



A high-resolution X chromosome copy-number variation map in fertile females and women with primary ovarian insufficiency

Svetlana A. Yatsenko, MD^{1,2,3,4}, Michelle Wood-Trageser, PhD¹, Tianjiao Chu, PhD³, Huaiyang Jiang, MD³ and Aleksandar Rajkovic, MD, PhD^{1,2,3,4,5,6}

Purpose: Sex-biased expression of genes on the X chromosome is accomplished by a complex mechanism of dosage regulation that leads to anatomical and physiological differences between males and females. Copy-number variations (CNVs) may impact the human genome by either affecting gene dosage or disturbing a chromosome structural and/or functional integrity.

Methods: We performed a high-resolution CNV profiling to investigate the X chromosome integrity in cohorts of 269 fertile females and 111 women affected with primary ovarian insufficiency (POI) and assessed CNVs impact into functional and nonfunctional genomic elements.

Results: In POI patients, we observed a 2.5-fold enrichment for rare CNVs comprising ovary-expressed genes, and genes implicated in autoimmune response and apoptotic signaling. Moreover, there was a higher prevalence of deletions encompassing genes that escape X inactivation, noncoding RNAs, and intergenic DNA

sequences among POI females, highlighting structural differences between X chromosomes of fertile and POI females. Furthermore, we discovered a ~4% carrier incidence for X-linked disorders among fertile women.

Conclusion: We constructed a high-resolution map of female-specific CNVs that provides critical insights into the spectrum of human genetic variation, sex-specific disease risk factors, and reproductive potential. We discovered novel CNVs associated with ovarian dysfunction and support polygenic models for POI.

Genetics in Medicine (2019) 21:2275–2284; <https://doi.org/10.1038/s41436-019-0505-2>

Keywords: X chromosome; primary ovarian insufficiency; high-resolution copy-number map; X-linked disease; fertility

INTRODUCTION

Sex chromosomes play a critical role in human fertility and sex determination. The X chromosome is enriched for sex-biased genes implicated in human reproduction, and pathogenic variants within *TEX11* (MIM *300311), *ADGRG2* (MIM *300572), *AR* (MIM *313700), and other X-linked genes are known to impact male fertility.^{1–3} The X chromosome is also critical in ovarian development and its disruption can lead to primary ovarian insufficiency (POI), which is clinically characterized by menstrual disturbance (amenorrhea or oligomenorrhea) with elevated follicle-stimulating hormone (FSH) and luteinizing hormone and low estradiol (hypergonadotropic hypogonadism).^{4–7} The POI spectrum is broad, affects 2–4% of women,^{4–6} and ranges from congenital absence of ovaries (gonadal dysgenesis), which clinically presents as primary amenorrhea, to individuals with irregular menstrual cycles whose ovarian function

ceases prior to age 40. The X chromosome aberrations are encountered in 8–10% of individuals with POI and include monosomy X, cytogenetically visible deletions and duplications, and balanced and unbalanced X-autosome rearrangements.⁷ Monosomy X (45,X or Turner syndrome) is characterized by loss of oocytes at the pachytene stage of meiotic prophase during female fetal development, resulting in streak gonads at birth.⁸ The presence of two structurally normal X chromosomes is essential for meiosis, and meiotic disturbance can lead to accelerated germ cell depletion, apoptosis of surrounding somatic granulosa cells, and in the most severe form causes streak ovaries.^{5,8,9} Although one of the X chromosomes is inactive in somatic cells, both X chromosomes are active in ovaries and biallelic expression is likely important for the completion of meiosis.^{8–11} Moreover, structural abnormalities, such as X chromosome/autosome translocations, and large X chromosome deletions on both the

¹Department of Pathology, University of Pittsburgh, Pittsburgh, PA, USA; ²Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA, USA; ³Magee-Womens Research Institute, Pittsburgh, PA, USA; ⁴Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ⁵Department of Pathology, University of California–San Francisco, San Francisco, CA, USA; ⁶Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA, USA. Correspondence: Svetlana A. Yatsenko (yatsenkosa@mail.magee.edu) or Aleksandar Rajkovic (Aleks.Rajkovic@ucsf.edu)

Submitted 5 December 2018; accepted: 20 March 2019

Published online: 5 April 2019

short and long arms, have been associated with POI and infertility.^{12,13} In the majority of POI patients with X chromosome translocations, the breakpoints do not contain genes.^{12–14} Despite clinical studies that associate large tracts of X chromosome with POI, only few genes on the X chromosome have been implicated in ovarian development, including *BMP15* (MIM *300247), *FMR1* (MIM *309550), and *PGRMC1* (MIM *300435). These observations suggest that failure in both the maintenance of the X chromosome dosage and integrity can cause POI.^{12–14} The X chromosome copy-number profile in fertile women and women with POI is unknown. We conducted a high-resolution analysis of copy-number variants (CNVs) on 269 fertile women and 111 women with POI to establish a fine-scale X chromosome architecture in these two populations, to evaluate their impact on the X chromosome integrity, and to assess CNVs association with POI and female reproductive success.

MATERIALS AND METHODS

The recruitment of participants and research study were approved by the University of Pittsburgh institutional review board (IRB) (PRO09080427, PRO12120172). Consent for genetic testing was received from all patients. A population of fertile females (269 unrelated participants) of European descent aged 23 to 37, who had at least one live birth, have regular menses and no family history of genetic disorders, was defined as healthy women. One hundred and eleven POI women were defined as individuals diagnosed with hypergonadotropic hypogonadism, less than 40 years of age, and two serum FSH concentrations greater than 40 IU/L. The POI cohort had four subgroups: 34 women with primary amenorrhea, 21 women with secondary amenorrhea diagnosed prior to age 25, 30 women with secondary amenorrhea diagnosed after age 25, and 26 women with POI, whose age of onset was prior to 40 but exact onset was uncertain. High-resolution (~0.5–3 kb) survey¹⁵ of the X chromosome CNVs was performed using a custom 180 K comparative genomic hybridization (CGH) microarray (Agilent Inc., Santa Clara, CA) on genomic DNA from 269 fertile women and 111 POI study participants. The CNV calls are provided in Supplementary Tables S1 and S7. Additional technical information including detailed description of study subjects, microarray design, bioinformatics and genomic tools, and associated references are available in the Supplementary Material and Methods. Raw data are available in the National Center for Biotechnology Information (NCBI) GEO database (accessions GSE126893 and GSE127453).

