



Communication

Discovery of Pathogenic Variants Associated with Idiopathic Recurrent Pregnancy Loss Using Whole-Exome Sequencing

Jeong Yong Lee ¹, JaeWoo Moon ², Hae-Jin Hu ² , Chang Soo Ryu ¹, Eun Ju Ko ¹, Eun Hee Ahn ³, Young Ran Kim ³, Ji Hyang Kim ^{3,*} and Nam Keun Kim ^{1,*}

¹ Department of Biomedical Science, College of Life Science, CHA University, Seongnam 13488, Republic of Korea; smilee3625@naver.com (J.Y.L.); regis2040@nate.com (C.S.R.); ejko05@naver.com (E.J.K.)

² Endomics, Inc., Seongnam-si 13595, Republic of Korea; masicer@goendomics.com (J.M.); haejinhu@goendomics.com (H.-J.H.)

³ Department of Obstetrics and Gynecology, CHA Bundang Medical Center, School of Medicine, CHA University, Seongnam 13596, Republic of Korea; bestob@cha.ac.kr (E.H.A.); happyiran@cha.ac.kr (Y.R.K.)

* Correspondence: bin0902@chamc.co.kr (J.H.K.); nkkim@cha.ac.kr (N.K.K.)

Abstract: Idiopathic recurrent pregnancy loss (RPL) is defined as at least two pregnancy losses before 20 weeks of gestation. Approximately 5% of pregnant couples experience idiopathic RPL, which is a heterogeneous disease with various causes including hormonal, chromosomal, and intrauterine abnormalities. Although how pregnancy loss occurs is still unknown, numerous biological factors are associated with the incidence of pregnancy loss, including genetic variants. Whole-exome sequencing (WES) was conducted on blood samples from 56 Korean patients with RPL and 40 healthy controls. The WES data were aligned by means of bioinformatic analysis, and the detected variants were annotated using machine learning tools to predict the pathogenicity of protein alterations. Each indicated variant was confirmed using Sanger sequencing. A replication study was also conducted in 112 patients and 114 controls. The Variant Effect Scoring Tool, Combined Annotation Dependent Depletion tool, Sorting Intolerant from Tolerant annotation tool, and various databases detected 10 potential variants previously associated with spontaneous abortion genes in patients by means of a bioinformatic analysis of WES data. Several variants were detected in more than one patient. Interestingly, several of the detected genes were functionally clustered, including some with a secretory function (mucin 4; *MUC4*; rs200737893 G>A and hyaluronan-binding protein 2; *HABP2*; rs542838125 G>T), in which growth arrest-specific 2 Like 2 (*GAS2L2*; rs140842796 C>T) and dynamin 2 (*DNM2*; rs763894364 G>A) are functionally associated with cell protrusion and the cytoskeleton. ATP Binding Cassette Subfamily C Member 6 (*ABCC6*) was the only gene with two variants. *HABP2* (rs542838125 G>T), *MUC4* (rs200737893 G>A), and *GAS2L2* (rs140842796 C>T) were detected in only the patient group in the replication study. The combination of WES and machine learning tools is a useful method to detect potential variants associated with RPL. Using bioinformatic tools, we found 10 potential variants in 9 genes. WES data from patients are needed to better understand the causes of RPL.

Keywords: bioinformatics; recurrent spontaneous abortion; miscarriage



Citation: Lee, J.Y.; Moon, J.; Hu, H.-J.; Ryu, C.S.; Ko, E.J.; Ahn, E.H.; Kim, Y.R.; Kim, J.H.; Kim, N.K. Discovery of Pathogenic Variants Associated with Idiopathic Recurrent Pregnancy Loss Using Whole-Exome Sequencing. *Int. J. Mol. Sci.* **2024**, *25*, 5447. <https://doi.org/10.3390/ijms25105447>

Academic Editor: Giovanni Tossetta

Received: 17 April 2024

Revised: 8 May 2024

Accepted: 15 May 2024

Published: 17 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recurrent pregnancy loss (RPL) or recurrent spontaneous abortion (RSA) is defined as at least two consecutive failed pregnancies within 20 weeks of gestation [1,2]. Approximately 5% of pregnant couples experience RPL [1,3,4]. Several factors are associated with RPL, including chromosomal, uterine, and hormonal abnormalities, infection, immune responses, and environmental conditions. However, the etiology of the disease is unexplained in more than 50% of couples diagnosed with RPL.

Many recent studies have reported single-nucleotide polymorphisms associated with the pathogenicity of various diseases including RPL [5–7]. Candidate variants with known ontology have previously been characterized into molecular function, biological process, or cellular component categories. Currently, advanced technologies, such as next-generation sequencing (NGS) and chromosome microarray analysis, are used to discover potential disease-associated variants, including those for an RPL diagnosis [8,9]. Whole-exome sequencing is an NGS method that allows the screening of many variants and the discovery of potential novel human disease variants. Therefore, it is a useful tool to identify protein-encoding and associated pathogenic variants.

