly been confined to use of WGS in combination with other sequencing approaches [26] or to identify non-coding mutations that have an effect on genes known previously to cause the specific phenotype [27]. Combination approaches allow refinement of the data analysis from tens of gigabytes to megabyte levels. The control datasets for non-coding variants are less mature and the functional assays to determine the potential phenotypic effects of non-coding variants are challenging to undertake and interpret, such that confident identification of pathogeneic mutations in the non-coding genome for rare diseases remains a formidable challenge. However, WGS presents considerable technological advantages over exome sequencing in that, because it is not based around biased capture-based enrichment approaches, it generates data on an entire genome, often with a consistent average coverage. Consequently, coverage of GC rich regions is improved and there is a considerably improved ability to determine rearrangements and copy number variants. Most recently this has been applied to a cohort of 50 individuals with a diagnosis of severe learning disability (LD) [28], a series of conditions that are associated with extraordinary genetic heterogeneity that are frequently undiagnosed. The conditions can be associated with macroscopic and/or submicroscopic chromosomal rearrangements as well as *de novo* copy number variations (CNVs) and single-nucleotide variations (SNVs). These are currently diagnosed using combined microarray/NGS (targeted panel or exome) sequencing approaches and has been demonstrated that WGS represents a single genetic test that can characterize the full range of genetic variants and enable a clinician to reach a genetic diagnosis in the majority of patients with severe LD.

However, WGS is yet to be introduced widely into routine clinical practice due in large part to the technological and practical hurdles presented by the technology. The generation of terabytes of sequence data that require massive computing capacity to analyse means that WGS is mainly confined to large-scale research or commercial laboratories where it has been applied in disease gene discovery studies. Advances in computing will ensure that WGS will be introduced rapidly over the coming years to supercede both gene panel and WES.

2.2.3. Methodological Considerations of Different NGS Approaches

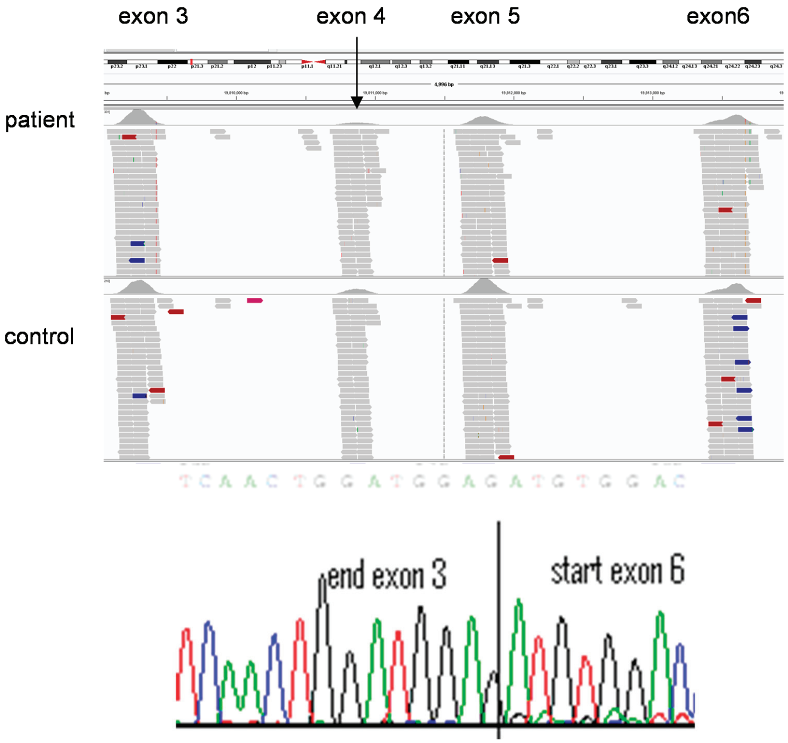
In adopting genomic technologies—from panel-based testing to WGA—the standardisation and full *clinical* validation of downstream processing will be essential. Here, a challenge is in ensuring that clinicians and clinical scientists are fully aware of the capabilities, limitations and overall design of analysis pipelines. For example, for WES we currently use a library preparation that results in, an average read depth of 140× across the exome which results in 94% coverage of the reference exome at 30× depth and generates approximately 13 Gb of data. Consequently, there is somewhat uneven coverage across many genes, a limitation that is important to stress to clinicians who may need to understand why a negative test may be received. For many capture technologies—that are used for both panel-based NGS approaches and for WES—the ability to assess dosage is limited and means that CNV analysis remains very challenging, potentially requiring reflex dosage testing. This is likely to be a limitation that WGS overcomes.

Analyses of raw data include data generation, collection and processing, followed by application-specific clustering, parsing and visualisation. Here there has been a pragmatic need to adopt research-designed bioinformatic analyses, which are often performed "in-house" using custom freeware designed pipelines for variant calling. Standardisation and full clinical stress testing will be key to ensuring that testing is of high quality, is reliably adopted and also to enabling effective data sharing across different diagnostic centres and platforms.

Variant interpretation remains extremely time-consuming and highly specialised. The process currently relies—in a diagnostic setting—on trained clinical scientists and has, to date, been far less automated than other areas of the NGS pipeline. *In silico* analyses determine whether sequence alterations are predicted to cause disruption of conserved residues. In diagnostic laboratories potential causal variants are often confirmed (currently, at least) by Sanger sequencing and segregation analyses, where possible, are undertaken to provide further evidence of pathogenicity. The definition of novel and pathogenic variants use sequence comparisons of sequences with (i) published data (themselves of highly variable reliability) (ii) databases of known mutations such as the Human Gene Mutation Database or publically available exome data resources such as Exome Variant Server [29] and (iii) the use of in-house databases of exome data. Such a labour intensive process remains important since a trained understanding of the technology and a high index of clinical suspicion can lead to re-evaluation of sequence data to define a causative mutation. For example, in a young child with severe triglyceridemia in whom a heterozygous, previously reported, mutation in *LPL* was identified. There was no sequence variant evident on the second allele to support a diagnosis of the autosomal recessive condition, lipoprotein lipase deficiency. However, the number of sequence reads was diminished across exons 4 and 5 of *LPL* (Figure 4) in comparison with an exome on the same sequence run. Subsequent, cDNA sequencing confirmed a heterozygous deletion of exons 4 and 5, confirming the diagnosis of lipoprotein lipase deficiency.

The limitations of current databasing are well understood amongst bioinformaticians and many clinical geneticists but will need to be more widely understood to enable secure variant interpretation across the clinical spectrum [30]. Of course, such resources are becoming more mature and informative as additional data is deposited and the ability to interpret exome-derived data is improving rapidly. By contrast, WGS will generate significant numbers of novel variants which will be difficult to interpret for pathogenicity and indeed it is likely that most early clinical analyses of WGS data will be focused on the *in silico* exome. For a thorough understanding of variant pathogenicity, high throughput functional studies including reporter assays, expression analyses, biochemical tests or *in vivo* assays will need to be developed to complement the emergent sequence data to allow full interpretation. Finally, the strategy around testing only the affected individual or in some scenarios WES or WGS of parents or other relatives (affected/unaffected) may be informative to refine the bioinformatics analysis and reduce the number of potential candidate causative gene variants. A successful example has been the application of a trio sequencing approach of affected child and unaffected parents to identify *de novo* pathogenic mutations, especially for severe congenital/developmental disorders [31].

Figure 4. Copy number variation detected by exome sequencing. Decreased numbers of sequence reads are present in exons 4 (e.g., see arrow) and 5 of *LPL* in the individual with lipoprotein lipase deficiency (top panel) compare to exons 3 and 6 (similar number of reads in patient DNA sequence and that of a control individual below). This indicates a heterozygous deletion of exons 4 and 5 of *LPL*, which is confirmed in the bottom panel by sequencing of cDNA generated from RNA extracted from lymphocytes from the affected individual.



3. Adoption of Clinical Genomics into Routine Clinical Practice

Next generation sequencing presents an exciting opportunity to revolutionise the diagnosis of rare disease and improve the effectiveness of healthcare delivery across all specialties. A number of specific areas will require focus if this is to be realised in a safe and effective manner:

3.1. Training

NGS is applicable across the healthcare spectrum—that is, it has been shown to be disease-agnostic. It is already proven to be a fundamental tool both clinical and research spheres and also—as recent studies have shown—relevant to both rare and common diseases. For example the next generation sequencing era introduces exciting new possibilities for singling out genetic variants of large effect that contribute to common disease in individuals as demonstrated for age-related macular degeneration [32]. A consequence of this broad relevance will be the opportunity to introduce NGS testing into the mainstream medical disciplines, including cardiology, neurology, and gastroenterology where to date genetic testing has been used less extensively and where the experience in delivering it remains more limited.

A recent survey of over 130 physicians at our Hospital across a number of specialties, including medicine (21%), surgery (13%), paediatrics (18%), anaesthetics (16%), and ophthalmology (7.5%) indicated enthusiasm for exome testing as a diagnostic aid. Over 11% of respondents had already requested an exome and over 53% envisaged requesting a test within the next five years. Limitations of current testing were availability (23%), difficulty with interpretation (47%) and concerns regarding identification of unexpected complex predictive data on cancer or neurodegenerative disease (23%). Such concerns emphasize the importance of clear guidance being established by national professional organisations in concert with patient support groups and other relevant stakeholders. However, experience from the practice of genetic medicine suggests that there is a need for an understanding of genetics, such as mutational mechanism and of genomic architecture and that this is aligned to experience in working closely with families and in delivering the counselling required to ensure effective and safe adoption of testing. Overall, therefore there is an urgent need for training to facilitate the adoption of the types of genomic technologies discussed above. This will need to be applied across all aspects of healthcare, including subspecialty clinicians and counsellors—potentially including those in primary care—who will need to be comfortable in understanding the nature and capability of the tests they order. Furthermore, this creates pressure to increase the numbers of scientists and bioinformatics experts who will be required to process the increasing number of tests.

The comparative youth of NGS is itself an inhibitor to widespread adoption in the clinical arena. In such a rapidly changing environment, the choices of technology and approach are fluid; exome capture technologies continue to improve, WGA costs are reducing and platforms rapidly maturing/becoming obsolete. Many healthcare-facing laboratories have until now been exercised with the decision to invest in the development of panel-based NGS tests or genome based (exome) approaches which are already considered by some to be out-dated. It is likely that the high cost of computing and of capitalisation/recapitalisation will either favour the larger healthcare organisations, or even lead to widespread outsourcing of sequencing. This is exemplified by the move by the 100000 Genome Project to a centralised and homogenised sequencing approach [33]. Both approaches will have a significant impact on how the technologies are introduced.

3.2. Standardised Phenotyping

The power of new genetic testing technologies to define the causes of rare inherited disorders has been remarkable. However, a limitation to further discovery has been the ability to share data generated on independent families with variants in the same gene with similar or different clinical phenotypes. Such data sharing will facilitate the definition of the ultra-rare conditions which, to date have remained undiagnosed [30]. Many research groups have identified potential causative genetic variants in single families where the burden of proof has not been satisfied to confirm causation as a mutation(s) in a second unrelated family has not been demonstrated. Many international efforts have been initiated to address this issue, including The Human Phenotype Ontology project [34] and databases that allow sharing of clinical and sequence data between clinical research groups, e.g., PhenomeCentral.

3.3. Ethical Issues

A range of complex ethical issues will influence a generalised introduction of genome-wide NGS. At present, clinical reports from such genomic testing are generated to provide feedback relevant to the presentation of the tested individual. Thus, despite the breadth of genetic information available many centres, including our own, have decided initially to apply a bioinformatic filter based on the phenotypic features of the patient that predefines the panel of genes that will be analysed [21]. Such an approach significantly reduces, but does not abolish, the likelihood of identifying co-incidental genetic variants and speeds up the data analysis.

However, the potential to generate data that identify predisposition to conditions that are not predicted from family history or current health is significant. The American College of Medical Genetics [35] and European Society of Human Genetics [36] have considered how extra information potentially generated from genome analysis should be fed back to individuals. Information about increased risks of cardiac disease, cancer and rare inherited disorders (such as Marfan syndrome) potentially lend themselves to targeted interventions with improved outcomes. However, concerns have been raised about individual autonomy, inappropriate use of this information to discriminate in terms of employment and insurance and the burden placed upon health professionals to feedback accurate information that can have a measurable benefit [35–37].

A key area of future debate will be whether only those genes that are relevant to a specific patient phenotype are assessed and information relating to these fed back to the patient from their clinical exome—and if not, then precisely which so-called "actionable variants" are reported. The use of WES and WGS is a rapidly evolving area of medicine with different views emerging as to how this should be delivered. Our local patient advocate group has indicated that patients are keen for supplemental information that is derived from such testing to be used for patient advantage. However, the anecdotal feedback from patients interviewed in a clinic setting where exome testing has been offered, has suggested more reluctance in this regard.

Lastly, it is important to note the cautionary tales from newborn screening programmes. Tandem mass spectrometry has revolutionized the number of inborn errors of metabolism that can potentially be identified in the newborn period from blood spot analysis. However, results should only be fed back to parents where there is clear evidence of benefit for the newborn child through treatment or altered clinical management, or information that may influence future parental reproductive choices. The natural history of the metabolic disorder should be known, reference should be made to histidinemia and the inappropriate adoption of newborn screening when some children were exposed the risk of liver biopsy despite the condition having a benign natural history [38]. The results of any genome test should be societally and individually acceptable and understandable.

3.4. Economic and Societal Issues

The adoption of NGS—and ultimately WGS—will happen only if the diagnostic yield is sufficient to offset the costs of adoption. The 100000 Genome Project in the UK and similar initiatives across the world will start to address the technical and interpretative challenges posed by WGS and allow comparison with WES. However, it is challenging to measure the benefit of NGS as introduced across a population. Many groups, including our own, have numerous case reports of benefit through the identification of a previously unknown diagnosis. Clinical testing has already been introduced and so undertaking studies to establish improvements in outcome is difficult in this context. Randomized control trials will potentially provide the most compelling evidence of benefit and may be possible for defined groups of conditions, but it will be very challenging to interpret the benefits across heterogeneous groups of rare disorders. Such studies will be increasingly difficult to conduct if genome testing becomes the standard of care. Furthermore there are no universally agreed outcome measures in Genetic Medicine. Standard outcome measures such as the EQ-5D are not likely to capture the potential benefit of genetic testing, as they do not often result in an alteration in any of the measured parameters, e.g., mobility [39]. An alternative to randomized trials will be to make comparisons against historical data to determine potential benefit, but such analyses are beset by potential bias.

The point at which a genome test should be used in the diagnostic pathway is yet to be defined. Should a standard suite of diagnostic tests be used initially and sequencing applied as a second line or for certain clinical indications? Should the NGS test be the first line investigation? Studies to define these pathways are urgently required to ensure appropriate use of resources and to maximise patient benefit. At present genomic tests are used with rather limited scope within medical practice. This may reflect limited education of health care professionals about their utility, a lack of a robust evidence base for their routine adoption into clinical practice; and limited evidence that some genetic tests alter the clinical management.

4. Conclusions

NGS has already transformed the landscape for individuals and families with rare inherited disorders. Conditions previously resistant to research or accurate diagnosis are now the focus of study and amenable to routine diagnosis through panel based approaches or clinical exomes. The advances in genomic sequencing technology and computing will mean that such sophisticated tests will become the standard of care for individuals with rare inherited disorders. The obligation for geneticists and healthcare professionals to harness this genomic revolution for maximum patient benefit is a real one. The ethical, legal and social implications are complex and require an open vibrant dialogue and engagement from all members of society.