Statistical analysis

Two-tailed Chi-square tests for the equality of proportions between two samples were performed to determine if a certain type of CNV is more/less likely to be carried in the POI versus the control (fertile females) group. To determine if the number of a certain type of CNVs carried by a person is higher/lower in the POI group than in the control group, we

performed two-tailed Wilcoxon rank sum tests. The values of the statistical tests are further adjusted using the Holm method to control the family-wise error rates for the simultaneous test of all types of CNVs.

RNA-Seq library preparation

The ovary RNA-Seq library was constructed using a 21-week-old human fetal ovary. The study was approved by the University of Pittsburgh IRB (PRO16070068).

RESULTS

X chromosome CNVs in a population of fertile women

A total of 4099 CNVs (Table S1) comprising 2523 losses/deletions (62%) and 1576 gains/duplications (38%) were detected in 601 distinct genomic intervals, meaning that a number of CNVs recurred (Table S2). CNVs analysis showed that 86% of the DNA losses were less than 10 kb in size (Fig. 1a).

We classified copy-number alterations into six different categories (Fig. 1b). In the nonrepetitive genomic regions, the CNVs were classified as 411 homozygous losses (DNA copy number = 0), 2074 heterozygous deletions (copy number = 1), 914 heterozygous duplications (copy number = 3), and 641 CNVs with gains of two or more DNA copies (copy number ≥ 4). In regions with segmental duplications and/or gene family clusters, we recorded 38 partial losses and 21 partial gains.

A total of 601 genomic intervals were defined by unique breakpoints observed among recorded CNVs. About 89.1% (3652/4099) of CNVs recurred with a population frequency of greater than 1%. A total of 447/4099 (10.9%) CNVs were rare. These 447 CNVs were detected in 189/269 (70.3%) of fertile women with an average of 2.36 (447/189) CNVs per individual, and an overall incidence of 1.66 (447/269) rare CNVs per fertile female. Four or more rare CNVs were observed in ~12% of fertile females (Fig. 1c). The majority (87%) of rare CNVs were heterozygous, with ~60% representing DNA losses (mean size 13.9 kb) and ~40% containing DNA gains (mean size 130.7 kb).

We examined all deletions and duplications intervals for the presence of genes and known regulatory elements. Whole genes or exons were present in 166/601 (~27.6%) genomic intervals. In fertile females, the majority of the gene-coding intervals are gains (67.4%) as opposed to losses (32.6%). The majority of “loss” intervals (~81%) span intergenic and intronic regions, locations that are less likely to be harmful. Only 30.5% of fertile females contain coding sequences among rare CNVs with a predominance of gains. Additional CNV characteristics in fertile females are provided in the online Supplemental Data (Fig. S1–4).

Prevalence of known X chromosome pathogenic CNVs among fertile females

Sequencing studies suggest that ~2% of women have disease-causing variants for X-linked disorders.¹⁶ We assessed the prevalence of known and likely pathogenic CNVs in our cohort of fertile women. We identified ten (10/269, 3.7%)

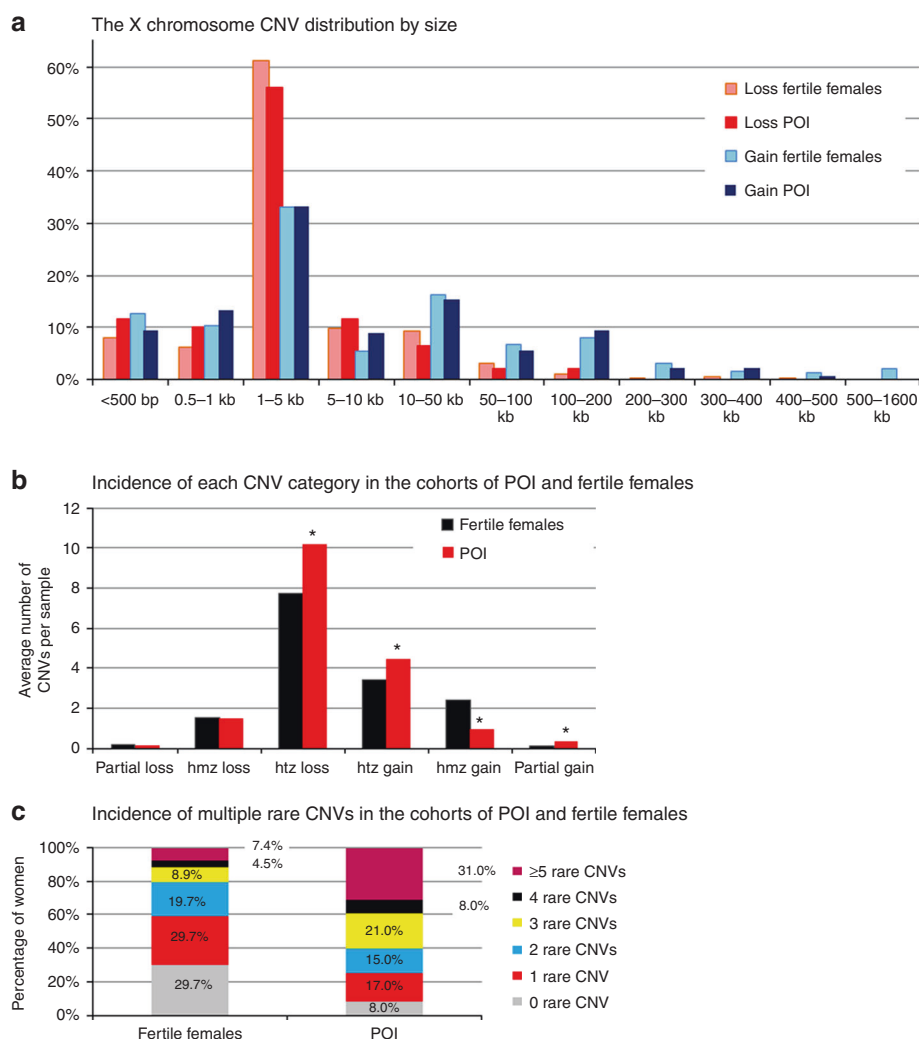


Fig. 1 Copy-number variant (CNV) differences between X chromosomes of fertile and primary ovarian insufficiency (POI) females. **a** CNVs distribution by size in the cohorts of fertile and POI females. The proportion of observed genomic CNV intervals with losses and gains is shown for each size interval. **b** Incidence of each CNV category in the cohorts of POI and fertile females. A statistically significant difference (indicated by asterisk) has been observed for heterozygous losses and gains, homozygous gains, as well as gains affecting repetitive regions/gene family clusters ($p < 0.05$). **c** Incidence of multiple rare CNVs in the cohorts of POI and fertile females.

fertile women who possess pathogenic CNVs, and are at risk to transmit X-linked disorder to their male offspring (Table 1).

