Original Article

Identification of genetic polymorphisms in unexplained recurrent spontaneous abortion based on whole exome sequencing

Jiang-Tao Mou[1#], Shi-Xing Huang[2#], Li-Li Yu[3], Jing Xu[3], Qiao-Ling Deng[4], Yi-Shan Xie[1], Kun Deng[1]

[1]Department of Laboratory Medicine, The Third Affiliated Hospital of Chongqing Medical University, Chongqing, China; [2]Department of Obstetrics and Gynecology, Three Gorges Hospital affiliated to Chongqing University, Chongqing, China; [3]Department of Obstetrics and Gynecology, The Third Affiliated Hospital of Chongqing Medical University, Chongqing, China; [4]Department of Obstetrics and Gynecology, The Second People's Hospital of Jiulongpo District, Chongqing, China

Contributions: (I) Conception and design: D Kun, JT Mou; (II) Administrative support: D Kun; (III) Provision of study materials or patients: SX Huang, LL Yu, J Xu, QL Deng; (IV) Collection and assembly of data: JT Mou, YS Xie; (V) Data analysis and interpretation: JT Mou, SX Huang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Kun Deng. Department of Laboratory Medicine, The Third Affiliated Hospital of Chongqing Medical University, 1st Shuanghu Branch Road, Yubei District, Chongqing 401120, China. Email: dengkun@hospital.cqmu.edu.cn.

Background: The precise etiology of approximately 50% of patients with recurrent spontaneous abortion (RSA) is unclear, known as unexplained recurrent spontaneous abortion (URSA). This study identified the genetic polymorphisms in patients with URSA.

Methods: Genomic DNA was extracted from 30 couples with URSA and 9 couples with normal reproductive history for whole exome sequencing. Variations in annotation, filtering, and prediction of harmfulness and pathogenicity were examined. Furthermore, predictions of the effects of changes in protein structure, Sanger validation, and functional enrichment analyses were performed. The missense mutated genes with significant changes in protein function, and genes with mutations of premature stop, splice site, frameshift, and in-frame indel were selected as candidate mutated genes related to URSA.

Results: In 30 unrelated couples with URSA, 50%, 20%, and 30% had 2, 3, and more than 4 miscarriages, respectively. Totally, 971 maternal and 954 paternal mutations were found to be pathogenic or possibly pathogenic after preliminary filtering. Total variations were not associated with age nor the number of miscarriages. In 28 patients (involving 23 couples), 22 pathogenic or possibly pathogenic variants of 19 genes were found to be strongly associated with URSA, with an abnormality rate of 76.67%. Among these, 12 missense variants showed obvious changes in protein functions, including *ANXA5* (c.949G>C; p.G317R), *APP* (c.1530G>C; p.K510N), *DNMT1* (c.2626G>A; p.G876R), *FN1* (c.5621T>C; p.M1874T), *MSH2* (c.1168G>A; p.L390F), *THBS1* (c.2099A>G; p.N700S), *KDR* (c.2440G>A; p.D814N), *POLR2B* (c.406G>T; p.G136C), *ITGB1* (c.655T>C; p.Y219H), *PLK1* (c.1210G>T; p.A404S), *COL4A2* (c.4808 A>C; p.H1603P), and *LAMA4* (c.3158A>G; p.D1053G). Six other genes with mutations of premature stop, splice site, frameshift, and in-frame indel were also identified, including *BUB1B* (c.1648C>T; p.R550*) and *MMP2* (c.1462_1464delTTC; p.F488del) from the father, and mutations from mother and/or father including *BPTF* (c.396_398delGGA; p.E138 del and c.429_431GGA; p.E148del), *MECP2* (c.21_23delCGC; p.A7del), *LAMA2* (HGVS: NA; Exon: NA; SPLICE_SITE, DONOR), and *SOX21* (c.640 _641insT; p. A214fs, c.644dupC; p. A215fs and c.644_645ins ACGCGTCTTCTTCCCGCAGTC; p. A215dup).

Conclusions: These pathogenic or potentially pathogenic mutated genes may be potential biomarkers for URSA and may play an auxiliary role in the treatment of URSA.

Keywords: Whole exome sequencing; unexplained recurrent spontaneous abortion (URSA); genetic polymorphisms; signaling pathway

Submitted Mar 17, 2022. Accepted for publication May 20, 2022.
doi: 10.21037/atm-22-2179
View this article at: https://dx.doi.org/10.21037/atm-22-2179

Introduction

Recurrent spontaneous abortion (RSA) is a major cause of infertility that represents a significant burden on human reproductive health. According to the guidelines of the European Society of Human Reproduction and Embryology (ESHRE) (published in 2018), RSA is defined as the loss of two or more pregnancies (1). The incidence of clinical spontaneous abortion ranges from 15–25% of all pregnancies, while the incidence of two or more abortions and three or more abortions is 5% and 1%, respectively (2).

The etiology of RSA is complex and remains to be fully elucidated. It may involve one single factor or multiple factors, with a high degree of heterogeneity. These factors include cytogenetic, antiphospholipid syndrome, anatomic factors, inherited thrombophilias, hormonal and metabolic factors, infections, factors from the father's side such as fragmentation of sperm DNA, psychological factors, alloimmune factors, lifestyle, environmental, and occupational factors (2). However, consensus guidelines do not support routine testing of some of the above etiological indicators of RSA (2). Chromosomal abnormalities are known to be one of the main genetic factors of RSA (3). In addition, in nearly 50% of RSA cases, the cause is unknown and these are referred to as unexplained recurrent spontaneous abortions (URSAs) (4). Studies have suggested that most of the URSAs are related to immune and genetic factors (5-8). The paucity of data related to the mechanisms of URSAs has hindered the development of effective treatments. Previous reports have mostly analyzed maternal causes, and studies investigating causes in both couples have mainly focused on chromosome abnormalities. In recent years, the pathogenesis of genetic polymorphism in RSA has been thoroughly studied. However, the results of some studies are inconsistent, especially in different populations. Therefore, it is necessary to increase the sample size, carry out in-depth studies in multiple regions and ethnic groups, and further identify other risk genes.

Chromosome microarray is a common clinical diagnostic tool, which is mainly divided into comparative genomic hybridization microarray and single nucleotide polymorphism microarray. It covers about 30% of the genome with a resolution of 5–10 Mb and can accurately detect CNV, most uniparental disomy (UPD) and chimera of >10% ratio (9). It is often used for the detection of stream products. However, there are defects in the detection of a balanced translocation, inversion and complex rearrangement of CMA chromosomes, single nucleotide changes and insert deletions below 5 Mb. Second-generation sequencing technology is a precise, sensitive, high-resolution, high-throughput assay that can determine the genetic changes that cause human diseases at the nucleotide level of the entire DNA sequence (whole genome sequencing) or the coding portion of the genome (exome) (10), with the latter being widely used. Second-generation sequencing technology has the ability to detect abnormalities and seek out new information, helping to identify the mechanisms by which genetic variations and transcriptional disorders lead to single-gene and multi-gene diseases. With the successful completion of Human Genome Project (HGP), Haplotype Mapping (HapMap) and 1000 Genome Project, genome-wide association study (GWAS) based on whole genome sequencing has developed rapidly. GWAS has been widely used in the study of susceptibility genes for complex and difficult diseases, providing new ideas and methods for searching the pathogenesis of complex and difficult diseases. Compared with whole genome sequencing, whole exome sequencing is more widely used. Whole exome sequencing detects sequence changes in protein-coding genes, which account for about 1% of the genome but contain about 85% of disease-causing mutations. Whole exome sequencing is easier to analyze and has great advantages in studying SNP, insertions and deletions mutations. Moreover, whole exome sequencing has higher sequencing depth and can identify mutations that are not detected in whole genome sequencing (11).

In recent years, it has been suggested that genetic polymorphism may play an important role in RSA. Tsurusaki *et al.* (12,13) performed whole exome sequencing analysis on a family with 7 pregnancy abnormalities and found that Intraflagellar Transport 122 (*IFT122*) mutation was the cause of repeated miscarriages. *IFT122* is closely related to the cilia system, which plays an important role in the development of human embryos. Filges *et al.* reported Phospholipase C Delta 4 (*PLCD4*) and Oxysterol Binding Protein Like 5

(*OSBPL5*) mutations in a woman who had 18 consecutive spontaneous abortions. These genes have been associated with oocyte maturation, activation, and fertilization (14). Qiao and colleagues conducted whole exome sequencing of peripheral blood DNA in 4 couples with RSA and their 7 fetal villus tissues with normal karyotype. Complex heterozygous mutations of the Dynein Cytoplasmic 2 Heavy Chain 1 (*DYNC2H1*) gene was found in the placental villus tissue of one family and mutations in the Arachidonate 15-Lipoxygenase (*ALOX15*) gene was detected in another family (15). *DYNC2H1* affects cilia biogenesis and *ALOX15* is involved in placental development. Quintero-Ronderos *et al.* conducted whole exome sequencing of 49 unrelated RSA cases and found 27 potentially harmful gene variants with 22 genes distributed in different molecular biological pathways closely related to embryo implantation and pregnancy maintenance (16). The study by Xiang *et al.* performed whole exome sequencing on 100 women with RSA and found 35 heterozygous variations including Forkhead Box A2 (*FOXA2*), Fibrinogen Alpha Chain (*FGA*), Coagulation Factor XIII A Chain (*F13A1*), and KH Domain Containing 3 Like, Subcortical Maternal Complex Member (*KHDC3L*) (17). For almost all human diseases, not just genetic diseases, an individual's susceptibility is affected to some extent by genetic variations (18). Compared to previous studies, in this study, a relatively large amount of data was collected from patients with URSA in Chongqing, China for whole exome sequencing. Candidate genes were screened in three directions according to gene function: (I) genetic abnormalities in both parents (factors associated with embryonic development); (II) maternal alone genetic abnormalities in couples (germ-cell, intrauterine environment-related factors); (III) paternal genetic abnormalities in couples (germ-cell related factors). Our study may provide a new field in understanding the pathology mechanism of URSA. We present the following article in accordance with the STREGA reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-2179/rc).

Methods

Participant recruitment and sample collection

In this study, a total of 30 couples with URSA and 9 couples with normal birth history were recruited from the Third Affiliated Hospital of Chongqing Medical University from May 2020 to December 2020. The inclusion criteria for RSA (two or more fetal loss before 24 weeks of gestation) were based on the ESHRE in 2018. All subjects signed informed consent forms and completed case report registration forms for the screening criteria for inclusion. Couples were excluded if they presented with chromosomal karyotype abnormalities, endocrine abnormalities, reproductive system malformations, reproductive tract infections, autoimmune diseases, factors indicative of a prethrombotic state, abnormal male DNA fragmentation index (DFI), varicocele in the male, abnormal male sex hormone, other internal/surgical diseases, abnormalities of liver and kidney function, and poor environmental and living habits. According to the above inclusion and exclusion criteria, couples with recurrent abortions of unknown cause were selected as the case group. A total of 5 mL whole blood was collected intravenously from patients using EDTA tubes and cryopreserved at –80 ℃ for whole exome sequencing. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Third Affiliated Hospital of Chongqing Medical University (No. 2019-9) and informed consent was taken from all the patients.

Whole exome sequencing

Whole genome DNA was extracted from 1 mL of whole blood to assess the concentration, integrity, and purity. Subsequently, 1 μg of DNA was broken into 250–300 bp using the Covaris ultrasonic fragmentation instrument for terminal repair and joint connection. Agilent SureSelect Human All Exon V6 was used to capture exons of amplified samples to prepare DNA nanospheres. The MGISEQ-2000 high-throughput sequencing platform based on DNBSEQTM technology was used for 100 bp double-terminal sequencing with sequencing depth of 100×.

Sequencing data preprocessing and genome alignment

The FastQC software (V0.11.9) was used to assess the quality of the raw sequencing data. Sequencing connectors and low-quality bases were removed using Trim Galore software (V0.6.7). Clean reads were compared with the human reference genome (GRCh38) using BWA-MEM software (V0.7.17). GATK4 Best Practices of Germline short variant Discovery (SNPs + Indels) was used for variants detection. MarkDuplicates command in GATK4 was used to label the repeated sequences (duplicates) generated by

sequencing construction. BaseRecalibrator and ApplyBQSR commands in GATK4 were used to recalibrate the base quality score. Subsequently, the gVCF (Genome Variant Call Format) files for each sample were obtained, which contains the mutation detection results for both mutation and non-mutation locations. Finally, the gVCF files of all the queue samples were combined and preliminarily filtered with the screening criteria of min-mean DP >3 and minGQ >3. After obtaining some candidate variants, most of the variants met the following conditions: mim-meanDP >20 and minGQ >10.