Acknowledgements

The authors thank the Manchester Biomedical Research Centre for support.

Author Contributions

The authors contributed equally to the writing of this manuscript.

Conflict of Interests

The authors have no conflicts to declare.

References

1. Chief Medical Officer. Rare is common. In *Chief Medical Officer. Annual Report of the Chief Medical Officer 2009*; Department of Health: London, UK, 2009; pp. 38–45.
2. National Institutes of Health. Office of Rare Disease Research (ORDR). Available online: http://rarediseases.info.nih.gov/AboutUs.aspx (accessed on 10 August 2014).
3. Genetic Testing Registry (GTR). Available online: http://www.ncbi.nlm.nih.gov/gtr/ (accessed on 10 August 2014).
4. Macpherson, J.; Sawyer, H. Best practice guidelines for molecular diagnosis of Fragile X Syndrome. Available online: http://www.acgs.uk.com/quality-committee/best-practice-guidelines/ (accessed on 10 August 2014).
5. Yang, Y.; Muzny, D.M.; Reid, J.G.; Bainbridge, M.N.; Willis, A.; Ward, P.A.; Braxton, A.; Beuten, J.; Xia, F.; Niu, Z.; *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **2013**, *369*, 1502–1511.
6. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **2009**, *461*, 272–276.
7. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; *et al*. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35.
8. Biesecker, L.G. Exome sequencing makes medical genomics a reality. *Nat. Genet.* **2010**, *42*, 13–14.
9. Rehm, H.L. Disease-targeted sequencing: A cornerstone in the clinic. *Nat. Rev. Genet.* **2013**, *14*, 295–300.
10. O'Sullivan, J.; Mullaney, B.G.; Bhaskar, S.S.; Dickerson, J.E.; Hall, G.; O'Grady, A.; Webster, A.; Ramsden, S.C.; Black, G.C. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J. Med. Genet.* **2012**, *49*, 322–326.
11. Gilissen, C.; Hoischen, A.; Brunner, H.G.; Veltman, J.A. Disease gene identification strategies for_exome sequencing. *Eur. J. Hum. Genet.* **2012**, *20*, 490–497.
12. Braun, T.A.; Mullins, R.F.; Wagner, A.H.; Andorf, J.L.; Johnston, R.M.; Bakall, B.B.; Deluca, A.P.; Fishman, G.A.; Lam, B.L.; Weleber, R.G.; *et al.* Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum. Mol. Genet.* **2013**, *22*, 5136–5145.

13. Harismendy, O.; Ng, P.C.; Strausberg, R.L.; Wang, X.; Stockwell, T.B.; Beeson, K.Y.; Schork, N.J.; Murray, S.S.; Topol, E.J.; Levy, S.; *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **2009**, *10*, doi:10.1186/gb-2009-10-3-r32.
14. Gillespie, R.L.; O'Sullivan, J.; Ashworth, J.; Bhaskar, S.; Williams, S.; Biswas, S.; Kehdi, E.; Ramsden, S.C.; Clayton-Smith, J.; Black, G.C.; *et al.* Personalized diagnosis and management of congenital cataract by next-generation sequencing. *Ophthalmology* **2014**, doi:10.1016/j.ophtha.2014.06.006.
15. Kocarnik, J.M.; Fullerton, S.M. Returning pleiotropic results from genetic testing to patients and research participants. *JAMA* **2014**, *311*, 795–796.
16. Black, G.C. University of Manchester, Manchester, UK. Personal Communication, 2014.
17. Clarke, A.J. Managing the ethical challenges of next-generation sequencing in genomic medicine. *Br. Med. Bull.* **2014**, *111*, 17–30.
18. Jones, M.A.; Rhodenizer, D.; da Silva, C.; Huff, I.J.; Keong, L.; Bean, L.J.; Coffee, B.; Collins, C.; Tanner, A.K.; He, M.; *et al.* Molecular diagnostic testing for congenital disorders of glycosylation (CDG): Detection rate for single gene testing and next generation sequencing panel testing. *Mol. Genet. Metab.* **2013**, *110*, 78–85.
19. Manyika, J.; Chui, M.; Bughin, J.; Dobbs, R.; Bisson, P.; Marrs, A. Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Global Institute, 2013. Available online: http://www.mckinsey.com/insights/business_technology/disruptive_technologies (accessed on 10 August 2014).
20. Jacob, H.J. Next-generation sequencing for clinical diagnostics. *N. Engl. J. Med.* **2013**, *369*, 1557–1558.
21. Saunders, C.J.; Miller, N.A.; Soden, S.E.; Dinwiddie, D.L.; Noll, A.; Alnadi, N.A.; Andraws, N.; Patterson, M.L.; Krivohlavek, L.A.; Fellis, J.; *et al.* Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.* **2012**, *4*, doi:10.1126/scitranslmed.3004041.
22. Dixon-Salazar, T.J.; Silhavy, J.L.; Udpa, N.; Schroth, J.; Bielas, S.; Schaffer, A.E.; Olvera, J.; Bafna, V.; Zaki, M.S.; Abdel-Salam, G.H.; *et al.* Exome sequencing can improve diagnosis and alter patient management. *Sci. Transl. Med.* **2012**, *4*, doi:10.1126/scitranslmed.3003544.
23. Worthey, E.A.; Mayer, A.N.; Syverson, G.D.; Helbling, D.; Bonacci, B.B.; Decker, B.; Serpe, J.M.; Dasu, T.; Tschannen, M.R.; Veith, R.L.; *et al.* Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **2011**, *13*, 255–262.
24. Bainbridge, M.N.; Wiszniewski, W.; Murdock, D.R.; Friedman, J.; Gonzaga-Jauregui, C.; Newsham, I.; Reid, J.G.; Fink, J.K.; Morgan, M.B.; Gingras, M.C.; *et al.* Whole-genome sequencing for optimized patient management. *Sci. Transl. Med.* **2011**, *3*, doi:10.1126/scitranslmed.3002243.
25. Lupski, J.R.; Reid, J.G.; Gonzaga-Jauregui, C.; Rio Deiros, D.; Chen, D.C.; Nazareth, L.; Bainbridge, M.; Dinh, H.; Jing, C.; Wheeler, D.A.; *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **2010**, *362*, 1181–1191.

26. Onoufriadis, A.; Shoemark, A.; Munye, M.M.; James, C.T.; Schmidts, M.; Patel, M.; Rosser, E.M.; Bacchelli, C.; Beales, P.L.; Scambler, P.J.; *et al.* Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm. *J. Med. Genet.* **2014**, *51*, 61–67.
27. Weedon, M.N.; Cebola, I.; Patch, A.M.; Flanagan, S.E.; de Franco, E.; Caswell, R.; Rodríguez-Seguí, S.A.; Shaw-Smith, C.; Cho, C.H.; Lango A.H.; *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **2014**, *46*, 61–64.
28. Gilissen, C.; Hehir-Kwa, J.Y.; Thung, D.T.; van de Vorst, M.; van Bon, B.W.; Willemsen, M.H.; Kwint, M.; Janssen, I.M.; Hoischen, A.; Schenck, A.; *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **2014**, *511*, 344–347.
29. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). Available online: http://evs.gs.washington.edu/EVS/ (accessed on 10 August 2014).
30. Gottlieb, B.; Beitel, L.K.; Trifiro, M. Changing genetic paradigms: Creating next-generation genetic databases as tools to understand the emerging complexities of genotype/phenotype relationships. *Hum. Genomics* **2014**, *8*, doi:10.1186/1479-7364-8-9.
31. Veltman, J.A.; Brunner, H.G. *De novo* mutations in human genetic disease. *Nat. Rev. Genet.* **2012**, *13*, 565–575.
32. Raychaudhuri, S.; Iartchouk, O.; Chin, K.; Tan, P.L.; Tai, A.K.; Ripke, S.; Gowrisankar, S.; Vemuri, S.; Montgomery, K.; Yu, Y.; *et al*. A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat. Genet.* **2011**, *43*, 1232–1236.
33. Torjesen, I. Genomes of 100,000 people will be sequenced to create an open access research resource. *Br. Med. J.* **2013**, *347*, doi:10.1136/bmj.f6690.
34. Köhler, S.; Doelken, S.C.; Mungall, C.J.; Bauer, S.; Firth, H.V.; Bailleul-Forestier, I.; Black, G.C.; Brown, D.L.; Brudno, M.; Campbell, J.; *et al.* The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **2014**, *42*, D966–D974.
35. Green, R.C.; Berg, J.S.; Grody, W.W.; Kalia, S.S.; Korf, B.R.; Martin, C.L.; McGuire, A.L.; Nussbaum, R.L.; O'Daniel, J.M.; Ormond, K.E.; *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **2013**, *15*, 565–574.
36. Van El, C.G.; Cornel, M.C.; Borry, P.; Hastings, R.J.; Fellmann, F.; Hodgson, S.V.; Howard, H.C.; Cambon-Thomsen, A.; Knoppers, B.M.; Meijers-Heijboer, H.; *et al.* Whole-genome sequencing in health care: Recommendations of the European Society of Human Genetics. *Eur. J. Hum. Genet.* **2013**, *21*, 580–584.
37. Green, R.C.; Lupski, J.R.; Biesecker, L.G. Reporting genomic sequencing results to ordering clinicians: Incidental, but not exceptional. *JAMA* **2013**, *310*, 365–366.
38. Levy, H.L.; Shih, V.E.; Madigan, P.M. Routine newborn screening for histidinemia. Clinical and biochemical results. *N. Engl. J. Med.* **1974**, *291*, 1214–1219.
39. Payne, K.; McAllister, M.; Davies, L.M. Valuing the economic benefits of complex interventions: When maximising health is not sufficient. *Health Econ.* **2013**, *22*, 258–271.

Somatic Mosaicism in the Human Genome

Donald Freed, Eric L. Stevens and Jonathan Pevsner

Abstract: Somatic mosaicism refers to the occurrence of two genetically distinct populations of cells within an individual, derived from a postzygotic mutation. In contrast to inherited mutations, somatic mosaic mutations may affect only a portion of the body and are not transmitted to progeny. These mutations affect varying genomic sizes ranging from single nucleotides to entire chromosomes and have been implicated in disease, most prominently cancer. The phenotypic consequences of somatic mosaicism are dependent upon many factors including the developmental time at which the mutation occurs, the areas of the body that are affected, and the pathophysiological effect(s) of the mutation. The advent of second-generation sequencing technologies has augmented existing array-based and cytogenetic approaches for the identification of somatic mutations. We outline the strengths and weaknesses of these techniques and highlight recent insights into the role of somatic mosaicism in causing cancer, neurodegenerative, monogenic, and complex disease.

Reprinted from *Genes*. Cite as: Freed, D.; Stevens, E.L.; Pevsner, J. Somatic Mosaicism in the Human Genome. *Genes* **2014**, *5*, 1064-1094.

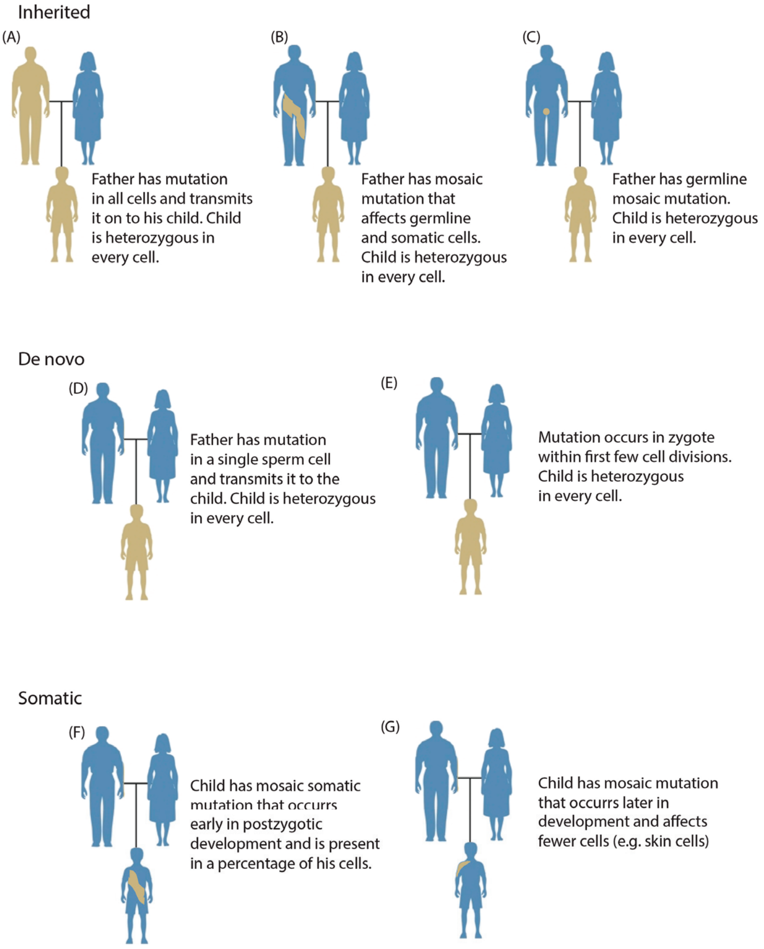
1. Introduction to Somatic Mosaicism

1.1. Early Studies of Mosaicism

Somatic mosaic mutations are defined as mutations that occur in some cells of the soma of a single individual (Figure 1) [1,2]. The mixture of mutation-positive cells with non-mutated cells results in an individual who is a mosaic, or contains different DNA within different cells of his or her body. Mosaic mutations may be present in the germline or soma; however, typically only mutations in the soma have phenotypic consequences or are detectable by current genotyping methods. Mosaic mutations in germ cells are usually only discovered when they lead to inherited conditions in multiple progeny. *De novo* mutations are operationally defined as mutations found in all cells of an individual but not detected in that individual's parents (Figure 1d,e) [3]. *De novo* mutations only present in the offspring may occur very early in development; however, this is rare and increasingly sensitive genetic assays are discovering low-level parental mosaicism in supposedly *de novo* cases (Figure 1b) [4,5].

The role of somatic genetic changes in human health has been considered at least since 1914 when Theodor Boveri recognized that cancers frequently have abnormal karyotypes [6]. Alfred Knudson built upon the work of Boveri and others and in 1971 described a two-hit model of cancer resulting from both an inherited germline mutation and a later somatic mutation [7]. The model of metastatic cancer occurring as a result of multiple mutations in a single cell lineage has remained largely unchanged for over 40 years [8,9].

Figure 1. Overview of categories of variation including inherited (panels A–C), *de novo* (panels D,E), and somatic variation (panels F,G). Inherited mutations are always transmitted through the germline (**A**); although a parent may also have a mosaic mutation (this combination of somatic and germline mosaicism is occasionally termed gonadal mosaicism) (**B**); In such cases, a child may inherit the variant as a heterozygous mutation with a more severe clinical phenotype. A parent may also have germline mosaicism that may be inherited by progeny (**C**); *De novo* mutations are operationally defined as genotypes observed in a child but not in either parent. They may originate in a parental germ cell (as may be inferred in a pedigree having multiple affected offspring) (**D**) or postzygotically (**E**); Somatic mutation may occur relatively early in development (**F**) or at any later time throughout the lifespan (**G**), generally affecting fewer cells.