Impact of X chromosome CNVs: analysis of gene function embraced by CNVs in fertile women

CNV intervals among fertile women encompassed a total of 217 protein-coding genes (Table S3), including 107 genes among rare CNVs. These comprise ~19.8% of 1098 known X-linked protein-coding genes.¹⁷ Thirty-six of these genes (3.3%) were found among losses, 125 genes (11.4%) were among gains, and 56 genes (5.1%) were among intervals containing both gain and loss CNVs. We examined 217 genes for their expression level in ovarian tissue. We sequenced RNA extracted from fetal human ovary, at the time when dynamic developmental changes occur in human ovarian development and include primordial follicle formation

(Table S4). Importantly, in fertile women, ovary-expressed (OE) genes affected by deletions comprise ~1.5% of all X-linked protein-coding genes, while duplications accounted for three-fold more OE genes. Among 217 genes encompassed by CNVs, 15 genes are known to escape X chromosome inactivation^{18,19} (Fig. S5).

Each individual woman was evaluated for all CNVs that contain coding sequences and whether genes in those CNVs are expressed in the ovary. We focused our analysis on deletions because these CNVs are more likely to be damaging variants. This analysis revealed that 44.2% of fertile females do not have a loss CNV encompassing OE genes, 36.8% have a single OE gene affected by a CNV loss, and ~19% of fertile females have 2–6 OE genes within a CNV loss. Only 5.9% of fertile women had rare deletion CNVs within OE genes.

All 217 genes were subjected to further ontology analyses using the PANTHER (Protein Analysis through Evolutionary

Table 1 Pathogenic CNVs and variants of uncertain clinical significance in a population of fertile and POI women

Affected loci

	Cohort	Chr band	Size (kb)	Coding region	CNV	Number of carriers	Gene OMIM	Disorder OMIM/ PMID	Disorder
<i>PLCXD1</i> , <i>GTPBP6</i> , <i>PPP2R3B</i> , <i>CRLF2</i> , <i>SHOX</i> , <i>IL3RA</i> , <i>CSF2RA</i> , <i>MIR3690</i> , <i>SLC25A6</i> , <i>ASMT</i> , <i>ASMTL</i> , <i>P2RY8</i> , <i>AKAP17A</i>	FF	Xp22.33	1688	Whole gene	Gain	1	*312865	PMID: 27073233	Xp22.33 microduplication
<i>ZNF81</i> , <i>ZNF182</i> , <i>SSX6</i> , <i>SSX5</i> , <i>SSX1</i> , <i>SPACA5</i> , <i>SPACA5B</i> , <i>SSX3</i> , <i>ZNF630</i> , <i>CXXC1P1</i> , <i>SSX4B</i> , <i>SSX4</i>	FF	Xp11.23	726	hole gene	Gain	1	*314998	PMID: 20662849	Xp11.22-p11.23 microduplication, mental retardation, X-linked 45; MRX45
<i>DMD</i>	FF	Xp21.2	58.5	Exons 50–51	Loss	1	*300377	300376	Muscular dystrophy, Becker type; BMD
<i>DMD</i>	FF	Xp21.2	1.2	Exon 23	Gain	2	*300377	300376	Muscular dystrophy, Becker type; BMD
<i>FAAH2</i>	FF	Xp11.21	119.5	Exons 1–7	Gain	1	*300654	PMID: 26239645	Age of menarche
<i>GRIA3</i>	FF	Xq25	2.2	Exon 3	Loss	1	*305915	300699	Mental retardation, X-linked, syndromic, Wu type; MRXSW
<i>KAL1</i>	FF	Xp22.32	109.4	Exons 1–2	Gain	1	*300836	308700	Hypogonadotropic hypogonadism
<i>OPN1LW</i>	FF	Xq28	1	Exon 5	Loss	1	*300822	303800/ 303900	Red/green color vision
<i>PCDH19</i>	POI	Xq22.1	7.3	Exon 6	Loss	1	*300460	300088	Epileptic encephalopathy, early infantile, 9
<i>PTCHD1</i>	FF	Xp22.11	0.8	Exon 1	Gain	2	*300828	300830	Susceptibility to autism, X-linked 4; AUTSX4
<i>RPGR</i>	FF	Xp11.4	1.0	Exons 14–15	Loss	1	*312610	300029	Retinitis pigmentosa 3; RP3
<i>SHOX</i>	POI	Xp22.33	16.9	Exon 6	Loss	1	*312865	127300; 300582	Leri–Weill dyschondrosteosis, idiopathic familial short stature
<i>TMLHE</i>	FF	Xq28	5.7	Exon 2	Loss	1	*300777	300872	Autism, susceptibility to, X-linked 6; AUTSX6
<i>TSPAN7</i>	FF	Xp11.4	144.5	Exons 2– 7	Gain	1	*300096	300210	Mental retardation, X-linked 58; MRX58

Critical gene within a microdeletion/duplication syndrome is underlined.
BMD Becker muscular dystrophy, CNV copy-number variant, FF fertile females, POI primary ovarian insufficiency.