NGS technology has become a new paradigm for the diagnosis of genetic diseases [10]. Advanced NGS technologies include various methods developed to screen pathogenic variants according to the effects of amino acid changes. Although predicting causal variants is difficult, various methods exist utilizing computer programs or machine learning algorithms which typically predict structural results after sequence alteration or analyze known phenotype data after alterations. In this study, we used four tools to predict the effects of amino acid changes, including Combined Annotation Dependent Depletion (CADD) [11,12], Polymorphism Phenotyping v2 (Polyphen-2) [13], Sorting Intolerant from Tolerant (SIFT) [14], and the Variant Effect Scoring Tool (VEST) [15,16]. CADD is a tool used to score the deleteriousness of single-nucleotide and insertion/deletion (indel) variants in the human genome [11]. PolyPhen-2 is a tool that predicts the possible impact of an amino acid substitution on the structure and function of a human protein [13]. The SIFT algorithm uses sequence homology to predict how an amino acid substitution will affect protein function and potentially alter its phenotype [14]. VEST is a random forest machine learning algorithm used to identify functional missense mutations and classify them into grades [16].

In this study, we analyzed whole-exome data from 56 patients with RPL and 40 controls to screen potential rare variants associated with RPL. After annotating protein-coding variants, we filtered the variants using various bioinformatics tools and identified RPL-associated genes using open databases. According to the bioinformatic results, we conducted a replication study, using Genotyping by Taq-man assay and SNaPshot Multiplex assay, with a large sample size including 112 patients and 114 controls, and discovered three potential pathogenic variants found only in patients.

2. Results

NGS methods were used to identify potential pathogenic variants in samples from 56 RPL patients and 40 controls (Figure 1). All of the selected genes were rare variants (minor allele frequency < 0.01) with the cutoff criteria of CADD under 20, VEST p-value under 0.05, SIFT > 0.05, and a PolyPhen v2 of “possibly damaging”. After filtering, we confirmed the variant allele frequency to identify false positives with a cutoff value of 30%. We screened 10 variants and 9 genes, and all variants were heterozygous (Table 1). The filtered variants consisted of eight missense variants and two stop-gain variants; therefore, Chr3:195488449 and Chr 20:48130848 did not have SIFT and PolyPhen-2 scores. The Chr16:16276686, rs758166222 A>T variant was detected in two samples. Two variants in the *ABCC6* gene (Chr16:16276686 and Chr16:16282707) were detected in three samples. The *MUC4*, Chr3:195488449 variant had the highest CADD score (PHRED-like score = 51) and the lowest VEST p-value (0.006). However, because this was a stop-gain variant, it had no PolyPhen-2 or SIFT score. The Chr19:10897280 rs763894364 G>A variant had the lowest minor allele frequency (0.0005) in the Korea1K database.

We conducted a replication study of 10 rare variants previously associated with RPL, using the RPL and control samples (Table 2). As a result, only the patients of group 2 (rs200737894 G>A, rs542838125 G>T, rs140842796 C>T) showed continuous bioinformatics, but other variants were detected in the controls.

