Clinical implications of next-generation sequencing for cancer medicine

A. Shuen HBSc MD and W.D. Foulkes MBBS PhD*†*

KEY WORDS

Next-generation sequencing, cancer genome, personalized medicine

~

In 1976, Peter Nowell, following observations of cytogenetic heterogeneity in a population of cancerous cells, proposed that this genetic diversity could be explained by hypothesizing that these cells are subject to evolutionary forces [1]. He posited that clonal expansion of a malignant cell with acquired genetic instability could give rise to various subpopulations, with subsequent selection for phenotypes having a proliferative advantage. This insight led to his prediction that therapy may need to be individually tailored according to the unique genetic mosaic that constitutes a tumour. Now, more than 30 years later, with the advent of next-generation sequencing (NGS), we are at the cusp of realizing personalized medicine for cancer treatment.

It was 1960 when Nowell and Hungerford used cytogenetic studies to identify a translocation between chromosomes 9 and 22 in chronic myelogenous leukemia (CML) cells [2]. More than two decades later, this translocation was discovered to result in a fusion protein, Bcr-Abl, a constitutively active tyrosine kinase driving cellular proliferation [3]. However, it was not until 1996 that a drug, imatinib, was found that could specifically bind to Bcr-Abl and inhibit its function [4]. Another 5 years elapsed before imatinib was finally approved for the treatment of CML, 41 years after Nowell's original discovery [5].

The remarkable therapeutic response of CML to its inhibitor heralded a quest for additional cancer-causing mutations as potential therapeutic targets. Since then, about 400 cancer-causing genes involved in cell proliferation, DNA repair, apoptotic signalling, angiogenesis, and metastasis have been identified. Nonetheless, our picture of the genomic landscape of cancer remains far from complete. Past efforts at identifying cancer-causing genes relied on hypothesis-driven experiments and the low-throughput Sanger method of sequencing DNA, limiting the ability to comprehensively detect cancer-causing mutations. It was not until advances were made in nanotechnology, biochemistry, imaging, and bioinformatics that second-generation massively parallel sequencing made screening for such mutations possible at reasonable cost and within a reasonable time period.

Unlike Sanger sequencing, NGS circumvents the rate-limiting step of separating randomly terminated polymerase chain reaction amplicons by capillary pore electrophoresis. Conceptually, NGS technology is based on the simultaneous sequencing of millions of randomly cleaved fragments of the genome by reading the nucleotides as they are incorporated *in situ,* and computationally reconstructing the DNA fragments, effectively producing a nucleotide resolution of the genome. This method is particularly suited for cancer research because it allows the frequency of somatic mutations within a population of cancer cells to be determined (Figure 1).

To put the power of NGS into perspective, the Human Genome Project, based on Sanger sequencing, required 5 years and $300 million to sequence a single genome. Today, the same task can be achieved in a matter of weeks for a mere $4800 [6]. That said, it is expected that the cost and time required to sequence a genome will plummet still further as newer "third generation" sequencers are introduced in near future [7]. Although costs are still too prohibitive to allow NGS to be applied routinely in clinical practice, it will be only a matter of time before the long sought-after goal of the $1000 genome set by the U.S. National Institutes of Health will be achieved.

An exciting development in NGS technology is that it can be applied to efficiently sequence not only the genome, but also the transcriptome [8] (the total transcribed sequences of the genome) and the methylome [9] (the total methylated sequences of the genome), to further the understanding of cancer biology.

Since its introduction in 2007, NGS has already made impressive advances in identifying cancer genes. In a landmark paper, Ley *et al.* reported the first human cancer genome ever to be sequenced [10]. Those authors analyzed the genome of tumour and matched