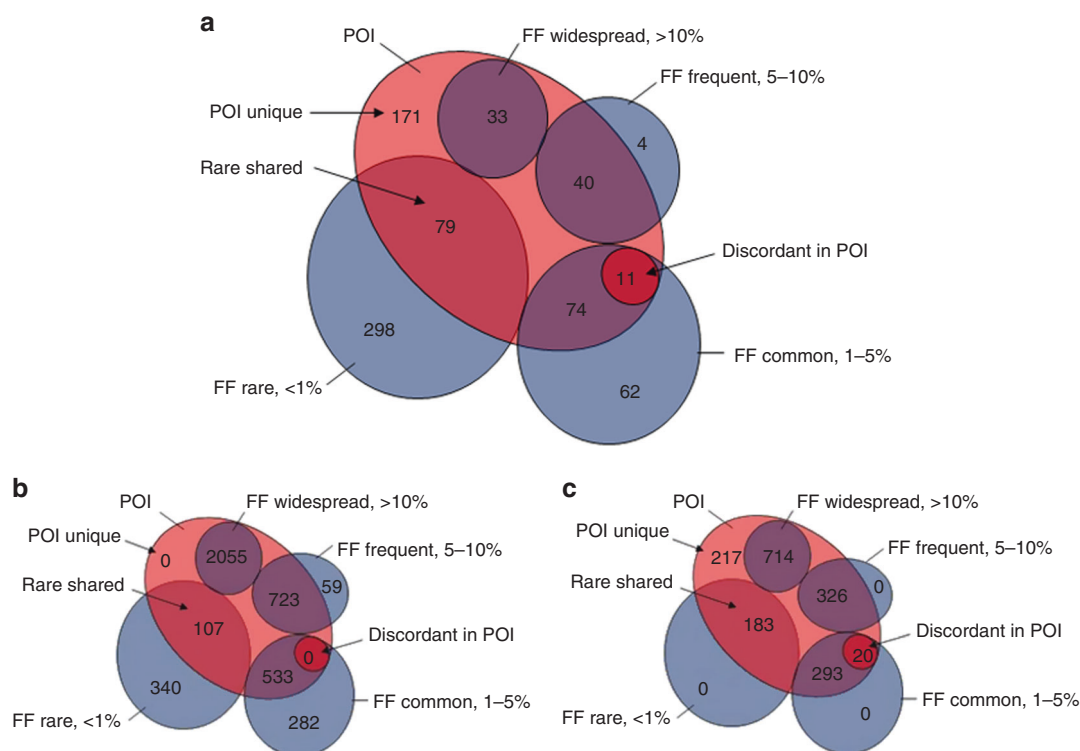


Fig. 2 The X chromosome copy-number variant (CNV) landscape in fertile and primary ovarian insufficiency (POI) females. **a** Number of genomic intervals. **b** Number of CNVs in 269 fertile females (FF). **c** Number of CNVs in 100 POI women.

Relationships) tool. This set of genes was categorized according to the Gene Ontology biological process, molecular function, cellular component enrichment, protein class, and an implicated pathway (Fig. S6) for a further comparison with the POI cohort.

Large X chromosome CNVs in women with POI

Similar to previous studies,^{4–8} ~10% (11/111) of women with POI had large X chromosome imbalances including 45,X; 47,XXX; unbalanced translocation t(X;1), and gross Xp and Xq deletions (Table S6, Fig. S7A), which likely explain their phenotype. These patients were excluded from further CNV analysis.

General characteristics of submicroscopic CNVs in idiopathic POI

We identified a total of 1753 submicroscopic CNVs (408 genomic intervals) in 100 women affected by POI (Table S7), ranging from 5 to 49 CNVs/individual, with a mean = 17.5 CNVs/female. The findings in POI patients are summarized in Table S8. Analysis of both populations of POI and fertile women revealed that more than 85% of losses and 60% of gains are below 10 kb in size (Fig. 1a). Overall, there was a statistically significant difference ($p < 0.01$) in CNV number in POI women versus our cohort of fertile women. The POI cohort showed a statistically significant increase in the number of heterozygous losses and gains and reduced number of homozygous gains as compared with fertile females (Fig. 1b).

The frequency of CNVs was compared between POI and fertile women. In women with POI, we identified 217 novel “POI-unique” CNVs (Table S9) in 171 genomic intervals that were not observed among the cohort of fertile females. Seventy-nine rare genomic intervals (named as rare-shared, Fig. 2), were observed in both POI and fertile female cohorts. Although they were rare (<1%) in fertile women, subsets of these rare intervals were more common among POI women. In POI women, CNVs in 17 genomic intervals that were seen in less than 1% in fertile women, were observed with 2–13% frequency (Table S10), which is ~6-fold higher than in fertile females and is statistically different. We also observed 20 CNVs in 11 genomic intervals that were discordant for copy-number state between POI and fertile females. For example, 13 losses and 7 gains among common CNVs in POI women were encountered as “gain only” or “loss only” in those intervals among fertile females (Table S11A, B). An overall incidence of rare CNVs is 2.5-fold higher among the POI cohort compared with fertile females. These results may have differential effect on X chromosome integrity in POI and fertile women.

We also examined population frequencies of common CNVs in fertile and POI females. CNVs in four genomic intervals (Table S11C) with a population frequency of 1–5% in fertile females were not observed in POI women. These intervals may represent regions that protect against ovarian failure. In contrast, six common genomic intervals (Table S10), which were seen with much higher frequency in POI, may represent POI risk regions.

Rare-shared and POI-unique CNVs in POI patients

We performed detailed analysis of 420 variants (217 POI-unique, 183 rare-shared, and 20 discordant CNVs) identified in POI women and compared with a cohort of fertile females (Fig. S8–10). These CNVs (238 losses and 182 gains) were considered to be rare/atypical events for fertile women. Importantly, rare and POI-unique CNVs were observed in 92/100 of POI patients and ~74% of POI women had two or more such CNVs on the X chromosome (Fig. 1c). This incidence was statistically different from 45% of fertile females carrying two or more rare CNVs. These findings may suggest a cumulative effect of CNV onto ovarian function. An overall average of 4.56 (420/92) for combined rare and POI-unique CNVs and incidence of 4.20 (420/100) per POI patient was significantly higher ($p < 3.51E-13$) than the average of 2.25 and incidence of 1.66 CNVs seen in fertile females (Table S20).

We evaluated if there was an enrichment of CNVs unique to POI depending on POI type and onset. Overall, the differences between POI subgroups did not reach statistical significance and analysis was limited by the small number of participants in POI subgroups.

Functional significance of rare and POI-unique CNVs to POI patients

We mapped a total of 261 genomic intervals (79 rare-shared between fertile and POI females, 11 discordant for copy-number state, and 171 POI-unique, Table S9) to the X chromosome and overlaid GENCODE and ENCODE data. These data were then used to characterize regions into six functional genomic categories: (1) CNVs involving at least one whole gene transcript, (2) CNVs comprising exonic sequences (coding exon[s] of a single gene), (3) CNVs affecting untranslated regions (UTRs) and/or noncoding exons of a gene, (4) CNVs containing noncoding RNA sequences, (5) CNVs affecting intronic regions of a gene, and (6) intergenic CNVs encompassing DNA sequences between the genes. The cohorts' comparison showed a significant enrichment for CNVs (particularly DNA losses) containing exonic sequences ($p < 3.00E-8$) and noncoding RNAs ($p < 1.15E-08$) in the POI group (Fig. S8).

