Novel variants in women with premature ovarian function decline identified via whole-exome sequencing

Ruiyi Tang¹ · Qi Yu¹

Received: 1 May 2020 / Accepted: 6 August 2020 / Published online: 13 August 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose To investigate the potential etiologies of premature ovarian insufficiency (POI) and diminished ovarian reserve (DOR). **Methods** Fourteen women with sporadic POI and 6 women with DOR were enrolled. We used whole-exome sequencing (WES) and bioinformatics analysis to identify variants in a subset of 599 selected POI candidate genes. The identified genes were subjected to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment and protein-protein interaction (PPI) network analyses to uncover key genes and pathways.

Results Among the 20 patients, 79 heterozygous variants were detected in 49 genes, which were classified as “likely pathogenic” or “variants of uncertain significance” according to the guidelines of the American College of Medical Genetics and Genomics. Most patients (17/20) carried two or more variants. Monoacylglycerol *O*-acyltransferase 1 mutations were found in six patients, and cytochrome P450 family 26 subfamily B member 1 and Bardet-Biedl syndrome 9 mutations were each found in four patients. Some variants were shared between DOR and POI. Enrichment analyses showed that the identified genes participate in key ovarian processes, such as follicular development, gonadal development, meiosis, Fanconi anemia, homologous recombination, and transforming growth factor β signaling. A PPI network revealed interactions between these proteins.

Conclusion Premature ovarian function decline may be polygenic, and overlap exists between the genetic backgrounds of DOR and POI. WES and in silico analyses may be a useful clinical tool for etiological diagnosis and risk prediction for high-risk women in the future.

Keywords Premature ovarian insufficiency · Diminished ovarian reserve · Whole-exome sequencing · Enrichment analysis

Introduction

The age of menopause depends on the complex balance of forces governing the underlying formation and progressive loss of ovarian follicles. Premature ovarian insufficiency (POI) is characterized by loss of ovarian function before the age of 40, and is a major cause of female infertility [1]. POI is common, affecting approximately 1–3% of women [2]. In fact, a very recent meta-

analysis calculated a global POI prevalence of 3.7% [3]. Young women with POI live a relatively longer portion of their lives in estrogen deficiency, significantly increasing their risk of cardiovascular disease and osteoporosis [1]. Therefore, accurate risk prediction and early diagnosis would greatly benefit these women.

Ovarian reserve decline occurs gradually [4] and the concept of diminished ovarian reserve (DOR) has been explored [5]. Women with DOR typically have regular menses; however, they have reduced numbers of ovarian follicles and reduced fecundity compared with unaffected women of similar age [6]. DOR may or may not progress into POI (depending on whether amenorrhea occurs before the age of 40).

A small pool of primordial follicles, premature follicle depletion due to accelerated atresia, and disturbances in follicle function are major factors in the development of ovarian reserve decline [4, 7]. The disorder is clinically and etiologically heterogeneous. Known causes of POI include genetic

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10815-020-01919-y>) contains supplementary material, which is available to authorized users.

✉ Qi Yu
yuqi2008001@sina.com

¹ Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Science, Beijing 100730, People's Republic of China

aberrations, autoimmune ovarian damage, and iatrogenic and environmental factors; however, the etiology of most cases remains unknown. Over 75 POI-associated genes have been identified, and genetic factors account for approximately 20–25% of cases [7–9]. Changes in fragile X mental retardation 1 (*FMR1*) are a leading cause of POI, with *FMR1* premutations occurring in an estimated 11–14% and 2–6% of familial and sporadic POI cases, respectively, in western countries [10]. Women with DOR may also harbor premutation-containing alleles [10]. In addition, mutations in genes related with gonadal development, hormonal signaling, meiosis, and DNA repair have been associated with POI [11]. However, the genetic causes of up to 90% of POI cases are undetermined [9], suggesting that additional causative genes remain to be discovered. Few studies have examined gene mutations associated with DOR, and the relationship between POI and DOR requires further evaluation.

Currently, only karyotyping and *FMR1* screening are recommended during clinical genetic testing for POI [12], with no recommendations to examine any other specific genes. Recently, studies have shown that next-generation sequencing (NGS) can efficiently map novel variants participating in POI etiology, which could be used in clinical diagnosis in the near future [13]. Many new genes associated with POI etiology have been identified via whole-genome sequencing or whole-exome sequencing (WES), especially in consanguineous and large POI pedigrees [7, 8]. Meanwhile, only a few studies have performed NGS on women with sporadic POI (i.e., affected women without a family history of POI) [14–16], and NGS have not been used to analyze women with DOR.

In this pilot study, WES was performed on women affected by sporadic POI or DOR. The analysis focused on 599 candidate POI-related genes, and bioinformatics analysis was used to identify promising candidate variants for future study. These analyses provide insight on the biological mechanisms associated with POI development, which may inform the development of tools for the early diagnosis of premature ovarian aging.

Materials and methods

Participants

Fourteen women affected by idiopathic POI and 6 women diagnosed with DOR were recruited from the Clinic of Gynecological Endocrinology and Reproductive Medicine of Peking Union Medical College Hospital (PUMCH) in the period between May 2018 and April 2019. All women with POI were amenorrhea before 40 years old and had elevated serum follicle-stimulating hormone (FSH) levels (> 25 IU/L on two occasions one month apart). The criteria for DOR

included regular menstrual cycles, FSH values > 10 IU/L on days 2–4 of the spontaneous menstrual cycle and/or anti-Müllerian hormone (AMH) values ≤ 1.1 ng/ml before the age of 40 years. All women in both groups have no family history of POI or DOR, and had a normal 46,XX karyotype. *FMR1* premutations were excluded. Women having a background of pelvic surgery, polycystic ovarian syndrome, endometriosis, ovarian infections, chemotherapy, radiotherapy, endocrine disorders, and/or autoimmune disease were excluded from the study.

This study protocol was granted by the Institutional Review Board (IRB) of PUMCH (No. JS-1604) and the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All individuals signed a written informed consent form.

Whole-exome sequencing and variant analysis

After searching public databases, such as PubMed, HGMD, Highwire, Genecards, and Geoprofiles, a 599 genes list (Supplementary Table S1) was selected. When searching the literature, the combinations of the following keywords were included: “primary ovarian insufficiency,” “premature ovarian insufficiency,” “premature ovarian failure,” “POI/POF genetics,” “mouse models of premature ovarian failure,” “diminished ovarian reserve,” “early menopause,” “gene expression and ovary,” “hypergonadotropic hypogonadism,” “genetics of sex determination,” “genetics of gametogenesis,” “molecular regulation of meiosis,” “genetics of folliculogenesis,” “ovulation genetics,” and “ovary and transcriptomics.” These genes were considered candidates as they had been reported as having expression/function during processes of sex determination, meiosis, folliculogenesis, and ovulation, or were identified in patients with POI.

Blood samples were obtained from each participant for molecular studies. Genomic DNA was extracted from peripheral blood leukocytes using the TIANamp Blood DNA Kit (TIANGEN Biotech, Beijing, China) following standard procedures. WES of the genomic DNA was performed, and all samples were sequenced using the Illumina HiSeq X Ten with 150 bp paired-end reads. All sequenced data were mapped to the human reference genome (human genome 19, *hg19*) using Burrows-Wheeler Alignment tool (BWA) version 0.7.12 (<http://bio-bwa.sourceforge.net/>). GATK software was used for the realignment correction nearby the insertions/deletions (InDels) and statistical analysis of the genome coverage depth. Based on the calibration results, GATK version 3.5.0 software was used for single nucleotide variations (SNVs)/InDel identification. The results were stored in Variant Call Format (VCF) files. The ANNOVAR tool was adopted to annotate single nucleotide polymorphisms (SNPs) and InDels and to

produce statistical analyses of genome-wide SNVs/InDels distribution. The VCF file was annotated using ANNOVAR software. Only variants located on the 599 genes list were extracted for further analysis.

The annotated SNVs fulfilling the following criteria were retained: (1) sequence variants reported as having minor allele frequencies (MAF) of 5% or less in the 1000 genomes project (<http://www.internationalgenome.org/>), gnomAD (<http://gnomad.broadinstitute.org/>), and ExAC (<http://exac.broadinstitute.org/>); (2) variants with potential effects at the protein sequence level (e.g., missense, nonsense, splice site, and frameshift mutations) for downstream analysis; (3) SNVs/InDels with average sequencing depths > 40; and (4) nonsynonymous SNVs predicted to be deleterious or damaging by analysis of amino acid substitutions in silico using three prediction algorithms (SIFT, PolyPhen-2, and PROVEAN) based on scores of < 0.05 (SIFT), > 0.85 (PolyPhen-2), and ≤ -2.5 (PROVEAN).