Variations annotation, filtering, and prediction of harm and pathogenicity

CADD (https://cadd.gs.washington.edu/score, GRCh38-v1.6) was used to predict potentially harmful mutations based on the American College of Medical Genetics and Genomics (ACMG) guide variation classification. CADD Score (PHREAD) >10, SIFT prediction is deleterious, or PolyPhen prediction is probably/possibly damaging were the screening criteria of missense mutations. For loss-of-function (LoF) variation, termination mutation (stop-gain), termination loss (stop-loss), frameshift insertions/deletions, canonical splice sites, and initiation loss (start-loss) were retained. Prediction of tolerance to genetic variation was performed using Residual Variation Intolerance Score (RVIS). RVIS sequenced genes according to their level of intolerance to functional genetic variants. The top 50% of genes that were least resistant to the variants were selected. The gnomAD database (https://gnomad.broadinstitute.org/), 1,000 Genomes database, native library (9 couples with normal reproductive history), ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/), and OMIM database (https://www.omim.org) were used to filter allele frequency (AF), disease citations, and other in silico attributes. Mutations with AF greater than 5% were excluded. The genes involved in the above mutations were analyzed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (by Metascape, http://metascape.org/gp). The pathogenicity of the mutation was evaluated by combining the above factors to obtain the candidate mutations and the genes affected.

Prediction of protein structure effects

The topological models of proteins were predicted using SMART to explore protein domain architecture. NetPhos 3.1 was used to investigate possible changes in phosphorylation events in mutant gene sequences. NetPhos 3.1 predicts serine, threonine, or tyrosine phosphorylation sites in amino acid sequences of eukaryotic proteins. Original and modified amino acid sequences were submitted to SWISS-MODEL (https://swissmodel.expasy.org/) for protein modeling and any changes in protein folding and structure were visualized using the PyMOL software.

Sanger validation

Sanger sequencing was used to verify pathogenic variants while Primer Premier 5.0 was used to design upstream and downstream primers for the genes to be tested. Genomic DNA was extracted and purified using magnetic beads. The integrity of the genomic DNA was identified and quantified by electrophoresis after polymerase chain reaction (PCR) under the following conditions: 14 cycles of denaturation at 95 ℃ for 5 minutes, 95 ℃ for 30 seconds, 67 ℃ for 30 seconds, 72 ℃ for 1 minute, and finally 72 ℃ extension for 7 minutes. The PCR products were used for Sanger sequencing on a 3730xl sequencer platform. The peak figure was checked using BioEdit software, by combining the Human Genome Variation Society (HGVS) naming and related database (https://www.ncbi.nlm.nih.gov/) and the searched mutation loci. SeqMan software in DNAstar software package was used for sequence splicing after Sanger sequencing. In addition, protein conservation analysis was performed in other species using a ClustalX tool.

Results

Description of the cohort

A total of 30 unrelated couples with URSA were included in this study, among which 50% (15/30), 20% (6/30), and 30% (9/30) of couples had 2, 3, and more than 4 miscarriages, respectively, ranging from 5 weeks to 24 weeks. There were couples with as many as 7 miscarriages. The age of the female participants ranged from 24 years to 42 years, with an average age of 31.03 years. The male participants ranged from 25–45 years old, with an average age of 33.13 years. Chromosome karyotypes in the peripheral blood of all patients were normal. There were no other common causes of RSA. After mutation filtering, a total of 971 maternal and 954 paternal pathogenic or possible pathogenic variants were identified (*Table 1*). According to the above data, the total number of mutations was not associated with age nor

Table 1 971 maternal and 954 paternal pathogenic/possible pathogenic variants

No.	Female				Male				No. of abortions	Gestational age at abortion (weeks)
	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations		
1	30	28	*BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *WNT1*: NM_005430.4 c.403G>A (p.V135I)	*BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	36	36	*SOX21*: ENST00000376945 c.747G>C (p.Q249H)		2	7/20
2	25	38	*APP*: NM_000484.4 c.1530G>C (p.K510N) *GGT2*: ENST00000401924 c.973G>A (p.A325T)	*APP*: NM_000484.4 c.1530G>C (p.K510N)	27	33	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	2	6/24
3	25	29	*BPTF*: NM_182641.4 c.3293C>T (p.S1098F) *GGT2*: ENST00000401924 c.973G>A (p.A325T)		27	24	*APOB*: NM_000384.3 c.581C>T (p.T194M) *MMP2*: NP_004521.1 c.1462_1464delTTC (p.F488del)	*MMP2*: NP_004521.1 c.1462_1464delTTC (p.F488del)	3	6/7/12
4	34	42			36	42	*SOX21*: ENST00000376945 c.640 _641insT (p. A214fs) *SOX21*: NM_007084.4 c.644dupC (p. A215fs) *BUB1B*: NM_001211.6 c.1648C>T (p.R550*)	*SOX21*: ENST00000376945 c.640 _641insT (p. A214fs) *SOX21*: NM_007084.4 c.644dupC (p. A215fs) *BUB1B*: NM_001211.6 c.1648C>T (p.R550*)	2	5/8
5	37	36			38	36	*ZIC2*: NM_007129.5 c.83C>G (p.A28G) *SIRT1*: ENST00000212015 c.17C>T (p.A6V)		7	8/9/9/10/10/11/11
6	36	19	*GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)		32	32	*APOB*: NM_000384.3 c.581C>T (p.T194M) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)		6	6/7/11/11/20/24
7	30	33	*ZIC2*: NM_007129.5 c.83C>G (p.A28G) *ANXA5*: NM_001154.4 c.949G>C (p.G317R) *VCAN*: NM_004385.5 c.9560G>A (p.R3187Q) *APOB*: NM_000384.3 c.581C>T (p.T194M)	*ANXA5*: NM_001154.4 c.949G>C (p.G317R)	36	29	*APOB*: NM_000384.3 c.581C>T (p.T194M) *GGT2*: ENST00000401924 c.973G>A (p.A325T)		4	12/18/21/24
8	33	36	*BPTF*: NM_182641.4 c.3278A>G (p.K1093R) *MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	*MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	35	36	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *MECP2*: ENST00000453960 c.21_23delCGC (p.A7del) *ZIC2*: NM_007129.5 c.83C>G (p.A28G) *SOX21*: ENST00000376945 c.747G>C (p.Q249H)	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	3	6/8/8
9	38	26	*FN1*: NM_212482.4 c.5621T>C (p.M1874T) *MSH2*: XM_011532867.2 c.1168G>A (p.L390F)	*FN1*: NM_212482.4 c.5621T>C (p.M1874T) *MSH2*: XM_011532867.2 c.1168G>A (p.L390F)	39	28	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *APOB*: NM_000384.3 c.581C>T (p.T194M) *APOB*: NM_000384.3 c.581C>T (p.T194M) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)		5	6/7/8/9/10
10	27	37	*COL11A1*: NM_001854.4 c.475A>G (p.I159V)		27	21	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I)(homoMut) *GGT2*: ENST00000401924 c.973G>A (p.A325T)		4	6/6/9/24
11	42	33	*COL4A2*: NM_001846 c.1328 C>G (p.P443R) *MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	*MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	45	28			4	9/9/23/24

Table 1 (*continued*)

Table 1 (*continued*)

No.	Female				Male				No. of abortions	Gestational age at abortion (W)
	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations		
12	30	22	*APOB*: NM_000384.3 c.581C>T (p.T194M)		33	39	*COL11A1*: NM_001854.4 c.2578 T>A (p.F860I) *MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	*MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	2	8/8
13	30	37	*MSH2*: XM_011532867.2 c.1168G>A (p.L390F)	*MSH2*: XM_011532867.2 c.1168G>A (p.L390F)	30	34	*COL4A2*: NM_001846 c.2102 A>G (p.K701R) *COL4A2*: NM_001846 c.1328 C>G (p.P443R) *COL11A1*: NM_001854.4 c.2578 T>A (p.F860I) *LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *THBS1*: NM_003246.4 c.1759G>C (p.D487H)		5	5/7/10/12/12
14	29	27	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *ZIC2*: NM_007129.5 c.83C>G (p.A28G) *ITGAM*: ENST00000648685 c.3056C>T (p.P1019L)	*BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	31	30	*COL4A2*: NM_001846 c.4808 A>C (p.H1603P) *ATM*: NM_001351834.2 c.6503C>T (p.S2168L)	*COL4A2*: NM_001846 c.4808 A>C (p.H1603P)	3	7/8/10
15	29	29			29	36	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *APOB*: NM_000384.3 c.581C>T (p.T194M)		2	8/8
16	27	40	*LAMA2*: NM_000426.4 c.6206A>G (p.Y2069C) *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	25	40	*COL4A2*: NM_001846 c.2102 A>G (p.K701R)		4	6/6/6/6
17	26	31	*LAMA2*: chr6:129514599,A>T,LAMA2	*LAMA2*: chr6:129514599, A>T	27	16	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I)		2	8/8
18	25	32	*COL11A1*: NM_001854.4 c.2578 T>A (p.F860I) *LAMA2*: NM_000426.4 c.2217G>T (p.W739C) *LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *SOX21*: ENST00000376945 c.640 _641insT (p. A214fs) *APOB*: NM_000384.3 c.7043A>G (p.Y2348C) *GGT2*: ENST00000401924 c.973G>A (p.A325T)	*SOX21*: ENST00000376945 c.640 _641insT (p. A214fs)	44	25	*COL11A1*: NM_001854.4 c.2578 T>A (p.F860I) *BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *KDR*: NM_002253.4 c.2440G>A (p.D814N) *ITGB1*: ENST00000396033 c.655T>C (p.Y219H)	*KDR*: NM_002253.4 c.2440G>A (p.D814N) *ITGB1*: ENST00000396033 c.655T>C (p.Y219H) *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	2	5/6
19	27	36	*COL11A1*: ENST00000370096 c.2519 C>T (p.P840L) BPTF: NM_182641.4 c.429_431GGA (p.E148del)	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	30	34	*COL4A2*: NM_001846 c.1328 C>G (p.P443R) *APOB*: NM_000384.3 c.581C>T (p.T194M) *KDR*: NM_002253.4 c.2516C>T (p.P839L) *GATA3*: NM_001002295.2 c.706C>A (p.P236T)		2	7/8
20	30	38	*COL4A2*: NM_001846 c.1328 C>G (p.P443R) *LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *THBS1*: NM_003246.4 c.70T>G (p.S24A)		32	30	*COL4A2*: NM_001846 c.2102 A>G (p.K701R)		2	8/9

Table 1 (*continued*)

Table 1 (*continued*)

No.	Female				Male				No. of abortions	Gestational age at abortion (W)
	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations		
21	24	33	*COL4A2*: NM_001846 c.4256 T>C (p.M1419T) *LAMA2*: NM_000426.4 c.4802C>T (p.P1601L) *ZIC5*: NM_033132.5 c.815G>A (p.G272D) *CREBBP*: NM_004380.3 c.833A>C (p.Q278P) *THBS1*: NM_003246.4 c.2099A>G (p.N700S) *KMT2D*: XM_011538772.2 c.15671G>A (p.R5224H) *APOB*: NM_000384.3 c.581C>T (p.T194M) *GGT2*: ENST00000401924 c.973G>A (p.A325T)	*THBS1*: NM_003246.4 c.2099A>G (p.N700S)	26	31	*MECP2*: ENST00000453960 c.21_23delCGC (p.A7del) *ZIC5*: NM_033132.5 c.815G>A (p.G272D) *ZIC5*: NM_033132.5 c.277C>T (p.P93S)	*MECP2*: ENST00000453960 c.21_23delCGC (p.A7del)	2	8/12
22	30	36	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I)		32	34	*COL4A2*: NM_001846 c.2102 A>G (p.K701R) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut) *PLK1*: NM_005030.6 c.1210G>T (p.A404S)	*PLK1*: NM_005030.6 c.1210G>T (p.A404S)	2	6/9
23	35	35	*THBS1*: NM_003246.4 c.70T>G (p.S24A)		38	36	*LAMA2*: NM_000426.4 c.4157A>T (p.Y1386F) *BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	*BPTF*: NM_182641.4 c.396_398delGGA *BPTF*: NM_182641.4 c.429_431GGA (p.E148del)	2	9/10
24	26	34	*DNMT1*: NM_001130823.3 c.2626G>A (p.G876R) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)	*DNMT1*: NM_001130823.3 c.2626G>A (p.G876R)	25	26			2	8/9
25	30	22	*LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *ZIC2*: NM_007129.5 c.83C>G (p.A28G)		33	35	*LAMA4*: NM_001105206.3 c.3158A>G (p.D1053G) *ZIC2*: NM_007129.5 c.83C>G (p.A28G) *POLR2B*: NM_000938.3 c.406G>T (p.G136C) *TOP2A*: NM_001067.4 c.4375A>G (p.K1459E)	*LAMA4*: NM_001105206.3 c.3158A>G (p.D1053G) *POLR2B*: NM_000938.3 c.406G>T (p.G136C)	2	8/10
26	39	26	*LAMA2*: NM_000426.4 c.6206A>G (p.Y2069C) *LAMA4*: NM_001105206.3 c.5443G>A (p.V1815I) *MSH2*: XM_011532867.2 c.1168G>A (p.L390F) *GGT2*: ENST00000401924 c.973G>A (p.A325T)	*MSH2*: XM_011532867.2 c.1168G>A (p.L390F)	39	39	*LAMA2*: NM_000426.4 c.9340G>T (p.V3114F)		3	10/11/12
27	32	34			29	37	*GGT2*: ENST00000401924 c.973G>A (p.A325T)		4	9/10/11/12
28	37	36			41	27	*GGT2*: ENST00000401924 c.973G>A (p.A325T)		3	8/9/10