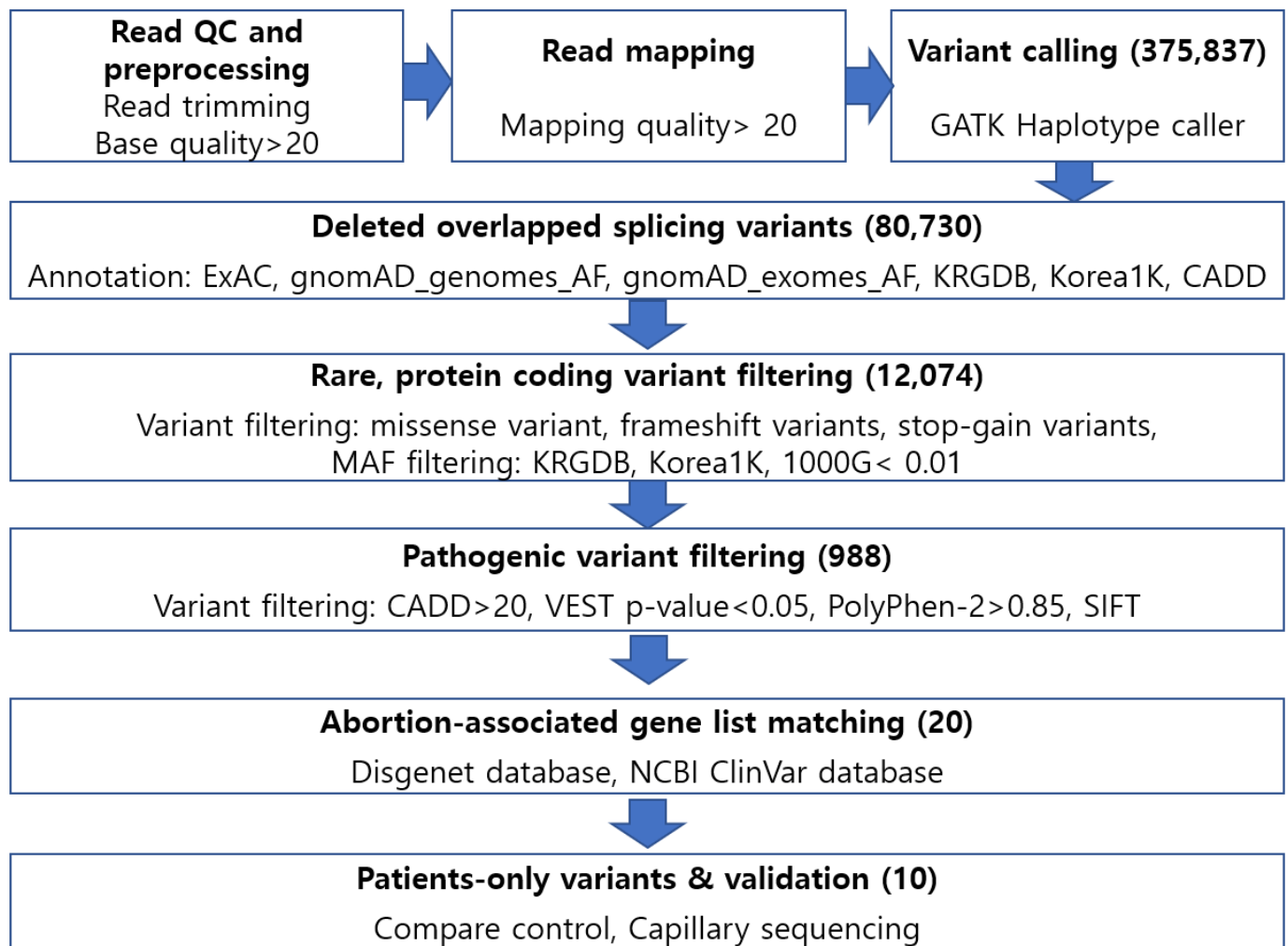


Figure 1. Whole-exome sequencing analysis pipeline. An abortion-associated, matching gene list was filtered using the Disgenet and ClinVar databases and the results were compared with those in the control sample to select patient-specific variants. Combined Annotation Dependent Depletion: CADD; Variant Effect Scoring Tool: VEST; Sorting Intolerant from Tolerant: SIFT.

Table 1. List of potential recurrent pregnancy loss candidate genes and information in 56 patients and 40 controls.

Position	rsID	Gene	Transcript_ID	HGVS.c	HGVS.p	MAF frequency							Pathogenicity				Detected Individual Count
						Replication Study	1000 Genomes	KRGDB	KOREA1K	gnomAD Genomes_AF	VEST <i>p</i> -Value	VEST Missense	VEST Stop-Gain	CADD	SIFT	Polyphen2	
chr2 233274367	rs575918099	ALPG	NM_031313.3	c.1384G>A	p.G462S	.	0.0002	0.0038	0.0027	.	0.0170	0.775	.	24.4	0	1	1
chr2 238283646	rs116238578	COL6A3	NM_057165.5	c.2470G>A	p.V824M	0.01327	0.0016	0.0013	0.0087	0.0003	0.0207	0.737	.	23.8	0	0.995	1
chr3 195488449	rs200737893	MUC4	NM_138297.5	c.1659G>A	p.W553 *	0.00443	0.0004	0.0013	0.0027	0.0000	0.0060	.	0.634	51	N/A	N/A	1
chr10 115334164	rs542838125	HABP2	NM_001177660.2	c.145G>T	p.D49Y	0.00443	.	0.0013	0.0005	0.0000	0.0150	0.799	.	32	0	0.987	1
chr16 16276686	rs758166222	ABCC6	NM_001171.5	c.2045A>T	p.K682M	.	.	0.0010	0.0011	.	0.0145	0.805	.	24.5	0	1	2
chr16 16282707	rs527236047	ABCC6	NM_001171.5	c.1760C>G	p.S587C	0.01327	.	0.0025	0.0022	0.0000	0.0398	0.613	.	24	0.03	0.979	1
chr17 34074883	rs140842796	GAS2L2	NM_139285.4	c.817C>T	p.R273C	0.00885	0.0018	0.0013	0.0044	0.0004	0.0075	0.93	.	28.7	0	1	1
chr19 10897280	rs763894364	DNM2	NM_001190716.2	c.890G>A	p.R297H	.	0.0002	0.0013	0.0005	0.0000	0.0129	0.827	.	28.6	0	0.942	1
chr20 48130848	rs13306027	PTGIS	NM_000961.4	c.940G>T	p.E314 *	0.00443	.	0.0013	0.0011	0.0000	0.0177	.	0.569	33	N/A	N/A	1
chr20 52789538	rs114476330	CYP24A1	NM_001128915.2	c.359G>A	p.R120H	.	0.0002	0.0013	0.0033	0.0000	0.0159	0.788	.	32	0	1	1

VEST analysis generates values between 0 and 1, and VEST *p*-values ≥ 0.5 were classified as pathogenic. Scores ≥ 0.5 were classified as pathogenic and those < 0.5 as benign. CADD analysis generates a PHRED-like scaled value. Scores ≥ 20 were classified as pathogenic variants and those < 20 as benign. SIFT analysis generates values between 0 and 1, and scores ≥ 0.5 were classified as pathogenic. Polyphen2 analysis generates value between 0 and 1, and scores are matched with the RPL-associated gene list. * is stop-gain alteration that indicate ‘stop’ after alteration of nucleotide.