All filtered candidate sequence variants obtained by WES were validated by Sanger sequencing of amplified genomic DNA from the original samples. Target regions containing candidate variations were amplified by polymerase chain reaction (PCR) using corresponding primers (the sequences of the primers are available upon request). The PCR products were purified by polyethylene glycol precipitation followed by labeling with an ABI-PRISM BigDye Terminator Ready Reaction Cycle Sequencing Kit (Applied Biosystems, USA) and sequencing on a 3730xl DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions. The sequencing data were analyzed using CHROMAS software version 1.62. Variants confirmed by Sanger sequencing were retained.

Since our population was exclusively Chinese, we used the EAS (East Asian) population variant frequencies in the gnomAD database for comparisons. Associations between identified variants and POI/DOR phenotypes were assessed by comparing the frequency of the variants in our study and in the EAS population. Allele frequencies were compared using Fisher's exact test, and $P < 0.05$ was considered statistically significant. Statistical analysis was performed in SPSS version 24 (IBM, USA).

Each variant was evaluated using the most recent guidelines from the American College of Medical Genetics and Genomics (ACMG) [17]. Variants were sorted into five classes: pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, or benign. We considered pathogenic and likely pathogenic variants to be causative. Variants with MAFs > 5% in the public databases were considered benign and excluded. Variants not present in the 1000 Genomes Project, ExAC, gnomAD, or Single Nucleotide Polymorphism database were considered novel. Novelty and very low population frequency were considered moderate evidence for pathogenicity (PM2). Nonsense, frameshift,

canonical ± 1 or 2 splice sites, and initiation codon variants in genes where loss of function are known mechanisms of disease were recognized as evidence for pathogenicity (PVS1). Significantly increased prevalence in our study compared with that in the general population ($P < 0.05$) was recognized as pathogenic (PS4). Nonsynonymous variants predicted to be deleterious or damaging in in silico analysis were assigned as pathogenic (PP3). If a variant could not be unanimously identified as damaging or benign, it was designated as a VUS.

Functional enrichment analysis and biological signaling pathway analysis

The Gene Ontology (GO, www.geneontology.org) [18] analysis is common genes and gene product annotating method. GO database can provide a functional classification for genomic data from three aspects: biological processes (BP), cellular component (CC), and molecular function (MF). The Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.ad.jp/kegg/) [19] database is a knowledge base for systematic analysis, annotation, and visualization of gene functions and relationships between genes. We used the R package clusterProfiler [20] to conduct the GO enrichment and KEGG pathway analysis of the gene list. The enrichment score was assessed by the P value, indicating the significance of the pathway associated with the conditions.

The Search Tool for the Retrieval of Interacting Genes (STRING, stringdb.org) is a biological database designed to predict protein-protein interaction (PPI) information [21]. We used STRING for constructing a PPI network. The selected genes were mapped to STRING to evaluate the interactive relationships, with a confidence score > 0.4 defined as significant. The predicted interaction includes both the physical interaction between proteins and the indirect functional interaction of proteins.

Results

We identified 109,557 sequence variants in the exomes of the 20 samples, including 1434 variants in the 599 POI genes (Fig. 1). After excluding synonymous variants and retaining variants in exon regions and splicing sites with MAFs < 0.05, 154 variants predicted to have deleterious effects by SIFT, PolyPhen-2, and PROVEAN were retained. After verification by Sanger sequencing, 79 variants in 49 genes were confirmed (Fig. 1). The 49 genes and the occurrence of each variant are listed in Table 1. These variants included 2 alternative splice sites, 11 frameshifts, 2 non-frameshift deletions, and 65 nonsynonymous variants. All variants were heterozygous.

Patient clinical characteristics and identified variants are shown in Table 2. Two women with POI displayed

Fig. 1 Screening of variants. Rectangle 1 with rounded edges encloses the 109,557 variants found throughout the exomes of 20 samples. The subset of premature ovarian insufficiency (POI) genes included 599 candidate genes. Rectangle 2 with rounded edges highlights the 1434 variants found among the 599 genes of the POI subset. Rectangle 3 with rounded edges represents the 970 variants detected in the exons and splicing sites, while rectangle 4 with rounded edges shows the 528 variants retained after excluding the synonymous variants and the variants having $MAF \geq 0.05$. In rectangle 5 with rounded edges, 154 variants were retained, while rectangle 6 includes 79 variants confirmed after verification using Sanger sequencing, and the last rectangle with rounded edges lists the numbers and types of identified variants. MAF: minor allele frequency

hypergonadotropic primary amenorrhea, and 12 exhibited secondary amenorrhea (mean menopausal age: 24.6 ± 9.3 years). All patients in the POI group had markedly elevated FSH (81.4 ± 24.6 IU/L) and luteinizing hormone (LH; 42.1 ± 15.4 IU/L) levels and low levels of estradiol ($22.9 \pm$

13.3 ng/mL). The six patients diagnosed with DOR had slightly increased FSH (16.4 ± 3.3 IU/L), normal LH (4.6 ± 1.3 IU/L), and relatively low estradiol (39.0 ± 28.8 ng/mL) levels. Monoacylglycerol O-acyltransferase 1 mutations were found in six patients, while cytochrome P450 family 26 subfamily B

Table 1 Identified genes selected in the 20 patients with ovarian function decline

Gene	Mutation/total	Gene	Mutation/total	Gene	Mutation/total
ABCC5	1/20	GATA6	1/20	PGRMC1	2/20
ADAMTSL3	1/20	GRIP1	1/20	PIWIL1	1/20
AIRE	1/20	HK3	1/20	PRDM1	2/20
ALMS1	1/20	HSD17B4	1/20	RAD54L	1/20
AOX1	1/20	KNTC1	2/20	SOHLH1	2/20
ARFGAP3	1/20	LATS1	1/20	SPIDR	2/20
BBS9	4/20	LEPR	1/20	SPO11	1/20
BMP8B	3/20	LHX9	1/20	STRA8	1/20
BMPR2	1/20	LMNA	1/20	TBP	1/20
CDC25B	1/20	MAP3K4	1/20	TCF21	1/20
CNOT6	1/20	MOGAT1	6/20	THBS1	1/20
CYP26B1	4/20	MSH6	2/20	TOP3B	1/20
DHRS11	1/20	MYO19	2/20	UMODL1	3/20
ERCC2	1/20	NLRP11	2/20	VCX	4/20
ERCC6	1/20	NOS3	1/20	ZNF729	3/20
FANCA	3/20	NOTCH2	1/20		
FANCL	1/20	PGAP3	2/20		

Table 2 Clinical and molecular findings of POI and DOR patients studied via whole-exome sequencing

Patient ID	Phenotype	Age of menarche	Age at diagnosis/menopause	FSH (IU/L)	LH (IU/L)	E ₂ (pg/ml)	AMH (ng/ml)	Gene	Accession number
D15	DOR	14	32	19.89	4.39	68.42	0.28	ABCC5	NM_001320032
								ADAMTSL3	NM_207517
								ALMS1	NM_015120
								CYP26B1	NM_019885
								ZNF729	NM_001242680
D18	DOR	14	30	15.17	6.05	78.15	0.25	AOX1	NM_0011159
								BBS9	NM_001348041
								GATA6	NM_005257
								TCF21	NM_198392
								UMODL1	NM_001199527
D23	DOR	14	30	17.07	3.95	38.06	0.25	BMP8B	NM_001720
								GRIPI	NM_021150
								HSD17B4	NM_001199291
								MAP3K4	NM_001291958
								MOGAT1	NM_058165
D29	DOR	12	32	12.33	2.44	25.07	0.14	MYO19	NM_001163735
								SPDR	NM_001080394
								VCX	NM_013452
								VCHX	NM_013452
								LHX9	NM_020204
D35	DOR	13	29	20.33	5.8	19.15	0.49	MOGAT1	NM_058165
								SPO11	NM_012444
								UMODL1	NM_001199527
								BMP8B	NM_001720
								MOGAT1	NM_058165
D36	DOR	14	32	13.58	4.99	5	0.24	PIWIL1	NM_004764
								SPDR	NM_001080394
								CDC25B	NM_021873
								MYO19	NM_001163735
								PRDM1	NM_001198
P39	PA	NA	14	75.57	28.97	11.8	0.07	ZNF729	NM_001242680
								LATS1	NM_001242680
								AIRE	NM_001350339
								BBS9	NM_000383
								CYP26B1	NM_001348041
P24	SA	15	16	110.7	41.48	10.1	<0.01	CYP26B1	NM_019885
								CNOT6	NM_001303241
								MOGAT1	NM_058165
								PGAP3	NM_033419
								RAD54L	NM_003579
P27	SA	11	15	64.6	47.87	18.86	<0.01	SOHLH1	NM_001286167
								FANCA	NM_001286167
								BBS9	NM_001348041
								FANCA	NM_001286167
								FANCA	NM_001286167