Inherited
(A)
Father has mutation in all cells and transmits it on to his child. Child is heterozygous in every cell.
(B)
Father has mosaic mutation that affects germline and somatic cells. Child is heterozygous in every cell.
(C)
Father has germline mosaic mutation. Child is heterozygous in every cell.
De novo
(D)
Father has mutation in a single sperm cell and transmits it to the child. Child is heterozygous in every cell.
(E)
Mutation occurs in zygote within first few cell divisions. Child is heterozygous in every cell.
Somatic
(F)
Child has mosaic somatic mutation that occurs early in postzygotic development and is present in a percentage of his cells.
(G)
Child has mosaic mutation that occurs later in development and affects fewer cells (e.g. skin cells)

The scientific community was slower to realize the importance of postzygotic mutational events outside of cancer. In the early 1950s, Barbara McClintock demonstrated the phenotypic importance of somatic transposition in *Zea mays*, and in 1959 Sir Macfarlane Burnet proposed a role for somatic mutation in disease [10,11]. Nonetheless, few studies indicated a role for somatic mosaicism in human health. This changed in the 1970s with the discovery that somatic gene rearrangement creates functional diversity of immunoglobulin and T-cell receptor genes [12–14]. Today, it is known that somatic mutations are ubiquitous [15] and have important roles in cancer [9], aging [16,17], neurodegeneration [18], monogenic disease [19–21], reversion of inherited disease [22–25], and numerous neurocutaneous disorders [26].

1.2. Categories of Somatic Variation

Somatic variation has been observed at all genomic scales from point mutations to aneuploidies. At the level of whole chromosomes and large chromosomal segments, complex genomic rearrangements occur somatically (as well as in the germline). The loss or gain of entire chromosomes is thought to be caused by errors in chromosomal segregation during anaphase, while non-allelic homologous recombination may cause the loss, gain, or rearrangement of large genomic regions [27,28]. The phenotypic consequences of these events vary considerably based on the size of the event and the genomic region involved.

In many instances, both copies of a chromosome pair (or of a chromosomal segment) are inherited from one parent, a phenomenon termed uniparental disomy (UPD) [29,30]. UPD may involve two copies from a parent that are identical (uniparental isodisomy) or different (uniparental heterodisomy). Either form may disrupt epigenetically imprinted regions (defined as undergoing differential expression depending on the parent of origin), while uniparental isodisomy may also expose two copies of a recessive mutation. One mechanism for the occurrence of UPD involves trisomic rescue in which an extra (third) copy of a chromosome is rejected, producing a diploid cell line in which one parent's monoploid copy is lost [31]. Frequently, the trisomic rescue is restricted to a fraction of cells in an individual resulting in mosaic trisomy/UPD [32]. UPD may also result from somatic recombination occurring from a reciprocal exchange during mitosis, leading to loss of heterozygosity.

RNA-templated DNA polymerases are another cause of genomic instability. While numerous types of repetitive elements are present in human genomes, only non-long terminal repeat retrotransposons are currently competent for retrotransposition [33]. Successful retrotransposition of these elements is dependent upon functional protein products from long interspersed elements (LINEs). In most somatic tissues, LINEs are epigenetically suppressed; however, these elements escape epigenetic repression during early embryonic development, and their integration into other functional genomic elements occasionally results in disease such as choroideremia (Online Mendelian Inheritance in Man [OMIM] #303100) [34]. Retrotransposition may also occur in somatic tissues with unusual epigenetic states [35].

Low complexity regions, including trinucleotide repeats, are scattered throughout the mammalian genome. Trinucleotide repeats can be hypervariable and expansions of some trinucleotide repeats are the causes of nearly 30 disorders [36,37]. The molecular mechanisms underlying expansion or contraction of these regions are complex and cause these regions to have variable length throughout the body of those afflicted with disease [38–44].

Small genetic aberrations may be caused by a number of mechanisms. Polymerase errors may result in nucleotide misincorporation or small insertions or deletions in the germline or soma. Over time, DNA will accumulate numerous lesions and DNA polymerization across these lesions is especially error-prone. DNA lesions may be detected and repaired prior to DNA polymerization, but lesion repair may also create single nucleotide variants, or small insertions or deletions [45,46].

In linear mammalian genomes, DNA replication starts at multiple origins with DNA polymerases ε and δ [47,48]. Polymerase ε moves with high processivity 5'–3' along the genome on the leading strand, moving in the same direction as the replication fork. On the lagging strand replication by polymerase δ also proceeds 5'–3' but in the opposite direction as the replication fork, causing replication of that strand to be iterative. This process works well for the majority of the genome, but replication of the lagging strand leads to loss of genetic information at the ends of the chromosome during every replication [49]. This end replication problem is solved in the germline because the ends of chromosomes, telomeres, are protected by repetitive DNA which is synthesized by a dedicated RNA-templated DNA polymerase called telomerase [49]. However, telomerase is not usually expressed in somatic tissues, likely as a method of protection against malignant transformation, and decreased telomere length is a form of somatic variation.

1.3. Mosaicism during Development

A defining characteristic of mosaic mutations is that they occur postzygotically and are inherited by all subsequent cells in their lineage (Figure 1). Somatic errors in chromosomal segregation in early development induce an extraordinarily high rate of aneuploidy. Fifteen to 20% of clinically recognized pregnancies result in spontaneous abortion, and half of these are attributed to aneuploidy [29]. A review of 36 published studies showed that of 815 human preimplantation embryos, only 177 (22%) were diploid while 73% were mosaic [50]. In most cases, these were diploid-aneuploid mosaic embryos, having one or more diploid cells as well as other cells that were haploid or polyploid for a particular chromosome. Mitotic errors could account for the high rate of chromosomal mosaicism.

Due to the exponential rate of growth during development, somatic mutations must occur early in development to have phenotypic effects over large portions of the body. Severe somatic mutations, which would be embryonic lethal if inherited, have a short window during development in which they must occur to be observed in adults [19]. If these severe mutations occur early in development, they will be embryonically or prenatally lethal; occurring later in development they may have little or no obvious phenotypic effect.

Mutations that alter cellular growth do not necessarily have to occur within such a short developmental window. Inactivating mutations in genes encoding tumor suppressors or activating mutations in oncogenes may have functional consequences regardless of when they occur, as evident from their known roles in cancer. On the other hand, somatic growth-retarding mutations, such as inactivating mutations in oncogenes or certain cyclins, are unlikely to have phenotypic effect in adults regardless of when they occur in development as the total number of cells containing the mutation is likely to be small.

Somatic mutations are thought to occur in all cells during replication. On average, 50 mutations occur in microsatellite regions during every mitotic division of a given cell [15]. Mutations in microsatellites and other regions of the genome, assessed by either single-cell or deep sequencing, can then be used to infer cell lineage trees [51]. To date, the most successful lineage tracing experiments have made use of increasingly sophisticated microscopy techniques [52]. However, microscopy-based approaches have practical and technological barriers such as the requirement that non-transgenic cells must be monitored over time. Recent advances in whole genome amplification (WGA) and second-generation sequencing offer genetic-based approaches that do not have the same limitations. Already, these techniques have been used to provide a detailed view of the genetics of cancer metastasis [53,54].

1.4. Mosaicism across the Body

By definition, somatic mosaic mutations affect only a subset of cells within an individual (Figure 1). This is most easily visible in monogenic mutations affecting pigmentation patterns. While such patterns may be mistaken for stochastic X chromosome inactivation or autoimmune response, somatic mutation is generally localized over a small portion of the body and in many cases occurs along lines of Blaschko [55]. To date, almost all non-cancerous somatic mutations characterized at the molecular level result in visible abnormalities, usually involving hypertrophy (cellular overgrowth) or abnormal pigmentation [26,55]. Some of our inability to identify mutations that do not result in visible phenotypes is practical; during dissection it is difficult to distinguish affected from unaffected tissue. However, due to the current emphasis on visible phenotypes, few data are available on the extent to which non-visible somatic mutations influence important biological processes.

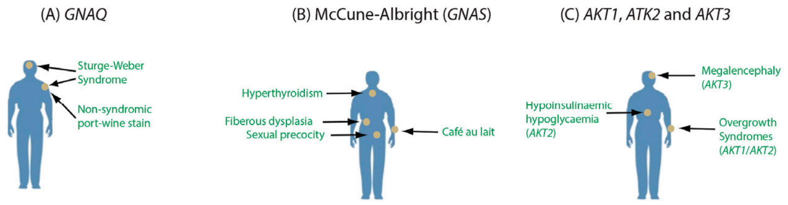
An important consideration is that somatic mutations occur in varying cell types and tissues as well as different developmental stages. This raises the possibility that a specific mutation may vary in its clinical importance depending on where the mutation occurs across the body. Mutations in *GNAQ* provide an example. We identified p.Arg183Gln mutations in *GNAQ*, encoding the G protein alpha subunit Gαq, as the cause of both Sturge-Weber syndrome (OMIM #185300) and port-wine stain birthmarks (OMIM #163000) [56]. Port-wine stains are non-syndromic vascular abnormalities, while the Sturge-Weber syndrome is a severe neurocutaneous disorder, although both conditions likely affect some of the same cell types (e.g., endothelial cells). The milder phenotype of the birthmarks could result from a later developmental origin of the mutation during fetal development. The identical p.Arg183Gln mutation in *GNAQ*, when occurring in melanocytes later in life, is a frequent driver mutation in uveal melanoma (OMIM #155720), highlighting the importance of both the location and timing of the mutation. p.Arg183Gln mutations in different cell types and developmental stages could have different phenotypic consequences, if any [57].

Other mosaic mutations also differ in their clinical importance based on cell or tissue-specific involvement. McCune-Albright syndrome (OMIM #174800) is characterized by increased function of endocrine glands, sexual precocity, café-au-lait macules, and fibrous dysplasia. These symptoms can vary considerably based, in part, on the bodily extent of the mutation [58]. Like Sturge-Weber syndrome, this disorder is caused by somatic activating mutations in a gene encoding a G protein alpha subunit (*GNAS* encoding Gαs). Expression of this gene highlights another dimension of mosaicism. *GNAS* is expressed biallelically through most of the body, but the maternal allele is imprinted in particular tissues such as the pituitary. The disorders progressive osseous heteroplasia (OMIM #166350) and pseudopseudohypoparathyroidism (OMIM #612463) result from loss of function mutations in the paternal allele of *GNAS* [59].

Somatic mutations in three *AKT* genes also have cell-specific effects [60–62]. Somatic *AKT1* mutations are associated with somatic breast cancer, colorectal cancer, and ovarian cancer as well as the Proteus syndrome. The *AKT2* gene is expressed selectively in insulin-responsive tissues and mutations are associated with diabetes. Somatic mutations in *AKT3* cause Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 2 (OMIM *615937). Given the localized nature of somatic mutations in *AKT* discovered to date, it is likely that mutations in these genes occurring outside of vulnerable cell types have few effects. These examples highlight the complex interaction of localized somatic mutation with tissue or cell-specific gene expression and signaling pathways (Figure 2).

Numerous studies have aimed to assess the prevalence of mosaic alterations in tissues of apparently normal individuals. Reanalysis of data from multiple large genome-wide association studies have determined that the number of detectable mosaic events rises sharply after age 50. Furthermore, individuals with increased numbers of mosaic events have higher risk for developing cancer [63,64]. While this measured increase of mosaicism may be due to increased rates mutation rates in elderly individuals, it is much more likely that these events are the result of clonal expansion and positive selection within the stem cell niche or decline in the total number of hematopoietic stem cell progenitors later in life. Notably, increased rates of mosaicism in apparently normal tissues have been linked to poorer prognosis in individuals with ovarian cancer [65].

Figure 2. Tissue-specific effects of mutations in *GNAQ* (**A**); *GNAS* (**B**); and *AKT1*, *AKT2*, and *AKT3* (**C**). Constitutively activating mutations in *GNAQ* may lead to either Sturge-Weber syndrome, nonsyndromic port-wine stains, or uveal melanoma (**A**). Somatic activating mutations in *GNAS* lead to McCune-Albright syndrome, which may involve variable hyperthyroidism, *café au lait* macules and sexual precocity (**B**). Activating mutations in all three of the *AKT* genes cause cellular overgrowth phenotypes with mutations in *AKT2* also implicated in abnormal insulin signaling (**C**).



Studies of twins have demonstrated that post-zygotic mutations may be phenotypically important. Notable examples are monozygotic twins who are discordant for phenotypic sex due to mosaic loss of chromosome Y [66,67]. Numerous examples of monozygotic twins exist where either the presence [68,69] or severity [70] of disease is discordant between twin pairs due to variable proportions of mosaic cells.

Studies of multiple tissues of apparently normal individuals have also found evidence for mosaic events. Analysis of CNVs using hybridization of DNA from multiple tissues of three apparently normal individuals to bacterial artifical chromosome arrays found evidence for six somatic CNVs [71]. Higher resolution examination of a total 33 tissues from six individuals using array comparative genomic hybridization found evidence for 73 high-confidence mosaic CNVs, although a majority of high-confidence events (54/73) were found in one of two particular tissues [72]. It has been noted that induced pluripotent stem cells (iPSC) frequently contain CNVs which may cause genomic instability inheritent to the process of immortalization. Abyzov *et al.* performed a detailed study of this phenomenon and concluded that almost half of CNVs present in iPSC lines can be found in the parental fibroblasts. Furthermore, they conclude that approximately 30% of all fibroblasts in their sample contain some mosaic CNVs [73].

While experimentation with bulk tissues has shown that somatic mosaicism occurs frequently in normal populations, the combination of DNA from many cells limits the ability of an assay to detect mosaic events unique to single or few cells. As a result, sequencing of single-cells has been recently used to assay mosaicism in normal tissues. These methods have been used to sensitively reexamine conclusions regarding the extent of mosaicism in the brain. Previous reports had indicated that up to 33% of neuroblasts were aneuploid while up 80 retrotransposon insertions occur per neuron [74–77]. Single-cell experiments of the same phenomena have shown that large copy-number variants occur in over 14% of neurons but whole chromosome aneuplodies and retrotransposition events are relatively rare [78–80].

Single-cell studies have also been used to investigate the extent to which mosaicism occurs in early development. It has been known since 1983 that chorionic villus sampling may indicate the presence of a trisomy, while the fetus is diploid without the presence of mosaicism, a condition termed confined placental mosaicism [81–83]. Single-cell studies of young embryos cultured *in vitro* also demonstrate that chromosomal aneuploidies are common and were found in 83% of tested embryos [84]. While it is likely that many aneuploid embryos are unlikley to result in viable pregnancies, recent advances in prenatal testing allow for the sensitive and specific detection of numerous trisomies by sequencing of circulating fetal DNA from maternal plasma [85].

2. Detection of Somatic Mosaicism

2.1. Technical Considerations

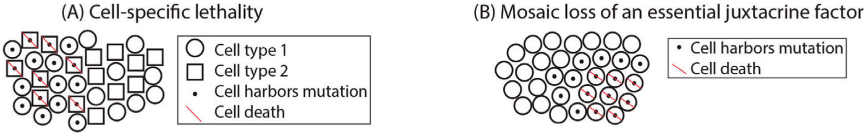
Almost every type of genetic variation has been implicated as a source of somatic variation including expansion of trinucleotide repeats, point mutation, copy-number loss/gain, uniparental disomy, mitotic recombination, aneuploidy, translocation, and retrotransposition [39–41,43,44,69,77,79,86–94]. The techniques summarized below vary widely in their ability to detect specific types of somatic variation and more specialized techniques exist for the sensitive detection of some types of variation.