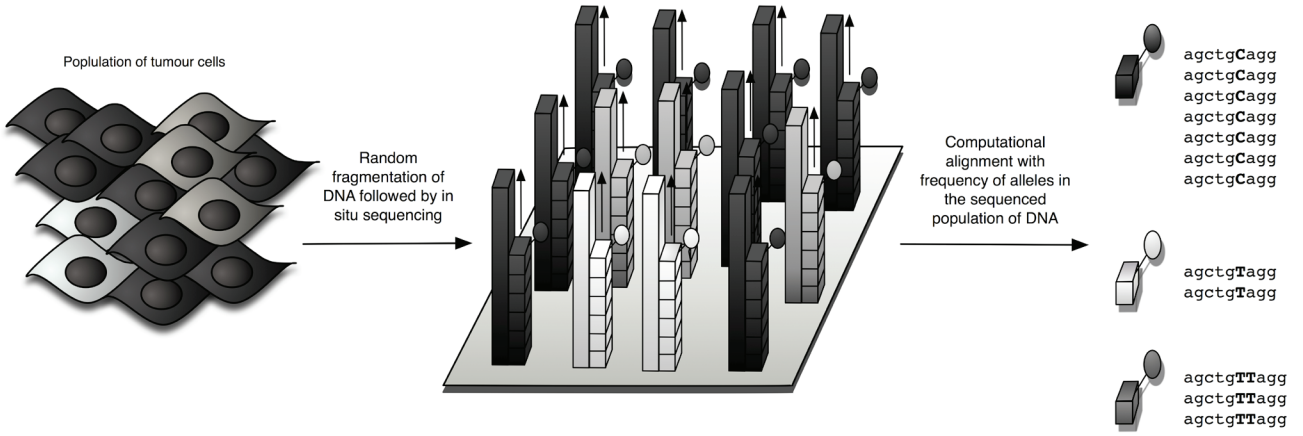

FIGURE 1 Genome sequencing of a tumour using next-generation sequencing. The genome of a heterogeneous population of tumour cells (represented by white, gray, and dark cells) is randomly fragmented. Each template is immobilized onto a solid support followed by 5′ to 3′ (vertical arrows) incorporation of a fluorescently labelled nucleotide. Simultaneously reading millions of fluorescent labels from each template, cleaving the fluorescent dye, and incorporating the next nucleotide yields the sequence of a random fragment of the genome. The final step, aligning the fragments, provides the full sequence and the relative frequency of the alleles.

normal skin cells from a patient with acute myeloid leukemia (AML). Comparing the genome of the tumour against that of the patient, they counted 63,000 somatic mutations, of which only 10 were confirmed to be non-synonymous changes in coding regions. Two of those genes, *FLT3* and *NPM1,* were already known to be associated with AML and are thought to contribute to cancer progression. However, the other 8 have never been previously described in AML.

The foregoing example underscores the ability of NGS to objectively and comprehensively identify potential "driver mutations" (genetic alterations that provide the cell with a survival advantage), while filtering out "passengers" (mutations that can be expected to have no effect on cell survival). Recurrence of these candidate driver mutations in other AML genomes, or even in other types of cancers, would suggest that they are being selected for and likely contributing to disease pathogenesis. Such discoveries potentially require that a great number, possibly thousands, of AML genomes be sequenced to obtain the statistical power necessary to differentiate drivers from passengers. Ultimately, functional studies of these mutations from cell and animal models will need to be performed to understand their true significance.

Sequencing a primary tumour yields only a "snapshot" of its genetic composition at one point in time. To fully understand how a population of tumour cells evolves, its genome must be sequenced at various stages in its progression. Shah *et al.* recently described the genome of a metastatic lobular breast tumour [11]. Filtering first against the reference human genome for aberrations in coding regions, and then filtering again for germline mutations, these authors were left with 30 validated candidate driver mutations. They then inspected the primary tumour (which had been removed 9 years earlier) for these 30 mutations and found that only 11 were present: 5 were shown to be prevalent, and 6 were rare (ranging in frequency from 1% to 13%) in the population of primary tumour cells. What is remarkable is that just a few mutations—*ABCB11, HAUS3, SLC24A4, SNX4,* and *PALB2*—in key pathways are enough to deregulate normal cell processes and transform cells into tumours. Interestingly, of those 5 genes, 2 are involved in genome integrity. The Haus3 protein is involved in centrosome and spindle stability [12], and Palb2 is involved in homologous recombination repair [13]. These results show that tumour evolution is associated with an increasing mutational burden; however, it is not known whether these changes reflect innate genetic instability or resulted as a consequence of radiation therapy.

In contrast with Shah's study, a recent analysis of the genome of a pre-treatment basal-like breast cancer and the subsequent cerebellar metastasis after chemotherapy revealed greater genetic heterogeneity in the primary tumour than in the metastasis, suggesting that selective pressures were at work on a subpopulation, enriching for mutations in the metastatic disease [14].

The effectiveness of NGS in systematically identifying the spectrum of genetic alterations in cancer, including single-nucleotide variants, indels, copy number variations, and large genomic rearrangements, has opened the doors for genome centres around the world to begin sequencing all the various types of cancers. The complete genomes of acute myeloid leukemia [10,15], breast cancer [11,14], lung cancer [16,17], melanoma [18], glioblastoma [19], and mesothelioma [20] have already been described (Table I). Many more cancer genomes will be uncovered as members of the International Cancer Genome Consortium seek to catalog more than 50 different cancers [21].