Table 1 (*continued*)

Table 1 (*continued*)

No.	Female				Male				No. of abortions	Gestational age at abortion (W)
	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations	Age (years)	Total number[a]	Nucleotide alteration	Protein structure with significant variations		
29	35	36	*COL4A2*: NM_001846 c.2102 A>G (p.K701R) *LAMA2*: NM_000426.4 c.6206A>G (p.Y2069C) *IL1B*: NM_000576.3 c.562G>T (p.V188L) *KMT2D*: ENST00000301067 c.6864G>T (p.K2288N) *KMT2D*: XM_011538772.2 c.13258C>T (p.R4420W) *CCNB3*: NM_033031.3 c.356C>A (p.P119Q)	*CCNB3*: NM_033031.3 c.356C>A (p.P119Q)	37	30			3	5/5/8
30	33	30	*APP*: ENST00000346798 c.2131G>T (p.V711F) *GGT2*: ENST00000401924 c.973G>A (p.A325T)(homoMut)		35	30	*BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *ZIC5*: NM_033132.5 c.176T>C (p.L59P) *SOX21*: NM_007084.4 c.644_645ins ACGCGTCTTCTTCCC *ATM*: NM_001351834.2 c.6679C>T (p.R2227C) *POLR2B*: NM_000938.3 c.406G>T (p.G136C)	*POLR2B*: NM_000938.3 c.406G>T (p.G136C) *BPTF*: NM_182641.4 c.429_431GGA (p.E148del) *SOX21*: NM_007084.4 c.644_645ins ACGCGTCTTCTTCCC	2	7/7

[a], total number of mutations after preliminary filtration.

the number of abortions.

Analysis of variants

In couples with recurrent miscarriages, one or both parties may have germline mutations, that is, DNA sequence mutations in the sperm or eggs. In general, all cells of an individual carry mutations, and germline mutations can be inherited. Therefore, this study used exome sequencing to determine whether both parents have pathogenic mutations. The mutant genes may be passed on to the fetus, resulting in abnormal fetal production and development, and subsequently, abortion. In addition, maternal coagulation, immunity, and other related gene mutations may result in abnormal fetal development *in utero*, resulting in an abortion. Mutated genes can also cause abnormal reproductive cells that result in defective fertilized eggs to miscarry. In this study, a unique analysis model was adopted to screen mutant genes. This involved the examination of gene function in the following three dimensions: (I) genetic abnormalities in the couple (factors related to embryo development); (II) genetic abnormalities in the mother alone (reproductive cells, intrauterine environment related factors such as thrombosis, immune factors, placental diseases, etc.); and (III) paternal genetic abnormalities (germ cell related factors). According to ACMG guidelines, 34 genes and 61 pathogenic or possible pathogenic variants were screened from 30 couples with URSA (*Figure 1*). Among these, 52 were missense variants, 1 was a splice site donor variant, 4 were in-frame deletions, 1 were in-frame insertions, 2 were frameshift variants, 1 was a stop gained mutation. After reviewing the literature and combining the report of the Human Gene Mutation Database in 2020, these 61variants were determined to be novel genetic mutations that may be associated with URSA.

Protein function prediction and conservative analysis of 12 missense variants

Among the 52 missense variants, 12 variants (in mother, father, and both couples) had obvious changes in their functions (*Figure 2*). In maternal-only variants, 6 mutations had obvious changes in their functions including *ANXA5* (c.949G>C; p.G317R), *APP* (c.1530G>C; p.K510N), *DNMT1* (c.2626G>A; p.G876R), *FN1* (c.5621T>C; p.M1874T), *MSH2* (c.1168G>A; p.L390F), and *THBS1* (c.2099A>G; p.N700S). In paternal-only variants, 4 mutations had obvious changes in their functions including *KDR* (c.2440G>A; p.D814N), *ITGB1* (c.655T>C; p.Y219H), *PLK1* (c.1210G>T; p.A404S), and *POLR2B* (c.406G>T; p.G136C). In mother and/or father variants, 2 mutations had obvious changes in their functions including *LAMA4* (c.3158A>G; p.D1053G) and *COL4A2* (c.4808 A>C; p.H1603P). For *ANXA5*, the hydrophilicity of G317R changed from hydrophobic amino acid to hydrophilic amino acid. For APP, the K510N mutation was located in the APP_E2 domain and phosphorylation site, resulting in the disappearance of the hydrogen bonds. For *COL4A2*, the H1603P mutation was located in the C4 (C-terminus of type 4 collagens) domain. Hydrophilicity and hydrophobicity changed, the hydrophilic amino acid became a hydrophobic amino acid, and the hydrogen bond disappeared. For *DNMT1*, the G876R mutation was located in the BAH (bromo adjacent homology) domain resulting in a hydrophobic amino acid change to a hydrophilic amino acid. For *FN1*, the M1874T mutation changed the hydrophilicity from hydrophilic amino acid to hydrophilic amino acid in the FN3 (fibronectin type 3) domain and phosphorylation site. For *ITGB1*, the Y219H mutation was located in the domain of the integrin beta subunits (INB), and the hydrogen bond disappeared. For *LAMA4*, D1053G changed the hydrophilic amino acid to a hydrophobic amino acid. For *MSH2*, the L390F mutation was located in the MUTSd (DNA-binding domain of DNA mismatch repair MUTS family) domain. For *POLR2B*, the G136C mutation was located in the RNA_pol_Rpb2_1 domain, and the hydrophilicity was changed from hydrophobic amino acid to hydrophilic amino acid. For *THBS1*, the N700S hydrogen bond disappeared. Similarly, for *KDR*, the D814N hydrogen bond disappeared. For *PLK1*, the hydrophilicity of A404S changed from hydrophobic amino acid to hydrophilic amino acid. In addition, the conserved analysis of amino acid changes caused by the variation of these sites in humans and other related species showed that the sequences were highly conserved.

Identification of single-gene causes of unexplained RSA

The foundation of genome-wide variant interpretation is the functional annotation of the genome. The resulting annotation of genes, transcripts, and regulatory elements has enabled high-quality predictions of the most severe classes of likely gene-disrupting variants. In each variant function annotation type, premature stop, splice site, frameshift indel, and splice region has serious effects on protein function (19). The mutated genes and missense

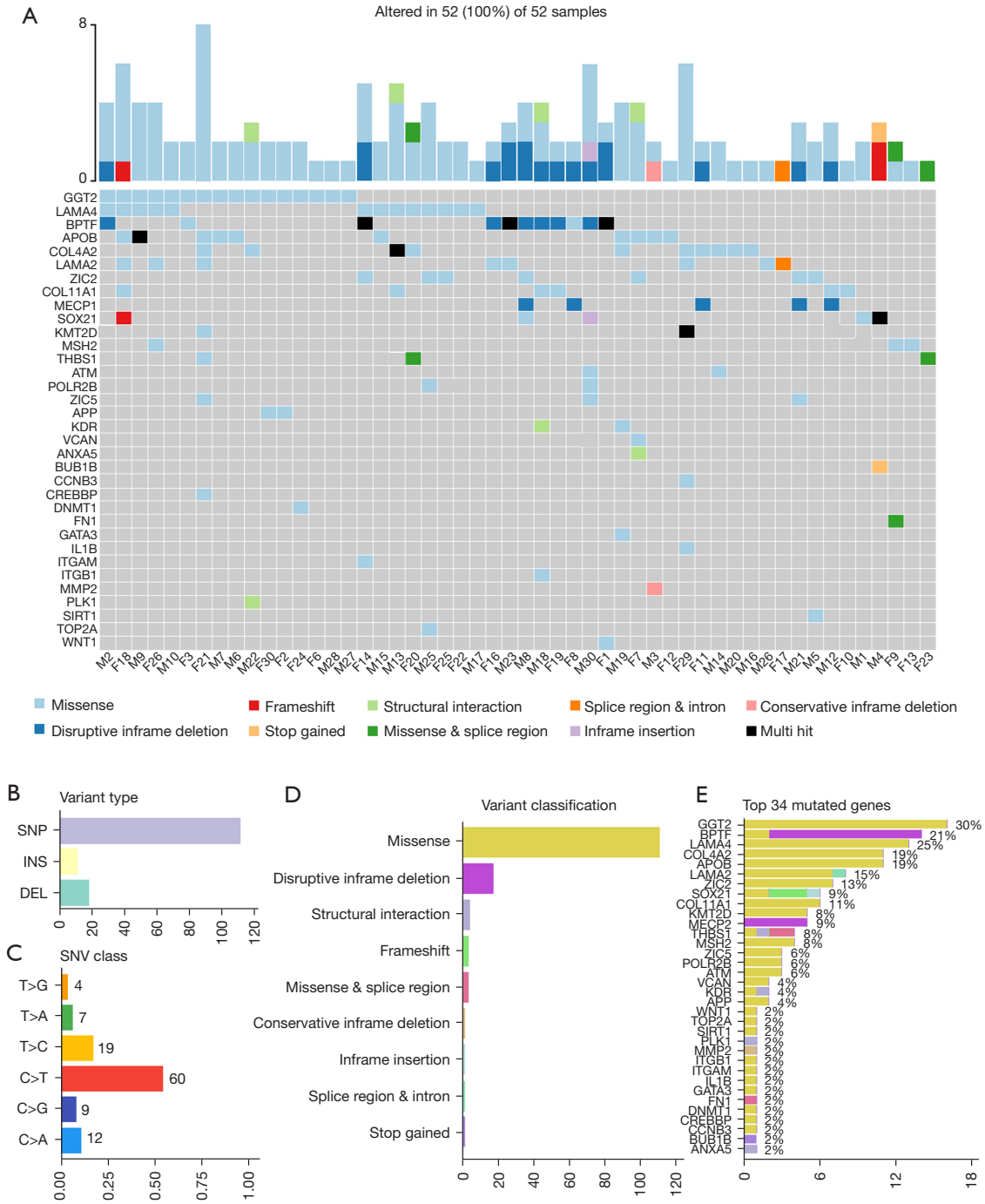
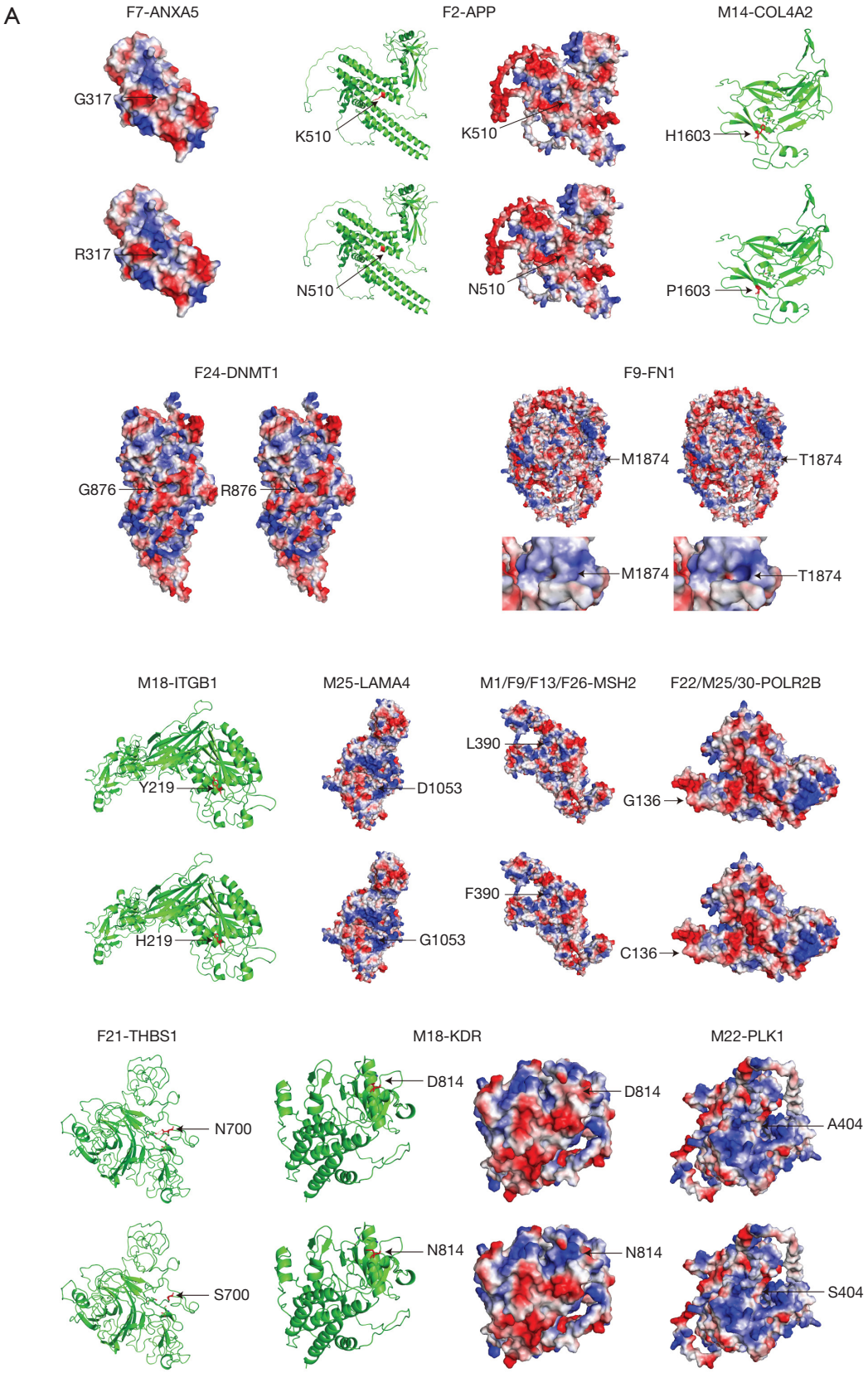