A primary consideration during the analysis of mosaic samples is the purity of the dissection from tissue samples. The presence of normal cells in affected tissue significantly decreases the ability of downstream analyses to detect mosaic alterations. This problem can be compounded by the prevalence of cellular migration during development in some tissues. Thus, in a tissue affected by a somatic mutation, two neighboring cells may both be affected if they share a common lineage from the mutated cell. Alternatively, cellular migration could cause neighboring cells to originate from distinct precursors with only one cell affected. Cellular migration can place an important biological constraint on the visible frequency of driver somatic mutations in affected tissues (e.g., in the brain) [3,95,96].

While contamination of normal cells is known to decrease the observed frequency of mosaic mutations, other mechanisms may decrease the detectable fraction of mosaic cells within a sample. Two possibilities are cell-type specific lethality and mosaic absence of essential juxtacrine or paracrine signaling factors. Cellular signaling pathways are known to have cell-type specific effects raising the possibility that a mosaic mutation may be lethal in only one type of cell within a tissue (Figure 3a). Furthermore, some paracrine or juxtacrine signaling factors are essential for cell viability [97–99]. Mosaic loss of these factors could result in affected tissue that is dependent upon surrounding normal tissue for survival, reducing the total number of mutant cells (Figure 3b).

In Sturge-Weber affected tissues, we detected *GNAQ* mutant allele frequencies between 1% and 18% [56]. Other studies using similar techniques have detected mutant allele frequencies of 1%–47% [61], 3%–30% [100], and 3%–35% [60] for causative mutations in individuals with Proteus syndrome (OMIM #176920), CLOVE (Congenital Lipomatous Overgrowth, Vascular anomalies, and Epidermal nevi) syndrome (OMIM #612918), and hemimegalencephaly (e.g., OMIM #615937), respectively. Such relatively low allele frequencies are likely explained by the presence of low proportions of affected cells in a given tissue due to cellular migration or impure dissection. However, the occurrence of mosaic cell death due to either cell-specific lethality or loss of essential signaling factors should be considered.

Figure 3. Cell death may reduce the total number of cells harboring somatic mutation. Mosaic mutations may cause cell-type-specific lethality (**A**); Mosaic loss of an essential juxtacrine signaling factor may cause localized death of cells that are not adjacent to unaffected tissues (**B**).



In second-generation sequencing experiments, sequencing and mapping errors are a major concern, as some portions of the genome are known to be prone to false-positive variant calls [101]. Recent improvements in sequencing chemistry have lowered the frequency of sequencing errors. However, biased errors in sequencing are still problematic for the detection of somatic variation, especially when the mutant allele frequency may be close to the technology's inherent error rate. Generally, ultra-high depth sequencing (>500 reads) of normal and affected tissues will permit detection of these errors. However, exploratory studies generally do not reach this level of depth. It is likely that without validation, these errors are a source of false positives in somatic variation databases. Comparing suspected somatic mutations across multiple tissue types from multiple individuals to estimate local error profiles may be a possible solution to this problem [102].

2.2. Cytogenetics

Microscopy-based methods allow for the detection of large mosaic events in single cells. Early cytogenetic methods for identifying extra or fewer chromosomes involved counting condensed metaphase chromosomes under a microscope [103]. Later methods using Giemsa staining and other dyes produced unique chromosomal bands allowing for the identification of intra- and interchromosomal translocations, duplications, deletions, and large structural rearrangements. However, banding techniques can only resolve aberrations larger than 3–10 Mb [104]. Other methods, such as fluorescent *in situ* hybridization (FISH), label a specific region of the genome by hybridization of a fluorescent probe allowing for the detection of deletions and some duplications [105]. Variations in this methodology exist using multiple probes of different color to detect several unique fragments at a time (e.g., multicolor FISH). These methods are able to achieve resolutions below 100 kilobases or, in some cases, as few as several kilobases [106]. Potential probe binding to off-target regions is a major consideration in most FISH experiments and adequate controls are required to confirm locus specificity [106]. Variants on classical FISH methods continue to be developed which promise to increase the ability of fluorescent probes to detect small chromosomal abnormalities across increasingly large portions of the genome [106,107]. In combination with high-throughput techniques, these approaches may be used to screen large numbers of cells from a single individual allowing for the detection of low levels of mosaicism.

2.3. Genome-Wide Arrays

Comparative genomic hybridization (CGH) is a technique in which fluorophore-labeled DNA from a control and test individual are hybridized to a metaphase reference chromosome [108]. The ratio of fluorescence emission is then measured to allow for the detection of duplication or deletions. A ratio of 1:1 indicates that both samples of DNA carry the same copy number while deviations from this ratio indicate a copy number variant [109].

Two principal array-based techniques that have emerged as alternatives to CGH are array CGH (aCGH) and single nucleotide polymorphism microarrays (SNP microarrays) [110–112]. Similar to CGH, both aCGH and SNP microarrays have the ability to detect changes in copy number over large regions of the genome. SNP microarrays further have the ability to genotype individuals at the probed sites, which may be useful in the detection of low-level somatic events [113]. Array-based approaches offer increased sensitivity over the entire genome for small CNVs relative to genome-wide microscopy-based approaches. aCGH and SNP microarray analysis can resolve regions less than 100 kb in size. However, the sensitivity of array-based approaches for somatic CNVs is dependent on having at least 5%–10% of the cells assayed containing the genetic variant. For larger CNVs affecting a smaller fraction of cells, microscopy-based approaches are more sensitive.

In both aCGH and SNP microarrays, deviations in relative probe intensities indicate deletion or insertion events. Normalized probe intensities are commonly reported as log-R ratios, with higher intensities indicating insertions while lower intensities indicate deletions. For SNP microarrays, the relative intensities of the two probes (one specific to each allele) at a locus is informative, and normalization of these intensities is measured as a B-allele frequency. For normal diploid tissues, B-allele frequencies approximate 0.0, 0.5, and 1.0 for AA, AB, and BB genotypes, respectively, while log-R ratios approximate 0 indicating no copy number change.

The hybridization of genomic DNA to microarrays is inherently noisy and can be subject to large batch effects [114]. Furthermore, individual probes or even whole arrays may have errors caused by faulty manufacture. Together these artifacts make the detection of statistically significant mosaic CNVs difficult, but many software packages detect these events. Numerous tools use hidden Markov Models (HMMs) to integrate B-allele frequency and log-R ratio information for the detection of mosaic events, including PennCNV-2, GPHMM and MixHMM [115–117]. gBPCR uses an approach similar to the Bayesian Piecewise Constant Regression for the detection of mosaic abnormalities but has a long runtime per sample [118]. We developed triPOD which uses multiple algorithms for the detection of mosaic events and is unique in that it utilizes parental genotypes allowing for more sensitive detection of haplotype-specific mosaic abnormalities [113].

2.4. Second-Generation Sequencing

Second-generation sequencing techniques have revolutionized human genetics in the last decade. Sequencing is performed either on single cells, a discrete number of cells, or bulk tissue. In the typical sequencing experiment, DNA is extracted from the input material and is fragmented, size-selected, and sequenced to produce strings of inferred nucleotides and their respective quality scores [119]. This information is used to align the sequencing reads to a reference genome. Differences between the aligned reads and the reference can be used to infer genetic variants including single-nucleotide variants or polymorphisms (SNVs or SNPs), insertions, deletions, translocations, and retrotransposition events. Furthermore, the total number of reads aligned to certain regions of the genome can be used to infer copy-number changes [120,121]. Numerous variations on this basic approach exist and here we will discuss the methods most applicable for the detection of mosaic events.

Somatic genetic variants have been discovered via whole-exome or whole-genome sequencing of bulk tissue from paired affected and unaffected portions of the body [56,60,61,100]. Whole-exome sequencing relies upon an oligonucleotide bead or array-based enrichment of DNA fragments corresponding to exonic regions to reduce the representation of sequence from noncoding regions of the genome [122–124]. At similar depth, exome and whole-genome sequencing are considered to have similar sensitivity for most pathogenic SNVs and small insertions or deletions. Whole-exome sequencing is considered less sensitive for the identification of medium to large insertions or deletions or the detection of copy-number changes by analysis of read depth due to introduced biases. However, exome sequencing experiments are typically performed at higher depth due to the lower cost of the method.

Numerous software packages allow the identification of somatic variants from these data. Somatic variant callers typically evaluate second-generation sequence data from paired tumor/normal (or other affected/unaffected) samples. Examples include VarScan2 [121], SomaticSniper [125], JointSNVMix [126], Strelka [127], and MuTect [128]. After removal of low-quality reads, sequences are aligned to a reference genome to generate aligned binary sequence alignment/map (BAM) files [129]. At least three approaches have been employed for the detection of somatic SNVs and small insertions or deletions. (1) Allele frequencies can be compared. For example, VarScan2 performs pairwise comparisons of base calls and normalized sequence depth at each position, accounting for factors such base quality scores, coverage and variant allele frequencies; (2) Bayesian comparison of joint diploid genotype likelihood can be estimated for both samples. The SomaticSniper algorithm calculates the statistical significance of all somatic variants at positions above a minimum threshold of coverage using this method; (3) Other Bayesian approaches have been applied. For example, Strelka models the normal sample as germline variation plus noise, while the affected sample includes noise along with germline and somatic variation. Other types of somatic variation may be detected from bulk sequencing. Tools such as VarScan2, ADTeX, Control-FREEC, SomatiCA, and LUMPY may be used for the detection of somatic CNVs or structural variants [121,130–133].

Besides variant identification, quantification of the fraction of cells affected by particular somatic changes provides a better understanding of the extent of the mosaic mutation and the period during development at which it occurred. Several tools have been developed to deconvolute somatic mutations into distinct populations as reviewed by Yadav and De and Ding *et al.* [134,135].

An alternative approach to sequencing bulk tissue is sequencing single cells or small numbers of cells. As single or hundreds of cells contain very little DNA, most experiments utilize multiple displacement amplification (MDA) or PCR based methods to amplify genomic DNA. Amplification can greatly increase the total amount of available DNA for sequencing at the expense of introduced biases such as allele dropout and chimeric amplification of genomic fragments [79,136–138]. Despite these introduced biases, amplification and subsequent second-generation sequencing or array-based analysis of single cells has been used to reliably find somatic copy number variation and retrotransposition events within the human brain as well as to map cell lineage within a bulk tumor dissection [53,54,79,139]. Numerous groups have also used single-cell techniques to discover SNVs or indels in single cells, however, allelic dropout and chimeric amplification are more problematic for these analyses as biases can be reduced for analysis of CNVs by increasing bin sizes but are more difficult to account for in analysis of SNVs [140–142].

3. Somatic Mosaicism in Disease

3.1. Cancer and Aging

The relationship between somatic mutation and cancer has been extensively reviewed elsewhere [9,17,87,143–145] and comprehensive lists of known oncogenes or tumor suppressors or genes significantly and recurrently mutated in cancer have been previously described [9,87]. Cancer has been described as having six hallmarks: proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, induction of angiogenesis, and inactivating invasion and metastasis [146]. Driver gene mutations are defined as conferring a selective growth advantage in tumor cells [9]. This may be achieved by elevating the activity of growth factors and/or their receptors, but more commonly driver mutations constitutively activate intracellular signal transduction cascades. Three of these are depicted in Figure 4 (in simplified form): Ras/Raf/MEK/ERK, Ras/PI3K/PTEN/Akt/mTOR [147], and *GNAQ*. These pathways contain both oncogenes (*RAS*, *RAF*, *MEK*, *PIK3CA*, *AKT*, *GNAQ*) and tumor suppressor genes (*NF1*, *PTEN*, *TSC1*, *TSC2*). For example the *RAS* family of oncogenes were the first oncogenes to be identified in cancer. Comprised of *HRAS*, *KRAS* and *NRAS*, activating mutations in these genes occur in approximately 20% of all cancers [148]. Germline variants are also well known to contribute to cancer morbidity [149–151]. Frequently, these variants affect proteins involved in DNA repair, highlighting the role of somatic mutations in tumorigenesis [152–155].

In common solid tumors, ~95% of protein altering mutations consist of single base substitutions, >90% of which are missense mutations, <8% are nonsense mutations, and <2% affect splice sites or untranslated regions [9]. Relatively large numbers of somatic mutations occur in tumors that are associated with mutagens such as ultraviolet light and cigarette smoke. For example, in non-small cell lung carcinomas the average mutation frequency is greater than ten-fold higher in smokers compared to those who never smoke [156].

Large-scale projects and databases have been developed to provide comprehensive catalogues of somatic mutations found in cancer [157,158]. COSMIC (Catalogue Of Somatic Mutations In Cancer) includes information on more than 1.6 million mutations from nearly 1 million cancer samples and includes various types of mutations (fusions, genomic rearrangements, whole genomes, and copy number variants) [157].

The combination of well-characterized somatic mutation databases and low-cost sequencing technologies may lead to improved patient outcomes in the near future. Biopsied tumors may be screened rapidly for putative driver mutations based on cancer type, informing treatment. Furthermore, once a cancer is in remission, tumor-specific DNA may be assayed at low cost with ultra-sensitive second-generation sequencing-based techniques [159]. These advances will likely improve prognosis for millions of cancer patients within the next decade.

The primary risk factor for cancer is age, and cancers offer insight into age or mutagen-associated mutational processes [160]. Somatic mutations have long been suspected to be an important part of the molecular mechanism of aging, and accumulation of DNA lesions and mutations occurs in both the germline and soma over time [63,64,161,162]. By chance, these mutations may result in malignant transformation, apoptosis, or otherwise hampered cellular function. As visible in cancers, the characteristics of acquired mutations differ by tissue type and are dependent upon environmental exposure [9]. Furthermore, frequently dividing stem cells and frequently transcribed genomic regions have different patterns of mutation that are cell-type specific.

In both mouse and human, increased rates of somatic mutation and numbers of DNA lesions due to either error-prone DNA polymerases or faulty DNA repair mechanisms cause cancer predisposition, early aging, and neurodegenerative phenotypes [17]. Increased rates of somatic mutation in the nuclear genome cause cancer predisposition, likely due to increased rates of mutation in somatic stem cell populations. This has been demonstrated in transgenic mice whose processive DNA polymerases lack proofreading. Notably, mice with mutated polymerases δ and ε develop distinct cancers but do not demonstrate premature aging phenotypes [163–165]. While these mice may not live long enough to demonstrate early aging phenotypes, their predisposition towards the development of cancer demonstrates a strong link between cancer and somatic mutation.

Mutations in genes affecting other pathways demonstrate a strong relationship between somatic mutations and aging. Mice with error-prone mitochondrial polymerases demonstrate a premature aging phenotype without cancer predisposition, although subsequent data by some of the same authors demonstrate that mitochondrial point mutations are unlikely the primary cause of aging in normal mice [166,167]. Individuals with defects in DNA repair also demonstrate symptoms of progeria. Cockayne syndrome (OMIM #216400) is caused by defects in transcription-coupled exonucleotide repair leading to an early aging phenotype combined with intellectual disability and neurodegeneration without noted predisposition to development of cancer [168]. Mutations in the genes encoding RecQ helicases cause Werner syndrome (OMIM #277700) and Rothmund-Thomson syndrome (OMIM #268400) [169]. The most prominent phenotype of individuals affected by these diseases is premature aging, although these individuals are also predisposed to developing cancer [169]. Bloom Syndrome (OMIM #210900) is notable in that it is also caused by mutations in a RecQ helicase-like protein and also increases cancer incidence, but does not appear to result in progeria. Mutations in numerous other genes are known to cause cancer predisposition. One such example is *BUB1B*. Loss of BUB1B protein function leads to premature chromatid separation and mosaic variegated aneuploidy syndrome 1 (OMIM #257300) typically resulting in cancer predisposition and intellectual disability [170].

