Whole-exome sequencing reveals SALL4 variants in premature ovarian insufficiency: an update on genotype–phenotype correlations

Qi Qi Wang1,2,3 · Da Li4 · Baozhu Cai1 · Qing Chen1,2,3 · Caihua Li5 · Yanhua Wu1 · Li Jin1 · Xiuxia Wang4 · Xiaojin Zhang1,3 · Feng Zhang1,2,3

Received: 7 July 2018 / Accepted: 20 November 2018 © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
Premature ovarian insufficiency (POI) is a severe female disorder characterized by primary or secondary amenorrhea before 40 years of age. Genetic factors have been implicated in the pathogenesis of POI, but known POI-associated genes account for only a small fraction of heritability. Here, we performed whole-exome sequencing (WES) to explore pathogenic genes in Han Chinese subjects with POI. Intriguingly, we identified novel or rare heterozygous missense variants of SALL4 (spalt-like transcription factor 4) in 3 (6%) of 50 POI subjects. The SALL4 c.541G>A and c.2279C>T variants were paternally inherited, while c.1790A>G was inherited from an affected mother with early menopause. SALL4 encodes a transcription factor that is highly expressed in oocytes and early embryos. Our in vitro functional assays suggested that all of these SALL4 missense variants had significantly increased SALL4 protein expression with enhanced regulatory activity in regard to its downstream target POU5F1 compared to that of wild-type SALL4. Notably, previous studies demonstrated the genetic involvement of SALL4 loss-of-function variants in Okihiro syndrome and related syndromic developmental disorders. Through our analysis of genotype–phenotype correlations, we suggest that different variation types of SALL4 might have different effects on SALL4 activity, resulting in phenotypic variability. Our findings highlight the genetic contribution of SALL4 missense variants with enhanced regulatory activities to POI and underscore the importance of variant classification and evaluation for molecular diagnosis and genetic counseling.

Introduction
Premature ovarian insufficiency (POI) is defined as the cessation or disturbance of the menses for four or more months in women under 40 years old that is accompanied by two measurements of elevated serum follicle-stimulating hormone (FSH) levels (> 25 IU/l) at least 4 weeks apart (The ESHRE Guideline Group on POI et al. 2016). POI affects approximately 1–5% of women worldwide (De Vos et al. 2010). This disorder occurs through two major pathogenic mechanisms: follicle dysfunction or follicle depletion (Nelson 2009). Though the etiology of POI remains unclear, genetic aberrations, autoimmune disorders, and iatrogenic and environmental factors have been confirmed to be involved in the pathogenesis of POI (Goswami and Conway 2005). Among them, genetic factors have been regarded as the primary causes of POI. Notably, 14–31% of subjects with POI have at least one affected relative with POI or early menopause, supporting a strong genetic predisposition (Tucker et al. 2016).
Recently, whole-exome sequencing (WES) studies have provided solid evidence for genetic contributions to POI pathogenesis (Qin et al. 2015). For example, NOBOX (Li et al. 2017b) and SOHLH1 (Bayram et al. 2015), two genes encoding transcription factors that regulate oocyte-specific gene expression during the early stages of folliculogenesis, have been reported to cause POI in humans. However, known POI-associated genes account for only a small fraction of POI pathogenesis (Qin et al. 2015). Other causative or susceptible genes for POI have yet to be identified.

In this study, we performed a genetic analysis using WES of 50 Han Chinese subjects with POI and identified novel or rare missense variants in SALL4 (spalt-like transcription factor 4) from three unrelated families. Further in vitro functional assays suggested that these SALL4 missense variants with enhanced regulatory activities contribute to the etiology of POI.