Table 2 (continued)

	Patient ID	Exon	Sequence variation	Protein position	1000G	gnomAD	ExAC	S/Po/Pr*	dbSNP
P35	SA	10	13	107.1	64.01	8	<0.01	FANCL	NM_001114636
								TOP3B	NM_001349852
								CYP26B1	NM_019885
								KNTC1	NM_014708
								MOGAT1	NM_058165
P37	SA	14	17	85.61	30.53	20.94	<0.01	NLRP11	NM_145007
								BBS9	NM_001348041
								ERCC6	NM_000124
								FANCA	NM_001286167
								PGRMC1	NM_006667
P61	SA	13	18	125.8	76.66	32.00	<0.01	VCX	NM_013452
								BMP8B	NM_001720
								NOTCH2	NM_024408
								PGAP3	NM_033419
								PGRMC1	NM_006667
P10	SA	13	30	55.34	28.50	34.00	0.11	MOGAT1	NM_058165
								MSH6	NM_000179
								STRA8	NM_182489
								TBP	NM_003194
								DHRS11	NM_024308
P15	SA	12	33	51.15	48.16	195	0.01	ERCC2	NM_000400
								HK3	NM_002115
								LMNA	NM_170707
								NOS3	NM_000603
								VCX	NM_013452
P1	SA	14	39	68.5	25.50	42	0.1	KNTC1	NM_014708
								LEPR	NM_001003679
								MSH6	NM_000179
								ARFGAP3	NM_014570
								SOHLH1	NM_003246
P31	SA	13	32	121	45.47	11.8	0.01	THBS1	NM_003246
								NLRP11	NM_145007
P32	SA	12	27	61.38	43.04	11.80	0.01	BMPR2	NM_001204
								CYP26B1	NM_019885
								PRDM1	NM_001198
D15		20 16 8 5 4	c.1478C>T c.1852C>T c.2351A>G c.967C>T c.2326G>C	p.T493M p.R618W p.E784G p.R323W p.A776P	0.0002 0.0016 0.00679 0.0026	0.0009 0.0008 0.0023 0.0014	0.00008 0.0007 0.0023 0.0018	D/D/D D/D/D -/D/D D/D/D	rs564676660
									rs139253965
									rs17848880
									rs138478634
									-
D18		21 21 2 1	c.2264C>T c.2470C>T c.43G>C c.10G>A	p.T755I p.R824C p.G15R p.G4S	0.002 0.00739 0.03355	0.0009 0.0026 0.0138	0.0009 0.0026 0.0137	D/D/D D/D/D D/D/D D/D/D	rs35217482
									rs146752751
									rs116262672
									-
									-

Table 2 (continued)

D23	16	c.3169_3170insCCAT	p.A1057fs	0.002	0.0021	-/-	rs20185268
	4	c.778C>T	p.R260C	0.0025	0.0032	D/D/D	rs199806017
	22	c.2846A>G	p.E949G			D/D/D	-
	24	c.2135C>T	p.T712I	0.0067	0.0067	D/P/D	rs28943592
D29	12	c.1301C>T	p.P434L	0.0003	0.0002	D/D/D	rs202168620
	3	c.415_416insATCTTCAC	p.Y139fs	0.0033	0.0032	-/-	rs147393916
	6	c.203C>G	p.A68G	0.0002	0.0002	-D/D	rs768388466
	7	c.815G>A	p.R272Q	0.00004	0.00004	D/D/D	rs182613962
D35	3	c.388_445del	p.V130fs	0.0064	0.0029	-/-	-
	3	c.385_386del	p.Q129fs	0.008	0.0027	-/-	rs752893781
	1	c.48delC	p.R16fs			-/-	-
	3	c.415_416insATCTTCAC	p.Y139fs	0.0033	0.0032	-/-	rs147393916
D36	1	c.38T>C	p.F13S	0.0002	0.0002	D/D/D	rs185741379
	16	c.3169_3170insCCAT	p.A1057fs	0.002	0.0021	-/-	rs20185268
	4	c.778C>T	p.R260C	0.0025	0.0032	D/D/D	rs199806017
	3	c.415_416insATCTTCAC	p.Y139fs	0.0033	0.0032	-/-	rs147393916
D37	15	c.1691G>C	p.R564P			D/D/D	-
	19	c.2569T>C	p.S857P	0.0004	0.0006	D/D/D	rs376975557
	10	c.945C>G	p.C315W	0.0014	0.0006	D/D/D	rs200104998
	6	c.203C>G	p.A68G	0.0002	0.0002	-D/D	rs768388466
D39	2	c.239A>T	p.E80V	0.0003	0.0002	D/D/D	rs202200155
	4	c.1007-1009GAA>CTG	p.336_337ArgLys>ThrGlu				
	4	c.1027A>G	p.T343A	0.0079	0.0002	D/D/D	rs764887966
	6	c.1766G>A	p.R589H	0.00001		D/D/D	-
P24	9	c.1051C>T	p.R351W	0.00005	0.0002	D/D/D	rs149078622
	21	c.2470C>T	p.R824C	0.0026	0.0026	D/D/D	rs146752751
	5	c.967C>T	p.R323W	0.0014	0.0018	D/D/D	rs138478634
	4	c.26C>T	p.P9L	0.0019	0.0019	D/D/D	rs75964385
P27	3	c.415_416insATCTTCAC	p.Y139fs	0.0033	0.0032	-/-	rs147393916
	6	c.683A>G	p.N228S	0.0021	0.0021	D/D/D	rs142596676
	5	c.354G>C	p.L118F	0.00002	0.00002	D/D/D	rs778622356
	4	c.346-1G>A		0.003		D/D/D	rs140132974
P28	33	c.3263C>T	p.S1088F	0.02336	0.051	D/P/D	rs17233497
	21	c.2470C>T	p.R824C	0.00739	0.0026	D/D/D	rs146752751
	33	c.3263C>T	p.S1088F	0.02336	0.051	D/P/D	rs17233497
	8	c.637G>A	p.D213N	0.0002	0.0001	D/D/D	rs199564543
P35	19	c.1597G>C	p.D533H			D/D/D	-
	5	c.967C>T	p.R323W	0.0026	0.0018	D/D/D	rs138478634
	44	c.4535T>G	p.V1512G	0.0016	0.0005	D/D/D	rs75696429
	3	c.415_416insATCTTCAC	p.Y139fs	0.0033	0.0032	-/-	rs147393916
P37	4	c.94C>T	p.R32C	0.0177	0.0104	-D/D	rs76935241
	21	c.2470C>T	p.R824C	0.00739	0.0026	D/D/D	rs146752751
	7	c.1639G>A	p.A547T	0.000004	0.000008	D/D/D	rs768810263
	33	c.3263C>T	p.S1088F	0.02336	0.051	D/P/D	rs17233497
P61	3	c.533C>T	p.T178I	0.00026	0.0003	D/D/D	rs201254642
	3	c.522_581del	p.174_194del	0.0023	0.0062	-/-	-
	7	c.1165C>T	p.R389C	0.0008	0.0003	D/D/D	rs138043926

Table 2 (continued)

P10	31	c.5683C>T	p.R1895C	0.0024	0.0021	0.0021	D/D/D	-
	6	c.683A>G	p.N228S	0.00026	0.0003	0.0003	D/D/D	rs142596676
	3	c.533C>T	p.T178I		0.0003	0.0003	D/D/D	rs201254642
	3	c.415_416insATCTTCAC	p.Y139fs		0.0033	0.0032	-/-	rs147393916
	9	c.3832C>G	p.P1278A				D/D/D	-
P15	2	c.164G>C	p.R55P		3E-05	0.00002	D/D/D	rs373793185
	6	c.767A>G	p.Q256R				D/D/D	-
	3	c.395C>G	p.A132G				-/D/D	-
	8	c.691G>A	p.V231M	0.0006	0.0001	0.0001	D/D/D	rs200895828
	17	c.2282G>A	p.R761H		0.0002	0.00007	D/D/D	rs143713331
P1	4	c.659G>A	p.R220H		0.0002	0.00002	D/D/D	rs780066296
	4	c.382C>T	p.R128W	0.0004	0.0003	0.0003	D/D/D	rs143324164
	3	c.522_581del	p.174_194del		0.0023	0.0062	-/-	-
	44	c.4535T>G	p.V1512G	0.0016	0.0005	0.0005	D/D/D	rs75696429
	13	c.1820A>G	p.Y607C		0.0002	0.00002	D/D/D	rs777224298
P29	9	c.3819T>G	p.N1273K		0.00001	0.00001	D/D/D	rs759642651
	15	c.1469A>G	p.Q490R	0.00619	0.0116	0.0115	D/P/D	rs11551619
	4	c.346-1G>A					D/P/D	rs140132974
	11	c.1690A>G	p.S564G	0.0002	0.0002	0.0002	D/P/D	rs200366954
	4	c.94C>T	p.R32C	0.01777	0.0107	0.0104	-/D/D	rs76935241
P32	11	c.1481C>T	p.A494V		6E-05	0.00004	D/D/D	rs2229778
	5	c.967C>T	p.R323W	0.0026	0.0014	0.0018	D/D/D	rs138478634
	5	c.1400C>T	p.P467L	0.0012	0.0005	0.0005	D/D/D	rs77256382