Cancer is associated with many genomic changes. Large chromosomal changes occur in a variety of noncancerous conditions. An example is Pallister-Killian syndrome (OMIM #601803) is a dysmorphic condition caused by mosaicism for tetrasomy 12p. Affected individuals display tissue mosaicism, typically with apparently normal karyotypes from lymphocytes but 47 chromosomes in skin fibroblasts and chorionic villus and amniotic fluid cells. The extra chromosome is an isochromosome for a portion of chromosome 12p. In several cases hexasomy of chromosome 12p has been observed.

3.2. Neurodegenerative Disease

Somatic mutation is suspected to have a role in neurodegenerative disease [17,18]. As in cancer, mutations in genes directly involved in DNA repair are implicated in neurodegenerative diseases such as ataxia-telangiectasia (OMIM #208900) and ataxia-ocular apraxia 1 (OMIM #208920) [16,169,171–174]. These neurodegenerative phenotypes are likely caused by an increase of somatic mutation in the nervous system leading to cellular dysfunction, indicating a possible role for somatic changes and DNA lesions in age-related related neurodegenerative disorders.

There is evidence that mosaic mutations or accumulated damage to other macromolecules play a role in Alzheimer’s disease (OMIM #104300) and Creutzfeldt-Jakob disease (CJD) (OMIM #123400). Alzheimer’s disease is characterized by the accumulation of β-amyloid (Aβ) plaques while CJD is caused by misfolded protein PRNP [175,176]. Significant incidence of both diseases is attributed to familial risk and causal mosaic mutations have been found in sporadic cases [177,178]. Aβ plaques have long been implicated in the formation of prions and introduction of Aβ plaques into the brains of mice overexpressing Aβ leads to disease progression [179–181]. Consistent with the link to prions, the pathology of inoculated mice displays phenotypes dependent upon the infecting host [180]. This has been corroborated by more recent experiments, which demonstrate that Aβ aggregates from distinct sources have unique biophysical characteristics depending on the seeding protein [182–184]. While it is possible that sporadic misfolded or damaged proteins act as seeds in Alzheimer's, this is unlikely given the steep increase in disease incidence later in life and the constant turnover of cellular proteins [185]. This steep rise in incidence mirrors the rise in incidence of CJD in individuals who have predisposing mutations [186]. It is possible that in both diseases misfolded proteins arising as a result of age-related somatic mutation or damage to other macromolecules in single cells act as seeds for the initial protein aggregates.

3.3. Monogenic Disease

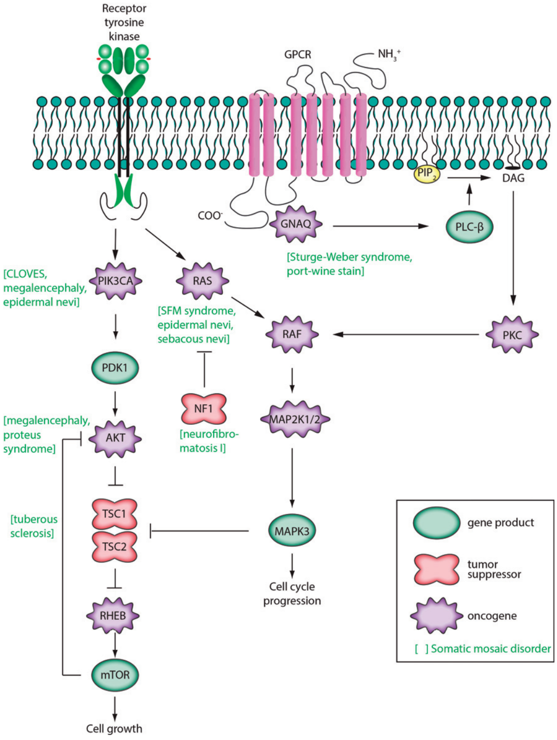
A list of diseases suspected to be caused by obligatory somatic mutations has been previously described [21] and subsequently updated [19,20]. We note that somatic mutation likely contributes significantly to nearly all Mendelian diseases.

We have described a series of oncogenes and tumor suppressor genes that undergo somatic mutation in cancer. These same genes can also acquire somatic mutations that result in neurocutaneous disorders or overgrowth syndromes, depending the particular cell type and developmental stage at which the mutation occurs. Mutations in *GNAQ* cause Sturge-Weber syndrome and port-wine stain birthmarks as well as uveal melanoma, as discussed above. Similarly, somatic mutations in *GNAS* can cause McCune-Albright syndrome or benign tumors such as adenomas. We next highlight several specific examples of such disorders affecting genes encoding intracellular signaling pathways (Figure 4).

Phosphatidylinositol 3-kinases (PIK3s) are lipid kinases that phosphorylate phosphatidylinositol and other phosphoinositides, catalyzing intracellular signaling pathways involving a PI3K/AKT/mTOR network (Figure 4). Somatic, mosaic, gain-of-function mutations in *PIK3CA* (OMIM *171834) are associated with several syndromes involving overgrowth of the brain or lipomatous body overgrowth [191]. These include CLOVE syndrome, megalencephaly-capillary malformation syndrome, fibroadipose hyperplasia, and hemimegalencephaly. These conditions are often characterized by early segmental overgrowth, abnormal vasculogenesis, digital anomalies, cortical brain malformations, and connective tissue dysplasia. Somatic gain-of-function mutations in *PIK3CA* are also found in a broad range of cancers (ovarian, breast, lung, stomach, colorectal, and brain). While over 100 activating mutations in *PIK3CA* are known, mutations in two domains of the protein account for 80% of cancer-associated somatic mutations, and these same sites can be mutated in overgrowth disorders [192].

Clinical presentation of Proteus syndrome (OMIM #176920) includes distorting, progressive overgrowth of various tissues including skin, skeleton, adipose, and central nervous system. In most patients it is caused by somatic mosaic mutation of *AKT1* involving c.49G > A (p.Glu17Lys) [61]. This identical mutation is associated with breast, colorectal and ovarian cancers [193]. Mutations in the homologs of *AKT1*, *AKT2* and *AKT3* are also known to cause somatic disorder. p.Glu17Lys mutations in *AKT3* cause hemimegalencephaly and other brain malformations, while the identical mutation in *AKT2* is causative for hypoglycemia [60,62,194,195].

Figure 4. Three intracellular signaling pathways are shown schematically. (**At left**), receptor tyrosine kinase activity leads to activation of PIK3CA, AKT, and mechanistic target of rapamycin (mTOR) [187,188]. mTOR participates in complexes (TORC1, activated by RHEB; TORC2, inhibited by RHEB) that regulate cell growth, proliferation, survival, and cell cycle progression. This pathway includes genes that are frequently mutated in tumors such as *PIK3CA* and *PTEN* (not shown); (**At center**), secreted growth factors bind to receptor tyrosine kinase receptors on the cell surface leading to activation of the low molecular weight G protein Ras and subsequent activation of Raf, MEK 1/2, and ERK 1/2 (official gene symbols *MAPK3*, *MAPK1*); (**At right**), a G-protein coupled receptor (GPCR) pathway is shown [189,190]. Ligands such as vasopressin, endothelin, glutamate, or norepinephrine bind to a GPCR. When bound by ligand, the receptor activates a G protein alpha subunit such as Gαq that binds and hydrolyzes GTP. This leads to activation of phospholipase Cβ producing inositol 1,4,5-triphosphate (IP$_3$) and membrane-associated diacylglycerol (DAG). DAG, through activation of protein kinase C, may activate the Raf/MEK/ERK pathway. IP$_3$ may bind to an IP$_3$ receptor activating calcium signaling pathways (not shown). Other G protein α subunits (such as Gαs encoded by *GNAS*) activate membrane-bound adenylate cyclase, producing cyclic AMP (cAMP) that activates protein kinase A (not shown).



Germline inactivating mutations in the *TSC1* gene encoding hamartin cause tuberous sclerosis 1 (OMIM #191100), while mutations in *TSC2* encoding tuberin cause tuberous sclerosis 2 (OMIM #613254). Hamartin and tuberin act as tumor suppressors by activating the GTPase function of RHEB [196]. Inactivating mutations in a single allele are sufficient to cause tuberous sclerosis. Rare somatic inactivating mutations, lack of expression of the second allele or mosaic UDP events give rise to the multiple benign tumors, tubers and macules characteristic of the disease [197,198].

Neurofibromatosis 1 (OMIM #162200) (NF1) is characterized by the occurrence of at least two (of a list of seven) features such as *café au lait* spots, cutaneous neurofibromas, Lisch nodules (hamartomas) of the iris, and inguinal freckles [199]. Clinical diagnosis requires a first-degree relative with the condition. It is inherited in an autosomal dominant manner (and is among the most common such disorders with a prevalence of 1:3000). Most cases of NF1 are caused by heterozygous loss-of-function mutations of the tumor suppressor gene encoding neurofibromin 1. Only 50% of NF1 individuals have an affected parent, with another 50% having a *de novo* mutation. Neurofibromin 1 is a negative regulator of the RAS signal transduction pathway, with loss of function mutations in neurofibromin 1 leading to RAS activation.

It is possible that mosaic variation occurring during development may result in disease across numerous tissues. One such example is somatic mutation of *IDH1* and *IDH2* that has been shown to cause Ollier disease and Maffucci syndrome. These syndromes are characterized by multiple enchondromas (benign bone tumors originating from cartilage). The causative variants for disease are typically not detectable outside of the tumors indicating that relatively few cells harbor the mutation [200].

The application of sensitive approaches for the detection of mosaicism to a smaller subset of genes based on a patient's phenotype may increase the likelihood of finding causative variants. Jamuar *et al.* applied this approach examining two sets of previously implicated genes in 158 individuals with cerebral cortical defects. Causal mutations were found in 27 individuals, eight of who harbored the causative variant in a mosaic fashion. Notably, causal mutations were only validated at extremely high read depth (>500×) highlighting both the importance of sequence coverage for the detection of mosaic variation and the utility of targeted approaches [201].

Somatic mutations are also known to cause reversion to normal mutations in individuals with monogenic disease [22,23,25,202,203]. Revertant mosaicism occurs when cells harboring a disease-causing mutation revert *in vivo* to a wild-type allele. The disease-causing mutation could be inherited from the germline or somatic. This has been observed for heritable skin diseases such as ichthyosis with confetti (OMIM #609165) and epidermolysis bullosa (OMIM #226650) [202,204] as well as rare blood disorders such as Fanconi anemia (OMIM #227646) and severe combined immunodeficiency resulting from adenosine deaminase deficiency (OMIM #102700) [205,206]. These somatic reversions to normal events may significantly ameliorate disease symptoms if the reversion occurs early enough in development.

For many other overgrowth syndromes somatic mutations have yet to be identified. Examples include Klippel-Trenaunay-Weber syndrome (OMIM %149,000), which involves cutaneous hemangiomata and clinically resembles Sturge-Weber syndrome; and Cobb syndrome (cutaneomeningospinal angiomatosis), which involves vascular cutaneous, muscular, osseous, or other lesions of spinal segments.

3.4. Complex Disease

Multiple recent papers have proposed that somatic mutation may play a role in the etiology of complex disease [3,207,208]. Studies of simplex autism probands have determined that *de novo* mutations account for 2%–15% of disease incidence and that at least 30% of *de novo* mutations can be causally implicated in simplex cases [209–212]. With *de novo* mutations playing such a large role, it is likely that post-zygotic somatic variation also contributes to disease in some individuals. To date, most genetic analysis has found few genetic variants to explain complex disease incidence, suggesting the occurrence of "missing heritability" [213]. A possible model is that somatic variation occurs in conjunction with common and rare inherited variation to cause disease. While this model is not directly supported by current evidence, recent experiments indicate that it warrants investigation. One surprising result from *in situ* hybridization experiments on postmortem brain tissue is the increased presence of patches of cortical disorganization in individuals with autism relative to controls [214]. The authors note that they examined only a small subsection of the brain and therefore cortical disorganization is likely widespread in individuals with autism. Furthermore, an interesting conclusion of recent large-scale examination of exonic *de novo* mutations in simplex autism is that most *de novo* variation implicated as causal occurs opposite wild type alleles [212]. Given that large CNVs are common in neurons of the cortex [78,80], we propose a model of brain-specific somatic mutation occurring opposite inherited *de novo* or rare mutation resulting in sporadic brain-specific loss of gene function and patches of cortical disorganization.

4. Conclusions

While the role of somatic mosaicism in disease is currently under active investigation, it is clear that functional somatic mosaicism has a significant role in human disease. In the last decade, major advances in both cytogenetic and second-generation sequencing techniques have enabled researchers to discover causative somatic mutations for an increasing number of diseases, and driver mutations in an increasing number of cancers. Furthermore, this increased understanding of the genetic underpinnings of disease is likely to lead to improved patient outcomes in the near future.

Acknowledgments

Donald Freed is supported in part by NIH grant 5T32GM007445-37. Jonathan Pevsner is supported by NIH grants R44 MH086192-02A1, P50 MH094268-01, and a grant from the Sturge-Weber Foundation.

Author Contributions

Donald Freed, Eric L. Stevens and Jonathan Pevsner wrote the manuscript and created the figures.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Edwards, J.H. Familiarity, recessivity and germline mosaicism. *Ann. Hum. Genet.* **1989**, *53*, 33–47.
2. Hartl, D.L. Recurrence risks for germinal mosaics. *Am. J. Hum. Genet.* **1971**, *23*, 124–134.
3. Poduri, A.; Evrony, G.D.; Xuyu, C.; Walsh, C.A. Somatic mutation, genomic variation, and neurological disease. *Science* **2013**, *341*, doi:10.1126/science.1237758.
4. Campbell, I.M.; Yuan, B.; Robberecht, C.; Pfundt, R.; Szafranski, P.; McEntagart, M.E.; Nagamani, S.C.; Erez, A.; Bartnik, M.; Wisniowiecka-Kowalnik, B.; *et al.* Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am. J. Hum. Genet.* **2014**, *95*, 173–182.
5. Van der Maarel, S.M.; Deidda, G.; Lemmers, R.J.L.F.; van Overveld, P.G.M.; van der Wielen, M.; Hewitt, J.E.; Sandkuijl, L.; Bakker, B.; van Ommen, G.-J.B.; Padberg, G.W.; *et al. De novo* facioscapulohumeral muscular dystrophy: Frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am. J. Hum. Genet.* **2000**, *66*, 26–35.
6. Boveri, T. *The Origin of Malignant Tumors*; The Williams and Wilkins Company: Baltimore, MD, USA, 1929.
7. Knudson, A.G. Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 820–823.
8. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **1976**, *194*, 23–28.
9. Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A., Jr.; Kinzler, K.W. Cancer genome landscapes. *Science* **2013**, *339*, 1546–1558.
10. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* **1951**, *16*, 13–47.
11. Burnet, M. *The Clonal Selection Theory of Acquired Immunity*; Vanderbilt University Press: Nashville, TN, USA, 1959.
12. Brack, C.; Hirama, M.; Lenhard-Schuller, R.; Tonegawa, S. A complete immunoglobulin gene is created by somatic recombination. *Cell* **1978**, *15*, 1–14.
13. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **1983**, *302*, 575–581.
14. Krangel, M.S. Mechanics of t cell receptor gene rearrangement. *Curr. Opin. Immunol.* **2009**, *21*, 133–139.
15. Frumkin, D.; Wasserstrom, A.; Kaplan, S.; Feige, U.; Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. *PLOS Comput. Biol.* **2005**, *1*, e50.
16. Hoeijmakers, J.H.J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **2009**, *361*, 1475–1485.
17. Kennedy, S.R.; Loeb, L.A.; Herr, A.J. Somatic mutations in aging, cancer and neurodegeneration. *Mech. Ageing Dev.* **2012**, *133*, 118–126.