Materials and methods

Study subjects

A total of 50 Han Chinese women with POI (48 sporadic cases and 2 familial cases) were enrolled at the Obstetrics and Gynecology Hospital of Fudan University and at Shengjing Hospital of China Medical University between September 2014 and February 2017. The inclusion criteria consisted of primary or secondary amenorrhea for at least 4 months before 40 years of age, along with two measurements of abnormal serum FSH levels (> 25 IU/l). All the subjects with POI in this study had a normal 46, XX karyotype. FMR1 premutations were excluded using a FragilEase™ PCR Reagent Kit (Chen et al. 2017). Women with ovarian surgery or radiotherapeutic or chemotherapeutic interventions were excluded. This study was approved by the institutional review boards at Fudan University and at Shengjing Hospital of China Medical University. Written informed consent was obtained from all participants.

Whole-exome sequencing and variant calling

Genomic DNA was extracted from peripheral blood samples using a Puregene Blood Core Kit B (QIAGEN, Hilden, Germany) according to standard procedures. WES was performed for the 50 POI subjects using a SureSelectXT Human All Exon V6 kit (Agilent Technologies, Santa Clara, USA). For each subject, 1.5 µg of genomic DNA was used to prepare a captured library that was then sequenced on a HiSeq X Ten platform (Illumina, San Diego, USA), generating 150 bp paired-end reads. Raw data of approximately 10 GB per exome were mapped to a human reference genome sequence (GRCh37/hg19) using the Burrows–Wheeler Alignment (BWA) tool (Li and Durbin 2010). Variant calling was performed using the Genome Analysis Toolkit (GATK) (McKenna et al. 2010). All variants were further annotated with ANNOVAR software (Wang et al. 2010).

Gene filtration and bioinformatic analysis

Rare and novel protein-altering variants (nonsense, frameshift, essential splicing-site and missense) were preferred in this study. Variant frequency was analyzed based on different ethnic subgroups from the Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD) (Lek et al. 2016) as well as the 1000 Genomes Project (The 1000 Genomes Project Consortium et al. 2012). Single-nucleotide variants (SNVs) were analyzed with the SIFT (Kumar et al. 2009), PolyPhen-2 (Adzhubei et al. 2010), MutationTaster (Schwarz et al. 2014), CADD (Kircher et al. 2014) and DANN tools (Quang et al. 2015) for functional prediction. Relevance to phenotype was considered on the basis of OMIM, gene ontology terms (Ashburner et al. 2000), KEGG pathway analysis (Kanehisa et al. 2017) and mouse model studies.

Variation validation

After filtering against multiple databases, Sanger sequencing was used to validate the novel and rare SALL4 gene variants. PCR products were sequenced on a 3730XL genetic analyzer (Applied Biosystems, Foster City, USA) according to the manufacturer’s instructions. The designed primers are shown in Table S1.

Array-based comparative genomic hybridization analysis

For three subjects with SALL4 missense variants, we further conducted a copy number variation (CNV) analysis using an Agilent SurePrint G3 Human 1 x 1M comparative genomic hybridization (CGH) microarray. Some of the experimental details were described previously (Boone et al. 2010). DNA processing, microarray handling and data analysis were conducted according to the Agilent oligonucleotide CGH protocol (version 6.3). CNV calling was performed using Agilent Genomic Workbench software.

Plasmid construction and mutagenesis

A luciferase reporter plasmid containing the POU5F1 promoter was constructed as described previously (Chew et al. 2005). A 3-kb fragment of the human POU5F1 promoter was cloned into the pGL3-Basic vector (Promega, Madison, USA). The human full-length SALL4 expression plasmid (RC213089) and pCMV6-Entry vector (PS100001) were
obtained from OriGene Technologies (Rockville, USA). Site-directed mutagenesis was performed to generate three missense variants (c.541G>A, c.1790A>G and c.2279C>T) and a null variant (c.2279delC) of SALL4 according to the instructions of a KOD-Plus-Mutagenesis Kit (Toyobo, Osaka, Japan). The relevant primers are shown in Table S2. The constructs were verified by direct Sanger sequencing prior to the functional studies.