*The results of SIFT/Polyphen2/Provean score; SIFT score < 0.05: D (damaging); Polyphen2 score > 0.909: D (probably damaging), Polyphen2 score 0.85–0.909: P (possibly damaging); Provean < -2.5: D (deleterious); - indicated not available

Abbreviation: *DOR*, diminished ovarian reserve; *PA*, primary amenorrhea; *SA*, secondary amenorrhea

1000G: the frequency of each variation in the 1000G database; gnomAD: the frequency of each variation in the gnomAD database; ExAC: the frequency of each variation in the ExAC database of east Asian population

member 1 and Bardet-Biedl syndrome 9 mutations were each identified in four patients. Some variants were found in both POI and DOR patients. Most patients (17/20) carried two or more genetic variants.

The identified variants are depicted in Table 3, and we compared their frequencies in our cohort and in the EAS population in the gnomAD database. Twenty variants were significantly associated with POI or DOR phenotypes compared with the reference values reported by the gnomAD database ($P < 0.05$). When the pathogenicity of each variant was evaluated according to ACMG guidelines, 19 were predicted to be likely pathogenic and 38 were considered VUS.

To acquire further understanding of the biological implications of the identified genes, and identify the most important genes and processes to focus gene list in future studies, we conducted enrichment analysis to identify significant GO categories and KEGG pathways. The top ten terms and pathways are shown in Fig. 2 and Supplementary Table S2. Enriched GO terms included the biological processes reproductive structure development ($P = 1.16\text{E}-08$), reproductive system development ($P = 1.24\text{E}-08$), gonad development ($P = 8.77\text{E}-08$), development of primary sexual characteristics ($P = 1.08\text{E}-07$), meiotic cell cycle ($P = 2.52\text{E}-07$), sex differentiation ($P = 4.66\text{E}-07$), cellular process involved in reproduction in multicellular organism ($P = 3.76\text{E}-06$), and female gonad development ($P = 5.16\text{E}-06$). The cellular components Fanconi anemia (FA) nuclear complex ($P = 0.00044498$) and spindle pole ($P = 0.00067514$) were enriched, as were the molecular functions DNA-dependent ATPase activity ($P = 3.93\text{E}-05$) and ATPase activity ($P = 0.00063824$). The most enriched KEGG pathways were the FA pathway ($P = 0.00092661$), the transforming growth factor (TGF) β signaling pathway ($P = 0.00454567$), and homologous recombination ($P = 0.00945354$). To identify core regulatory genes for further study, we entered the 49 variant-containing genes into the STRING database to obtain a PPI network. Proteins such as ERCC excision repair 2, TFIIH core complex helicase subunit; ERCC excision repair 6, chromatin remodeling factor; mutS homolog (MSH)6; RAD54 like (RAD54L); SPO11 initiator of meiotic double-stranded breaks; stimulated by retinoic acid 8; and dehydrogenase/reductase 11 were important central nodes in the PPI network (Fig. 3).

Discussion

This study is the first to report the results of WES of Chinese women affected by POI or DOR, and describes 79 variants in 49 genes that are potentially related to premature ovarian aging. These variants can be considered strong etiological candidates for ovarian aging, as they have been selected using innovative bioinformatics analysis and stringent filtering.

These genes participate in key biological processes in the ovaries, such as meiosis, follicular development, and DNA repair.

Although it is unclear whether the onset of puberty, normal reproductive aging, POI, and menopause share the same genetic mechanisms, they may have a somewhat shared genetic etiology, with overlap between the genes involved [22, 23]. Additionally, it is highly plausible that DOR and POI represent different stages of the same pathological phenomenon and have similar underlying genetic backgrounds [24]. The additive effects of environmental factors also cannot be ruled out [7]. Over the past decades, many genes have emerged as POI candidates, including those involved in primordial germ cell migration and proliferation (such as nanos C2HC-type zinc finger 3 (*NANOS3*)) [25], oocyte specific transcription factors (such as spermatogenesis and oogenesis specific basic helix-loop-helix (*SOHLH1*), *SOHLH2*, folliculogenesis specific bHLH transcription factor (*FIGLA*), and NOBOX oogenesis homeobox (*NOBOX*)) [26–29], folliculogenesis (nuclear receptor subfamily 5 group A member 1 (*NR5A1*), WT1 transcription factor (*WT1*), and forkhead box L2 (*FOXL2*)) [30–32], meiosis (stromal antigen 3 (*STAG3*), helicase for meiosis 1 (*HFM1*), *MSH4*, *MSH5*, and synaptonemal complex central element protein 1 (*SYCE1*)) [33–37], DNA damage repair (BRCA1 DNA repair associated (*BRCA1*); BRCA2 DNA repair associated (*BRCA2*); *ERCC6*; abraxas 1, BRCA1 A complex subunit (*ABRAXAS1*); scaffold protein involved in DNA repair (SPIDR); minichromosome maintenance 8 homologous recombination repair factor (*MCM8*); and minichromosome maintenance 9 homologous recombination repair factor (*MCM9*)) [8, 23, 38–41], members of the TGF β superfamily (bone morphogenetic protein 15 (*BMP15*), growth differentiation factor 2 (*GDF9*), and *AMH*) [42, 43], and hormones and their receptors (follicle-stimulating hormone receptor (*FSHR*) and anti-Müllerian hormone receptor type 2 (*AMHR2*)) [43, 44]. However, each of these genes is affected in only a few women, and most patients lack a clear genetic diagnosis [8]. There are still candidate genes to be discovered and further genetic studies of POI are warranted.

Although DOR is more common than POI, fewer studies have been conducted on it. Some recent studies support a role for genetic factors in DOR, and many candidate genes are known genetic causes of POI, including *FMRI* [45], *FSHR* [46], *GDF9* [47], and *NR5A1* [48]. Women with *BRCA1/2* mutations also present with DOR and accelerated primordial follicle loss [49]. Several genes, including *AMH*, luteinizing hormone/choriogonadotropin receptor (*LHCGR*), insulin-like growth factor (*IGF1*), *IGF2*, IGF1 receptor (*IGF1R*), IGF2 receptor (*IGF2R*), and gremlin 1, DAN family BMP antagonist (*GREM1*) [50–52] are differentially expressed between women with DOR and controls, representing candidates for further research. In mouse studies, a number of genes, including autoimmune regulator (*Aire*) [53]; forkhead box L2 (*Foxl2*) [54]; *Gdf9* [55]; wingless-type MMTV integration site

Table 3 Frequency of candidate variants and their pathogenicity according to the ACMG guidelines