18. Jeppesen, D.K.; Bohr, V.A.; Stevnsner, T. DNA repair deficiency in neurodegeneration. *Prog. Neurobiol.* **2011**, *94*, 166–200.

19. Erickson, R.P. Somatic gene mutation and human disease other than cancer: An update. *Mutat. Res.* **2010**, *705*, 96–106.

20. Erickson, R.P. Recent advances in the study of somatic mosaicism and diseases other than cancer. *Curr. Opin. Genet. Dev.* **2014**, *26*, 73–78.

21. Erickson, R.P. Somatic gene mutation and human disese other than cancer. *Mutat. Res.* **2003**, *543*, 125–136.

22. Hirschhorn, R. *In vivo* reversion to normal of inherited mutations in humans. *J. Med. Genet.* **2003**, *40*, 721–728.

23. Jonkman, M.F.; Castellanos Nuijts, M.; van Essen, A.J. Natural repair mechanisms in correcting pathogenic mutations in inherited skin disorders. *Clin. Exp. Dermatol.* **2003**, *28*, 625–631.

24. Lai-Cheong, J.E.; McGrath, J.A.; Uitto, J. Revertant mosaicism in skin: Natural gene therapy. *Trends Mol. Med.* **2011**, *17*, 140–148.

25. Jonkman, M.F. Revertant mosaicism in human genetic disorders. *Am. J. Med. Genet.* **1999**, *85*, 361–364.

26. Happle, R. Lethal genes surviving by mosaicism: A possible explanation for sporadic birth defects involving the skin. *J. Am. Acad. Dermatol.* **1987**, *16*, 899–906.

27. Liu, P.; Carvalho, C.M.; Hastings, P.J.; Lupski, J.R. Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.* **2012**, *22*, 211–220.

28. Cimini, D.; Howell, B.; Maddox, P.; Khodjakov, A.; Degrassi, F.; Salmon, E.D. Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J. Cell Biol.* **2001**, *153*, 517–528.

29. Robinson, W.P. Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* **2000**, *22*, 452–459.

30. Kotzot, D. Complex and segmental uniparental disomy updated. *J. Med. Genet.* **2008**, *45*, 545–556.

31. Conlin, L.K.; Thiel, B.D.; Bonnemann, C.G.; Medne, L.; Ernst, L.M.; Zackai, E.H.; Deardorff, M.A.; Krantz, I.D.; Hakonarson, H.; Spinner, N.B. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum. Mol. Genet.* **2010**, *19*, 1263–1275.

32. Liehr, T. Cytogenetic contribution to uniparental disomy (UPD). *Mol. Cytogenet.* **2010**, *3*, 8.

33. Hancks, D.C.; Kazazian, H.H., Jr. Active human retrotransposons: Variation and disease. *Curr. Opin. Genet. Dev.* **2012**, *22*, 191–203.

34. van den Hurk, J.A.; Meij, I.C.; Seleme, M.C.; Kano, H.; Nikopoulos, K.; Hoefsloot, L.H.; Sistermans, E.A.; de Wijs, I.J.; Mukhopadhyay, A.; Plomp, A.S.; *et al.* L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* **2007**, *16*, 1587–1592.

35. Lee, E.; Iskow, R.; Yang, L.; Gokcumen, O.; Haseley, P.; Luquette, L.J.; Lohr, J.G.; Harris, C.C.; Ding, L.; Wilson, R.K.; *et al.* Landscape of somatic retrotransposition in human cancers. *Science* **2012**, *337*, 967–971.

36. Mirkin, S.M. Expandable DNA repeats and human disease. *Nature* **2007**, *447*, 932–940.

37. Kim, J.C.; Mirkin, S.M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **2013**, *23*, 280–288.

38. McMurray, C.T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **2010**, *11*, 786–799.

39. Hellenbroich, Y.; Schwinger, E.; Zühlke, C.H. Limited somatic mosaicism for friedreich's ataxia GAA triplet repeat expansions identified by small pool PCR in blood leukocytes. *Acta Neurol. Scand.* **2001**, *103*, 188–192.

40. Hashida, H.; Goto, J.; Suzuki, T.; Jeong, S.-Y.; Masuda, N.; Ooie, T.; Tachiiri, Y.; Tsuchiya, H.; Kanazawa, I. Single cell analysis of cag repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA). *J. Neurol. Sci.* **2001**, *190*, 87–93.

41. Kahlem, P.; Djian, P. The expanded CAG repeat associated with juvenile Huntington disease shows a common origin of most or all neurons and glia in human cerebrum. *Neurosci. Lett.* **2000**, *286*, 203–207.

42. Møllersen, L.; Rowe, A.D.; Larsen, E.; Rognes, T.; Klungland, A. Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. *PLOS Genet.* **2010**, *6*, e1001242.

43. Ueno, S.-i.; Kondoh, K.; Komure, Y.; Komure, O.; Kuno, S.; Kawai, J.; Hazama, F.; Sano, A. Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (drpla). *Hum. Mol. Genet.* **1995**, *4*, 663–666.

44. Montermini, L.; Kish, S.J.; Jiralerspong, S.; Lamarche, J.B.; Pandolfo, M. Somatic mosaicism for friedreich's ataxia GAA triplet repeat expansions in the central nervous system. *Neurology* **1997**, *49*, 606–610.

45. Lindahl, T.; Wood, R.D. Quality control by DNA repair. *Science* **1999**, *286*, 1897–1905.

46. Hoeijmakers, J.H.J. Genome maintenance mechanisms for preventing cancer. *Nature* **2001**, *411*, 366–374.

47. Gilbert, D.M. Making sense of eukaryotic DNA replication origins. *Science* **2001**, *294*, 96–100.

48. Gilbert, D.M. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.* **2010**, *11*, 673–684.

49. Levy, M.Z.; Allsopp, R.C.; Futcher, A.B.; Greider, C.W.; Harley, C.B. Telomere end-replication problem and cell aging. *J. Mol. Biol.* **1992**, *225*, 951–960.

50. Van Echten-Arends, J.; Mastenbroek, S.; Sikkema-Raddatz, B.; Korevaar, J.C.; Heineman, M.J.; van der Veen, F.; Repping, S. Chromosomal mosaicism in human preimplantation embryos: A systematic review. *Hum. Reprod. Update* **2011**, *17*, 620–627.

51. Shapiro, E.; Biezuner, T.; Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **2013**, *14*, 618–630.

52. Amat, F.; Lemon, W.; Mossing, D.P.; McDole, K.; Wan, Y.; Branson, K.; Myers, E.W.; Keller, P.J. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nat. Methods* **2014**, doi:10.1038/nmeth.3036.

53. Navin, N.; Kendall, J.; Troge, J.; Andrews, P.; Rodgers, L.; McIndoo, J.; Cook, K.; Stepansky, A.; Levy, D.; Esposito, D.; *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **2011**, *472*, 90–94.
54. Wang, Y.; Waters, J.; Leung, M.L.; Unruh, A.; Roh, W.; Shi, X.; Chen, K.; Scheet, P.; Vattathil, S.; Liang, H.; *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **2014**, *512*, 155–160.
55. Bolognia, J.L.; Orlow, S.J.; Glick, S.A. Lines of blaschko. *J. Am. Acad. Dermatol.* **1994**, *31*, 157–190.
56. Shirley, M.D.; Tang, H.T.; Gallione, B.A.; Baugher, J.D.; Frelin, L.P.; Cohen, B.; North, P.E.; Marchuk, D.A.; Comi, A.M.; Pevsner, J. Sturge-weber syndrome and port-wine stains caused by somatic mutation in *GNAQ*. *N. Engl. J. Med.* **2013**, *368*, 1971–1979.
57. Van Raamsdonk, C.D.; Griewank, K.G.; Crosby, M.B.; Garrido, M.C.; Vemula, S.; Wiesner, T.; Obenauf, A.C.; Wackernagel, W.; Green, G.; Bouvier, N.; *et al.* Mutations in GNA11 in uveal melanoma. *N. Engl. J. Med.* **2010**, *363*, 2191–2199.
58. Collins, M.T.; Singer, F.R.; Eugster, E. Mccune-albright syndrome and the extraskeletal manifestations of fibrous dysplasia. *Orphanet J. Rare Dis.* **2012**, *7* (*Suppl 1*), S4.
59. Bastepe, M.; Juppner, H. Gnas locus and pseudohypoparathyroidism. *Horm. Res.* **2005**, *63*, 65–74.
60. Poduri, A.; Evrony, G.D.; Cai, X.; Elhosary, P.C.; Beroukhim, R.; Lehtinen, M.K.; Hills, L.B.; Heinzen, E.L.; Hill, A.; Hill, R.S.; *et al.* Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **2012**, *74*, 41–48.
61. Lindhurst, M.J.; Sapp, J.C.; Teer, J.K.; Johnston, J.J.; Finn, E.M.; Peters, K.; Turner, J.; Cannons, J.L.; Bick, D.; Blakemore, L.; *et al.* A mosaic activating mutation in AKT1 associated with the proteus syndrome. *N. Engl. J. Med.* **2011**, *365*, 611–619.
62. Hussain, K.; Challis, B.; Rocha, N.; Payne, F.; Minic, M.; Thompson, A.; Daly, A.; Scott, C.; Harris, J.; Smillie, B.J.; *et al.* An activating mutation of AKT2 and human hypoglycemia. *Science* **2011**, *334*, 474.
63. Jacobs, K.B.; Yeager, M.; Zhou, W.; Wacholder, S.; Wang, Z.; Rodriguez-Santiago, B.; Hutchinson, A.; Deng, X.; Liu, C.; Horner, M.-J.; *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **2012**, *44*, 651–658.
64. Laurie, C.C.; Laurie, C.A.; Rice, K.; Doheny, K.F.; Zelnick, L.R.; McHugh, C.P.; Ling, H.; Hetrick, K.N.; Pugh, E.W.; Amos, C.; *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **2012**, *44*, 642–650.
65. Aghili, L.; Foo, J.; DeGregori, J.; De, S. Patterns of somatically acquired amplifications and deletions in apparently normal tissues of ovarian cancer patients. *Cell Rep.* **2014**, *7*, 1310–1319.
66. Costa, T.; Lambert, M.; Teshima, I.; Ray, P.N.; Richer, C.-L.; Dallaire, L. Monozygotic twins with 45, X/46, XY mosaicism discordant for phenotypic sex. *Am. J. Med. Genet.* **1998**, *75*, 40–44.
67. Fujimoto, A.; Boelter, W.D.; Sparkes, R.S.; Lin, M.S.; Battersby, K. Monozygotic twins of discordant sex both with 45,X/46,X,idic(Y) mosaicism. *Am. J. Med. Genet.* **1991**, *41*, 239–245.

68. Kaplan, L.; Foster, R.; Shen, Y.; Parry, D.M.; McMaster, M.L.; O'Leary, M.C.; Gusella, J.F. Monozygotic twins discordant for neurofibromatosis 1. *Am. J. Med. Genet. A* **2010**, *152A*, 601–606.

69. Zeng, S.; Patil, S.R.; Yankowitz, J. Prenatal detection of mosaic trisomy 1q due to an unbalanced translocation in one fetus of a twin pregnancy following in vitro fertilization: A postzygotic error. *Am. J. Med. Genet. A* **2003**, *120A*, 464–469.

70. Helderman-van den Enden, A.T.J.M.; Maaswinkel-Mooij, P.D.; Hoogendoorn, E.; Willemsen, R.; Maat-Kievit, J.A.; Losekoot, M.; Oostra, B.A. Monozygotic twin brothers with the fragile X syndrome: Different CGG repeats and different mental capacities. *J. Med. Genet.* **1999**, *36*, 253–257.

71. Piotrowski, A.; Bruder, C.E.; Andersson, R.; Diaz de Stahl, T.; Menzel, U.; Sandgren, J.; Poplawski, A.; von Tell, D.; Crasto, C.; Bogdan, A.; *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum. Mutat.* **2008**, *29*, 1118–1124.

72. O'Huallachain, M.; Karczewski, K.J.; Weissman, S.M.; Urban, A.E.; Snyder, M.P. Extensive genetic variation in somatic human tissues. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 18018–18023.

73. Abyzov, A.; Mariani, J.; Palejev, D.; Zhang, Y.; Haney, M.S.; Tomasini, L.; Ferrandino, A.F.; Rosenberg Belmaker, L.A.; Szekely, A.; Wilson, M.; *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **2012**, *492*, 438–442.

74. Rehen, S.K.; McConnell, M.J.; Kaushal, D.; Kingsbury, M.A.; Yang, A.H.; Chun, J. Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13361–13366.

75. Baillie, J.K.; Barnett, M.W.; Upton, K.R.; Gerhardt, D.J.; Richmond, T.A.; de Sapio, F.; Brennan, P.M.; Rizzu, P.; Smith, S.; Fell, M.; *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **2011**, *479*, 534–537.

76. Coufal, N.G.; Garcia-Perez, J.L.; Peng, G.E.; Yeo, G.W.; Mu, Y.; Lovci, M.T.; Morell, M.; O'Shea, K.S.; Moran, J.V.; Gage, F.H. L1 retrotransposition in human neural progenitor cells. *Nature* **2009**, *460*, 1127–1131.

77. Muotri, A.R.; Chu, V.T.; Marchetto, M.C.; Deng, W.; Moran, J.V.; Gage, F.H. Somatic mosaicism in neuronal precursor cells mediated by l1 retrotransposition. *Nature* **2005**, *435*, 903–910.

78. Cai, X.; Evrony, G.D.; Lehmann, H.S.; Elhosary, P.C.; Mehta, B.K.; Poduri, A.; Walsh, C.A. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **2014**, *8*, 1280–1289.

79. Evrony, G.D.; Cai, X.; Lee, E.; Hills, L.B.; Elhosary, P.C.; Lehmann, H.S.; Parker, J.J.; Atabay, K.D.; Gilmore, E.C.; Poduri, A.; *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **2012**, *151*, 483–496.

80. McConnell, M.J.; Lindberg, M.R.; Brennand, K.J.; Piper, J.C.; Voet, T.; Cowing-Zitron, C.; Shumilina, S.; Lasken, R.S.; Vermeesch, J.R.; Hall, I.M.; *et al.* Mosaic copy number variation in human neurons. *Science* **2013**, *342*, 632–637.

81. Kalousek, D.K.; Vekemans, M. Confined placental mosaicism. *J. Med. Genet.* **1996**, *33*, 529–533.

82. Kalousek, D.K.; Dill, F.J. Chromosomal mosaicism confined to the placenta in human conceptions. *Science* **1983**, *221*, 665–667.

83. Taylor, T.H.; Gitlin, S.A.; Patrick, J.L.; Crain, J.L.; Wilson, J.M.; Griffin, D.K. The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans. *Hum. Reprod. Update* **2014**, *20*, 571–581.

84. Vanneste, E.; Voet, T.; Le Caignec, C.; Ampe, M.; Konings, P.; Melotte, C.; Debrock, S.; Amyere, M.; Vikkula, M.; Schuit, F.; *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.* **2009**, *15*, 577–583.