Figure 1 The general mutational landscape of 30 couples with unexplained recurrent spontaneous abortion. (A) The distribution of the top 34 mutated genes in 30 couples with URSA. Different colors represented different mutation types. Y-axis represents total number of variations in all genes involved in a sample. (B) The different variant types. The x-axis and y-axis represent the number of variants and the variant type, respectively. (C) The SNV class. The x-axis and y axis represent the ratio and the SNV class, respectively. The number on the right represents the number of SNV class. (D) The variant classification. The x-axis and y-axis represent variant classification and number of variant classifications, respectively. (E) The top 34 mutated genes. The x-axis and y-axis represent the number of variants and the mutated genes, respectively. The number on the right represents the percentage of mutated genes. URSA, unexplained recurrent spontaneous abortion; SNV, single nucleotide variant.



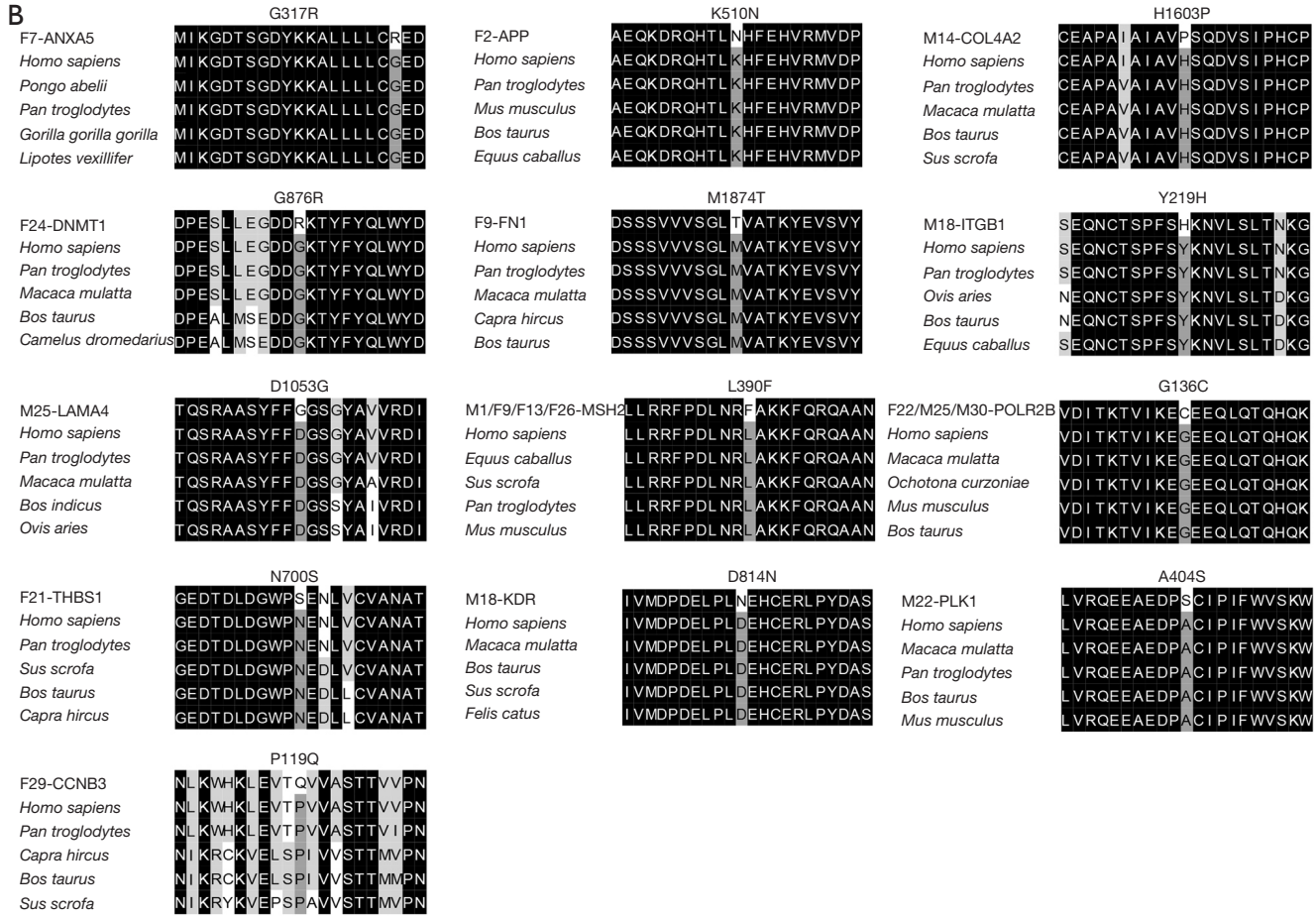


Figure 2 Prediction of protein function and conservative analysis of missense variants. (A) The predicted protein function of the 12 missense variants. (B) A conservative analysis of the 12 missense variants.

mutated genes with predicted altered protein function were strongly associated with URSA. There were 22 pathogenic or potentially pathogenic mutations of 19 genes (*Table 2*) (involving 12 missense mutant genes with altered protein function prediction) in 28 patients (23 couples with URSA) (*Figure 3*). Besides the 12 missense variants with predicted altered protein function, the remaining 10 heterozygous variants were *CCNB3* (c.356C>A; p.P119Q) from the mother, *BUB1B* (c.1648C>T; p.R550*) and *MMP2* (c.1462_1464delTTC; p.F488del) from the father, and mutations from mother and/or father included *BPTF* (c.396_398delGGA; p.E138 del and c.429_431GGA; p.E148del), *LAMA2* (HGVS: NA; Exon: NA; SPLICE_SITE, DONOR), *MECP2 (*c.21_23delCGC; p.A7del) and *SOX21* (c.640 _641insT; p. A214fs, c.644dupC; p. A215fs and c.644_645ins ACGCGTCTTCTTCCCGCAGTC; p. A215dup). Among these, *CCNB3* (c.356C>A; P.p119q) was a missense mutation with no predicted change in protein function. However, *CCNB3* has been reported to be associated with URSA. Therefore, the mutation may also be the cause of URSA in this patient. These variants are potentially relevant to the various phenotypes of RSA.

Functional analysis of the variants

GO analysis of the mutated genes in both couples revealed that they were significantly enriched in DNA-binding transcription activator activity, RNA polymerase II-specific activity, endoderm development, blood vessel development, extracellular matrix organization, neuron migration, post-embryonic development, negative regulation of response to external stimulus, and regulation of cell adhesion

Table 2 The 22 pathogenic or potentially pathogenic mutations in 19 genes in patients with unexplained recurrent spontaneous abortion

Gene	Variation	HGVS	CADD v1.6	Sift	PolyPhen	RVIS%	Exon	ConsDetail	Frequency	Inheritance
COL4A2	chr13:110508148, A>C	NM_001846 c.4808 A>C (p.H1603P)	24.8	Deleterious	Probably_damaging	45.49364614	47/48	Missense	–	AD
LAMA4	chr6:112139244, T>C	NM_001105206.3 c.3158A>G (p.D1053G)	29.9	Deleterious	Probably_damaging	15.11241447	24/39	Missense	–	AD
FN1	rs199684110, A>G	NM_212482.4 c.5621T>C (p.M1874T)	25.6	Tolerated	Probably_damaging	11.4173998	34/46	Missense-SpD (Dst,2)	0.000099	AD
APP	rs767201930, C>G	NM_000484.4 c.1530G>C (p.K510N)	24.7	Deleterious	Probably_damaging	12.69794721	12/18	Missense	0.000004	AD
DNMT1	rs62621087, C>T	NM_001130823.3 c.2626G>A (p.G876R) NM_001379.4 c.2578G>A (G860R)	23.6	Deleterious	Probably_damaging	2.443792766	27/41	Missense	0.000481	AD
THBS1	rs2228262, A>G	NM_003246.4 c.2099A>G (p.N700S)	27.7	Deleterious	Probably_damaging	26.33431085	13/22	Missense	0.079344	NA
MSH2	rs17224367, C>T	XM_011532867.2 c.1168G>A (p.L390F)	23.4	Deleterious	Benign	10.76246334	7/16	Missense	0.001547	AD
ANXA5	rs750064390, C>G	NM_001154.4 c.949G>C (p.G317R)	26.5	Deleterious	Probably_damaging	42.6686217	13/13	Missense	0.000008	AD
KDR	rs35603373, C>T	NM_002253.4 c.2440G>A (p.D814N)	20.2	Deleterious	Benign	28.45552297	17/30	Missense	0.000115	AD
PLK1	rs201491344, G>T	NM_005030.6 c.1210G>T (p.A404S)	24.8	Tolerated	Possibly_damaging	42.89345064	7/10	Missense	0.000084	NA
POLR2B	rs140288467, G>T	NM_000938.3 c.406G>T (p.G136C)	29.5	Deleterious	Possibly_damaging	5.630498534	6/26	Missense	0.000873	NA
ITGB1	chr10:32926002, A>G	ENST00000396033 c.655T>C (p.Y219H)	29.6	Deleterious	Probably_damaging	21.46627566	6/16	Missense	–	NA
CCNB3	rs190721449, C>A	NM_033031.3 c.356C>A (p.P119Q)	16.89	Deleterious	Possibly_damaging	23.19749216	6/13	Missense	0.000719	NA
BPTF	rs773299098, CGAG>C	NM_182641.4 c.396_398delGGA (p.E138 del)	19.51	NA	NA	0.410557185	1/28	Inframe_deletion	0.00628	AD
BPTF	rs751039972, CGAG>C	NM_182641.4c.429_431GGA (p.E148del)	21.0	NA	NA	0.410557185	1/28	Inframe_deletion	0.000136	AD
MECP2	chrX:154097642, CGCG>C	ENST00000453960 c.21_23delCGC (p.A7del)	21.1	NA	NA	27.5862069	1/3	Inframe_deletion	–	XLR
SOX21	chr13:94711409, G>GA	ENST00000376945 c.640 _641insT (p. A214fs)	28.8	NA	NA	NA	1/1	Frameshift	–	NA
SOX21	rs187532628, C>CG	NM_007084.4 c.644dupC (p. A215fs)	30.0	NA	NA	NA	1/1	Frameshift	0.000016	NA
SOX21	rs187532628, C>CGAC TGCGGGAAGAAGACGCGT	NM_007084.4 c.644_645ins ACGCGTCTTCTTCCCGCAGTC (p. A215dup)	18.28	NA	NA	NA	1/1	Inframe_insertion	0.000016	NA
MMP2	rs745677721, ATCT>A	NP_004521.1 c.1462_1464delTTC (p.F488del)	21.7	NA	NA	13.15738025	9/13	Inframe_deletion	0.000008	AR
BUB1B	rs767213728, C>T	NM_001211.6 c.1648C>T (p.R550*)	35.0	NA	NA	21.86705767	14/23	Stop_gained	0.000004	AD
LAMA2	chr6:129514599, A>T	NA	18.9	NA	NA	24.75073314	NA	SPLICE_SITE, DONOR	–	AR

HGVS, Human Genome Variation Society; RVIS, Residual Variation Intolerance Score; AD, autosomal dominant; NA, not applicable; XLR, X-Linked Recessive; AR, autosomal recessive.