Gene	Exon	Sequence variation	Protein position	dbSNP	MAF reported by gnomAD_EAS	Cases (no. alleles/no. total alleles)	Patient ID	Fisher's exact test (<i>P</i> value)	ACMG evidence	Variant classification
ABCC5	20	c.1478C>T	p.T493M	rs564676660	0.00006 (1/17972)	0.025 (1/40)	D15	0.004	PS4, PM2, PP3	Likely pathogenic
ADAMTSL3	16	c.1852C>T	p.R618W	rs139253965	0.00833 (153/18366)	0.025 (1/40)	D15	0.286	PM2, PP3	VUS
AIRE	9	c.1051C>T	p.R351W	rs149078622	0.00035 (5/14172)	0.025 (1/40)	P40	0.017	PS4, PM2, PP3	Likely pathogenic
ALMS1	8	c.2351A>G	p.E784G	rs17848880	0.0321 (577/17976)	0.025 (1/40)	D15	1.00	PP3	VUS
AOX1	21	c.2264C>T	p.T755I	rs35217482	0.01187 (218/18364)	0.025 (1/40)	D18	0.381	PM2, PP3	VUS
AREGAP3	15	c.1469A>G	p.Q490R	rs11551619	0.00011 (2/18394)	0.025 (1/40)	P29	0.006	PS4, PP3	VUS
BBS9	21	c.2470C>T	p.R824C	rs146752751	0.03366 (619/18392)	0.1 (4/40)	D18, P28, P37, P40	0.045	PS4, PP3	VUS
BMP8B	4	c.778C>T	p.R260C	rs199806017	0.03302 (597/18078)	0.05 (2/40)	D23, D35	0.383	PM2, PP3	VUS
BMP8B	7	c.1165C>T	p.R389C	rs138043926	0.0019 (35/18394)	0.025 (1/40)	P61	0.075	PM2, PP3	VUS
BMPR2	11	c.1481C>T	p.A494V	rs2229778	0.00092 (17/18394)	0.025 (1/40)	P32	0.038	PS4, PM2, PP3	Likely pathogenic
CDC25B	10	c.945C>G	p.C315W	rs200104998	0.00705 (129/18310)	0.025 (1/40)	D36	0.248	PP3	VUS
CNOT6	4	c.26C>T	p.P9L	rs75964385	0.01594 (293/18380)	0.025 (1/40)	P24	0.475	PM2, PP3	VUS
CYP26B1	5	c.967C>T	p.R323W	rs138478634	0.01786 (324/18144)	0.1 (4/40)	D15, P32, P35, P40	0.006	PS4, PP3	Likely pathogenic
DHRS11	3	c.395C>G	p.A132G	-	0.0169 (31/18376)	0.025 (1/40)	P15	0.067	PM2, PP3	VUS
ERCC2	8	c.691G>A	p.V231M	rs200895828	0.00169 (31/18376)	0.025 (1/40)	P15	0.013	PS4, PM2, PP3	Likely pathogenic
ERCC6	7	c.1639G>A	p.A547T	rs768810263	0 (0/3128)	0.025 (1/40)	P37	0.069	PS3, PP3	VUS
FANCA	33	c.3263C>T	p.S1088F	rs17233497	0.02361 (434/18384)	0.075 (3/40)	P27, P28, P37	0.043	PS4, PM2, PP3	Likely pathogenic
FANCL	8	c.637G>A	p.D213N	rs199564543	0.00103 (19/18392)	0.025 (1/40)	P28	1.00	PM2, PP3	VUS
GATA6	2	c.43G>C	p.G15R	rs116262672	0.04629 (813/17564)	0.025 (1/40)	D18	0.036	PM2, PP3	Likely pathogenic
GRIP1	22	c.2846A>G	p.E949G	-	0.00087 (16/18392)	0.025 (1/40)	D23	0.372	PP3	VUS
HK3	17	c.2282G>A	p.R761H	rs143713331	0.00087 (16/18392)	0.025 (1/40)	P15	0.032	PS4, PM2, PP3	Likely pathogenic
HSD17B4	24	c.2135C>T	p.T712I	rs28943592	0.07997 (1470/18382)	0.025 (1/40)	D23	0.002	PS4, PM2, PP3	Likely pathogenic
KNTC1	44	c.4535T>G	p.V1512G	rs75696429	0.00695 (125/17976)	0.05 (2/40)	P1, P35	0.002	PM2, PP3	VUS
LATS1	6	c.1766G>A	p.R589H	-	0 (0/18394)	0.025 (1/40)	P39	0.002	PS4, PM2, PP3	Likely pathogenic
LEPR	13	c.1820A>G	p.Y607C	rs777224298	0 (0/18394)	0.025 (1/40)	P1	0.002	PVS1, PM2	Likely pathogenic
LHX9	1	c.48delC	p.R16fs	-	0 (0/18394)	0.025 (1/40)	D29	0.002	PS4, PM2, PP3	Likely pathogenic
LMNA	4	c.659G>A	p.R220H	rs780066296	0 (0/18394)	0.025 (1/40)	P15	0.141	PM2, PP3	VUS
MAP3K4	12	c.1301C>T	p.P434L	rs202168620	0.00375 (69/18394)	0.025 (1/40)	D23	0.003	PS4, PM2	VUS
MOGAT1	3	c.415_416insATCTTCAC	p.Y139fs	rs147393916	0.03627 (652/17976)	0.15 (6/40)	D23, D29, D35, P10, P24, P35	0.006	PM2, PP3	VUS
MSH6	9	c.3832C>G	p.P1278A	-	0.00011 (2/18394)	0.025 (1/40)	P10	0.006	PS4, PM2, PP3	Likely pathogenic
MSH6	9	c.3819T>G	p.N1273K	rs759642651	0.00011 (2/18394)	0.025 (1/40)	P1	0.005	PS4, PM2, PP3	Likely pathogenic
MYO19	6	c.203C>G	p.A68G	rs768388466	0.00245 (44/17968)	0.05 (2/40)	D23, D36	0.771	PP3	VUS
NLRP11	4	c.94C>T	p.R32C	rs76935241	0.08296 (1526/18394)	0.05 (2/40)	P31, P35	0.078	PM2, PP3	VUS
NOS3	4	c.382C>T	p.R128W	rs143324164	0.00197 (36/18238)	0.025 (1/40)	P15	0.192	PM2, PP3	VUS
NOTCH2	31	c.5683C>T	p.R1895C	-	0.02007 (369/18390)	0.025 (1/40)	P61	0.011	PM2, PP3	VUS
PGAP3	6	c.683A>G	p.N228S	rs142596676	0.02007 (369/18390)	0.05 (2/40)	P24, P61	0.143	PM2, PP3	VUS
PGRMCI	3	c.533C>T	p.T178I	rs201254642	0.0039 (54/13860)	0.05 (2/40)	P36, P61	0.224	PS4, PM2, PP3	Likely pathogenic
PIWIL1	15	c.1691G>C	p.R564P	-	0.00381 (70/18394)	0.025 (1/40)	D35	0.013	PP3	VUS
PRDM1	2	c.239A>T	p.E80V	rs202200155	0.00626 (115/18372)	0.025 (1/40)	D36	0.059	PS4, PM2, PP3	Likely pathogenic
PRDM1	5	c.1400C>T	p.P467L	rs77256382	0.00027 (5/18394)	0.025 (1/40)	P32	0.016	PM2, PP3	VUS
RAD54L	5	c.354G>C	p.L118F	rs778622356	0.00975 (179/18362)	0.05 (2/40)	P24, P29	0.157	PS4, PM2, PP3	Likely pathogenic
SOHLH1	4	c.346-I>G>A	p.L118F	rs140132974	0.00975 (179/18362)	0.05 (2/40)	D23	0.003	PS4, PM2, PP3	VUS
SPDR	7	c.815G>A	p.R272Q	rs182613962	0.00034 (6/17804)	0.025 (1/40)	D35	0.006	PM2, PP3	VUS
SPDR	19	c.2569T>C	p.S857P	rs376975557	0.00421 (70/16630)	0.025 (1/40)	D35	0.006	PM2, PP3	VUS

Table 3 (continued)

Gene	Exon	Sequence variation	Protein position	dbSNP	MAF reported by gnomAD_EAS	Cases (no. alleles/no. total alleles)	Patient ID	Fisher's exact test (<i>P</i> value)	ACMG evidence	Variant classification
SPO11	1	c.38T>C	p.F13S	rs185741379	0.00261 (48/18368)	0.025 (1/40)	D29	0.101	PM2, PP3	VUS
STRA8	2	c.164G>C	p.R55P	rs73793185	0.00044 (8/18382)	0.025 (1/40)	P10	0.019	PS4, PM2, PP3	Likely pathogenic
TBP	6	c.767A>G	p.Q256R	-	-	0.025 (1/40)	P10	-	PM2, PP3	VUS
TCF21	1	c.10G>A	p.G4S	-	-	0.025 (1/40)	D18	-	PM2, PP3	VUS
THBS1	11	c.1690A>G	p.S564G	rs200366954	0.00332 (61/18394)	0.025 (1/40)	P29	0.126	PM2, PP3	VUS
TOP3B	19	c.1597G>C	p.D533H	-	-	0.025 (1/40)	P28	-	PM2, PP3	VUS
UMODL1	16	c.3169_3170insCCAT	p.A1057fs	rs201855268	0.02297 (72/3134)	0.05 (2/40)	D18, D29	0.239	PVS1, PM2	Likely pathogenic
VCX	3	c.388_445del	p.V130fs	rs1569092849	0.00324 (75/23181)	0.025 (1/40)	D23	0.122	PVS1	VUS
VCX	3	c.385_386del	p.Q129fs	rs752893781	0.00395 (93/23559)	0.025 (1/40)	D23	0.147	PVS1	VUS
VCX	3	c.522_581del	p.I74_194del	-	0.0019 (60/31593)	0.05 (2/40)	P15, P37	0.003	PS4	VUS
ZNF729	4	c.2326G>C	p.A776P	-	-	0.025 (1/40)	D15	-	PM2, PP3	VUS
ZNF729	4	c.1007-1009GAA>CTG	p.336_337ArgLys>ThrGlu	-	-	0.025 (1/40)	D36	-	-	VUS
ZNF729	4	c.1027A>G	p.T343A	rs764887966	0.00764 (135/17670)	0.025 (1/40)	D36	0.266	PP3	VUS

Abbreviation: MAF, minor allele frequencies; ACMG, the American College of Medical Genetics and Genomics; EAS, East Asian population; VUS, variant of unknown significance. Significant *P* values (*P* < 0.05) are presented in italic

family, member 4 (*Wnt4*) [56]; G protein–coupled receptor 3 (*Gpr3*) [57]; and *Bmp15* [58], result in a DOR-like phenotype. The previous studies demonstrate the overlap between the known DOR- and POI-associated genes [47, 48]. In this study, some variants were shared between women with POI and DOR, providing further evidence of this overlap.