Table 2. Replication of rare variants in 112 patients and 114 controls.

Groups	Gene (rsID) (# of Patients/# of Controls)			
Group 1	ABCC6 (rs758166222) 0/0	DNM2 (rs763894364) 0/0	ALPG (rs575918099) 0/0	CYP24A1 (rs114476330) 0/0
Group 2	MUC4 (rs200737893) 1/0	HABP2 (rs542838125) 1/0	GAS2L2 (rs140842796) 2/0	
Group 3	ABCC6 (rs527236047) 0/3	PTGIS (rs13306027) 0/1		
Group 4	COL6A3 (rs116238578) 2/1			

Group 1: Not detected in both controls and patients. Group 2: Detected only in patients and not in controls. Group 3: Detected only in controls and not in patients. Group 4: Detected in both controls and patients.

3. Discussion

In this study, we recruited 96 participants, including 56 RPL patients and 40 controls, and conducted whole-exome sequencing to screen potential causal variants for RPL. Ten variants were selected as potentially pathogenic variants that were previously reported with RPL-related genes. Among these genes, Chr16:16276686 (rs758166222 A>T; *ABCC6*; p.K682M) was the only variant found in two individuals, and the *ABCC6* gene was the only gene for which two variants were detected (Chr16:16276686 and Chr16:16282707) (Table 1). The *ABCC6* gene encodes ATP binding cassette subfamily C member 6, which is associated with transporter activity, and is a known causal gene of pseudoxanthoma elasticum [17,18]. Previous reports have described pseudoxanthoma elasticum as an autosomal recessive disorder that causes skin abnormalities. Additionally, it can occur after various pregnancy complications and result in miscarriage [18,19]. Among the nine detected genes, the variants found in our group were different from those reported previously in the same genes. Mucin 4 belongs to the *MUC* gene series (*MUC1* to *22*) and is associated with the endometrial epithelium. It has also been reported to be involved in recurrent implantation failure [20].

In our previous report, the rs882605 G>T and rs1104760A>G variants were associated with a subgroup of patients who experienced pregnancy loss more than twice [21]. Alkaline phosphatase, germ cell (*ALPG*), also known as Alkaline phosphatase, placental like-2 (*ALPPL2*), is highly involved in trophoblast expression in the human placenta and is associated with early human development and the first cellular differentiation in the inner cell mass. A single-nucleotide polymorphism of this gene (rs11678251) is associated with placental reduction [22,23]. Prostaglandin I2 synthase (*PTGIS*) is involved in fetal development because prostaglandin synthesis is important in early pregnancy to establish implantation and placentation. Additionally, the overexpression of *PGE2* and other inflammatory cytokines in endometrial tissue results in a risk of spontaneous abortion [24]. Dynamin 2 (*DNM2*) regulates the invagination and constriction of vesicles at the membrane in endothelial and epithelial cells. Additionally, *DNM2* GTPase is ubiquitously expressed and regulates riboflavin absorption in the human placental trophoblast model cell line BeWo [25]. *CYP24A1* encodes vitamin D-24-hydroxylase, which is very important clinically and physiologically because it regulates vitamin D metabolism [26]. *CYP24A1* can be stimulated hormonally, and its mutation results in hypercalcemia in pregnancy [26]. In a previous study, the placental and decidual expression of *CYP24A1* was increased in patients with spontaneous miscarriage compared with that in healthy controls [27].

In the replication study, variants in the hyaluronan-binding protein 2 (*HABP2*), *COL6A3*, and *GAS2L2* genes were only detected in the patient group. *HABP2* and *COL6A3* are associated with cell adhesion, especially *HABP2*, which was recently reported to be associated with recurrent implantation failure [28]. Reduced *HABP2* expression in the endometrium results in a decrease in endometrium receptivity and in vitro fertilization failure [28,29].

COL6A3 is a collagen VI-encoding gene that is an important component of the extracellular matrix and is usually found in interstitial tissue such as muscle, skin, and tendons [30]. An abnormally increased deposition of collagen disrupts uterine function, resulting in interrupted vascularization and delayed remodeling during early pregnancy [31]. The growth arrest-specific 2-like 2 (*GAS2L2*) protein is localized in actin stress fibers and microtubules, and the absence of *GAS2L2* in mice resulted in neonatal death [32]. It is also associated with a causal gene of primary ciliary dyskinesia, which is associated with miscarriage [33]. The rs758166222 A>T (*ABCC6*), rs763894364 G>A (*DNM2*), rs575918099 G>A (*ALPG*), and rs114476330 G>A (*CYP24A1*) variants were not detected in either patient or control groups. Thus, additional patients and controls are needed to examine these four variants.

We discovered possibly causative rare variants by means of whole-exome sequencing analysis in 56 non-related Korean RPL patients and 40 controls. In the process, we used bioinformatics tools and population data and these tools helped us find rare pathogenic variants. The prediction of disease risk using variants is difficult, but using whole-exome sequencing and bioinformatics tools is a good approach to finding pathogenic variants.