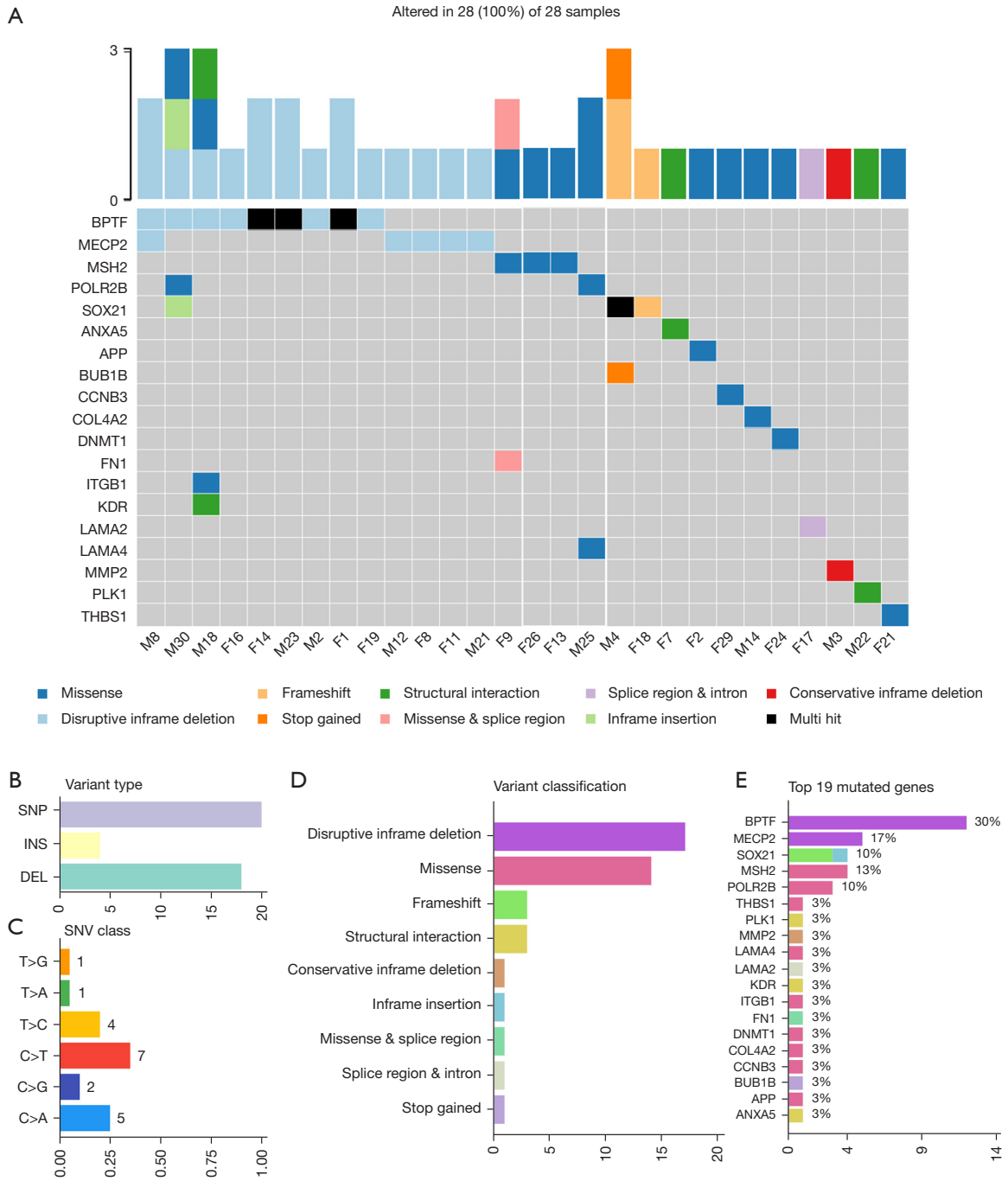


Figure 3 The 19 strongly associated mutant genes detected in 23 couples with unexplained recurrent spontaneous abortion. (A) The distribution of the top 19 mutated genes in 23 couples with URSA. Different colors represent different mutation types. Y-axis represents total number of variations in all genes involved in a sample. (B) The variant types. The x axis and y axis represent the number of variants and the variant type, respectively. (C) The SNV classes. The x-axis and y-axis represent the ratio and the SNV class, respectively. The number on the right represents the number of SNV classes. (D) The variant classification. The x-axis and y-axis represent the variant classification and the number of variant classifications, respectively. (E) The 19 mutated genes that are strongly associated with URSA. The x-axis and y-axis represent the number of variants and the mutated genes, respectively. The number on the right represents the percentage of mutated genes. URSA, unexplained recurrent spontaneous abortion; SNV, single nucleotide variant.

(*Figure 4A*). GO analysis showed that the mutated maternal genes (related to maternal germ cells and intrauterine environment) were significantly enriched in gamete generation, reproductive structure development, chromatin binding, regulation of cell cycle process and meiotic cell cycle (involved *CCNB3*), fertilization, regulation of immune effector process, response to steroid hormone, and regulation of body fluid levels (involved *ANXA5*) (*Figure 4B*). GO analysis demonstrated that the paternal mutated genes (associated with paternal germ cells) were significantly enriched in gamete generation, meiotic cell cycle, chromatin binding, regulation of histone modification, male meiotic nuclear division, and DNA methylation involved in gamete generation (*Figure 4C*). In addition, KEGG analysis of the mutated genes in both couples, as well as the maternal and paternal mutated genes (*Table 3*), showed that *CCNB3* was involved in the cell cycle signaling pathway. *LAMA4*, *LAMA2*, *FN1*, *THBS1*, and *KDR* were involved in the PI3K-Akt and focal adhesion signaling pathways.

Sanger validation

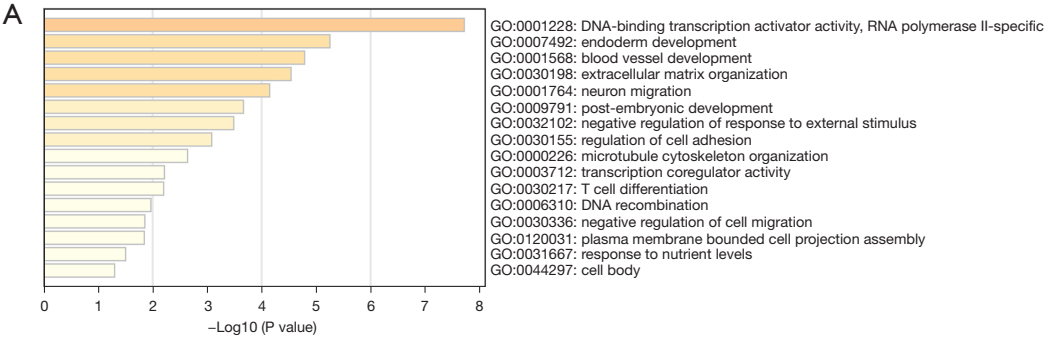
To validate the results of whole exome sequencing, Sanger validation of 13 variants was performed, including *ANXA5*, *APP*, *COL4A2*, *DNMT1*, *FN1*, *ITGB1*, *LAMA4*, *MSH2*, *POLR2B*, *THBS1*, *KDR*, *PLK1*, and *CCNB3* (*Figure 5*). The validation results were consistent with the whole exome sequencing analysis, which suggested that the results of the whole exome sequencing were reliable.

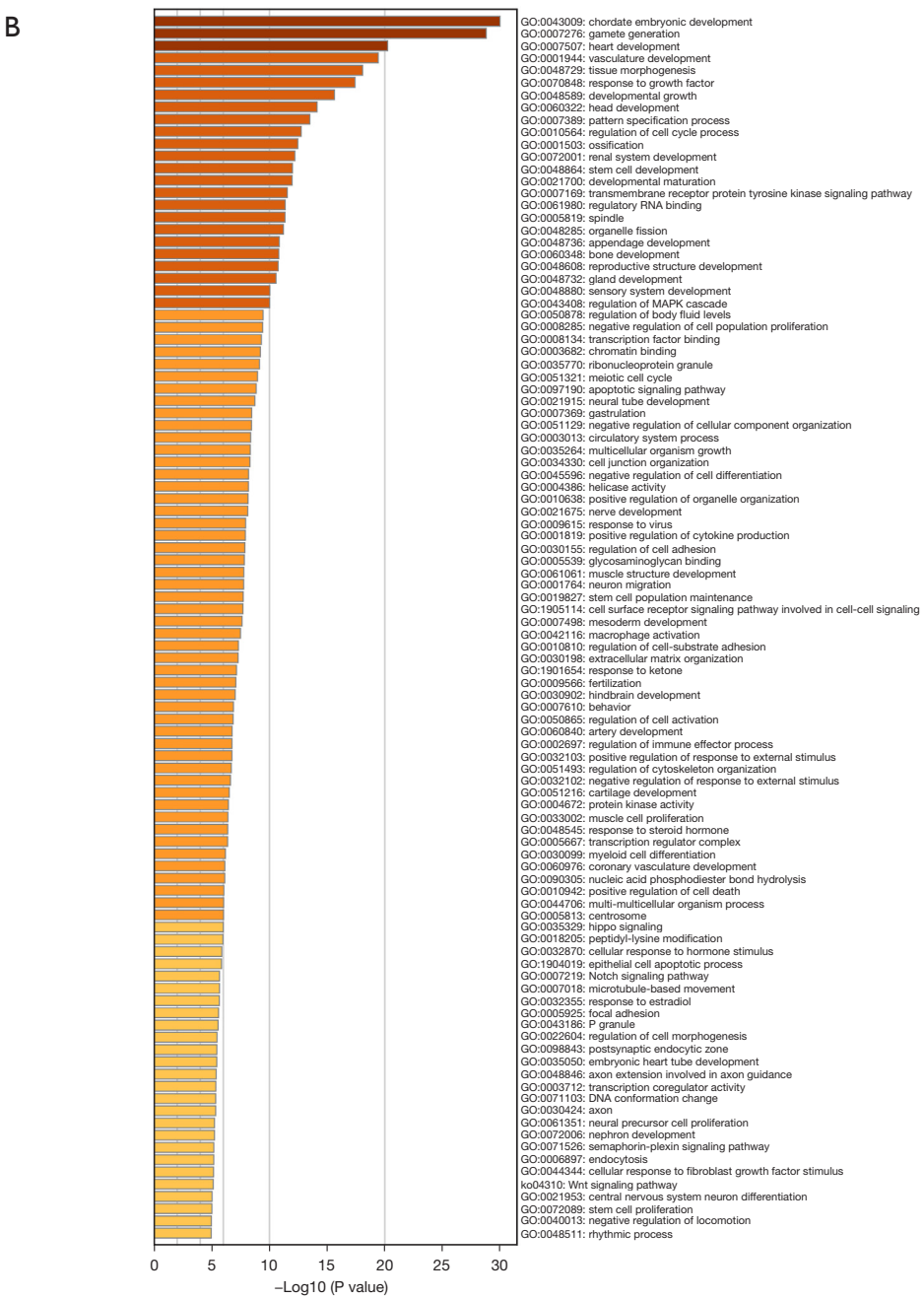
Discussion

Recently, whole-exome sequencing has been widely used in adverse pregnancy outcomes including RSA. Based on whole-exome sequencing, some mutant genes were detected in RSA, such as IFT122, DYNC2H1, ALOX15, FOXA2, FGA, FGA and KHDC3L (12,15,17). In this study, a total of 971 maternal and 954 paternal mutations were found to be pathogenic or possibly pathogenic in 30 couples with URSA through whole-exome sequencing analysis. The incidence of mutations that were strongly associated with URSA was 53.33% in females and 40% in males. The total number of mutations and the abnormality rate of mutations that were strongly associated with URSA were slightly higher in females than in males. Thus, both maternal and paternal factors play important roles in the development of URSA.