85. Chen, E.Z.; Chiu, R.W.K.; Sun, H.; Akolekar, R.; Chan, K.C.A.; Leung, T.Y.; Jiang, P.; Zheng, Y.W.L.; Lun, F.M.F.; Chan, L.Y.S.; *et al.* Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLOS ONE* **2011**, *6*, e21791.

86. Youssoufian, H.; Pyeritz, R.E. Mechanisms and consequences of somatic mosaicism in humans. *Nat. Rev. Genet.* **2002**, *3*, 748–758.

87. Watson, I.R.; Takahashi, K.; Futreal, P.A.; Chin, L. Emerging patterns of somatic mutations in cancer. *Nat. Rev. Genet.* **2013**, *14*, 703–718.

88. Ito, Y.; Tanaka, F.; Yamamoto, M.; Doyu, M.; Nagamatsu, M.; Riku, S.; Mitsuma, T.; Sobue, G. Somatic mosaicism of the expanded cag trinucleotide repeat in mrnas for the responsible gene of machado-joseph disease (MJD), dentatorubral-pallidoluysian atrophy (DRPLA), and spinal and bulbar muscular atrophy (SBMA). *Neurochem. Res.* **1998**, *23*, 25–32.

89. James, C.D.; Carlbom, E.; Nordenskjold, M.; Collins, V.P.; Cavenee, W.K. Mitotic recombination of chromosome 17 in astrocytomas. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2858–2862.

90. Kleczkowska, A.; Fryns, J.P.; Van den Berghe, H. On the variable effect of mosaic normal/balanced chromosomal rearrangements in man. *J. Med. Genet.* **1990**, *27*, 505–507.

91. Kotzot, D.; Schmitt, S.; Bernasconi, F.; Robinson, W.P.; Lurie, I.W.; Ilyina, H.; Méhes, K.; Hamel, B.C.J.; Otten, B.J.; Hergersberg, M.; *et al.* Uniparental disomy 7 in silver-russell syndrome and primordial growth retardation. *Hum. Mol. Genet.* **1995**, *4*, 583–587.

92. Rodríguez-Santiago, B.; Malats, N.; Rothman, N.; Armengol, L.; Garcia-Closas, M.; Kogevinas, M.; Villa, O.; Hutchinson, A.; Earl, J.; Marenne, G.; *et al.* Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am. J. Hum. Genet.* **2010**, *87*, 129–138.

93. Slatter, R.E.; Elliott, M.; Welham, K.; Carrera, M.; Schofield, P.N.; Barton, D.E.; Maher, E.R. Mosaic uniparental disomy in beckwith-wiedemann syndrome. *J. Med. Genet.* **1994**, *31*, 749–753.

94. Zori, R.T.; Gray, B.A.; Bent-Williams, A.; Driscoll, D.J.; Williams, C.A.; Zackowski, J.L. Preaxial acrofacial dysostosis (nager syndrome) associated with an inherited and apparently balanced X;9 translocation: Prenatal and postnatal late replication studies. *Am. J. Med. Genet.* **1993**, *46*, 379–383.

95. Walsh, C.; Cepko, C.L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **1992**, *255*, 434–440.

96. Pleasure, S.J.; Anderson, S.; Hevner, R.; Bagri, A.; Marin, O.; Lowenstein, D.H.; Rubenstein, J.L. Cell migration from the ganglionic eminences is required for the development of hippocampal gabaergic interneurons. *Neuron* **2000**, *28*, 727–740.
97. Hohn, A.; Leibrock, J.; Bailey, K.; Barde, Y.-A. Identification and characterization of a novel member of the nerve growth factor/brain-derived neurotrophic factor family. *Nature* **1990**, *344*, 339–341.
98. Leibrock, J.; Lottspeich, F.; Hohn, A.; Hofer, M.; Hengerer, B.; Masiakowski, P.; Thoenen, H.; Barde, Y.-A. Molecular cloning and expression of brain-derived neurotrophic factor. *Nature* **1989**, *341*, 149–152.
99. Levi-Montalcini, R. Growth control of nerve cells by a protein factor and its antiserum: Discovery of this factor may provide new leads to understanding of some neurogenetic processes. *Science* **1964**, *143*, 105–110.
100. Kurek, K.C.; Luks, V.L.; Ayturk, U.M.; Alomari, A.I.; Fishman, S.J.; Spencer, S.A.; Mulliken, J.B.; Bowen, M.E.; Yamamoto, G.L.; Kozakewich, H.P.; *et al.* Somatic mosaic activating mutations in PIK3CA cause cloves syndrome. *Am. J. Hum. Genet.* **2012**, *90*, 1108–1115.
101. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **2014**, *30*, 2843–2851.
102. Gerstung, M.; Papaemmanuil, E.; Campbell, P.J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics* **2014**, *30*, 1198–1204.
103. Crotwell, P.L.; Hoyme, H.E. Advances in whole-genome genetic testing: From chromosomes to microarrays. *Curr. Probl. Pediatr. Adolesc. Health Care* **2012**, *42*, 47–73.
104. Bushman, D.M.; Chun, J. The genomically mosaic brain: Aneuploidy and more in neural diversity and disease. *Semin. Cell Dev. Biol.* **2013**, *24*, 357–369.
105. Notini, A.J.; Craig, J.M.; White, S.J. Copy number variation and mosaicism. *Cytogenet. Genome Res.* **2008**, *123*, 270–277.
106. Vorsanova, S.G.; Yurov, Y.B.; Iourov, I.Y. Human interphase chromosomes: A review of available molecular cytogenetic technologies. *Mol. Cytogenet.* **2010**, *3*, 1.
107. Imataka, G.; Arisaka, O. Chromosome analysis using spectral karyotyping (sky). *Cell Biochem. Biophys.* **2012**, *62*, 13–17.
108. Kallioniemi, A.; Kallioniemi, O.P.; Sudar, D.; Rutovitz, D.; Gray, J.W.; Waldman, F.; Pinkel, D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **1992**, *258*, 818–821.
109. Pinkel, D.; Albertson, D.G. Array comparative genomic hybridization and its applications in cancer. *Nat. Genet.* **2005**, *37*, S11–S17.
110. Alkan, C.; Coe, B.P.; Eichler, E.E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **2011**, *12*, 363–376.
111. Bignell, G.R.; Huang, J.; Greshock, J.; Watt, S.; Butler, A.; West, S.; Grigorova, M.; Jones, K.W.; Wei, W.; Stratton, M.R.; *et al.* High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* **2004**, *14*, 287–295.

112. Mohr, S.; Leikauf, G.D.; Keith, G.; Rihn, B.H. Microarrays as cancer keys: An array of possibilities. *J. Clin. Oncol.* **2002**, *20*, 3165–3175.

113. Baugher, J.D.; Baugher, B.D.; Shirley, M.D.; Pevsner, J. Sensitive and specific detection of mosaic chromosomal abnormalities using the parent-of-origin-based detection (POD) method. *BMC Genom.* **2013**, *14*, 367.

114. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739.

115. Chen, G.K.; Chang, X.; Curtis, C.; Wang, K. Precise inference of copy number alterations in tumor samples from SNP arrays. *Bioinformatics* **2013**, *29*, 2964–2970.

116. Li, A.; Liu, Z.; Lezon-Geyda, K.; Sarkar, S.; Lannin, D.; Schulz, V.; Krop, I.; Winer, E.; Harris, L.; Tuck, D. GPHMM: An integrated hidden Markov model for identification of copy number alteration and loss of heterozygosity in complex tumor samples using whole genome SNP arrays. *Nucleic Acids Res.* **2011**, *39*, 4928–4941.

117. Liu, Z.; Li, A.; Schulz, V.; Chen, M.; Tuck, D. Mixhmm: Inferring copy number variation and allelic imbalance using SNP arrays and tumor samples mixed with stromal cells. *PLOS ONE* **2010**, *5*, e10909.

118. Rancoita, P.; Hutter, M.; Bertoni, F.; Kwee, I. An integrated bayesian analysis of LOH and copy number data. *BMC Bioinform.* **2010**, *11*, 321.

119. Cock, P.J.; Fields, C.J.; Goto, N.; Heuer, M.L.; Rice, P.M. The sanger FASTQ file format for sequences with quality scores, and the solexa/illumina FASTQ variants. *Nucleic Acids Res.* **2010**, *38*, 1767–1771.

120. Yoon, S.; Xuan, Z.; Makarov, V.; Ye, K.; Sebat, J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* **2009**, *19*, 1586–1592.

121. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. Varscan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576.

122. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **2009**, *461*, 272–276.

123. Choi, M.; Scholl, U.I.; Ji, W.; Liu, T.; Tikhonova, I.R.; Zumbo, P.; Nayir, A.; Bakkaloğlu, A.; Özen, S.; Sanjad, S.; *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci.* **2009**, *106*, 19096–19101.

124. Hodges, E.; Xuan, Z.; Balija, V.; Kramer, M.; Molla, M.N.; Smith, S.W.; Middle, C.M.; Rodesch, M.J.; Albert, T.J.; Hannon, G.J.; *et al.* Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* **2007**, *39*, 1522–1527.

125. Larson, D.E.; Harris, C.C.; Chen, K.; Koboldt, D.C.; Abbott, T.E.; Dooling, D.J.; Ley, T.J.; Mardis, E.R.; Wilson, R.K.; Ding, L. Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **2012**, *28*, 311–317.

126. Roth, A.; Ding, J.; Morin, R.; Crisan, A.; Ha, G.; Giuliany, R.; Bashashati, A.; Hirst, M.; Turashvili, G.; Oloumi, A.; *et al.* Jointsnvmix: A probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* **2012**, *28*, 907–913.

127. Saunders, C.T.; Wong, W.S.; Swamy, S.; Becq, J.; Murray, L.J.; Cheetham, R.K. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **2012**, *28*, 1811–1817.

128. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **2013**, *31*, 213–219.

129. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and samtools. *Bioinformatics* **2009**, *25*, 2078–2079.

130. Amarasinghe, K.C.; Li, J.; Halgamuge, S.K. Convex: Copy number variation estimation in exome sequencing data using HMM. *BMC Bioinform.* **2013**, *14*, S2.

131. Boeva, V.; Popova, T.; Bleakley, K.; Chiche, P.; Cappo, J.; Schleiermacher, G.; Janoueix-Lerosey, I.; Delattre, O.; Barillot, E. Control-freec: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **2012**, *28*, 423–425.

132. Layer, R.M.; Chiang, C.; Quinlan, A.R.; Hall, I.M. Lumpy: A probabilistic framework for structural variant discovery. *Genome Biol.* **2014**, *15*, R84.

133. Chen, M.; Gunel, M.; Zhao, H. Somatica: Identifying, characterizing and quantifying somatic copy number aberrations from cancer genome sequencing data. *PLOS ONE* **2013**, *8*, e78143.

134. Ding, L.; Wendl, M.C.; McMichael, J.F.; Raphael, B.J. Expanding the computational toolbox for mining cancer genomes. *Nat. Rev. Genet.* **2014**, *15*, 556–570.

135. Yadav, V.K.; De, S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief. Bioinform.* **2014**, doi:10.1093/bib/bbu002.

136. Dean, F.B.; Hosono, S.; Fang, L.; Wu, X.; Faruqi, A.F.; Bray-Ward, P.; Sun, Z.; Zong, Q.; Du, Y.; Du, J.; *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5261–5266.

137. Hosono, S.; Faruqi, A.F.; Dean, F.B.; Du, Y.; Sun, Z.; Wu, X.; Du, J.; Kingsmore, S.F.; Egholm, M.; Lasken, R.S. Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* **2003**, *13*, 954–964.

138. Pugh, T.J.; Delaney, A.D.; Farnoud, N.; Flibotte, S.; Griffith, M.; Li, H.I.; Qian, H.; Farinha, P.; Gascoyne, R.D.; Marra, M.A. Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acids Res.* **2008**, *36*, e80.

139. Baslan, T.; Kendall, J.; Rodgers, L.; Cox, H.; Riggs, M.; Stepansky, A.; Troge, J.; Ravi, K.; Esposito, D.; Lakshmi, B.; *et al.* Genome-wide copy number analysis of single cells. *Nat. Protoc.* **2012**, *7*, 1024–1041.

140. Gundry, M.; Li, W.; Maqbool, S.B.; Vijg, J. Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic Acids Res.* **2012**, *40*, 2032–2040.

141. Hou, Y.; Song, L.; Zhu, P.; Zhang, B.; Tao, Y.; Xu, X.; Li, F.; Wu, K.; Liang, J.; Shao, D.; *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **2012**, *148*, 873–885.
142. Xu, X.; Hou, Y.; Yin, X.; Bao, L.; Tang, A.; Song, L.; Li, F.; Tsang, S.; Wu, K.; Wu, H.; *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **2012**, *148*, 886–895.
143. Sjoblom, T.; Jones, S.; Wood, L.D.; Parsons, D.W.; Lin, J.; Barber, T.D.; Mandelker, D.; Leary, R.J.; Ptak, J.; Silliman, N.; *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **2006**, *314*, 268–274.
144. Stephens, P.J.; Tarpey, P.S.; Davies, H.; van Loo, P.; Greenman, C.; Wedge, D.C.; Nik-Zainal, S.; Martin, S.; Varela, I.; Bignell, G.R.; *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **2012**, *486*, 400–404.
145. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719–724.
146. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674.
147. McCubrey, J.A.; Steelman, L.S.; Chappell, W.H.; Abrams, S.L.; Montalto, G.; Cervello, M.; Nicoletti, F.; Fagone, P.; Malaponte, G.; Mazzarino, M.C.; *et al.* Mutations and deregulation of Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR cascades which alter therapy response. *Oncotarget* **2012**, *3*, 954–987.
148. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **2003**, *3*, 11–22.
149. Liaw, D.; Marsh, D.J.; Li, J.; Dahia, P.L.M.; Wang, S.I.; Zheng, Z.; Bose, S.; Call, K.M.; Tsou, H.C.; Peacoke, M.; *et al.* Germline mutations of the pten gene in cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* **1997**, *16*, 64–67.
150. Malkin, D.; Li, F.; Strong, L.; Fraumeni, J.; Nelson, C.; Kim, D.; Kassel, J.; Gryka, M.; Bischoff, F.; Tainsky, M.; *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **1990**, *250*, 1233–1238.
151. Morin, P.J.; Sparks, A.B.; Korinek, V.; Barker, N.; Clevers, H.; Vogelstein, B.; Kinzler, K.W. Activation of β-catenin-TCF signaling in colon cancer by mutations in β-catenin or APC. *Science* **1997**, *275*, 1787–1790.
152. Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.; Gregory, S.; Gumbs, C.; Micklem, G.; *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **1995**, *378*, 789–792.
153. Moynahan, M.E.; Chiu, J.W.; Koller, B.H.; Jasin, M. Brca1 controls homology-directed DNA repair. *Mol. Cell* **1999**, *4*, 511–518.
154. Miyaki, M.; Konishi, M.; Tanaka, K.; Kikuchi-Yanoshita, R.; Muraoka, M.; Yasuno, M.; Igari, T.; Koike, M.; Chiba, M.; Mori, T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **1997**, *17*, 271–272.
155. Cleaver, J.E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **1968**, *218*, 652–656.

156. Govindan, R.; Ding, L.; Griffith, M.; Subramanian, J.; Dees, N.D.; Kanchi, K.L.; Maher, C.A.; Fulton, R.; Fulton, L.; Wallis, J.; *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **2012**, *150*, 1121–1134.