4. Materials and Methods

4.1. Subjects

The study group consisted of 168 women with at least two consecutive pregnancy losses before 20 weeks of gestation, who were diagnosed with idiopathic RPL according to the definition of the American Society for Reproductive Medicine, and 154 healthy controls. The patients were enrolled in the study at the Infertility Medical Center of CHA Bundang Medical Center from March 1999 to February 2010. Women in the control group were recruited from CHA Bundang Hospital and met the following criteria: history of at least one spontaneous pregnancy; current pregnancy; regular menstrual cycles; karyotype 46, XX; and no history of miscarriage. The study was approved by the Institutional Review Board of CHA Bundang Medical Center (IRB approval no. BD2010-123D), and written informed consent was obtained from all participants. Pregnancy loss was diagnosed based on the results of human chorionic gonadotropin tests, ultrasound, and/or physical examination before 20 weeks of gestation. None of the participants had a history of smoking or alcohol use. The following parameters were also measured using participant blood samples: activated partial thromboplastin time, body mass index, blood urea nitrogen, creatinine, estradiol, follicle-stimulating hormone, luteinizing hormone, platelet count, and prothrombin time.

Patients with the following conditions were excluded from the study: RPL or implantation failure due to hormonal, genetic, anatomic, infectious, autoimmune, or thrombotic causes. Anatomic causes were evaluated using hysterosalpingogram, hysteroscopy, computed tomography, and magnetic resonance imaging to detect intrauterine adhesions, septate uterus, and uterine fibroids. Hormonal causes, including hyperprolactinemia, luteal insufficiency, and thyroid disease, were evaluated using blood analyses. Infectious causes, such as the presence of *Ureaplasma urealyticum* or *Mycoplasma hominis*, were evaluated by means of bacterial culture. Autoimmune causes, including antiphospholipid syndrome or lupus, were evaluated using lupus anticoagulant and anticardiolipin antibodies. Thrombotic causes, such as thrombophilia, were evaluated through the identification of protein C and S deficiencies and the detection of β -2-glycoprotein 1 antibodies.

4.2. Whole-Exome Sequencing

Whole-exome sequencing was conducted on 56 patients and 40 controls. Genomic DNA was extracted from anticoagulated peripheral blood using a G-DEX blood extraction kit (Intron). A minimum of 500 ng of genomic DNA per sample was used to construct sequencing libraries. Whole-exome sequencing was conducted by MacroGen, Inc. SureSelect V4+UTR-post libraries were constructed according to the manufacturer's recommendations (Agilent Technologies, Santa Clara, CA, USA) and were sequenced using a Novaseq system according to the manufacturer's suggested protocol (Illumina Inc., San Diego, CA, USA).

Trim_galore (version 0.6.4) was used to assess the quality of reads. The read sequences were aligned to the human reference genome (GRCh37/hg19) using BWA (version 0.7.17-r1188). The sorting and indexing of BAM files were performed using SAMtools, and the BaseRecalibrator tool of GATK (version 4.3.0) and dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>) (accessed on 21 September 2022) were used to annotate functionality, single-nucleotide variants, and deduplication. Variant calling after GATK analysis and functional annotation was performed using SnpEff tools (version 5.1).

4.3. Annotation and Variant Filtering

Before screening for predicted pathogenic single-nucleotide variants, only protein-coding gene variants, such as nonsynonymous, splicing variant, stop-gain, stop-loss, frameshift, and indel mutants, were selected. Variants were filtered according to a minor allele frequency > 0.01 in the 1000 Genomes database [34,35], and Korean-specific rare variants were selected using the Korean Reference Genome Database (KRGDB) [36] and Korea1K [37]. To select RPL-related genes, we collected data from various databases, including Disgenet, ClinVar, and Monarch, and excluded non-RPL genes (Supplementary Table S1).

4.4. Additional Annotation for Potential Variants

Pathogenic prediction annotation was used to select potential pathogenic variants. Common variants (minor allele frequency > 0.01) were filtered using population databases including the KRGDB and Korea1K databases (Table 1). Additionally, pathogenic annotation was conducted using the CADD, VEST, SIFT, and PolyPhen-2 tools to select pathogenic variants. The CADD tool integrates information contained in diverse annotations into a single score and prioritizes functional and pathogenic variants [38]. Variants with a CADD PHRED-like score ≥ 20 are classified as pathogenic [11]. VEST is a computational tool for the identification of specific variants that contribute to human disease [16]. It characterizes variants into categories (missense, frameshift, indel, stop-gain, stop-loss, and splice site), and the pathogenicity of genes is integrated into a p -value (non-silent). Variants with scores ≥ 0.5 in each category and an integrated p -value of ≤ 0.05 are classified as pathogenic. These tools are machine learning-based and suggest a score for each variant, including nonsynonymous, indel, and splice variants, regardless of whether they are deleterious or pathogenic. The SIFT algorithm predicts the potential effects of amino acid changes on protein function (<http://sift-dna.org>) (accessed on 31 March 2022). It characterizes missense and indel variants using a score ranging from 0.0 to 1.0. Variants with a score close to 0.00 are deleterious, but variants with a score ≥ 0.05 are predicted to be tolerated (benign) [14,39]. The PolyPhen-2 tool predicts the effects of amino acid substitutions on protein structure and function (<http://genetics.bwh.harvard.edu/pph2/>) (accessed on 21 June 2021). The score ranges from 0.0 to 1.0 which are classified into three categories: benign < 0.15; possibly damaging = 0.15–1.0; and probably damaging > 0.85 [13]. Function-related gene analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [40,41], and string interaction network analysis [42].