In 28 patients from 23 couples, 22 pathogenic or potentially pathogenic variants of 19 genes were found to be strongly associated with URSA, with an abnormality rate of 76.67%. Among these, 12 missense variants resulted in obvious changes in protein function, including 6 maternal mutations, namely, *ANXA5* (c.949G>C; p.G317R), *THBS1* (c.2099A>G; p.N700S), *DNMT1* (c.2626G>A; p.G876R), *FN1* (c.5621T>C; p.M1874T), *MSH2* (c.1168G>A; p.L390F), and *APP* (c.1530G>C; p.K510N); 4 paternal mutations, namely, *KDR* (c.2440G>A; p.D814N), *ITGB1* (c.655T>C; p.Y219H), *PLK1* (c.1210G>T; p.A404S), and *POLR2B* (c.406G>T; p.G136C); and 2 mutations from mother and/or father, namely, *COL4A2* (c.4808 A>C; p.H1603P) and *LAMA4* (c.3158A>G; p.D1053G). In addition to the above 12 missense variants with altered predicted protein function, 10 other variants (involving 7 genes) were detected in URSA patients, including *CCNB3* (c.356C>A; p.P119Q) from the mother, *BUB1B* (c.1648C>T; p.R550*) and *MMP2* (c.1462_1464delTTC; p.F488del) from the fathers, and mutations from the mother and/or father, including *LAMA2* (HGVS: NA; Exon: NA; SPLICE_SITE, DONOR), *BPTF* (c.396_398delGGA; p.E138 del and c.429_431GGA; p.E148del), *MECP2* (c.21_23delCGC; p.A7del) and *SOX21* (c.640 _641insT; p. A214fs, c.644dupC; p. A215fs and c.644_645ins ACGCGTCTTCTTCCCGCAGTC; p. A215dup).

Annexin A5 (*ANXA5*) is a thrombophilia-related protein that is abundantly expressed in normal placenta. It plays a role in promoting membrane repair (20) and functions as an inhibitor of coagulation through its ability to bind to the anionic phospholipids exposed on the surface of platelets (21). Studies have suggested that *ANXA5* is related to RSA (22-25). Lan *et al.* found that reduced *ANXA5* expression was associated with a higher risk of RSA (26). Furthermore, the haplotype in the proximal core promoter region of *ANXA5* gene, including rs28717001, rs112782763, rs113588187, and rs28651243, is associated with URSA (23). In this study, rs750064390 (c.949G>C; p.G317R) in *ANXA5* was identified in a female patient who had 4 URSAs. This missense mutation has not been previously reported in other studies. Multiple sequence comparison showed that the mutation occurred in the conserved region. The surface charge of the protein was significantly changed. Hydrophobic amino acids were changed into hydrophilic amino acids, and the protein function was changed after mutation. In addition, the thrombospondin 1 (*THBS1*) mutation (c.2099A>G; p.N700S) was also detected in URSA patients. The protein encoded by *THBS1* is an adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix





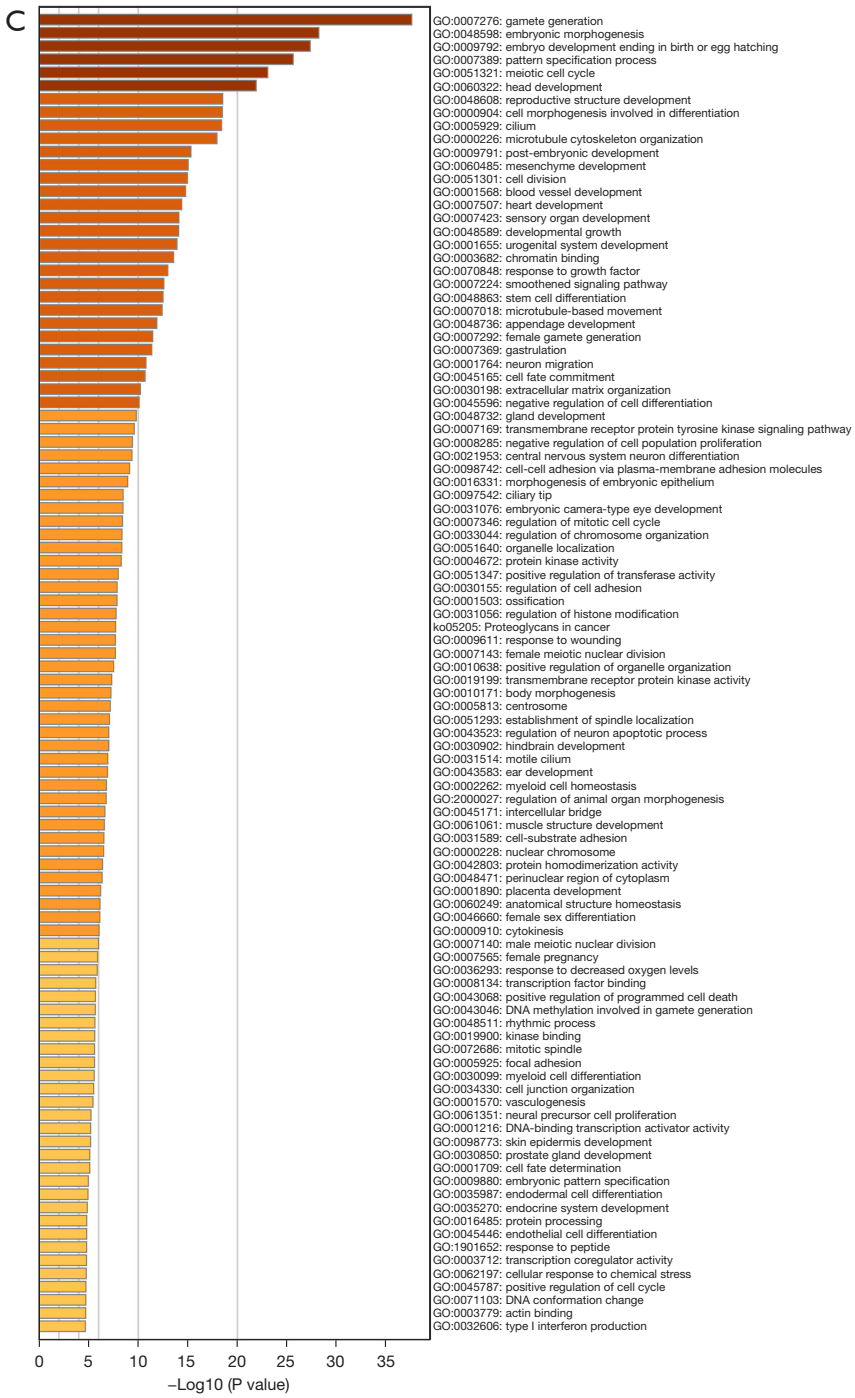


Figure 4 Gene Ontology analysis. (A) Gene Ontology analysis of the mutated genes in both couples. (B) The top 100 Gene Ontology analysis results of mutated genes in the mother. (C) The top 100 Gene Ontology analysis results of mutated genes in the father.

Table 3 KEGG analysis of variants detected both couples, in the mother only, and in the father only

Patients	Terms	P value	Genes
Female			
KEGG_2019_Human	PI3K-Akt signaling pathway	0.006516106	*ANGPT2*; *FN1*; *IGF2*; *AREG*; *THBS1*; *COMP*; *RBL2*; *RELN*; *CCNE1*; *DDIT4*; *TEK*; *FGFR3*; *MET*
KEGG_2019_Human	Cell cycle	0.017459303	*CCNB3*; *RBL2*; *CREBBP*; *CCNE1*; *PRKDC*; *ANAPC5*
KEGG_2019_Human	Focal adhesion	0.048813012	*COMP*; *RELN*; *ROCK2*; *FN1*; *DOCK1*; *MET*; *THBS1*
Male			
KEGG_2019_Human	Focal adhesion	0.00012928	*ITGB1*; *COL2A1*; *FLT1*; *ROCK1*; *ITGA2*; *KDR*; *ITGA8*; *FLNA*; *BRAF*; *FLNB*; *ITGAV*; *ITGA5*
KEGG_2019_Human	PI3K-Akt signaling pathway	0.00030498	*PHLPP2*; *ITGB1*; *NGFR*; *FLT1*; *CHUK*; *ITGA2*; *FGF7*; *COL2A1*; *RPS6KB1*; *KDR*; *ITGA8*; *ITGAV*; *ITGA5*; *JAK3*; *FGFR2*; *FGFR1*
Couples			
KEGG_2019_Human	Focal adhesion	0.008939194	*LAMA2*; *COL4A2*; *LAMA4*
KEGG_2019_Human	PI3K-Akt signaling pathway	0.040311836	*LAMA2*; *COL4A2*; *LAMA4*

KEGG, Kyoto Encyclopedia of Genes and Genomes.

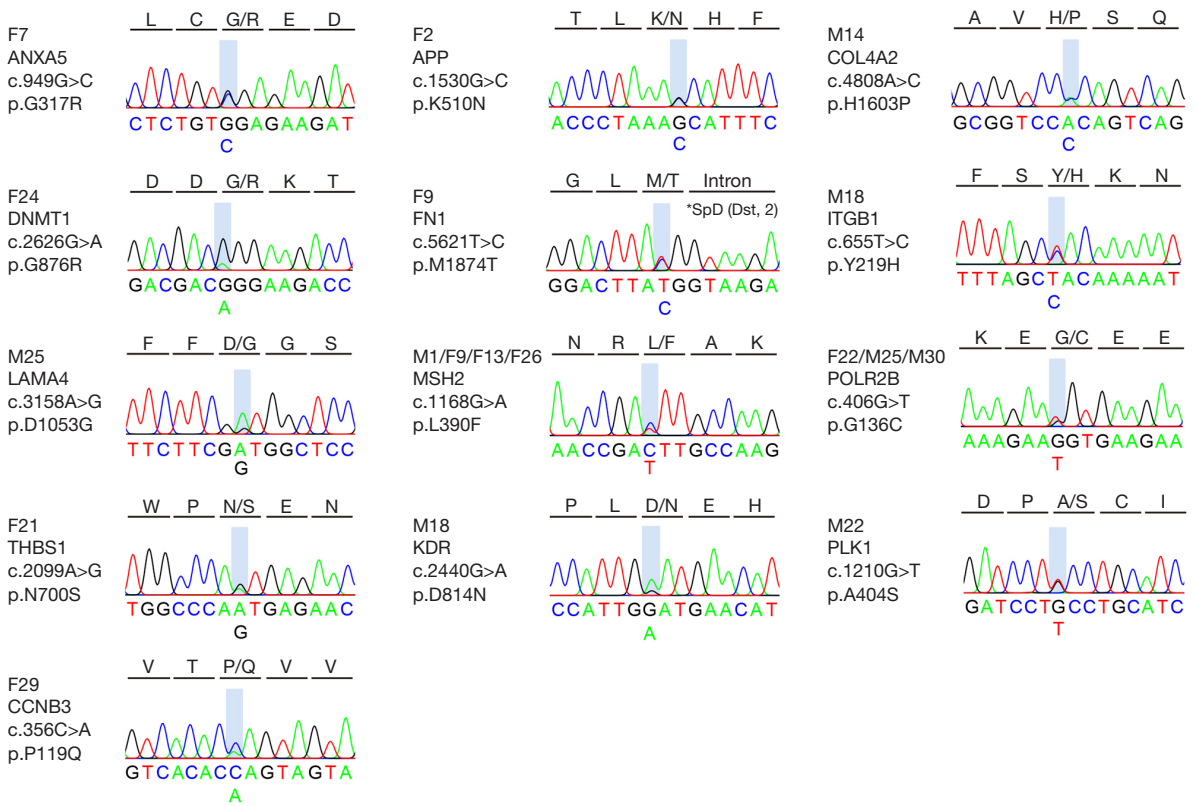


Figure 5 Sanger validations of the *ANXA5*, *APP*, *COL4A2*, *DNMT1*, *FN1*, *ITGB1*, *LAMA4*, *MSH2*, *POLR2B*, *THBS1*, *KDR*, *PLK1*, and *CCNB3* variants.

interactions. *THBS1* has been shown to play roles in platelet aggregation and angiogenesis (27). The results suggested that the mutations of *ANXA5* and *THBS1* may be disease susceptibility genes in URSA, and may be a novel diagnostic and anticoagulant therapy biomarker in URSA patients.

The enzyme encoded by DNA methyltransferase 1 (*DNMT1*) transfers methyl groups to cytosine nucleotides of genomic DNA, which is responsible for maintaining methylation patterns following DNA replication. Based on the GO dataset, *DNMT1* was shown to be involved in muscle structure development, positive regulation of cellular component biogenesis, muscle cell differentiation, muscle cell proliferation, and peptidyl-lysine modification. *DNMT1* is found in the decidua and chorionic villus of normal early pregnancy. The presence of *DNMT1* is important for normal embryogenesis (28). Interference of *DNMT1* impairs embryonic development and inhibits the attachment between endometrial cells and embryos (29). In addition, downregulation of *DNMT1* leads to URSA during the first trimester (28). The placenta is an important temporary tissue that supports fetal growth by regulating the transport of nutrients, gas exchange, and hormone production, all of which are necessary to maintain fetal viability during pregnancy (30,31). It is the organ of tissue binding between the mother and the fetus, which is formed from the membrane of the embryo and the endometrium of the mother. Fibronectin 1 (*FN1)* encodes fibronectin, which is involved in cell adhesion and migration, including embryogenesis, wound healing, blood coagulation, host defense, and metastasis. Based on the GO dataset, *FN1* is involved in embryonic morphogenesis, heart development, blood vessel development, and developmental growth. Lin *et al.* found that *FN1* plays an important role in the placenta and is highly expressed in placental tissues (32). In addition, *FN1* is closely associated with the invasion of human trophoblast. The overexpression of *FN1* inhibits the apoptosis of human trophoblasts by activating the PI3K/Akt signaling pathway (33).