157. Forbes, S.A.; Bindal, N.; Bamford, S.; Cole, C.; Kok, C.Y.; Beare, D.; Jia, M.; Shepherd, R.; Leung, K.; Menzies, A.; *et al.* Cosmic: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **2011**, *39*, D945–D950.

158. Collins, F.S.; Barker, A.D. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.* **2007**, *296*, 50–57.

159. Debeljak, M.; Freed, D.N.; Welch, J.A.; Haley, L.; Beierl, K.; Iglehart, B.S.; Pallavajjala, A.; Gocke, C.D.; Leffell, M.S.; Lin, M.-T.; *et al.* Haplotype counting by next-generation sequencing for ultrasensitive human DNA detection. *J. Mol. Diagn.* **2014**, *16*, 495–503.

160. Armitage, P.; Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **1954**, *8*, 1–12.

161. Szilard, L. On the nature of the aging process. *Proc. Natl. Acad. Sci. USA* **1959**, *45*, 30–45.

162. Curtis, H.J. Biological mechanisms underlying the aging process. *Science* **1963**, *141*, 686–694.

163. Goldsby, R.E.; Hays, L.E.; Chen, X.; Olmsted, E.A.; Slayton, W.B.; Spangrude, G.J.; Preston, B.D. High incidence of epithelial cancers in mice deficient for DNA polymerase δ proofreading. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15560–15565.

164. Goldsby, R.E.; Lawrence, N.A.; Hays, L.E.; Olmsted, E.A.; Chen, X.; Singh, M.; Preston, B.D. Defective DNA polymerase-[delta] proofreading causes cancer susceptibility in mice. *Nat. Med.* **2001**, *7*, 638–639.

165. Albertson, T.M.; Ogawa, M.; Bugni, J.M.; Hays, L.E.; Chen, Y.; Wang, Y.; Treuting, P.M.; Heddle, J.A.; Goldsby, R.E.; Preston, B.D. DNA polymerase ε and δ proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17101–17104.

166. Vermulst, M.; Bielas, J.H.; Kujoth, G.C.; Ladiges, W.C.; Rabinovitch, P.S.; Prolla, T.A.; Loeb, L.A. Mitochondrial point mutations do not limit the natural lifespan of mice. *Nat. Genet.* **2007**, *39*, 540–543.

167. Trifunovic, A.; Wredenberg, A.; Falkenberg, M.; Spelbrink, J.N.; Rovio, A.T.; Bruder, C.E.; Bohlooly-Y, M.; Gidlof, S.; Oldfors, A.; Wibom, R.; *et al.* Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **2004**, *429*, 417–423.

168. Marteijn, J.A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J.H.J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell. Biol.* **2014**, *15*, 465–481.

169. Mohaghegh, P.; Hickson, I.D. DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. *Hum. Mol. Genet.* **2001**, *10*, 741–746.

170. Hanks, S.; Coleman, K.; Reid, S.; Plaja, A.; Firth, H.; Fitzpatrick, D.; Kidd, A.; Mehes, K.; Nash, R.; Robin, N.; *et al.* Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in bub1b. *Nat. Genet.* **2004**, *36*, 1159–1161.

171. Date, H.; Onodera, O.; Tanaka, H.; Iwabuchi, K.; Uekawa, K.; Igarashi, S.; Koike, R.; Hiroi, T.; Yuasa, T.; Awaya, Y.; *et al.* Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nat. Genet.* **2001**, *29*, 184–188.
172. Moreira, M.-C.; Barbot, C.; Tachi, N.; Kozuka, N.; Uchida, E.; Gibson, T.; Mendonca, P.; Costa, M.; Barros, J.; Yanagisawa, T.; *et al.* The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nat. Genet.* **2001**, *29*, 189–193.
173. Niedernhofer, L.J. Tissue-specific accelerated aging in nucleotide excision repair deficiency. *Mech. Ageing Dev.* **2008**, *129*, 408–415.
174. Monnat Jr, R.J. Human RECQ helicases: Roles in DNA metabolism, mutagenesis and cancer biology. *Semin. Cancer Biol.* **2010**, *20*, 329–339.
175. Burdick, D.; Soreghan, B.; Kwon, M.; Kosmoski, J.; Knauer, M.; Henschen, A.; Yates, J.; Cotman, C.; Glabe, C. Assembly and aggregation properties of synthetic Alzheimer's A4/beta amyloid peptide analogs. *J. Biol. Chem.* **1992**, *267*, 546–554.
176. Goldfarb, L.G.; Brown, P.; McCombie, W.R.; Goldgaber, D.; Swergold, G.D.; Wills, P.R.; Cervenakova, L.; Baron, H.; Gibbs, C.J.; Gajdusek, D.C. Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. *Proc. Natl. Acad. Sci.* USA **1991**, *88*, 10926–10930.
177. Beck, J.A.; Poulter, M.; Campbell, T.A.; Uphill, J.B.; Adamson, G.; Geddes, J.F.; Revesz, T.; Davis, M.B.; Wood, N.W.; Collinge, J.; *et al.* Somatic and germline mosaicism in sporadic early-onset alzheimer's disease. *Hum. Mol. Genet.* **2004**, *13*, 1219–1224.
178. Alzualde, A.; Moreno, F.; Martínez-Lage, P.; Ferrer, I.; Gorostidi, A.; Otaegui, D.; Blázquez, L.; Atares, B.; Cardoso, S.; Martínez de Pancorbo, M.; *et al.* Somatic mosaicism in a case of apparently sporadic Creutzfeldt-Jakob disease carrying a *de novo* D178n mutation in the PRNP gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2010**, *153B*, 1283–1291.
179. Eikelenboom, P.; Bate, C.; Van Gool, W.A.; Hoozemans, J.J.M.; Rozemuller, J.M.; Veerhuis, R.; Williams, A. Neuroinflammation in alzheimer's disease and prion disease. *Glia* **2002**, *40*, 232–239.
180. Meyer-Luehmann, M.; Coomaraswamy, J.; Bolmont, T.; Kaeser, S.; Schaefer, C.; Kilger, E.; Neuenschwander, A.; Abramowski, D.; Frey, P.; Jaton, A.L.; *et al.* Exogenous induction of cerebral β-amyloidogenesis is governed by agent and host. *Science* **2006**, *313*, 1781–1784.
181. Kane, M.D.; Lipinski, W.J.; Callahan, M.J.; Bian, F.; Durham, R.A.; Schwarz, R.D.; Roher, A.E.; Walker, L.C. Evidence for seeding of beta-amyloid by intracerebral infusion of alzheimer brain extracts in beta-amyloid precursor protein-transgenic mice. *J. Neurosci.* **2000**, *20*, 3606–3611.
182. Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C.D.; Meredith, S.C.; Tycko, R. Molecular structure of β-amyloid fibrils in alzheimer's disease brain tissue. *Cell* **2013**, *154*, 1257–1268.
183. Stöhr, J.; Condello, C.; Watts, J.C.; Bloch, L.; Oehler, A.; Nick, M.; DeArmond, S.J.; Giles, K.; DeGrado, W.F.; Prusiner, S.B. Distinct synthetic aβ prion strains producing different amyloid deposits in bigenic mice. *Proc. Natl. Acad. Sci.* **2014**, *111*, 10329–10334.

184. Watts, J.C.; Condello, C.; Stöhr, J.; Oehler, A.; Lee, J.; DeArmond, S.J.; Lannfelt, L.; Ingelsson, M.; Giles, K.; Prusiner, S.B. Serial propagation of distinct strains of aβ prions from alzheimer's disease patients. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10323–10328.

185. Ott, A.; Breteler, M.M.B.; van Harskamp, F.; Claus, J.J.; van der Cammen, T.J.M.; Grobbee, D.E.; Hofman, A. Prevalence of alzheimer's disease and vascular dementia: Association with education. The Rotterdam study. *BMJ* **1995**, *310*, 970.

186. Chapman, J.; Ben-Israel, J.; Goldhammer, Y.; Korczyn, A.D. The risk of developing creutzfeldt-jakob disease in subjects with the PRNP gene codon 200 point mutation. *Neurology* **1994**, *44*, 1683–1686.

187. Fruman, D.A.; Rommel, C. PI3K and cancer: Lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* **2014**, *13*, 140–156.

188. Hay, N. The AKT-mtor tango and its relevance to cancer. *Cancer Cell* **2005**, *8*, 179–183.

189. Rozengurt, E. Mitogenic signaling pathways induced by G protein-coupled receptors. *J. Cell. Physiol.* **2007**, *213*, 589–602.

190. Dorsam, R.T.; Gutkind, J.S. G-protein-coupled receptors and cancer. *Nat. Rev. Cancer* **2007**, *7*, 79–94.

191. Mirzaa, G.; Conaway, R.; Graham, J.M., Jr.; Dobyns, W.B. PIK3CA-related segmental overgrowth. In *GeneReviews*; Pagon, R.A.; Adam, M.P.; Ardinger, H.H.; *et al.*, Eds. University of Washington Seattle: Seattle, WA, USA, 2013.

192. Samuels, Y.; Wang, Z.; Bardelli, A.; Silliman, N.; Ptak, J.; Szabo, S.; Yan, H.; Gazdar, A.; Powell, S.M.; Riggins, G.J.; *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **2004**, *304*, 554.

193. Carpten, J.D.; Faber, A.L.; Horn, C.; Donoho, G.P.; Briggs, S.L.; Robbins, C.M.; Hostetter, G.; Boguslawski, S.; Moses, T.Y.; Savage, S.; *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **2007**, *448*, 439–444.

194. Riviere, J.B.; Mirzaa, G.M.; O'Roak, B.J.; Beddaoui, M.; Alcantara, D.; Conway, R.L.; St-Onge, J.; Schwartzentruber, J.A.; Gripp, K.W.; Nikkel, S.M.; *et al. De novo* germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **2012**, *44*, 934–940.

195. Lee, J.H.; Huynh, M.; Silhavy, J.L.; Kim, S.; Dixon-Salazar, T.; Heiberg, A.; Scott, E.; Bafna, V.; Hill, K.J.; Collazo, A.; *et al. De novo* somatic mutations in components of the PI3K-AKT3-mtor pathway cause hemimegalencephaly. *Nat. Genet.* **2012**, *44*, 941–945.

196. Zhang, Y.; Gao, X.; Saucedo, L.J.; Ru, B.; Edgar, B.A.; Pan, D. RHEB is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nat. Cell. Biol.* **2003**, *5*, 578–581.

197. Curatolo, P.; Bombardieri, R.; Jozwiak, S. Tuberous sclerosis. *Lancet* **2008**, *372*, 657–668.

198. Henske, E.P.; Wessner, L.L.; Golden, J.; Scheithauer, B.W.; Vortmeyer, A.O.; Zhuang, Z.; Klein-Szanto, A.J.; Kwiatkowski, D.J.; Yeung, R.S. Loss of tuberin in both subependymal giant cell astrocytomas and angiomyolipomas supports a two-hit model for the pathogenesis of tuberous sclerosis tumors. *Am. J. Pathol.* **1997**, *151*, 1639–1647.

199. Tsang, E.; Birch, P.; Friedman, J.M. Valuing gene testing in children with possible neurofibromatosis 1. *Clin. Genet.* **2012**, *82*, 591–593.

200. Pansuriya, T.C.; van Eijk, R.; d'Adamo, P.; van Ruler, M.A.; Kuijjer, M.L.; Oosting, J.; Cleton-Jansen, A.M.; van Oosterwijk, J.G.; Verbeke, S.L.; Meijer, D.; *et al.* Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in ollier disease and maffucci syndrome. *Nat. Genet.* **2011**, *43*, 1256–1261.
201. Jamuar, S.S.; Lam, A.T.; Kircher, M.; D'Gama, A.M.; Wang, J.; Barry, B.J.; Zhang, X.; Hill, R.S.; Partlow, J.N.; Rozzo, A.; *et al.* Somatic mutations in cerebral cortical malformations. *N. Engl. J. Med.* **2014**, *371*, 733–743.
202. Choate, K.A.; Lu, Y.; Zhou, J.; Choi, M.; Elias, P.M.; Farhi, A.; Nelson-Williams, C.; Crumrine, D.; Williams, M.L.; Nopper, A.J.; *et al.* Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. *Science* **2010**, *330*, 94–97.
203. Pasmooij, A.M.; Jonkman, M.F.; Uitto, J. Revertant mosaicism in heritable skin diseases: Mechanisms of natural gene therapy. *Discov. Med.* **2012**, *14*, 167–179.
204. Pasmooij, A.M.; Pas, H.H.; Bolling, M.C.; Jonkman, M.F. Revertant mosaicism in junctional epidermolysis bullosa due to multiple correcting second-site mutations in LAMB3. *J. Clin. Invest.* **2007**, *117*, 1240–1248.
205. Hirschhorn, R.; Yang, D.R.; Puck, J.M.; Huie, M.L.; Jiang, C.K.; Kurlandsky, L.E. Spontaneous *in vivo* reversion to normal of an inherited mutation in a patient with adenosine deaminase deficiency. *Nat. Genet.* **1996**, *13*, 290–295.
206. Soulier, J.; Leblanc, T.; Larghero, J.; Dastot, H.; Shimamura, A.; Guardiola, P.; Esperou, H.; Ferry, C.; Jubert, C.; Feugeas, J.-P.; *et al.* Detection of somatic mosaicism and classification of fanconi anemia patients by analysis of the FA/BRCA pathway. *Blood* **2005**, *105*, 1329–1336.
207. De, S. Somatic mosaicism in healthy human tissues. *Trends Genet.* **2011**, *27*, 217–223.
208. Insel, T.R. Brain somatic mutations: The dark matter of psychiatric genetics [quest]. *Mol. Psychiatry* **2014**, *19*, 156–158.
209. Krumm, N.; O'Roak, B.J.; Shendure, J.; Eichler, E.E. A *de novo* convergence of autism genetics and molecular neuroscience. *Trends Neurosci.* **2014**, *37*, 95–105.
210. Gaugler, T.; Klei, L.; Sanders, S.J.; Bodea, C.A.; Goldberg, A.P.; Lee, A.B.; Mahajan, M.; Manaa, D.; Pawitan, Y.; Reichert, J.; *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **2014**, *46*, 881–885.
211. De Rubeis, S.; He, X.; Goldberg, A.P.; Poultney, C.S.; Samocha, K.; Ercument Cicek, A.; Kou, Y.; Liu, L.; Fromer, M.; Walker, S.; *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **2014**, *515*, 209–215.
212. Iossifov, I.; O'Roak, B.J.; Sanders, S.J.; Ronemus, M.; Krumm, N.; Levy, D.; Stessman, H.A.; Witherspoon, K.T.; Vives, L.; Patterson, K.E.; *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **2014**, *515*, 216–221.
213. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; *et al.* Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753.

214. Stoner, R.; Chow, M.L.; Boyle, M.P.; Sunkin, S.M.; Mouton, P.R.; Roy, S.; Wynshaw-Boris, A.; Colamarino, S.A.; Lein, E.S.; Courchesne, E. Patches of disorganization in the neocortex of children with autism. *N. Engl. J. Med.* **2014**, *370*, 1209–1219.

MDPI AG
Klybeckstrasse 64
4057 Basel, Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
http://www.mdpi.com/

Genes Editorial Office
E-mail: genes@mdpi.com
http://www.mdpi.com/journal/genes

MDPI • Basel • Beijing • Wuhan
ISBN 978-3-03842-172-6
www.mdpi.com