Some researchers have proposed the polygenicity of POI [15, 16]. Bouilly et al. reported that 36% of patients in their cohort harbored mutations in two different loci, suggesting a digenic nature for POI pathogenesis [15]. A WES study of a cohort of Caucasian patients with POI showed similar oligogenicity/polygenicity [16]. In the present study, 17 patients (85%) carried more than two variants, supporting a polygenic origin for ovarian aging. It is possible that multiple genetic variants influence ovarian function, and while individually, each may have a relatively low effect, their combinations may contribute substantially to the wide variation in reproductive aging observed in healthy populations and, thus, determine the age of menopause. Multifactorial heredity may partly explain the variable phenotypes observed in patients with POI, such as different ages at amenorrhea onset [22].

In recent decades, women have been increasingly postponing childbirth due to work or study. This tendency has consequently led to an increase in age-related infertility and unwanted childlessness. Declined ovarian reserve is one of the main causes of female infertility. Patients diagnosed with POI have only a 5–10% chance of natural pregnancy [12], while women with DOR tend to have decreased fertility and a higher risk of spontaneous abortion [59]. However, these diseases are often diagnosed too late, with severe and irreversible consequences for fertility. Therefore, it would be of great significance to be able to screen women at high risk for these diseases in the early stages of ovarian function decline and accurately predict the end of their reproductive lifespans. This would enable such individuals to acquire timely guidance regarding family planning and fertility treatment as early as possible, improving their quality of life.

The etiology of ovarian insufficiency remains poorly understood in most cases, and genetic factors appear to play an important role [7]. However, identifying precise causative genes has been challenging. Recently, through new technologies, a prodigious leap has occurred in our understanding of POI genetics [13]. NGS is a powerful tool for the study of genetic diseases and has great potential in future clinical applications [13]. However, previously, NGS has been mainly used in POI pedigrees [7], and only a few studies have used it to analyze women with sporadic POI to identify new putative candidate genes [14–16]. In this study, WES was performed in Chinese women with sporadic ovarian function decline, including both POI and DOR samples.

It was interesting to note that 19 variants were statistically associated with the POI or DOR phenotype. However, functional studies and additional reports on affected women will

Fig. 2 GO terms and KEGG pathway analyses of candidate genes. The top ten pathways were retained in the bar plot. **a** GO analysis show enrichment in the following three terms: biological process (BP), cellular component (CC), and molecular function (MF). **b** Enrichment of KEGG pathway

be required to confirm the predicted deleterious effects of the variants on ovarian function. Pathway analyses of groups of candidate genes can implicate etiological mechanisms. In this study, GO and KEGG enrichment revealed that the identified genes play important roles in meiosis, DNA damage repair, and reproductive system development, which may influence

folliculogenesis and ovarian function, promoting POI or DOR.

Meiosis is a universal cellular process in eukaryotes that is required for gametogenesis. Mutations in meiotic genes usually impair meiotic progression and trigger oocyte death, resulting in POI [60]. During meiosis, large amounts of programmed DNA double-strand breaks are introduced and repaired through homologous recombination [61]. The homologous recombination pathway is required during physiological cell cycle progression to repair replication-associated DNA damage [61]. Variants in *RAD54L* and DNA topoisomerase III beta (*TOP3B*), which encode components of the

Fig. 3 Protein-protein interaction network generated by STRING software for selected proteins. Each node in a PPI network represents a protein, and an edge between two nodes represents an interactional relationship between them. Colored lines display known and predicted

interactions. Light blue: curated databases; pink: experimentally determined; green: gene neighborhood; red: gene fusions; blue: gene co-occurrence; yellow: textmining; black: co-expression; purple: protein homology

homologous recombination pathway, were selected as candidates in this study. The structure of RAD54L is similar to that of RAD54, which is a core component of the homologous recombination machinery and plays a key role in maintaining the genomic integrity of developing germ cells [62]. TOP3B also plays a role in maintaining genome stability. In vitro, disruption of *TOP3B* causes genomic instability, including increased DNA damage and chromosome bridging [63]. In addition, male *Top3b* knockout mice display chromosome instability in their spermatocytes as well as infertility [64].

In the present study, we also found variants in FA complementation group A (*FANCA*) and FA complementation group L (*FANCL*) in women with POI. The FA pathway is another important DNA repair pathway, which is mainly involved in repairing interstrand crosslinks [65]. This pathway involves at least 20 proteins, including those encoded by *FANCA* and *FANCL*. Mutations in both genes have been associated with POI [66, 67]; however, the variants at these loci differed from those identified in this study. A recent study identified the *FANCA* heterozygous missense variants c.1772G > A (p.R591Q) and c.3887A > G (p.E1296G) in patients with POI, and in vitro functional studies and mice models indicated that the two heterozygous variants may be pathogenic [66]. Another study found that *FANCL* mutations can potentially cause POI by disrupting DNA damage repair. They identified two novel heterozygous frameshift mutations, c.1048_1051delGTCT (p.Gln350Valfs*18) and c.739dupA (p.Met247Asnfs*4) in the *FANCL* genes of patients with POI, and verified the deleterious effect of the variants via in vitro studies [67]. Both studies were conducted in Chinese women, indicating that the FA pathway plays an important role in the pathogenesis of POI in Han Chinese women.

There are several strengths to this study. First, this study is the first to conduct WES in patients with DOR and evaluate variants in women with POI and DOR simultaneously to confirm the overlap between their genetic backgrounds. Second, our study performed WES in women with sporadic premature ovarian decline and used innovative bioinformatics analysis. Considering the strong genetic background of this disorder, with a standardized selective approach, NGS may become a useful tool for the etiological diagnosis and risk prediction of high-risk women in the future.

Our study's limitations also need to be considered. First, all variants found in this study were evaluated using in silico tools, and some would have to be reported clinically as VUS [17]. Therefore, the findings of this study are unverified and should not be directly used in clinical practice until additional studies have been performed. Further validation will be necessary to understand the functional significance of identified variants. Second, this study only examined the coding regions and exon-intron boundaries of genes, and the functions of noncoding regulatory regions were not studied. Third, we could not determine the presence of compound heterozygous variants in

our participants, or whether variants occurred in *cis* or in *trans*, as their parents' DNA samples were not available for genotype analysis. Fourth, the prevalence and genetic heterogeneity of POI may be related to ethnicity [3]. Therefore, the pathophysiology of POI in a Chinese population might differ from that in other ethnicities. For example, fewer patients with POI in China and other Asian countries carry *FMR1* premutations compared with patients in western countries [68]. These variations may or may not extend to other genes/variants, and the results of this study may or may not apply to other ethnic groups.

In summary, we used WES and in silico analyses to identify variants and genes involved in sporadic POI and DOR. The results indicate that ovarian function decline is polygenic and that there is overlap between the genetic backgrounds of DOR and POI. In the future, WES and in silico analyses may become useful tools in the genetic diagnosis of patients with ovarian dysfunction and in risk prediction for high-risk women.

Acknowledgments The authors are deeply grateful to all participants involved in this study and all the doctors and researchers who participated in the study.

Authors' contributions All authors contributed to the study conception and design, material preparation, data collection, and analysis. The first draft of the manuscript was written by R. T, and Q. Y revised it. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding information This study was supported by the National Key Research and Development Program [grant number 2018YFC1002105]; CAMS Innovation Fund for Medical Sciences (CIFMS) [grant number 2017-I2M-1-002].

Compliance with ethical standards

This study was approved by the Ethics Committee of our hospital.