Amyloid precursor protein (*APP*), a protein coding gene, encodes a cell surface receptor and transmembrane precursor protein that are cleaved by secretases to form a number of peptides. *APP* functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion, and axonogenesis. Additionally, *APP* is associated with cell maturation, positive regulation of cytokine production, positive regulation of stress-activated MAPK cascade, positive regulation of stress-activated protein kinase signaling cascade, and sex differentiation. Interaction between APP molecules and neighboring cells can promote synaptogenesis (34). Herein, the mutation of *APP* (c.1530G>C; p.K510N) was detected in URSA patients, suggesting that *APP* may be involved in neurogenesis in fetal development. MutS homolog 2 (*MSH2*) is a component of the post-replicative DNA mismatch repair system. *MSH2* is involved in chordate embryonic development, embryo development ending in birth or egg hatching, gamete generation, cellular process involved in reproduction in multicellular organism, and germ cell development. Cyclin B3 (*CCNB3*), a maternal mutated gene, is highly conserved in eukaryotic evolution and plays an important role in female meiosis. Lack of *CCNB3* will affect meiosis I and lead to female infertility (35-37). GeneCard analyses showed that *CCNB3* is a protein-coding gene, and the encoded proteins belong to the highly conserved cyclin family, whose members are characterized by significant periodicity of protein abundance throughout the cell cycle. Cyclins function as positive regulators of cyclin-dependent kinase and thus play an important role in cell cycle control, particularly through their disruption during cell division. Its tissue specificity suggests that *CCNB3* may be required during early meiotic prophase I. Several studies have linked this gene to RSA (38,39). In addition, the *CCNB3*-V1251D variant is involved in the pathogenesis of RSA associated with fetus triploidy (38). It can be seen that mutations of maternal genes may affect anticoagulation, trophoblast invasion, placental formation, embryonic neurogenesis, embryonic morphogenesis and gametogenesis.

Kinase insert domain receptor (*KDR*) acts as a cell-surface receptor for vascular endothelial growth factor (*VEGF)A*, *VEGFC*, and *VEGFD* and plays an essential role in the regulation of angiogenesis, vascular development, vascular permeability, and embryonic hematopoiesis. *KDR* is crucial in placentation, oocyte maturation, embryo implantation, and fetal development (40,41). These reports support the roles of *KDR* in human early pregnancy. The integrin subunit beta 1 (*ITGB1*) integrin family members are membrane receptors involved in cell adhesion and recognition in a variety of processes including embryogenesis, hemostasis, tissue repair, immune response, and metastatic diffusion of tumor cells. Based on GO dataset, *ITGB1* is believed to be involved in embryonic morphogenesis, tissue morphogenesis, microtubule cytoskeleton organization, blood vessel development, and

vasculature development. In the present study, mutations of *KDR* (c.2440G>A; p.D814N) and *ITGB1* (c.655T>C; p.Y219H) were found in URSA patients, suggesting that *KDR* and *ITGB1* may play key roles in fetal development.

The polo-like kinases (PLKs), a family of four serine/threonine protein kinases, are important regulators of cell cycle progression, mitosis, cytokinesis, and the DNA damage response. *PLK1* is recruited to the kinetochore and is required for proper chromosomal segregation during cell division (42). In addition, *PLK1* is required for centrosome maturation and separation prior to metaphase I, leading to the formation of bipolar metaphase I spindles. *PLK1* is vital for proficient centrosome biogenesis and accurate chromosome segregation during spermatogenesis (43). Herein, pathogenic mutations of *PLK1* (c.1210G>T; p.A404S) were detected in the husband of a patient who had two miscarriages. The mutation frequency in the population is 0.000084. Indeed, *PLK1* may be involved in the process of spermatogenesis.

RNA polymerase II subunit B (*POLR2B*) encodes the second largest subunit of RNA polymerase II (Pol II), a DNA-dependent RNA polymerase that catalyzes the transcription of DNA into precursors of mRNA, small nuclear RNA (snRNA), and microRNA (miRNA). Based on analysis of the GeneCards dataset, *POLR2B* is involved in chromatin binding, response to growth factor, cellular response to growth factor stimulus, chromosomal region, and the transmembrane receptor protein tyrosine kinase signaling pathway. In this study, a mutation of *POLR2B* (c.406G>T; p.G136C) was found in URSA patients, suggesting that *POLR2B* may be involved in chromatin binding. BUB1 Mitotic Checkpoint Serine/Threonine Kinase B(*BUB1B*)encodes a kinase involved in spindle checkpoint function. It is an essential component of the mitotic checkpoint. Matrix metallopeptidase 2 (*MMP2*) is a member of the matrix metalloproteinase (*MMP*) gene family, that are zinc-dependent enzymes capable of cleaving components of the extracellular matrix and molecules involved in signal transduction. Based on analysis by GO dataset, *MMP2* is involved in blood vessel development, vasculature development, blood vessel morphogenesis, gastrulation, skeletal system development and skeletal system development. *MMP2* and acrosin are major proteinases associated with the inner acrosomal membrane and may cooperate in sperm penetration of the zona pellucida during fertilization(44). Placental trophoblast invasion involves not only the regulation of cellular adhesion but also the remodeling and degradation of the extracellular matrix (ECM). This process involves the coordinated regulation of matrix metalloproteinases (MMPs) as well as their endogenous inhibitors, tissue inhibitors of metalloproteinases (TIMPs) (45). Matrix metalloproteinases play important regulatory roles in implantation and placentation to ensure a successful pregnancy (46). It can be seen that mutations of paternal genes may affect fetal development, neurogenesis, spermatogenesis and fertilization, placenta formation and placental trophoblast invasion.

Collagen type IV alpha 2 (*COL4A2*) is a major constituent of the trophoblast basement membrane. GO dataset analyses have suggested that *COL4A2* is involved in endoderm development, blood vessel development, formation of primary germ layer, extracellular matrix organization, and extracellular structure organization. It has also been shown to be important for placentation (47). Indeed, *COL4A2* is believed to be a preeclampsia susceptibility gene (48). Herein, mutations of *COL4A2* (c.4808 A>C; p.H1603P) were found in URSA patients, supporting the role of *COL4A2* in placentation. Laminin subunit alpha 4 (*LAMA4*) belongs to the family of extracellular matrix glycoproteins, and these are the major noncollagenous constituents of basement membranes. *LAMA4* is implicated in a wide variety of biological processes, including cell adhesion, differentiation, migration, signaling, neurite outgrowth, metastasis, blood vessel development, and skeletal system development. Laminin encoded by *LAMA2* is an extracellular protein, is a major component of the basement membrane. It is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components. *LAMA2* is an important adhesion molecule induced by the presence of canine pre-implantation embryos (49). Bromodomain PHD finger transcription factor (*BPTF*) is a histone-binding component of nucleosome remodeling factor (NURF), a complex which catalyzes ATP-dependent nucleosome sliding and facilitates transcription of chromatin. Based on analysis by GO dataset, *BPTF* is involved in brain development, chromatin remodeling, embryonic placenta development, endoderm development, negative regulation of transcription by RNA polymerase II. *BPTF* is a chromatin remodeling protein that is essential for embryonic stem cell differentiation and embryonic development (50). In addition, *BPTF* is a transcription factor of Heterogeneous Nuclear

Ribonucleoprotein L Like (hnRNPLL). HnRNPLL is a major global regulator of alternative splicing in embryonic stem cells, and deletion of hnRNPLL in mice leads to differentiation defects, developmental defects and growth disorders (51). Methyl CpG binding protein 2 (*MECP2*), a chromosomal protein that binds to methylated DNA, is an epigenetic factor involved in chromatin folding and transcriptional regulation. *MECP2* has been found to be critical for embryonic development (52). Maternal *MECP2*_e2 mutations lead to placental abnormalities that eventually lead to premature death in the carrier of the mutation (53). *MECP2* is associated with early neuronal maturation and may play a role in the first stage of neuronal progenitor cell proliferation and neurogenesis that gives rise to early neurons (54). *MECP2* loss leads to reduced responses to stimuli in early neuron maturation through malregulated expression of ion channels and glutamate receptors (55). SRY-box transcription factor 21 (*SOX21*) is associated with DNA-binding transcription activator activity, as well as RNA polymerase II-specific and stem cell differentiation. Heterogeneous gene expression of *SOX21* in the 4-cell stage contributes to cell-fate decisions (56). In embryonic stem cells, overexpression of *SOX21* triggers phenotypic differentiation of these cells into mesoderm and neuroectoderm, suggesting that *SOX21* plays a key role in neurogenesis regulation (57,58). Trophoblast Stem cells (TSCs) are multipotent progenitor cells that can develop into different cells in the embryonic part of the placenta. *Sox21* is found to be a TSCs and chorion-specific marker gene (59). Proper self-renewal and differentiation of TSCs are key factors for normal placental development and function. The level of *SOX21* is high in undifferentiated TSCs, and the imbalance of *SOX21* expression may lead to abnormal differentiation of TSCs and affect the development and growth of embryos in utero (60). *BPTF*, *MECP2* and *SOX21* were the top three strongly correlated mutations in the study population, the three mutations came from paternal, maternal and the same couple, which play an important role in stem cell differentiation and embryo development. These paternal and/or maternal mutations play important roles in embryo implantation, placenta formation, stem cell differentiation and embryo development.

Based on KEGG analysis of the variants, *CCNB3* is thought to be involved in cell cycle signaling pathways. *LAMA4*, *LAMA2*, *FN1*, *THBS1*, and *KDR* are involved in the PI3K-Akt and focal adhesion signaling pathways. Signaling pathways of cell cycle, PI3K-Akt, and focal adhesion are believed to be involved in RSA (61,62). This suggested that the mutation in these genes may be involved in URSA by regulating the above signaling pathways.

Genetic structure includes the total number of functional variations that affect a trait or disease, as well as the frequency and effect size of each variation. Many genetic effects work in an additive manner towards a predisposition to underlying disease, which is caused by a cumulative burden of risk factors (63). In recent years, GWASs have found that most variations are common (18). From the clinical point of view, RSA is a common and complex disease, in other words, it is a multi-factor or multi-gene disease. In the two couples who had six or more miscarriages, no gene mutations that resulted in abnormal protein function were detected. These couples may have multiple small effect abnormal genes that have not been screened. Each small effect gene has a small effect on URSA, which alone will not cause URSA, while multiple gene abnormalities may cause URSA.

During the study period, one-third of the patients had a normal pregnancy and birth. In URSA women with the *ANXA5* gene mutation, normal prethrombotic status was noted, in addition to a normal pregnancy and delivery after anticoagulant therapy, supporting the role of *ANXA5* in anticoagulation. In addition, patients with *THBS1* mutations also had normal pregnancy and delivery after anticoagulant therapy. In some patients, the reason for the repeated abortion may be an abnormal gamete gene, but there is a probability of normal gamete formation. However, some patients failed to achieve pregnancy and give birth despite anticoagulant therapy or immunotherapy, suggesting that the root cause in such patients was not abnormal coagulation and immunity, but other genetic abnormalities. Phenotypes are the result of interaction between genes and genes, or genes and environment. Couples with the dominant disease-causing mutation had normal phenotypes. However, by giving their faulty genes to gametes, the resulting zygote failed to grow and develop properly. Therefore, it is a combination of mutated genes from both parents.

There are some limitations to this study. First, the GO/KEGG results of pathogenic or possibly pathogenic genes after the initial filtration identified many genes related to reproduction. Only the most significant genes were further analyzed, and some genes that were not significant but are important in RSA may have been missed. Second, the

functional effect of candidate mutations should be evaluated by *in vitro* cell model and *in vivo* animal model experiments to determine whether the mutations are causative. The RSA-associated genes identified in this study may be potential biomarkers for the detection of patients with recurrent miscarriages. In the future, more and more studies have confirmed the relationship between genes and RSA. We can conduct targeted treatments according to gene functions through the detection of related gene Pannel, adopt *in vitro* fertilization, embryo transfer and its derivative technologies to reduce the burden of medical resources and patients.