Among the rare-shared and POI-unique CNVs (Table S9), 111 potentially functional genomic loci were of specific interest due to the presence of exonic or whole-gene sequences (74 and 15 intervals, respectively) within a CNV, 5'-UTRs (8 intervals), and 14 intervals spanning the 3'-UTRs. A total of 138 genes (Table S14) were identified within these 111 genomic intervals and subjected to detailed analysis including gene expression profile and classification using the PANTHER and STRING bioinformatics tools.

The OE genes were present among 84/111 (75.7%) of POI-unique intervals, including 30% of DNA losses, affecting 89/100 (89%) POI women (Fig. S5F). In contrast, among fertile women, OE genes were present in 13.8% of rare losses (Figs. S5C and S10A). We also observed differences for a number of OE genes between POI diagnostic groups

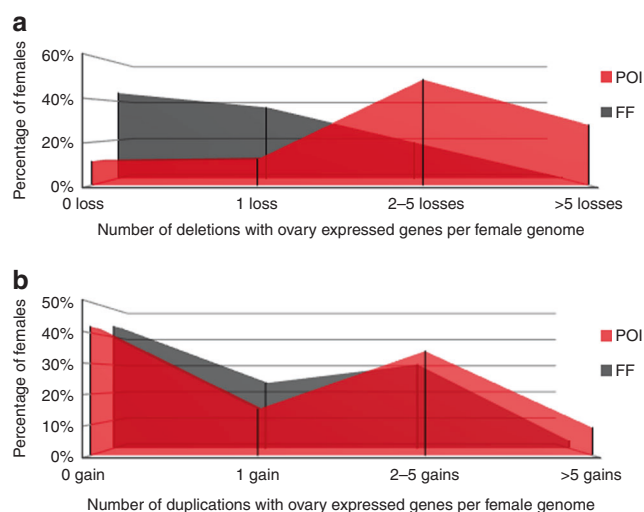


Fig. 3 Number of ovary-expressed genes affected by losses and gains in the individual genomes of primary ovarian insufficiency (POI) and fertile females (FF). **a** The number of losses with ovary-expressed genes and proportion of individuals carrying CNV losses are statistically different between the cohorts of POI and fertile females ($p < 1.78E-08$). **b** The proportion of individuals with the gains in POI cohort is not different from a cohort of fertile women ($p < 0.83886$). Likewise, the number of gains with ovary-expressed genes is similar between cohorts ($p < 0.10183$).

(Fig. S10B). Patients with primary amenorrhea have lower number of OE affected genes, while more OE genes were detected in subgroups of patients with secondary amenorrhea. This may suggest a more damaging effect of CNVs in patients with primary POI and a possible polygenic cause in patients with secondary POI due to a cumulative effect of less harmful CNVs.

Overall, individual genomes from the POI cohort contained a higher number of CNV losses that encompassed OE genes, compared with our fertile cohort ($p < 3.69E-25$) (Fig. 3), as well as a higher number of genes that escape X inactivation (Fig. S5D, E).

The ontology analysis of 138 genes encompassed by rare CNVs in POI patients showed enrichment in genes implicated into several physiological processes including inflammation mediated by chemokine and cytokine signaling; T-cell activation; cytoskeletal regulation by Rho GTPase and synaptic vesicle trafficking; cadherin, WNT, and FGF signaling; and RAS pathway as well as apoptosis activation (Fig. S6F–J). Some of these pathways, such as WNT-, FGF-, and RAS-signaling pathways, have been previously shown to affect growth, differentiation, and maintenance of gonadal tissue in a variety of species.^{20–22}

ENCODE annotations of the noncoding genomic intervals CNVs in the regions that do not contain protein-coding DNA or gene regulatory sequences were annotated into three interval categories: those that contain known noncoding RNA transcripts, intergenic intervals that do not comprise any transcripts or known regulatory elements with the

enhancer- or promoter-like capacity, and CNVs that comprise intronic segments. CNVs in the noncoding genomic intervals are commonly viewed as benign changes. Our analysis showed that the incidence of noncoding intergenic CNVs on the X chromosome was statistically different ($p < 1.00E-06$) and much higher in the POI cohort compared with fertile women (Fig. S8–9). Taking into account the distinctive heterochromatic and functional properties of the X chromosome between somatic and germline cells in females, we examined intergenic and intronic CNVs for the presence of predicted transcription factor-binding sites (TF-BS) and the epigenetic chromatin signatures (Table S12). Notably, noncoding CNVs seen in POI patients spanned methylated forms of histone H3 such as H3K27me3, a hallmark of topologically inactive chromatin domains H3K9me3 and H4K20me1, which are associated with the X chromosome silencing (Fig. S11A). In addition, in POI females, ~53% of noncoding CNVs encompassed at least one TF-BS. A total of 102 TF-BS (Table S16) were found within the rare-shared and POI-unique genomic intervals. Interestingly, proteins (ATF2, ATF3, CREB1, EGR1, ELK1, EP300, FOS, GATA2, JUN, JUNB, JUND, NFYA, NFYB, NR3C1, RELA, SP1, SRF, STAT3) implicated in the gonadotropin-releasing hormone receptor (P06664), CCKR signaling (P06959), and PDGF signaling (P00047) were the most common transcription factors bound to the CNV affected genomic regions in POI females. In addition, POI-unique CNVs were also enriched in TF-BS associated with TP53-apoptosis signaling (P00006), RAS pathway (P04393), and oxidative stress response (P00046) (Fig. S11B). Similarly to alterations in coding CNVs, POI-unique noncoding CNVs appear to affect multiple pathways implicated in gonadal development, autoimmune response, and apoptotic signaling in POI females.