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee of Peking Union Medical College Hospital and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

1. Nelson LM. Clinical practice. Primary ovarian insufficiency. *N Engl J Med*. 2009;360(6):606–14. <https://doi.org/10.1056/NEJMc0808697>.
2. Webber L, Davies M, Anderson R, Bartlett J, Braat D, Cartwright B, et al. ESHRE Guideline: management of women with premature

- ovarian insufficiency. *Hum Reprod*. 2016;31(5):926–37. <https://doi.org/10.1093/humrep/dew027>.
3. Golezar S, Ramezani Tehrani F, Khazaei S, Ebadi A, Keshavarz Z. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric*. 2019;22(4):403–11. <https://doi.org/10.1080/13697137.2019.1574738>.
 4. Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clin Endocrinol*. 2008;68(4):499–509.
 5. Cohen J, Chabbert-Buffet N, Darai E. Diminished ovarian reserve, premature ovarian failure, poor ovarian responder—a plea for universal definitions. *J Assist Reprod Genet*. 2015;32(12):1709–12. <https://doi.org/10.1007/s10815-015-0595-y>.
 6. Medicine ASR. Testing and interpreting measures of ovarian reserve: a committee opinion. *Fertil Steril*. 2015;103(3):e9–e17. <https://doi.org/10.1016/j.fertnstert.2014.12.093>.
 7. Jiao X, Ke H, Qin Y, Chen ZJ. Molecular genetics of premature ovarian insufficiency. *Trends Endocrinol Metab*. 2018;29(11):795–807. <https://doi.org/10.1016/j.tem.2018.07.002>.
 8. Franca MM, Mendonca BB. Genetics of primary ovarian insufficiency in the next-generation sequencing era. *J Endocr Soc*. 2020;4(2):bvz037. <https://doi.org/10.1210/jendso/bvz037>.
 9. Qin Y, Jiao X, Simpson J, Chen Z. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update*. 2015;21(6):787–808.
 10. Man L, Lekovich J, Rosenwaks Z, Gerhardt J. Fragile X-associated diminished ovarian reserve and primary ovarian insufficiency from molecular mechanisms to clinical manifestations. *Front Mol Neurosci*. 2017;10:290. <https://doi.org/10.3389/fnmol.2017.00290>.
 11. Tucker EJ, Grover SR, Bachelot A, Touraine P, Sinclair AH. Premature ovarian insufficiency: new perspectives on genetic cause and phenotypic Spectrum. *Endocr Rev*. 2016;37(6):609–35. <https://doi.org/10.1210/er.2016-1047>.
 12. Committee opinion no. 605: primary ovarian insufficiency in adolescents and young women. *Obstet Gynecol* 2014;124(1):193–197. doi:<https://doi.org/10.1097/01.AOG.0000451757.51964.98>.
 13. Laissue P. The molecular complexity of primary ovarian insufficiency aetiology and the use of massively parallel sequencing. *Mol Cell Endocrinol*. 2018;460:170–80. <https://doi.org/10.1016/j.mce.2017.07.021>.
 14. Fonseca DJ, Patino LC, Suarez YC, de Jesus Rodriguez A, Mateus HE, Jimenez KM et al. Next generation sequencing in women affected by nonsyndromic premature ovarian failure displays new potential causative genes and mutations. *Fertil Steril*. 2015;104(1):154–62.e2. doi:<https://doi.org/10.1016/j.fertnstert.2015.04.016>.
 15. Bouilly J, Beau I, Barraud S, Bernard V, Azibi K, Fagart J, et al. Identification of multiple gene mutations accounts for a new genetic architecture of primary ovarian insufficiency. *J Clin Endocrinol Metab*. 2016;101(12):4541–50. <https://doi.org/10.1210/jc.2016-2152>.
 16. Patino LC, Beau I, Carlosama C, Buitrago JC, Gonzalez R, Suarez CF, et al. New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Hum Reprod*. 2017;32(7):1512–20. <https://doi.org/10.1093/humrep/dex089>.
 17. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–24. <https://doi.org/10.1038/gim.2015.30>.
 18. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4(5):P3.
 19. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
 20. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16(5):284–7. <https://doi.org/10.1089/omi.2011.0118>.
 21. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362–D8. <https://doi.org/10.1093/nar/gkw937>.
 22. Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. *Clin Genet*. 2017;91(2):183–98. <https://doi.org/10.1111/cge.12921>.
 23. Xu X, Zhang Y, Zhao S, Bian Y, Ning Y, Qin Y. Mutational analysis of the FAM175A gene in patients with premature ovarian insufficiency. *Reprod BioMed Online*. 2019;38(6):943–50.
 24. Zhao M, Feng F, Chu C, Yue W, Li L. A novel EIF4ENIF1 mutation associated with a diminished ovarian reserve and premature ovarian insufficiency identified by whole-exome sequencing. *J Ovarian Res*. 2019;12(1):119. <https://doi.org/10.1186/s13048-019-0595-0>.
 25. Qin Y, Zhao H, Kovanci E, Simpson JL, Chen ZJ, Rajkovic A. Mutation analysis of NANOS3 in 80 Chinese and 88 Caucasian women with premature ovarian failure. *Fertil Steril*. 2007;88(5):1465–7. <https://doi.org/10.1016/j.fertnstert.2007.01.020>.
 26. Zhao S, Li G, Dalgleish R, Vujovic S, Jiao X, Li J, et al. Transcription factor SOHLH1 potentially associated with primary ovarian insufficiency. *Fertil Steril*. 2015;103(2):548–53.e5. <https://doi.org/10.1016/j.fertnstert.2014.11.011>.
 27. Qin Y, Jiao X, Dalgleish R, Vujovic S, Li J, Simpson JL, et al. Novel variants in the SOHLH2 gene are implicated in human premature ovarian failure. *Fertil Steril*. 2014;101(4):1104–9.e6. <https://doi.org/10.1016/j.fertnstert.2014.01.001>.
 28. Tosh D, Rani HS, Murty US, Deenadayal A, Grover P. Mutational analysis of the FIGLA gene in women with idiopathic premature ovarian failure. *Menopause*. 2015;22(5):520–6. <https://doi.org/10.1097/gme.0000000000000340>.
 29. Li L, Wang B, Zhang W, Chen B, Luo M, Wang J, et al. A homozygous NOBOX truncating variant causes defective transcriptional activation and leads to primary ovarian insufficiency. *Hum Reprod*. 2017;32(1):248–55. <https://doi.org/10.1093/humrep/dew271>.
 30. Philibert P, Paris F, Lakhal B, Audran F, Gaspari L, Saad A, et al. NR5A1 (SF-1) gene variants in a group of 26 young women with XX primary ovarian insufficiency. *Fertil Steril*. 2013;99(2):484–9. <https://doi.org/10.1016/j.fertnstert.2012.10.026>.
 31. Settas N, Anapliotou M, Kanavakis E, Fryssira H, Sofocleous C, Dacou-Voutetakis C, et al. A novel FOXL2 gene mutation and BMP15 variants in a woman with primary ovarian insufficiency and blepharophimosis-ptosis-epicanthus inversus syndrome. *Menopause*. 2015;22(11):1264–8. <https://doi.org/10.1097/gme.0000000000000473>.
 32. Laissue P. Aetiological coding sequence variants in non-syndromic premature ovarian failure: from genetic linkage analysis to next generation sequencing. *Mol Cell Endocrinol*. 2015;411:243–57. <https://doi.org/10.1016/j.mce.2015.05.005>.
 33. He WB, Banerjee S, Meng LL, Du J, Gong F, Huang H, et al. Whole-exome sequencing identifies a homozygous donor splice-site mutation in STAG3 that causes primary ovarian insufficiency. *Clin Genet*. 2018;93(2):340–4. <https://doi.org/10.1111/cge.13034>.
 34. de Vries L, Behar DM, Smirin-Yosef P, Lagovsky I, Tzur S, Basel-Vanagaite L. Exome sequencing reveals SYCE1 mutation associated with autosomal recessive primary ovarian insufficiency. *J Clin Endocrinol Metab*. 2014;99(10):E2129–32. <https://doi.org/10.1210/jc.2014-1268>.
 35. Carlosama C, Elzaia M, Patino LC, Mateus HE, Veitia RA, Laissue P. A homozygous donor splice-site mutation in the meiotic gene MSH4 causes primary ovarian insufficiency. *Hum Mol Genet*. 2017;26(16):3161–6. <https://doi.org/10.1093/hmg/ddx199>.
 36. Guo T, Zhao S, Zhao S, Chen M, Li G, Jiao X, et al. Mutations in MSH5 in primary ovarian insufficiency. *Hum Mol Genet*. 2017;26(8):1452–7. <https://doi.org/10.1093/hmg/ddx044>.

