Human Reproduction, Vol.34, No.3 pp. 574–583, 2019
Advanced Access publication on January 28, 2019 doi:10.1093/humrep/dey389

human reproduction

ORIGINAL ARTICLE ***Reproductive genetics***

High-resolution array-CGH analysis on 46,XX patients affected by early onset primary ovarian insufficiency discloses new genes involved in ovarian function

I. Bestetti[1,2], C. Castronovo[1], A. Sironi[1,2], C. Caslini[2], C. Sala[3], R. Rossetti[4], M. Crippa[1,2], I. Ferrari[4], A. Pistocchi[2], D. Toniolo[3], L. Persani[4,5], A. Marozzi[2], and P. Finelli[1,2,*]

[1]Istituto Auxologico Italiano, IRCCS, Lab of Medical Cytogenetics and Molecular Genetics, Milan 20145, Italy [2]Department of Medical Biotechnology and Translational Medicine, University of Milan, Segrate, Milan 20090, Italy [3]Division of Genetics and Cell Biology, San Raffaele Research Institute and Vita Salute University, Milan 20132, Italy [4]Istituto Auxologico Italiano, IRCCS, Division of Endocrine and Metabolic Diseases and Lab of Endocrine and Metabolic Research, Milan 20149, Italy [5]Department of Clinical Sciences and Community Health, University of Milan, Milan 20121, Italy

*Corresponding address. Medical Cytogenetics and Molecular Genetics Lab, IRCCS Istituto Auxologico Italiano, via Ariosto 13, Milan 20145, Italy. E-mail: finelli@auxologico.it, palma.finelli@unimi.it

Submitted on September 7, 2018; resubmitted on October 23, 2018; accepted on December 27, 2018

STUDY QUESTION: Can high resolution array-CGH analysis on a cohort of women showing a primary ovarian insufficiency (POI) phenotype in young age identify copy number variants (CNVs) with a deleterious effect on ovarian function?

SUMMARY ANSWER: This approach has proved effective to clarify the role of CNVs in POI pathogenesis and to better unveil both novel candidate genes and pathogenic mechanisms.

WHAT IS KNOWN ALREADY: POI describes the progression toward the cessation of ovarian function before the age of 40 years. Genetic causes are highly heterogeneous and despite several genes being associated with ovarian failure, most of genetic basis of POI still needs to be elucidated.

STUDY DESIGN, SIZE, DURATION: The current study included 67 46,XX patients with early onset POI (<19 years) and 134 control females recruited between 2012 and 2016 at the Medical Cytogenetics and Molecular Genetics Lab, IRCCS Istituto Auxologico Italiano.

PARTICIPANTS/MATERIALS, SETTING, METHODS: High resolution array-CGH analysis was carried out on POI patients' DNA. Results of patients and female controls were analyzed to search for rare CNVs. All variants were validated and subjected to a gene content analysis and disease gene prioritization based on the present literature to find out new ovary candidate genes. Case-control study with statistical analysis was carried out to validate our approach and evaluate any ovary CNVs/gene enrichment. Characterization of particular CNVs with molecular and functional studies was performed to assess their pathogenic involvement in POI.

MAIN RESULTS AND THE ROLE OF CHANCE: We identified 37 ovary-related CNVs involving 44 genes with a role in ovary in 32 patients. All except one of the selected CNVs were not observed in the control group. Possible involvement of the CNVs in POI pathogenesis was further corroborated by a case-control analysis that showed a significant enrichment of ovary-related CNVs/genes in patients ($P = 0.0132$; $P = 0.0126$). Disease gene prioritization identified both previously reported POI genes (e.g. *BMP15*, *DIAPH2*, *CPEB1*, *BNC1*) and new candidates supported by transcript and functional studies, such as *TP63* with a role in oocyte genomic integrity and *VLDLR* which is involved in steroidogenesis.

LARGE SCALE DATA: ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/); accession numbers SCV000787656 to SCV000787743.

LIMITATIONS, REASONS FOR CAUTION: This is a descriptive analysis for almost all of the CNVs identified. Inheritance studies of CNVs in some non-familial sporadic cases was not performed as the parents' DNA samples were not available. Addionally, RT-qPCR analyses were carried out in few cases as RNA samples were not always available and the genes were not expressed in blood.