**Cell culture and transfection**

HeLa cells were cultured in Dulbecco’s modified eagle’s medium (DMEM) (Gibco, Grand Island, USA) supplemented with 10% fetal bovine serum (FBS) (Gibco, Grand Island, USA) and 1% Plasmocin (InvivoGen, San Diego, USA) at 37 °C with 5% CO2. To evaluate the effect of the variants on SALL4, HeLa cells were transfected with the wild-type or mutated SALL4 constructs using Lipofectamine 3000 (Invitrogen, Carlsbad, USA) according to the manufacturer’s instructions.

**Western blotting**

Whole-cell lysates were separated by SDS–PAGE and transferred onto PVDF membranes. After being blocked with non-fat milk, each membrane was incubated with specific antibodies against different proteins at 4 °C overnight, followed by incubation with an HRP-conjugated secondary antibody. Antibodies against the MYC Tag (Abmart, Arlington, USA) were used at a 1:2000 dilution, and an anti-GAPDH antibody (1:5000 dilution, Sigma, St. Louis, USA) was used as an internal control. The secondary antibody was a goat anti-mouse IgG (1:10,000 dilution, Abmart, Arlington, USA).

**Luciferase reporter assay**

HeLa cells that were replated in 24-well plates were cotransfected with 200 ng of the POU5FI promoter luciferase reporter and 200 ng of wild-type or mutated SALL4 plasmid per well. The pRL-TK plasmid (10 ng/well, Promega, Madison, USA) was used as an internal control. After 24 h of transfection, luciferase activity was measured using the Dual-Luciferase reporter assay system (Promega, Madison, USA) according to the manufacturer’s protocol. Assays were performed for three independent experiments, and each experiment was assayed in quadruplicate.

**Results**

**Identification of novel and rare SALL4 variants by WES**

The mean coverage depth of the target regions in our exome sequencing data was 106x, with more than 93% of the target bases covered at more than 20x. First, we investigated genetic variants with minor allele frequencies (MAF) ≤ 0.1% according to the public population genomic databases (1000 Genomes Project, ExAC and gnomAD). Furthermore, deleterious variants predicted using no fewer than four in silico tools were chosen for further analysis. We then investigated recurrently mutated genes among the study subjects. Interestingly, we identified three heterozygous missense variants in exon 2 of SALL4 in three unrelated subjects with POI (Fig. 1). In these subjects with SALL4 variants, no rare variants were identified in known POI-associated genes, according to the OMIM database (Tables S3–S5). Furthermore, no potential causative CNVs were identified in these cases (Tables S6–S8).

The SALL4 variants were all confirmed by Sanger sequencing (Fig. 1a). Among them, the SALL4 c.541G>A (p.V181M) and c.2279C>T (p.T760I) variants were paternally inherited (Fig. 1a). Both variants have extremely low frequencies in human populations (Table 1). The SALL4 c.1790A>G (p.K597R) variant was inherited from the affected mother of the subject and is a novel missense variant (Fig. 1a; Table 1). We also found that the coding positions of these three SALL4 variants were evolutionarily conserved among species (Fig. 1c). For example, the novel SALL4 variant c.1790A>G (p.K597R) is located in a highly conserved C2H2 zinc finger motif. We conducted functional prediction analyses for the novel and rare SALL4 variants using the SIFT (Kumar et al. 2009), PolyPhen-2 (Adzhubei et al. 2010), MutationTaster (Schwarz et al. 2014), CADD (Kircher et al. 2014) and DANN tools (Quang et al. 2015). Notably, the vast majority of these bioinformatic tools predicted that all three SALL4 variants were deleterious (Table 1). According to the American College of Medical Genetics (ACMG) guidelines, the c.541G>A and c.2279C>T variants are considered to be variants of uncertain significance (VUS), while the c.1790A>G variant is classified as likely pathogenic (Richards et al. 2015). In view of these heterogeneous results by in silico prediction, we performed further functional studies to assess the biological impact of the novel and rare SALL4 missense variants.