4.5. Validation

Variants identified using whole-exome sequencing that were found in both controls and patients were excluded. Before capillary sequencing, we calculated variant allele frequency and excluded false positives with a cutoff of 30%. We validated candidate variants by means of capillary sequencing, using an Applied Biosystems 3730xl DNA analyzer. All primers were designed using Primer3Plus web software (<https://www.primer3plus.com/primer3plusAbout.html>) (ver 3.3.0) [43]. Each primer sequence is shown in Supplementary Table S2. Data obtained by means of capillary sequencing were analyzed using Chromas (version 2.6.6 Technelysium Pty Ltd., Brisbane, Australia).

4.6. Replication of Patient and Control Samples

After the annotation and filtering of whole-exome sequencing data, to see if this result could also be verified in a larger sample size, a replication study was conducted using RPL and control samples (112 patients and 114 controls). Genotyping was performed using the SNaPshot multiplex system for rs140842796 C>T, rs13306027 G>T, and rs114476330 G>A; a Taqman assay for rs575918099 G>A, rs200737893 G>A, rs542838125 G>T, rs758166222 A>T, rs527236047 C>G, and rs763894364 G>A; and PCR-restriction fragment length polymorphism for rs116238578 G>A.

5. Conclusions

This is the first report of a case–control study utilizing bioinformatic tools. Most studies, especially case–control studies, target specific already known genes, SNPs, or pathways; however, we used whole-exome sequencing data and discovered novel potential pathogenic variants. These results provide a strategy for discovering novel pathogenic variants. Idiopathic RPL is a heterogeneous disorder that is affected by various factors. We conducted whole-exome sequencing of 56 patients and discovered 10 variants according to population frequency, using various annotation tools including CADD, VEST, SIFT, and PolyPhen-2, that were previously reported to be associated with RPL. In a replication study, only the *HABP2*, *GAS2L2*, and *MUC4* variants were detected in the patient group. To increase the accuracy of our analysis, additional population information should be collected and validated using databases, and the study should be replicated with additional patients and controls.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25105447/s1>.

Author Contributions: Conceptualization, N.K.K. and J.Y.L.; methodology, J.Y.L.; software, J.Y.L., J.M., H.-J.H.; validation, J.Y.L., C.S.R. and E.J.K.; formal analysis, J.Y.L.; investigation, J.Y.L.; resources, E.H.A., J.H.K.; data curation, Y.R.K.; writing—original draft preparation, J.Y.L.; writing—review and editing, J.Y.L.; visualization, J.Y.L.; supervision, N.K.K.; project administration, N.K.K.; funding acquisition, N.K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HR22C1605). This research was supported by National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT; grant numbers: 2022R1F1A1064169).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the CHA Bundang Medical Center (IRB number: 2010-01-123).

Informed Consent Statement: Informed con-sent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to another publication and personal information, but are available from the corresponding author on reasonable request.

Acknowledgments: This study was conducted with bioresources from the National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (NBK-2022-093). Korean Genome Project data were provided by the Korean Genomics Center at the Ulsan National Institute of Science and Technology, Republic of Korea.