Acknowledgments

Funding: This study was funded by the Science and Technology Research Project of Chongqing Education Commission (No. KJQN202100427) and the Young and Middle-Aged Medical Talents Project of Chongqing Municipal Health Commission/Chongqing Science and Technology Bureau (No. 2022GDRC001).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-2179/rc

Data Sharing Statement: Available at https://atm.amegroups.com/article/view/10.21037/atm-22-2179/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-2179/coif). All authors report that this study was funded by the Science and Technology Research Project of Chongqing Education Commission (No. KJQN202100427) and the Young and Middle-Aged Medical Talents Project of Chongqing Municipal Health Commission/Chongqing Science and Technology Bureau (No. 2022GDRC001). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Third Affiliated Hospital of Chongqing Medical University (No. 2019-9) and informed consent was taken from all the patients.

References

1. ESHRE Guideline Group on RPL; Bender Atik R, Christiansen OB, et al. ESHRE guideline: recurrent pregnancy loss. Hum Reprod Open 2018;2018:hoy004.
2. Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. Fertil Steril 2012;98:1103-11.
3. Suzumori N, Sugiura-Ogasawara M. Genetic factors as a cause of miscarriage. Curr Med Chem 2010;17:3431-7.
4. Saravelos SH, Regan L. Unexplained recurrent pregnancy loss. Obstet Gynecol Clin North Am 2014;41:157-66.
5. Christiansen OB, Nielsen HS, Pedersen B. Active or passive immunization in unexplained recurrent miscarriage. J Reprod Immunol 2004;62:41-52.
6. Peng L, Chelariu-Raicu A, Ye Y, et al. Prostaglandin E2 Receptor 4 (EP4) Affects Trophoblast Functions via Activating the cAMP-PKA-pCREB Signaling Pathway at the Maternal-Fetal Interface in Unexplained Recurrent Miscarriage. Int J Mol Sci 2021;22:9134.
7. Shi X, Xie X, Jia Y, et al. Maternal genetic polymorphisms and unexplained recurrent miscarriage: a systematic review and meta-analysis. Clin Genet 2017;91:265-84.
8. Wang WJ, Hao CF, Qu QL, et al. The deregulation of regulatory T cells on interleukin-17-producing T helper cells in patients with unexplained early recurrent miscarriage. Hum Reprod 2010;25:2591-6.
9. Manning M, Hudgins L; Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med 2010;12:742-5.

10. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011;12:745-55.
11. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet 2014;59:5-15.
12. Tsurusaki Y, Yonezawa R, Furuya M, et al. Whole exome sequencing revealed biallelic IFT122 mutations in a family with CED1 and recurrent pregnancy loss. Clin Genet 2014;85:592-4.
13. Walczak-Sztulpa J, Eggenschwiler J, Osborn D, et al. Cranioectodermal Dysplasia, Sensenbrenner syndrome, is a ciliopathy caused by mutations in the IFT122 gene. Am J Hum Genet 2010;86:949-56.
14. Filges I, Manokhina I, Peñaherrera MS, et al. Recurrent triploidy due to a failure to complete maternal meiosis II: whole-exome sequencing reveals candidate variants. Mol Hum Reprod 2015;21:339-46.
15. Qiao Y, Wen J, Tang F, et al. Whole exome sequencing in recurrent early pregnancy loss. Mol Hum Reprod 2016;22:364-72.
16. Quintero-Ronderos P, Mercier E, Fukuda M, et al. Novel genes and mutations in patients affected by recurrent pregnancy loss. PLoS One 2017;12:e0186149.
17. Xiang H, Wang C, Pan H, et al. Exome-Sequencing Identifies Novel Genes Associated with Recurrent Pregnancy Loss in a Chinese Cohort. Front Genet 2021;12:746082.
18. Claussnitzer M, Cho JH, Collins R, et al. A brief history of human disease genetics. Nature 2020;577:179-89.
19. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. Science 2021;373:1464-8.
20. Krikun G, Lockwood CJ, Wu XX, et al. The expression of the placental anticoagulant protein, annexin V, by villous trophoblasts: immunolocalization and in vitro regulation. Placenta 1994;15:601-12.
21. Thiagarajan P, Tait JF. Binding of annexin V/placental anticoagulant protein I to platelets. Evidence for phosphatidylserine exposure in the procoagulant response of activated platelets. J Biol Chem 1990;265:17420-3.
22. Ang KC, Kathirgamanathan S, Ch'ng ES, et al. Genetic analysis of the M2/ANXA5 haplotype as recurrent pregnancy loss predisposition in the Malay population. J Assist Reprod Genet 2017;34:517-24.
23. Bogdanova N, Horst J, Chlystun M, et al. A common haplotype of the annexin A5 (ANXA5) gene promoter is associated with recurrent pregnancy loss. Hum Mol Genet 2007;16:573-8.
24. Danisik H, Bogdanova N, Markoff A. Micromolar Zinc in Annexin A5 Anticoagulation as a Potential Remedy for RPRGL3-Associated Recurrent Pregnancy Loss. Reprod Sci 2019;26:348-56.
25. Rai RS, Regan L, Clifford K, et al. Antiphospholipid antibodies and beta 2-glycoprotein-I in 500 women with recurrent miscarriage: results of a comprehensive screening approach. Hum Reprod 1995;10:2001-5.
26. Lan Y, Wang J, Zhang Q, et al. Genetic variations and haplotypes in the annexin A5 gene are associated with the risk of recurrent pregnancy loss. J Cell Physiol 2019;234:18308-16.
27. Seif K, Alidzanovic L, Tischler B, et al. Neutrophil-Mediated Proteolysis of Thrombospondin-1 Promotes Platelet Adhesion and String Formation. Thromb Haemost 2018;118:2074-85.
28. Fatima N, Ahmed SH, Chauhan SS, et al. Structural equation modelling analysis determining causal role among methyltransferases, methylation, and apoptosis during human pregnancy and abortion. Sci Rep 2020;10:12408.
29. Yin LJ, Zhang Y, Lv PP, et al. Insufficient maintenance DNA methylation is associated with abnormal embryonic development. BMC Med 2012;10:26.
30. Chen Q, Pang PC, Cohen ME, et al. Evidence for Differential Glycosylation of Trophoblast Cell Types. Mol Cell Proteomics 2016;15:1857-66.
31. Burton GJ, Fowden AL. The placenta: a multifaceted, transient organ. Philos Trans R Soc Lond B Biol Sci 2015;370:20140066.
32. Lin P, Lai X, Wu L, et al. Network analysis reveals important genes in human placenta. J Obstet Gynaecol Res 2021;47:2607-15.
33. Ji J, Chen L, Zhuang Y, et al. Fibronectin 1 inhibits the apoptosis of human trophoblasts by activating the PI3K/Akt signaling pathway. Int J Mol Med 2020;46:1908-22.
34. Baumkötter F, Schmidt N, Vargas C, et al. Amyloid precursor protein dimerization and synaptogenic function depend on copper binding to the growth factor-like domain. J Neurosci 2014;34:11159-72.
35. Jacobs HW, Knoblich JA, Lehner CF. Drosophila Cyclin B3 is required for female fertility and is dispensable for mitosis like Cyclin B. Genes Dev 1998;12:3741-51.
36. Karasu ME, Bouftas N, Keeney S, et al. Cyclin B3 promotes anaphase I onset in oocyte meiosis. J Cell Biol 2019;218:1265-81.
37. Li Y, Wang L, Zhang L, et al. Cyclin B3 is required for

metaphase to anaphase transition in oocyte meiosis I. J Cell Biol 2019;218:1553-63.
38. Fatemi N, Salehi N, Pignata L, et al. Biallelic variant in cyclin B3 is associated with failure of maternal meiosis II and recurrent digynic triploidy. J Med Genet 2021;58:783-8.
39. Rezaei M, Buckett W, Bareke E, et al. A protein-truncating mutation in CCNB3 in a patient with recurrent miscarriages and failure of meiosis I. J Med Genet 2021;59:568-70.
40. Zygmunt M, Herr F, Münstedt K, et al. Angiogenesis and vasculogenesis in pregnancy. Eur J Obstet Gynecol Reprod Biol 2003;110 Suppl 1:S10-8.
41. Jackson MR, Carney EW, Lye SJ, et al. Localization of two angiogenic growth factors (PDECGF and VEGF) in human placentae throughout gestation. Placenta 1994;15:341-53.
42. Hood EA, Kettenbach AN, Gerber SA, et al. Plk1 regulates the kinesin-13 protein Kif2b to promote faithful chromosome segregation. Mol Biol Cell 2012;23:2264-74.
43. Wellard SR, Zhang Y, Shults C, et al. Overlapping roles for PLK1 and aurora A during meiotic centrosome biogenesis in mouse spermatocytes. EMBO Rep 2021;22:e54106.
44. Ferrer M, Rodriguez H, Zara L, et al. MMP2 and acrosin are major proteinases associated with the inner acrosomal membrane and may cooperate in sperm penetration of the zona pellucida during fertilization. Cell Tissue Res 2012;349:881-95.
45. Cohen M, Meisser A, Bischof P. Metalloproteinases and human placental invasiveness. Placenta 2006;27:783-93.
46. Yan Y, Fang L, Li Y, et al. Association of MMP2 and MMP9 gene polymorphisms with the recurrent spontaneous abortion: A meta-analysis. Gene 2021;767:145173.
47. Furuhashi N, Kimura H, Nagae H, et al. Serum collagen IV and laminin levels in preeclampsia. Gynecol Obstet Invest 1994;37:250-3.
48. Johnson MP, Fitzpatrick E, Dyer TD, et al. Identification of two novel quantitative trait loci for pre-eclampsia susceptibility on chromosomes 5q and 13q using a variance components-based linkage approach. Mol Hum Reprod 2007;13:61-7.
49. Graubner FR, Gram A, Kautz E, et al. Uterine responses to early pre-attachment embryos in the domestic dog and comparisons with other domestic animal species. Biol Reprod 2017;97:197-216.
50. Landry J, Sharov AA, Piao Y, et al. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. PLoS Genet 2008;4:e1000241.
51. Wang X, Ping C, Tan P, et al. hnRNPLL controls pluripotency exit of embryonic stem cells by modulating alternative splicing of Tbx3 and Bptf. EMBO J 2021;40:e104729.
52. Tate P, Skarnes W, Bird A. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat Genet 1996;12:205-8.
53. Itoh M, Tahimic CG, Ide S, et al. Methyl CpG-binding protein isoform MeCP2_e2 is dispensable for Rett syndrome phenotypes but essential for embryo viability and placenta development. J Biol Chem 2012;287:13859-67.
54. Cobolli Gigli C, Scaramuzza L, De Simone M, et al. Lack of Methyl-CpG Binding Protein 2 (MeCP2) Affects Cell Fate Refinement During Embryonic Cortical Development. Cereb Cortex 2018;28:1846-56.
55. Bedogni F, Cobolli Gigli C, Pozzi D, et al. Defects During Mecp2 Null Embryonic Cortex Development Precede the Onset of Overt Neurological Symptoms. Cereb Cortex 2016;26:2517-29.
56. Goolam M, Scialdone A, Graham SJL, et al. Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. Cell 2016;165:61-74.
57. Mallanna SK, Ormsbee BD, Iacovino M, et al. Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. Stem Cells 2010;28:1715-27.
58. Sandberg M, Källström M, Muhr J. Sox21 promotes the progression of vertebrate neurogenesis. Nat Neurosci 2005;8:995-1001.
59. Kuales G, Weiss M, Sedelmeier O, et al. A Resource for the Transcriptional Signature of Bona Fide Trophoblast Stem Cells and Analysis of Their Embryonic Persistence. Stem Cells Int 2015;2015:218518.
60. Moretto Zita M, Soncin F, Natale D, et al. Gene Expression Profiling Reveals a Novel Regulatory Role for Sox21 Protein in Mouse Trophoblast Stem Cell Differentiation. J Biol Chem 2015;290:30152-62.
61. Wang JM, Gu Y, Zhang Y, et al. Deep-sequencing identification of differentially expressed miRNAs in decidua and villus of recurrent miscarriage patients. Arch Gynecol Obstet 2016;293:1125-35.
62. Cui Y, He L, Yang CY, et al. iTRAQ and PRM-based

quantitative proteomics in early recurrent spontaneous abortion: biomarkers discovery. Clin Proteomics 2019;16:36.
63. Visscher PM, Yengo L, Cox NJ, et al. Discovery and implications of polygenicity of common diseases. Science 2021;373:1468-73.

(English Language Editor: J. Teoh)

Cite this article as: Mou JT, Huang SX, Yu LL, Xu J, Deng QL, Xie YS, Deng K. Identification of genetic polymorphisms in unexplained recurrent spontaneous abortion based on whole exome sequencing. Ann Transl Med 2022;10(10):603. doi: 10.21037/atm-22-2179