**Clinical characteristics of the affected individuals with SALL4 variants**

Clinical information of the POI subjects affected by SALL4 variants is summarized in Table 2. Subject F003 had normal puberty and secondary amenorrhea. Her menarche
occurred at 13 years of age and was followed by regular cycles until 22 years of age. Pelvic ultrasonography revealed a normal uterus but small ovaries with few antral follicles. Further consecutive hormonal measurements revealed elevated FSH levels (75.8 IU/l and 110.4 IU/l) with decreased estradiol concentrations (5 pg/ml and 4 pg/ml). Table 2 shows data from a later measurement for subject F003.

![Fig. 1](image-url) Identification of SALL4 variants in POI. a Novel/rare SALL4 missense variants (M1, M2 and M3) were identified in three families with POI. Sanger sequencing confirmed that the SALL4 c.541G>A (M1) and c.2279C>T (M3) variants were both paternally inherited, while the c.1790A>G variant (M2) was inherited from the affected mother with early menopause. The red arrows indicate the positions of the heterozygous variants. Reference DNA sequences in the wild-type (WT) individuals are also shown. The black arrows in the pedi-gree plots indicate the probands with POI. b Schematic representation of the SALL4 gene and protein. The red arrows indicate the variant positions. ZF, zinc finger motif. c All three novel/rare SALL4 mis-sense variants identified in this study were located at amino acid residues conserved among different species.
Subject F024 was 29 years old at examination due to primary infertility. Her menses became irregular and completely stopped at 18 years of age. She was diagnosed with POI and was placed on sex steroid replacement therapy. Physical examination showed a normal body mass index and breast development. However, pelvic ultrasound imaging revealed small ovaries with no visible antral follicles. Her mother presented with early menopause at an age of 44 without known cause. Family history has been suggested to be considered ‘positive’ when a first- or second-degree relative had either POI or early menopause (Jiao et al. 2017). Therefore, the mother was considered to be an affected individual related to subject F024. No other known chronic diseases, endocrinopathies or autoimmune disorders were observed for subject F024 or her affected mother. After oestrogen treatment, subject F024 had visible follicles.

Subject L030 had normal puberty and established regular menses, with her menarche at 14 years old. She was diagnosed with POI at the age of 28 years when she developed amenorrhea. Subject L030 exhibited normal growth and development, and no other significant clinical or dysmorphic features were identified. Her basal serum gonadotropin levels were elevated (FSH = 52.0 IU/l and LH = 41.1 IU/l), whereas her serum estradiol was low (< 20 pg/ml).

**Genotype–phenotype correlations for the SALL4 variants**

To date, 24 pathogenic variants of SALL4 have been identified in subjects with typical Okihiro syndrome (also known as Duane-radial ray syndrome, DRRS) and Holt–Oram syndrome (HOS) (Al-Baradie et al. 2002; Alves et al. 2016; Borozdin et al. 2004; Brassington et al. 2003; Chacon-Camacho et al. 2016; Kohlhase et al. 2002b, 2003, 2005; Terhal et al. 2006). We summarized these previously reported syndrome-associated SALL4 variants and conducted a comparative analysis to the POI-associated SALL4 variants identified in this study (Table S9). Furthermore, we briefly predicted the possible effects of the SALL4 variants using multiple bioinformatic tools (Table S9). Notably, all heterozygous SALL4 null variants resulted in syndromic developmental disorders (DRRS and HOS), whereas the subjects with pathogenic missense variants of SALL4 herein presented with POI (Table S9). Based on the differential distribution of SALL4 variants between POI and developmental syndromes, a statistically significant correlation was observed using the two-tailed Fisher’s exact test ($P = 3 \times 10^{-4}$) (Table S10).