Conflicts of Interest: JaeWoo Moon and Hae-Jin Hu were employed by Endomics, Inc. The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Practice Committee of the American Society for Reproductive Medicine. Definitions of infertility and recurrent pregnancy loss: A committee opinion. *Fertil. Steril.* **2013**, *99*, 63. [\[CrossRef\]](#) [\[PubMed\]](#)
- El Hachem, H.; Crepau, V.; May-Panloup, P.; Descamps, P.; Legendre, G.; Bouet, P.E. Recurrent pregnancy loss: Current perspectives. *Int. J. Womens Health* **2017**, *9*, 331–345. [\[CrossRef\]](#)
- Stirrat, G.M. Recurrent miscarriage I: Definition and epidemiology. *Lancet* **1990**, *336*, 673–675. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stirrat, G.M. Recurrent miscarriage II: Clinical associations, causes, and management. *Lancet* **1990**, *336*, 728–733. [\[CrossRef\]](#)
- Lee, J.Y.; Ahn, E.H.; Kim, J.O.; Park, H.S.; Ryu, C.S.; Kim, J.H.; Kim, Y.R.; Lee, W.S.; Kim, N.K. Associations between microRNA (miR-25, miR-32, miR-125, and miR-222) polymorphisms and recurrent implantation failure in Korean women. *Hum. Genom.* **2019**, *13*, 68. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cho, S.H.; Kim, J.H.; Park, H.W.; Park, H.S.; An, H.J.; Kim, Y.R.; Ahn, E.H.; Lee, W.S.; Kim, N.K. Associations between HOTAIR polymorphisms rs4759314, rs920778, rs1899663, and rs7958904 and risk of primary ovarian insufficiency in Korean women. *Maturitas* **2021**, *144*, 74–80. [\[CrossRef\]](#)
- Hu, Y.; Liu, C.M.; Qi, L.; He, T.Z.; Shi-Guo, L.; Hao, C.J.; Cui, Y.; Zhang, N.; Xia, H.F.; Ma, X. Two common SNPs in pri-miR-125a alter the mature miRNA expression and associate with recurrent pregnancy loss in a Han-Chinese population. *RNA Biol* **2011**, *8*, 861–872. [\[CrossRef\]](#) [\[PubMed\]](#)
- Qiao, Y.; Wen, J.; Tang, F.; Martell, S.; Shomer, N.; Leung, P.C.K.; Stephenson, M.D.; Rajcan-Separovic, E. Whole exome sequencing in recurrent early pregnancy loss. *MHR Basic Sci. Reprod. Med.* **2016**, *22*, 364–372. [\[CrossRef\]](#) [\[PubMed\]](#)
- Quintero-Ronderos, P.; Laissue, P. Genetic Variants Contributing to Early Recurrent Pregnancy Loss Etiology Identified by Sequencing Approaches. *Reprod. Sci.* **2020**, *27*, 1541–1552. [\[CrossRef\]](#)
- Sawyer, S.L.; Hartley, T.; Dymont, D.A.; Beaulieu, C.L.; Schwartzentruber, J.; Smith, A.; Bedford, H.M.; Bernard, G.; Bernier, F.P.; Brais, B.; et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: Time to address gaps in care. *Clin. Genet.* **2016**, *89*, 275–284. [\[CrossRef\]](#)
- Rentsch, P.; Schubach, M.; Shendure, J.; Kircher, M. CADD-Splice—Improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **2021**, *13*, 31. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schubach, M.; Maass, T.; Nazaretyan, L.; Röner, S.; Kircher, M. CADD v1.7: Using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Res.* **2024**, *52*, D1143–D1154. [\[CrossRef\]](#)
- Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [\[CrossRef\]](#)
- Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [\[CrossRef\]](#)
- Douville, C.; Masica, D.L.; Stenson, P.D.; Cooper, D.N.; Gyax, D.M.; Kim, R.; Ryan, M.; Karchin, R. Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-Indel). *Hum. Mutat.* **2016**, *37*, 28–35. [\[CrossRef\]](#)
- Carter, H.; Douville, C.; Stenson, P.D.; Cooper, D.N.; Karchin, R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. *BMC Genom.* **2013**, *14*, S3. [\[CrossRef\]](#) [\[PubMed\]](#)
- Combrinck, M.; Gilbert, J.D.; Byard, R.W. Pseudoxanthoma Elasticum and Sudden Death. *J. Forensic Sci.* **2011**, *56*, 418–422. [\[CrossRef\]](#) [\[PubMed\]](#)
- Camacho, M.; Rengel, C.; López-Herrero, E.; Carrillo, J.L.; Eslava, A.J.; Valdivielso, P. Approach to the management of pregnancy in patients with pseudoxanthoma elasticum: A review. *J. Obstet. Gynaecol.* **2016**, *36*, 1061–1066. [\[CrossRef\]](#)
- Bercovitch, L.; Leroux, T.; Terry, S.; Weinstock, M.A. Pregnancy and obstetrical outcomes in pseudoxanthoma elasticum. *Br. J. Dermatol.* **2004**, *151*, 1011–1018. [\[CrossRef\]](#)
- Koscinski, I.; Viville, S.; Porchet, N.; Bernigaud, A.; Escande, F.; Defosse, A.; Buisine, M.-P. MUC4 gene polymorphism and expression in women with implantation failure. *Hum. Reprod.* **2006**, *21*, 2238–2245. [\[CrossRef\]](#)
- Kim, J.-H.; Park, H.-S.; Lee, J.-Y.; Ko, E.-J.; Kim, Y.-R.; Cho, H.-Y.; Lee, W.-S.; Ahn, E.-H.; Kim, N.-K. Association Study between Mucin 4 (MUC4) Polymorphisms and Idiopathic Recurrent Pregnancy Loss in a Korean Population. *Genes* **2022**, *13*, 937. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kikas, T.; Rull, K.; Beaumont, R.N.; Freathy, R.M.; Laan, M. The Effect of Genetic Variation on the Placental Transcriptome in Humans. *Front. Genet.* **2019**, *10*, 550. [\[CrossRef\]](#)
- Bai, Q.; Assou, S.; Haouzi, D.; Ramirez, J.-M.; Monzo, C.; Becker, F.; Gerbal-Chaloin, S.; Hamamah, S.; De Vos, J. Dissecting the First Transcriptional Divergence During Human Embryonic Development. *Stem Cell Rev. Rep.* **2012**, *8*, 150–162. [\[CrossRef\]](#)
- Velez Edwards, D.R.; Edwards, T.L.; Bray, M.J.; Torstenson, E.; Jones, S.; Shrubsole, M.J.; Muff, H.J.; Hartmann, K.E. Nonsteroidal Anti-inflammatory Drug Interaction with Prostacyclin Synthase Protects from Miscarriage. *Sci. Rep.* **2017**, *7*, 9874. [\[CrossRef\]](#)
- Foraker, A.B.; Ray, A.; Da Silva, T.C.; Bareford, L.M.; Hillgren, K.M.; Schmittgen, T.D.; Swaan, P.W. Dynamin 2 Regulates Riboflavin Endocytosis in Human Placental Trophoblasts. *Mol. Pharmacol.* **2007**, *72*, 553–562. [\[CrossRef\]](#)
- Carpenter, T.O. CYP24A1 loss of function: Clinical phenotype of monoallelic and biallelic mutations. *J. Steroid Biochem. Mol. Biol.* **2017**, *173*, 337–340. [\[CrossRef\]](#)
- Hou, H.; Zhang, J.Y.; Chen, D.; Deng, F.; Morse, A.N.; Qiu, X.; He, P.; Lash, G.E. Altered decidual and placental catabolism of vitamin D may contribute to the aetiology of spontaneous miscarriage. *Placenta* **2020**, *92*, 1–8. [\[CrossRef\]](#)