37. Wang J, Zhang W, Jiang H, Wu BL. Mutations in HFM1 in recessive primary ovarian insufficiency. *N Engl J Med*. 2014;370(10):972–4. <https://doi.org/10.1056/NEJMc1310150>.
38. Qin Y, Guo T, Li G, Tang TS, Zhao S, Jiao X, et al. CSB-PGBD3 mutations cause premature ovarian failure. *PLoS Genet*. 2015;11(7):e1005419. <https://doi.org/10.1371/journal.pgen.1005419>.
39. Yilmaz NK, Karagin PH, Terzi YK, Kahyaoglu I, Yilmaz S, Erkaya S, et al. BRCA1 and BRCA2 sequence variations detected with next-generation sequencing in patients with premature ovarian insufficiency. *J Turk Ger Gynecol Assoc*. 2016;17(2):77–82. <https://doi.org/10.5152/jtgga.2016.16035>.
40. Weinberg-Shukron A, Rachmiel M, Renbaum P, Gulsuner S, Walsh T, Lobel O, et al. Essential role of BRCA2 in ovarian development and function. *N Engl J Med*. 2018;379(11):1042–9. <https://doi.org/10.1056/NEJMoa1800024>.
41. Desai S, Wood-Trageser M, Matic J, Chipkin J, Jiang H, Bachelot A, et al. MCM8 and MCM9 nucleotide variants in women with primary ovarian insufficiency. *J Clin Endocrinol Metab*. 2017;102(2):576–82. <https://doi.org/10.1210/jc.2016-2565>.
42. Renault L, Patino LC, Magnin F, Delemer B, Young J, Laissue P, et al. BMPR1A and BMPR1B missense mutations cause primary ovarian insufficiency. *J Clin Endocrinol Metab*. 2019. <https://doi.org/10.1210/clinem/dgz226>.
43. Alvaro Mercadal B, Imbert R, Demeestere I, Gervy C, De Leener A, Englert Y, et al. AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Hum Reprod*. 2015;30(5):1196–202. <https://doi.org/10.1093/humrep/dev042>.
44. Liu H, Xu X, Han T, Yan L, Cheng L, Qin Y, et al. A novel homozygous mutation in the FSHR gene is causative for primary ovarian insufficiency. *Fertil Steril*. 2017;108(6):1050–5.e2. <https://doi.org/10.1016/j.fertnstert.2017.09.010>.
45. Gleicher N, Weghofer A, Barad DH. Ovarian reserve determinations suggest new function of FMR1 (fragile X gene) in regulating ovarian ageing. *Reprod BioMed Online*. 2010;20(6):768–75. <https://doi.org/10.1016/j.rbmo.2010.02.020>.
46. Ghezelayagh Z, Totonchi M, Zarei-Moradi S, Asadpour O, Maroufizadeh S, Eftekhari-Yazdi P, et al. The impact of genetic variation and gene expression level of the follicle-stimulating hormone receptor on ovarian reserve. *Cell J*. 2018;19(4):620–6. <https://doi.org/10.22074/cellj.2018.4183>.
47. Wang TT, Ke ZH, Song Y, Chen LT, Chen XJ, Feng C, et al. Identification of a mutation in GDF9 as a novel cause of diminished ovarian reserve in young women. *Hum Reprod*. 2013;28(9):2473–81. <https://doi.org/10.1093/humrep/det291>.
48. Warman DM, Costanzo M, Marino R, Berenshtein E, Galeano J, Ramirez PC, et al. Three new SF-1 (NR5A1) gene mutations in two unrelated families with multiple affected members: within-family variability in 46,XY subjects and low ovarian reserve in fertile 46,XX subjects. *Horm Res Paediatr*. 2011;75(1):70–7. <https://doi.org/10.1159/000320029>.
49. Lin W, Titus S, Moy F, Ginsburg ES, Oktay K. Ovarian aging in women with BRCA germline mutations. *J Clin Endocrinol Metab*. 2017;102(10):3839–47. <https://doi.org/10.1210/jc.2017-00765>.
50. Skiadas CC, Duan S, Correll M, Rubio R, Karaca N, Ginsburg ES, et al. Ovarian reserve status in young women is associated with altered gene expression in membrana granulosa cells. *Mol Hum Reprod*. 2012;18(7):362–71. <https://doi.org/10.1093/molehr/gas008>.
51. Greenseid K, Jindal S, Hurwitz J, Santoro N, Pal L. Differential granulosa cell gene expression in young women with diminished ovarian reserve. *Reprod Sci*. 2011;18(9):892–9. <https://doi.org/10.1177/1933719111398502>.
52. Jindal S, Greenseid K, Berger D, Santoro N, Pal L. Impaired gremlin 1 (GREM1) expression in cumulus cells in young women with diminished ovarian reserve (DOR). *J Assist Reprod Genet*. 2012;29(2):159–62. <https://doi.org/10.1007/s10815-011-9684-8>.
53. Jasti S, Warren BD, McGinnis LK, Kinsey WH, Petroff BK, Petroff MG. The autoimmune regulator prevents premature reproductive senescence in female mice. *Biol Reprod*. 2012;86(4):110. <https://doi.org/10.1095/biolreprod.111.097501>.
54. Tran S, Zhou X, Lafleur C, Calderon MJ, Ellsworth BS, Kimmins S, et al. Impaired fertility and FSH synthesis in gonadotropespecific Foxl2 knockout mice. *Mol Endocrinol*. 2013;27(3):407–21. <https://doi.org/10.1210/me.2012-1286>.
55. Dong J, Albertini DF, Nishimori K, Kumar TR, Lu N, Matzuk MM. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature*. 1996;383(6600):531–5.
56. Boyer A, Lapointe E, Zheng X, Cowan RG, Li H, Quirk SM, et al. WNT4 is required for normal ovarian follicle development and female fertility. *FASEB J*. 2010;24(8):3010–25. <https://doi.org/10.1096/fj.09-145789>.
57. Ledent C, Demeestere I, Blum D, Petermans J, Hämmäläinen T, Smits G, et al. Premature ovarian aging in mice deficient for Gpr3. *Proc Natl Acad Sci U S A*. 2005;102(25):8922–6.
58. McMahon HE, Hashimoto O, Mellon PL, Shimasaki S. Oocyte-specific overexpression of mouse bone morphogenetic protein-15 leads to accelerated folliculogenesis and an early onset of acyclicity in transgenic mice. *Endocrinology*. 2008;149(6):2807–15. <https://doi.org/10.1210/en.2007-1550>.
59. Lyttle Schumacher BM, Jukic AMZ, Steiner AZ. Antimüllerian hormone as a risk factor for miscarriage in naturally conceived pregnancies. *Fertil Steril*. 2018;109(6):1065–71.e1. <https://doi.org/10.1016/j.fertnstert.2018.01.039>.
60. Dou X, Guo T, Li G, Zhou L, Qin Y, Chen ZJ. Minichromosome maintenance complex component 8 mutations cause primary ovarian insufficiency. *Fertil Steril*. 2016;106(6):1485–9.e2. <https://doi.org/10.1016/j.fertnstert.2016.08.018>.
61. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*. 2018;127(2):187–214. <https://doi.org/10.1007/s00412-017-0658-1>.
62. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016;17(6):337–49. <https://doi.org/10.1038/nrm.2016.48>.
63. Zhang T, Wallis M, Petrovic V, Challis J, Kalitsis P, Hudson DF. Loss of TOP3B leads to increased R-loop formation and genome instability. *Open Biol*. 2019;9(12):190222. <https://doi.org/10.1098/rsob.190222>.
64. Kwan KY, Moens PB, Wang JC. Infertility and aneuploidy in mice lacking a type IA DNA topoisomerase III. *beta Proc Natl Acad Sci U S A*. 2003;100(5):2526–31.
65. Räschele M, Knipscheer P, Knipscheer P, Enou M, Angelov T, Sun J, et al. Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell*. 2008;134(6):969–80. <https://doi.org/10.1016/j.cell.2008.08.030>.
66. Yang X, Zhang X, Jiao J, Zhang F, Pan Y, Wang Q, et al. Rare variants in FANCA induce premature ovarian insufficiency. *Hum Genet*. 2019;138(11–12):1227–36. <https://doi.org/10.1007/s00439-019-02059-9>.
67. Yang Y, Guo T, Liu R, Ke H, Xu W, Zhao S, et al. FANCL gene mutations in premature ovarian insufficiency. *Hum Mutat*. 2020;41(5):1033–41. <https://doi.org/10.1002/humu.23997>.
68. Pastore L, Manichaikul A, Wang X, Finkelstein J. FMR1 CGG repeats: reference levels and race-ethnic variation in women with normal fertility (study of women's health across the nation). *Reprod Sci*. 2016;23(9):1225–33.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.