**In vitro functional characteristics of the SALL4 variants**

The putative impact of the p.V181M, p.K597R and p.T760I variants on SALL4 function was investigated using
luciferase reporter assays (Fig. 2a) and western blotting (Fig. 2b). Previous experimental evidence suggested that Sall4 is a transcriptional activator of Pou5f1 (Zhang et al. 2006). Here, we showed that the wild-type SALL4 protein had obvious regulatory activity and activated the POU5F1 promoter compared with the empty vector (Fig. 2a). The SALL4 loss-of-function (LOF) variant p.N761Tfs*6 had remarkably suppressed regulatory activity, with substantially less activation of its downstream target POU5F1. Intriguingly, the SALL4 variants identified in our POI subjects (p.V181M, p.K597R and p.T760I) exhibited significantly enhanced regulatory activities compared with that of the wild-type SALL4. Consistently, western blotting analysis further revealed that all three variants were expressed at higher levels than wild-type SALL4 (Fig. 2b), which might at least partially explain their enhanced regulatory activity observed in the luciferase reporter assays (Fig. 2a). Given these findings, we presumed that the POI SALL4 variants identified in this study might be gain-of-function variants. Our experimental observations illustrate the differential biological consequences of the SALL4 LOF variants and the three SALL4 missense variants identified in our subjects with POI.

**Discussion**

The SALL4 gene (NM_020436.3) is located on human chromosome 20q13.2 and contains four exons. The SALL4 protein (NP_065169.1) belongs to the spalt transcription factor family, which plays an important role in regulating the embryonic development of many organisms (Al-Baradie et al. 2002; Sweetman and Munsterberg 2006). Spalt was initially identified in Drosophila as a homeotic gene required for the early development of the posterior head and anterior tail regions (Kuhnlein et al. 1994). Murine Sall4 is ubiquitously expressed in embryos, especially in the primitive inner cell mass. In mouse embryonic stem cells, Sall4 binds to the highly conserved regulatory region of the pluripotent master gene Pou5f1 and activates its expression, suggesting that Sall4 is essential for early embryonic cell-fate decisions (Zhang et al. 2006). Mice with Sall4 heterozygous LOF variants display increased postnatal lethality, as well as anencephaly, deafness, renal agenesis, anorectal malformations, and skeletal defects (Warren et al. 2007). In humans, heterozygous LOF variants of SALL4 have been previously reported to cause Okihiro syndrome/Duane-radial ray syndrome, which is characterized by Duane anomaly and radial ray defects involving the thumbs and other parts of limbs, anorectal and renal malformations, and deafness (Kohlhase et al. 2002b). SALL4 is also involved in Holt–Oram syndrome, which is characterized by radial forelimb malformations and congenital heart abnormalities and associated features (Brassington et al. 2003). No reproductive phenotypes have been reported for these syndromic cases with SALL4 LOF variants.

SALL4 is pleotropic, and it plays important roles in oogenesis. During germ cell development, murine Sall4 is involved in ensuring the correct specification and migration of primordial germ cells (PGCs) (Yamaguchi et al. 2015). In adult mice, Sall4 expression is restricted to the germline (Kohlhase et al. 2002a). The expression patterns of SALL4 in human embryos and adult tissues are basically consistent with those that have been described in mice (Miettinen et al. 2014). Human SALL4 is highly expressed in mitotic PGCs and fetal germ cells (Guo et al. 2015; Li et al. 2017a).
Moreover, *Sall4* is highly expressed in oocytes at different developmental stages and could regulate oocyte meiosis resumption in female mice, indicating that *Sall4* is indispensable for follicular development and the completion of oogenesis (Xu et al. 2017). According to the Human Protein Atlas (Ponten et al. 2008), immunohistochemical analysis of human ovaries showed significant SALL4 protein expression in follicle cells.

In this study, we recruited 50 Han Chinese subjects clinically diagnosed with POI and performed WES to identify pathogenic variants. After stringent filtration and validation, our WES analysis revealed novel and rare missense variants of *SALL4* in three (6%) unrelated subjects. All three probands presented with normal puberty, secondary amenorrhea, and elevated FSH levels. According to the guidelines of ACMG, the *SALL4* c.541G>A and c.2279C>T variants are considered to be VUS, and the c.1790A>G variant is likely pathogenic (Richards et al. 2015). Therefore, we conducted in-vitro functional assays on these three *SALL4* missense variants for further functional characterization. Our experimental observations suggested that compared to wild-type SALL4, all three *SALL4* missense variants identified in our POI subjects significantly increased the SALL4 protein expression levels, leading to enhanced SALL4 regulatory activity. Since no significant differences were found in the *SALL4* mRNA levels, as investigated by quantitative RT-PCR (Fig. S1), we presumed that these missense variants contribute to protein stability regulation, with posttranslational regulation at the protein level possibly being the key to the molecular mechanism underlying POI. Our findings suggest the important genetic involvement of *SALL4* in the etiology of POI.