28. Pathare, A.D.S.; Hinduja, I. Endometrial Expression of Cell Adhesion Genes in Recurrent Implantation Failure Patients in Ongoing IVF Cycle. *Reprod. Sci.* **2022**, *29*, 513–523. [[CrossRef](#)] [[PubMed](#)]
29. Suhorutshenko, M.; Kukushkina, V.; Velthut-Meikas, A.; Altmäe, S.; Peters, M.; Mägi, R.; Krjutškov, K.; Koel, M.; Codoñer, F.M.; Martinez-Blanch, J.F.; et al. Endometrial receptivity revisited: Endometrial transcriptome adjusted for tissue cellular heterogeneity. *Hum. Reprod.* **2018**, *33*, 2074–2086. [[CrossRef](#)]
30. Bushby, K.M.D.; Collins, J.; Hicks, D. *Collagen Type VI Myopathies BT—Progress in Heritable Soft Connective Tissue Diseases*; Halper, J., Ed.; Springer: Dordrecht, The Netherlands, 2014; pp. 185–199, ISBN 978-94-007-7893-1.
31. Diao, H.; Aplin, J.D.; Xiao, S.; Chun, J.; Li, Z.; Chen, S.; Ye, X. Altered Spatiotemporal Expression of Collagen Types I, III, IV, and VI in Lpar3-Deficient Peri-Implantation Mouse Uterus1. *Biol. Reprod.* **2011**, *84*, 255–265. [[CrossRef](#)] [[PubMed](#)]
32. Bustamante-Marin, X.M.; Yin, W.-N.; Sears, P.R.; Werner, M.E.; Brotslaw, E.J.; Mitchell, B.J.; Jania, C.M.; Zeman, K.L.; Rogers, T.D.; Herring, L.E.; et al. Lack of GAS2L2 Causes PCD by Impairing Cilia Orientation and Mucociliary Clearance. *Am. J. Hum. Genet.* **2019**, *104*, 229–245. [[CrossRef](#)] [[PubMed](#)]
33. Vanaken, G.J.; Bassinet, L.; Boon, M.; Mani, R.; Honoré, I.; Papon, J.-F.; Cuppens, H.; Jaspers, M.; Lorent, N.; Coste, A.; et al. Infertility in an adult cohort with primary ciliary dyskinesia: Phenotype–gene association. *Eur. Respir. J.* **2017**, *50*, 1700314. [[CrossRef](#)] [[PubMed](#)]
34. Chen, S.; Francioli, L.C.; Goodrich, J.K.; Collins, R.L.; Kanai, M.; Wang, Q.; Alföldi, J.; Watts, N.A.; Vittal, C.; Gauthier, L.D.; et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **2024**, *625*, 92–100. [[CrossRef](#)] [[PubMed](#)]
35. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)] [[PubMed](#)]
36. Jung, K.S.; Hong, K.-W.; Jo, H.Y.; Choi, J.; Ban, H.-J.; Cho, S.B.; Chung, M. KRGDB: The large-scale variant database of 1722 Koreans based on whole genome sequencing. *Database* **2020**, *2020*, baz146. [[CrossRef](#)]
37. Jeon, S.; Bhak, Y.; Choi, Y.; Jeon, Y.; Kim, S.; Jang, J.; Jang, J.; Blazyte, A.; Kim, C.; Kim, Y.; et al. Korean Genome Project: 1094 Korean personal genomes with clinical information. *Sci. Adv.* **2023**, *6*, eaaz7835. [[CrossRef](#)]
38. Kircher, M.; Witten, D.M.; Jain, P.; O’Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310–315. [[CrossRef](#)]
39. Sim, N.-L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [[CrossRef](#)]
40. Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **2022**, *50*, W216–W221. [[CrossRef](#)]
41. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)]
42. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [[CrossRef](#)]
43. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.