TABLE I Summary of sequenced cancer genomes

Cancer	*Source*	*Platform*	*SNV*	*CNV*	*GR*	*Highlight of genetic findings*
AML M1 [10]	Primary tumour	Illumina[a]	10	0	0	Confirmation of 2 recurrent somatic mutations in *FLT3* and *NPM1*.
AML M1 [15]	Primary tumour	Illumina	12	0	0	*IDH1* mutations present in 16% of AML genomes.
Lobular breast cancer [11]	Lung metastasis	Illumina	32	NA	NA	RNA editing events contribute to the transcriptional variation of lobular breast cancer. Only 5 mutations (in *ABCB11, HAUS3, SLC24A4, SNX4,* and *PALB2*) were prevalent in both the primary and the metastatic tumour.
Basal-like breast cancer [14]	Primary tumour Brain metastasis Mouse xenograft	Illumina	200 225 328	155 101 97	34 34 34	Deletion of *CTNNA1* present in all 3 tumour samples.
Small-cell lung cancer [16]	NCI H209 cell line	SOLiD[b]	134	334	58	Recurrent rearrangement of *CHD7* found in 2 other small-cell lung cancer cell lines.
Non-small-cell lung cancer [17]	Primary tumour	Complete Genomics[c]	530	ND	43	Mutational imprint left by tobacco carcinogens.
Malignant melanoma [18]	COLO-829 cell line	Illumina	292	41	51	Mutational imprint left by ultraviolet damage.
Glioblastoma [19,d]	Primary tumours (n=21)	SOLiD	993	281	NA	Mutations in *IDH1* associated with poorer survival found in 12% of patients.
Malignant pleural mesothelioma [20]	Primary tumour	Illumina and 454 sequencing[e]	44	14	30	*DPP10* deletion identified. Resequencing of this deletion in 53 tumour samples showed that loss of *DPP10* is associated with poorer survival.

[a] Illumina, San Diego, CA, U.S.A.
[b] Applied Biosystems, Carlsbad, CA, U.S.A.
[c] Complete Genomics, Mountain View, CA, U.S.A.
[d] Exome sequencing was performed.
[e] 454 Life Sciences (a Roche company), Branford, CT, U.S.A.
SNV = single-nucleotide variants in coding regions; CNV = copy number variant; GR = genomic rearrangement; AML = acute myeloid leukemia; NA = not assessed; NCI = U.S. National Cancer Institute; ND = not described.

With the recent introduction of NGS, an explosion of information will occur as increasing numbers of cancer genomes, transcriptomes, and methylomes become unravelled. The likely medium-term effect of integrating these data in practice will be to more accurately diagnose, predict, and monitor treatment outcome. However, the ultimate and far more challenging goal will be to determine the effect that these mutations have at a functional level.

The hope is that a complete molecular understanding of cancer will provide the tools necessary to develop novel biomarkers for screening and drugs that target relevant cellular pathways. We anticipate that NGS will be used routinely to contribute clinically useful information to the diagnosis and monitoring of cancer. With the $1000 genome tantalizingly close at hand, we could be just a few base pairs away from making personalized cancer medicine a reality.

REFERENCES

1. Nowell PC. The clonal evolution of tumour cell populations. *Science* 1976;195:23–8.
2. Nowell PC, Hungerford D. A minute chromosome in chronic granulocytic leukemia. *Science* 1960;132:1497.
3. Ben-Neriah Y, Daley GQ, Mes–Masson AM, Witte ON, Baltimore D. The chronic myelogenous leukemia–specific P210 protein is the product of the *BCR/ABL* hybrid gene. *Science* 1986;233:212–14.
4. Buchdunger E, Zimmermann J, Mett H, *et al.* Inhibition of the Abl protein–tyrosine kinase *in vitro* and *in vivo* by a 2-phenylaminopyrimidine derivative. *Cancer Res* 1996;56:100–4.
5. Savage DG, Antman KH. Imatinib mesylate—a new oral targeted therapy. *N Engl J Med* 2002;346:683–93.
6. Drmanac R, Sparks AB, Callow MJ, *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78–81.

7. Munroe DJ, Harris TJ. Third-generation sequencing fireworks at Marco Island. *Nat Biotech* 2010;28:426–8.
8. Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* 2010;464:1012–17.
9. Flusberg BA, Webster DR, Lee JH, *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 2010;6:461–5.
10. Ley TJ, Mardis ER, Ding L, *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66–72.
11. Shah SP, Morin RD, Khattra J, *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009;461:809–13.
12. Lawo S, Bashkurov M, Mullin M, *et al.* Haus, the 8-subunit human augmin complex, regulates centrosome and spindle integrity. *Curr Biol* 2009;19:816–26.
13. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 2010;11:196–207.
14. Ding L, Ellis MJ, Li S, *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 2010;464: 999–1005.
15. Mardis ER, Ding L, Dooling D, *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009;361:1058–66.
16. Pleasance ED, Stephens PJ, O'Meara S, *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010;463:184–90.
17. Lee W, Jiang Z, Liu J. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;465:473–7.
18. Pleasance ED, Cheetham RK, Stephens PJ, *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010;463:191–6.
19. Parsons DW, Jones S, Zhang X, *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–12.
20. Bueno R, De Rienzo A, Dong L, *et al.* Second generation sequencing of the mesothelioma tumor genome. *PLoS One* 2010;5:e10612.
21. International Cancer Genome Consortium. International network of cancer genome projects. *Nature* 2010;464:993–8.

Correspondence to: William D. Foulkes, Department of Medical Genetics and Research Institute, McGill University Health Centre, Room L10-116, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4.
E-mail: william.foulkes@mcgill.ca

* Department of Medical Genetics, McGill University Health Centre, McGill University, Montreal, QC.
† Segal Cancer Centre and Lady Davis Institute, and the Program in Cancer Genetics, Departments of Oncology and Human Genetics, McGill University, Montreal, QC.