Possible target genes in idiopathic POI

Eleven genes—*FMR1* (MIM *309550), *STAG3* (MIM *608489), *MCM8* (MIM *608187), *MCM9* (MIM *610098), *POF1B* (MIM *300603), *DIAPH2* (MIM *300108), *BMP15* (MIM *300247), *MSH5* (MIM *603382), *SYCE1* (MIM *611486), *NOBOX* (MIM *610934), and *NR5A1* (MIM *184757)—were previously associated with POI. Using an in silico approach (STRING)²³ we evaluated for the known and predicted protein-protein interactions (<https://string-db.org/cgi/network.pl?taskId=HTEn9j4Qqpam>) between POI causing genes and the set of 138 genes found among POI-unique and rare-shared CNVs (Fig. S12). Of the CNVs unique to the POI cohort, several genes have been implicated in ovarian function and/or POI, including *DIAPH2*, *LDLOC1* (MIM *300402), *SPANXA1* (MIM *300305), *SPANXA2* (MIM *300493), *FANCB* (MIM *300515), *GPR101* (MIM *300393), and *PGRMC1* (MIM *300435) (Table S17A).

Based on comparative genomics and function-dependent approaches utilized in this study, results of previous studies of mouse models, and case reports on women affected by POI, we propose novel X-linked gene candidates, such as *PAK3* (MIM *300142), *H2AFB1* (MIM *300445), *STAG2* (MIM

*300826), *TIMP1* (MIM *305370), *TENM1* (MIM *300588), *MAP7D3* (MIM *300930), *SYN1* (MIM *313440), and *NAPIL2* (MIM *300026) among others that may cause or contribute to POI etiology (Table S17B). We also outlined noncoding RNA transcripts identified among POI CNVs (Table S17C) that may play a role in regulation of inactivation/reactivation processes²⁴ of the X chromosome in primordial germ cells.

Distribution of rare-shared and POI-unique CNVs along the X chromosome in POI patients

Copy-number alterations of a few genes on the X chromosome are known to be implicated in sex differentiation, such as duplications and deletions surrounding the *NR0B1* (MIM *300473) gene.^{25–27} We hypothesized that fertile women are less likely to have CNVs in the regions that are involved in sexual development and gonadal differentiation. We constructed a high-resolution CNV map to assess distribution of observed gains and losses along the X chromosome in fertile females (Fig. 4). We calculated a total number of intervals with unique breakpoints in each 2-Mb window on the X chromosome and identified genomic segments in which gains or losses were present at a lower rate than expected for a random distribution (average 3.3 intervals per 2-Mb segment for gains and 4.6 for losses). Alterations in a cohort of POI patients were compared with fertile females, particularly within these CNV-poor segments. We plotted 91 genomic intervals defined above that have the highest potential to cause POI to the X chromosome map. Remarkably, these 91 genomic intervals were clustered on the X chromosome at the Xp11.2, Xq13, Xq22, Xq26.3, and Xq28 regions, and also correspond to the OE gene clusters. The Xq13 and Xq22 chromosomal regions were previously implicated to POI based on multiple reports on POI females with the X chromosome deletions and balanced rearrangements.^{12,13} We conducted a comprehensive literature search and identified published reports on 75 women with POI and balanced X-autosome translocations (Table S18). Breakpoint mapping analyses showed that the vast majority of these translocations do not involve a gene-coding region on the X chromosome but are clustered within Xq13, Xq21, Xq22, and Xq24 (Fig. 4). The cumulative findings suggest that the genomic integrity within the Xq13 and Xq22 chromosomal loci involving both coding and noncoding DNA sequences may be essential for gonadal development or maintenance. Significantly, a higher number of POI women (40%) had rare CNVs in the Xq13 ($p < 1.29E-07$) and/or Xq22 ($p < 0.004179$) regions compared with the fertile females (13.8%). Also, the CNV frequency in the Xq13 ($p < 2.98E-08$) and Xq22 ($p < 0.002322$) regions were significantly higher in POI women compared with fertile females.

DISCUSSION

Previous studies on POI patients with structural rearrangements show that the X chromosome architecture and integrity are essential to maintain its functionality and efficiently