Considering the pleiotropism of *SALL4* variants in both developmental syndromes and POI (McKusick 2007), we further investigated genotype–phenotype correlations for the *SALL4* variants identified in the POI subjects in this study as well as in other disorders. All 24 previously reported *SALL4* variants in typical Okihiro syndrome and Holt–Oram syndrome were frameshift or nonsense variants (Table S9). Variants of unclear pathogenicity or variants in atypical Okihiro and related syndromic developmental disorders were excluded. The heterozygous null alleles of *SALL4* are thought to cause these syndromic phenotypes via a haploinsufficiency mechanism (Kohlhase et al. 2005).

In contrast, only heterozygous missense variants of *SALL4* were identified in the nonsyndromic POI subjects in this study (Table S9). Interestingly, the *SALL4* c.541G>A variant has been recurrently reported in POI (Wang et al. 2009). Although it was reported as a VUS in 2017 in ClinVar, our *in vitro* functional evidence supports the pathogenicity of the c.541G>A variant. Furthermore, all three *SALL4* missense variants identified in our POI subjects showed enhanced SALL4 regulatory activity. Therefore, we speculate that *SALL4*-associated POI does not share the same molecular mechanism as that of *SALL4* haploinsufficiency in Okihiro syndrome and Holt–Oram syndrome. Future efforts may elucidate the detailed mechanism underlying *SALL4*-associated POI.

In summary, our genetic analysis using WES of Chinese POI subjects illustrated the important genetic contribution of *SALL4* to POI and broadened the phenotypic range of heterozygous *SALL4* variants. We also described for the first time the genotype–phenotype correlations for *SALL4* variants in POI and DRRS/HOS based on in-vitro functional experimentation.

**Fig. 2** Functional evaluation of the *SALL4* variants. **a** Luciferase reporter assays with the *POU5F1* promoter in HeLa cells. The cells were transfected with equal amounts of different *SALL4* expression constructs, individually: empty vector (pCMV6-Entry), WT (wild-type), LOF variant (p.N761Tfs*6) or rare missense variants (p.V181M, p.K597R, or p.T760I). The relative luciferase activity was normalized to Renilla activity. Error bars indicate the SEM in quadruplicate. *P*<0.05; **P*<0.01; ***P*<0.001; ****P*<0.0001. **b** Western blotting analysis of the SALL4 protein expression levels in HeLa cells transfected with equal amounts of different *SALL4* expression constructs, individually: empty vector, WT, missense variants or LOF variant. GAPDH was used as a loading control. Band density was quantified with ImageJ software. Representative images of three independent experiments are shown. IntDen, integrated density.
assays. These updated genotype–phenotype correlations enhance the current knowledge of SALL4-related disorders.

Web resources

The URLs for the data presented herein are as follows:
DANN, http://cbcl.ics.uiuc.edu/public_data/DANN.

Acknowledgements

We are grateful to the subjects and their families who participated in this study. We also thank Shuxia Chen for experimental support. This work was supported by National Key Research and Development Program of China (2017YFC1001100), National Natural Science Foundation of China (31625015 and 31521003), Shanghai Medical Center of Key Programs for Female Reproductive Diseases (2017ZZ01016), and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01).

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

Al-Baradie R, Yamada K, St-Hilaire C, Chan WM, Andrews C, McIn-}
Borozdin W, Wright MJ, Hennekam RC, Hannibal MC, Crow YJ, Neu-}
Kohlhase J, Heinrich M, Liebers M, Frohlich Archangelo L, Rear-}