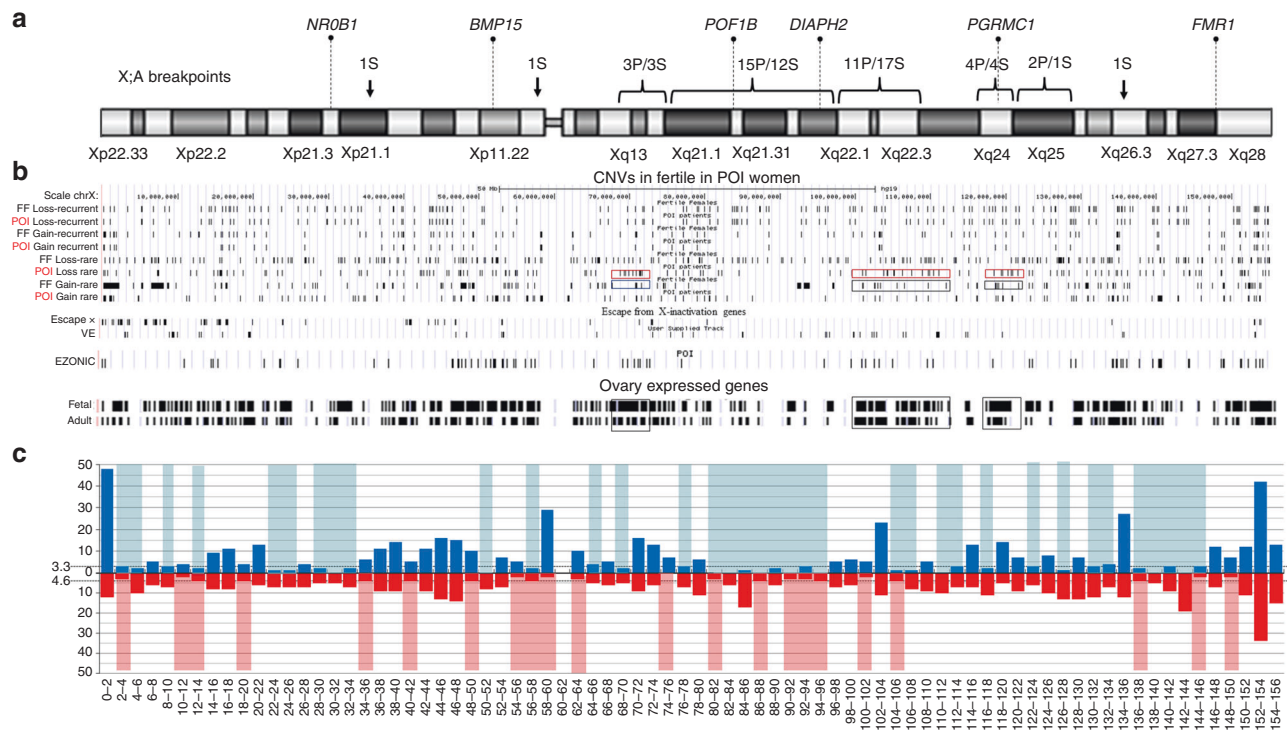


Fig. 4 The map of observed gains and losses on the X chromosome in fertile and primary ovarian insufficiency (POI) females. **a** The X chromosome ideogram. Location of genes implicated to POI are shown on the top. Breakpoints of the X chromosome–autosome balanced translocations (X; A) reported in POI women (Table S18) are clustered within Xq13, Xq21, Xq22, and Xq24. Number of patients with primary (P) and secondary (S) amenorrhea is listed for each cluster. **b** All observed CNVs detected in this study were uploaded to the University of California–Santa Cruz (UCSC) Genome Browser and displayed as Gain or Loss custom tracks. A list of genes that completely escape (Escape track) or variably escape (Var-Esc track) X inactivation was composed from previously published studies.¹⁸ Custom tracks are aligned below the X chromosome ideogram. **c** Breakpoint clustering analysis in control females. Number of unique CNV breakpoints for gains (blue bars) and losses (red bars) were calculated for each 2-Mb window along the X chromosome length (155.3 Mb). CNV-poor intervals, defined as segments with less than 5 of either gain or loss CNVs per 2-Mb window, are outlined as light blue (gain-poor) or pink (loss-poor) shaded areas. CNV copy-number variant, FF fertile females.

balance the dosage of X-linked genes with the genes expressed from autosomes. DNA copy-number variation is a substantial source of genetic diversity in humans and also contributes to single-gene and complex trait disorders via altered gene dosage or its regulation. Our analysis of CNVs and their functional impact suggest that POI is a heterogeneous condition in which the X chromosome CNVs play a more substantial role than has been anticipated. First, we discovered that in POI females CNVs are significantly enriched with ovary-expressed genes, genes that escape X inactivation, and noncoding RNA. All of this is likely to result in gene dosage imbalance. Second, POI females have 2.5-fold higher number of rare CNVs than fertile females. The number of deletion events was significantly higher among women with POI. These microdeletions may impact the X chromosome spatial structure essential for a successful meiosis, disturb boundaries of topologically active/inactive chromatin domains and sex-biased gene expression, or disorganize long-range interactions.^{28–31}

Our results suggest a polygenic contribution of genes to the etiology of secondary amenorrhea. Interestingly, 79 rare genomic intervals for which CNVs were seen in both cohorts are enriched for exonic and noncoding RNAs elements supporting the hypothesis that these CNVs may confer POI

susceptibility loci, while additional genomic changes at other loci (inherited and/or de novo) may compel the cumulative effect above an inheritance “threshold” causing a spectrum phenotype and leading to primary or secondary POI. POI patients also have an increased number of CNVs encompassing genes implicated in immune response and inflammatory processes, which supports the longstanding view that a subset of women may have autoimmune etiology of POI and a coexisting susceptibility for autoimmune diseases.^{32,33} Autoimmunity may lead to increased cell death in the ovary and result in follicle depletion, causing POI.^{34,35} Recent GWAS and exome studies in a subset of patients have identified DNA damage response pathways as important modulators of reproductive aging and gonadal dysgenesis.^{36–38} It is possible that differences observed between CNVs on the X chromosome in POI and fertile women are the consequences of deficiencies in DNA repair. This remains to be studied in the future. Another important difference between the content of CNVs in fertile and POI females is that in POI women CNVs involving noncoding intergenic and intronic sequences commonly affect gonadotropin-releasing hormone receptor binding sites.³⁹ Thus, genome structural alterations may result in an inefficient hormonal signaling or desensitization of gonadotropin secretion leading to a dysregulation of sex

steroid hormone synthesis and, as a consequence, impaired ovarian cycle and gametogenesis.⁴⁰

Our study discovered an unprecedented number of CNVs in fertile females. Surprisingly, X-linked pathogenic CNVs were found in ~3.7% of females with no family history of a genetic disorder. This rate is about two-fold higher than the estimated frequency of pathogenic sequence variants in known X-linked disease-causing genes, and was detected in our study due to exceptionally high resolution of our X chromosome microarray platform. Moreover, ~85% of X chromosome CNVs are less than 10 kb in size and likely to be missed by current diagnostic microarray designs.

In conclusion, we conducted a high-resolution copy-number analysis of the X chromosome in fertile women and individuals affected with POI. For the first time our study provided a comprehensive characterization and map of cryptic CNV and outlined key differences in the X chromosome architecture between two cohorts, providing evidence that the X chromosome may play a superb role in POI etiology, and future functional studies will be needed to assess the role of POI-unique and associated CNVs.

URLS

GENCODE <https://www.encodegenes.org/>, ENCODE <https://www.genome.gov/10005107/encode-project/>, PANTHER <http://www.pantherdb.org/>, STRING www.string-db.org

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-019-0505-2>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We are grateful to the staff of the Pittsburgh Cytogenetics Laboratory for the technical assistance in aCGH studies, and Yu Ren for assistance in RNA sequencing. This work was supported by the National Institute of Child Health and Human Development (R01HD070647, R21HD074278, A.R.).

DISCLOSURE

The authors declare no conflicts of interest.

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

REFERENCES

1. Yatsenko AN, Georgiadis AP, Röpke A, et al. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N Engl J Med*. 2015;372:2097–2107.
2. Patat O, Pagin A, Siegfried A, et al. Truncating mutations in the adhesion G protein-coupled receptor G2 gene ADGRG2 cause an X-linked congenital bilateral absence of vas deferens. *Am J Hum Genet*. 2016;99:437–442.
3. Hiort O, Holterhus PM, Horter T, et al. Significance of mutations in the androgen receptor gene in males with idiopathic infertility. *J Clin Endocrinol Metab*. 2000;85:2810–2815.
4. Nelson LM. Clinical practice. Primary ovarian insufficiency. *N Engl J Med*. 2009;360:606–614.
5. Simpson JL, Rajkovic A. Ovarian differentiation and gonadal failure. *Am J Med Genet*. 1999;89:186–200.
6. Wood MA, Rajkovic A. Genomic markers of ovarian reserve. *Semin Reprod Med*. 2013;31:399–415.
7. Qin Y, Jiao X, Simpson JL, Chen ZJ. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update*. 2015;21:787–808.
8. Heard E, Turner J. Function of the sex chromosomes in mammalian fertility. *Cold Spring Harb Perspect Biol*. 2011;3:a002675.
9. Pellestor F, Anahory T, Lefort G, et al. Complex chromosomal rearrangements: origin and meiotic behavior. *Hum Reprod Update*. 2011;17:476–494.
10. Epstein CJ. Mammalian oocytes: X chromosome activity. *Science*. 1969;163:1078–1079.
11. Deng X, Berletch JB, Nguyen DK, Distèche CM. X chromosome regulation: diverse patterns in development, tissues and disease. *Nat Rev Genet*. 2014;15:367–378.
12. Portnoi MF, Aboura A, Tachdjian G, et al. Molecular cytogenetic studies of Xq critical regions in premature ovarian failure patients. *Hum Reprod*. 2006;21:2329–2334.
13. Lakhal B, Braham R, Berquigua R, et al. Cytogenetic analyses of premature ovarian failure using karyotyping and interphase fluorescence in situ hybridization (FISH) in a group of 1000 patients. *Clin Genet*. 2010;78:181–185.
14. Moysés-Oliveira M, Guilherme Rdos S, Dantas AG, et al. Genetic mechanisms leading to primary amenorrhea in balanced X-autosome translocations. *Fertil Steril*. 2015;103:1289–e2.
15. Yatsenko SA, Bakos HA, Vitullo K, et al. High-resolution microarray analysis unravels complex Xq28 aberrations in patients and carriers affected by X-linked blue cone monochromacy. *Clin Genet*. 2016;89:82–87.
16. Martin J, Asan YiY, et al. Comprehensive carrier genetic test using next-generation deoxyribonucleic acid sequencing in infertile couples wishing to conceive through assisted reproductive technology. *Fertil Steril*. 2015;104:1286–1293.
17. Ross MT, Grafham DV, Coffey AJ, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434:325–337.
18. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005;434:400–404.
19. Tukiainen T, Villani AC, Yen A, et al. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017;550:244–248.
20. Hernandez Gifford JA. The role of WNT signaling in adult ovarian folliculogenesis. *Reproduction*. 2015;150:R137–148.
21. Chaves RN, de Matos MH, Buratini J Jr, de Figueiredo JR. The fibroblast growth factor family: involvement in the regulation of folliculogenesis. *Reprod Fertil Dev*. 2012;24:905–915.
22. Fan HY, Liu Z, Mullany LK, Richards JS. Consequences of RAS and MAPK activation in the ovary: the good, the bad and the ugly. *Mol Cell Endocrinol*. 2012;356:74–79.
23. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45:D362–D368.
24. Yang F, Deng X, Ma W, et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol*. 2015;16:52.
25. Bardoni B, Zanaria E, Guioli S, et al. A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nat Genet*. 1994;7:497–501.
26. Smyk M, Berg JS, Pursley A, et al. Male-to-female sex reversal associated with an approximately 250 kb deletion upstream of NR0B1 (DAX1). *Hum Genet*. 2007;122:63–70.
27. Dangle P, Touzon MS, Reyes-Múgica M, et al. Female-to-male sex reversal associated with unique Xp21.2 deletion disrupting genomic regulatory architecture of the dosage-sensitive sex reversal region. *J Med Genet*. 2017;54:705–709.
28. Arnold AP, Reue K, Eghbali M, et al. The importance of having two X chromosomes. *Philos Trans R Soc Lond B Biol Sci*. 2016;371:20150113.
29. Crane E, Bian Q, McCord RP, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*. 2015;523:240–244.
30. Wood AJ, Bickmore WAA. TAD closer to understanding dosage compensation. *Dev Cell*. 2015;33:498–499.
31. Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JLA. TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat Commun*. 2018;9:1444.

32. Persani L, Rossetti R, Cacciato C, Bonomi M. Primary ovarian insufficiency: X chromosome defects and autoimmunity. *J Autoimmun.* 2009;33:35–41.
33. La Marca A, Brozzetti A, Sighinolfi G, Marzotti S, Volpe A, Falorni A. Primary ovarian insufficiency: autoimmune causes. *Curr Opin Obstet Gynecol.* 2010;22:277–282.
34. Modi DN, Sane S, Bhartiya D. Accelerated germ cell apoptosis in sex chromosome aneuploid fetal human gonads. *Mol Hum Reprod.* 2003;9:219–225.
35. Forges T, Monnier-Barbarino P, Faure GC, Bene MC. Autoimmunity and antigenic targets in ovarian pathology. *Hum Reprod Update.* 2004;10:163–175.
36. Katari S, Aarabi M, Kintigh A, Mann S, Yatsenko SA, Sanfilippo JS, Zeleznik AJ, Rajkovic A. Chromosomal instability in women with primary ovarian insufficiency. *Hum Reprod.* 2018;33:531–538.
37. Wood-Trageser MA, Gurbuz F, Yatsenko SA, Jeffries EP, Kotan LD, Surti U, Ketterer DM, Matic J, Chipkin J, Jiang H, Trakselis MA, Topaloglu AK, Rajkovic A. MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am J Hum Genet.* 2014;95:754–762.
38. AlAsiri S, Basit S, Wood-Trageser MA, Yatsenko SA, Jeffries EP, Surti U, Ketterer DM, Afzal S, Ramzan K, Faiyaz-Ul Haque M, Jiang H, Trakselis MA, Rajkovic A. Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J Clin Invest.* 2015;125:258–262.
39. Metallinou C, Asimakopoulos B, Schröer A, Nikolettos N. Gonadotropin-releasing hormone in the ovary. *Reprod Sci.* 2007;14:737–749.
40. Perrett RM, McArdle CA. Molecular mechanisms of gonadotropin-releasing hormone signaling: integrating cyclic nucleotides into the network. *Front Endocrinol (Lausanne).* 2013;4:180.