WIDER IMPLICATIONS OF THE FINDINGS: Our array-CGH screening turned out to be efficient in identifying different CNVs possibly implicated in disease onset, thus supporting the extremely wide genetic heterogeneity of POI. Since almost 50% of cases are negative rare ovary-related CNVs, array-CGH together with next generation sequencing might represent the most suitable approach to obtain a comprehensive genetic characterization of POI patients.

STUDY FUNDING/COMPETING INTEREST(S): Supported by Italian Ministry of Health grants 'Ricerca Corrente' (08C203_2012) and 'Ricerca Finalizzata' (GR-2011-02351636, BIOEFFECT) to IRCCS Istituto Auxologico Italiano.

Key words: primary ovarian insufficiency / array-CGH / ovarian dysgenesis / *TP63* / *VLDLR*

Introduction

Primary ovarian insufficiency (POI) is a heterogeneous group of disorders whose incidence increases with increasing age (1:10,000 women by age 20, and 1:100 by age 40) (Goswami and Conway, 2005). POI describes the progression toward the cessation of ovarian function and can occur as either primary (PA) or secondary amenorrhea (SA), in syndromic or non-syndromic phenotypes (Rossetti *et al.*, 2017). While PA represents the most severe manifestation of the disorder, characterized by absent pubertal development and/or ovarian dysgenesis (OD), SA is generally associated with milder phenotypes with a post-pubertal onset (Beck-Peccoz and Persani, 2006; Cox and Liu, 2014). Genetic causes are highly heterogeneous and might explain at least some of the sporadic idiopathic cases, which comprise 50–90% of cases (Chapman *et al.*, 2015). Both numerical and structural chromosomal anomalies of which Turner's syndrome (45,X) is the most frequent, as well as 46,XX non-syndromic patients' genomic alterations have been so far reported as causative of POI (Fonseca *et al.*, 2015; Laissue, 2015; Huhtaniemi *et al.*, 2018).

Recently, thanks to genome-wide investigation based on both single-nucleotide polymorphism array and array comparative genomic hybridization (array-CGH) technologies, several X-linked and autosomal copy number variants (CNVs) affecting loci with a possible role in female fertility have been identified (Aboura *et al.*, 2009; Dudding *et al.*, 2010; Ledig *et al.*, 2010; Quilter *et al.*, 2010; Knauff *et al.*, 2011; Zhen *et al.*, 2013; Castronovo *et al.*, 2014). A few high resolution studies have analyzed small groups of 46,XX non-syndromic POI patients, either phenotypically mixed (i.e. PA, SA and infertile patients) (McGuire *et al.*, 2011; Norling *et al.*, 2014; Jaillard *et al.*, 2016) or homogeneously affected by early menopause (Tsuiko *et al.*, 2016). Besides genes already known to be involved in POI pathogenesis, other novel genes with an important role in ovarian development have been identified and proposed as candidates. However, the low detection rate («50%) suggests that most of the genetic basis of POI still needs to be elucidated, possibly by a complementary high throughput genome-wide approach such as whole exome sequencing (Tsuiko *et al.*, 2016). In addition, the extreme degree of genetic heterogeneity as well as the presence of multiple rare rearrangements/single nucleotide variants carried by single individuals (McGuire *et al.*, 2011; Norling *et al.*, 2014; Tsuiko *et al.*, 2016) seem to be constant features of POI genetic architecture.

Here, we performed the first high resolution array-CGH-based case-control investigation on a cohort of young 46,XX patients (POI onset <19 years) homogeneously affected by PA and/or OD to identify genetic defects caused by highly penetrant variants with an earlier deleterious effect on ovarian function. This approach has proved effective to clarify the role of CNVs in POI pathogenesis and to better unveil both novel candidate genes and pathogenic mechanisms.

Materials and Methods

Study design

A cohort of 46,XX patients affected by severe primary ovarian insufficiency was collected and screened by high resolution array-CGH analysis to identify novel ovary-related genes involved in the POI phenotype, which were then analyzed by disease gene prioritization to highlight the most likely candidates. To validate our approach, a case-control study using a cohort of fertile females was also carried out and the pathogenic effect of some identified CNVs containing candidate genes was further investigated at the genomic and transcript level (Fig. 1).

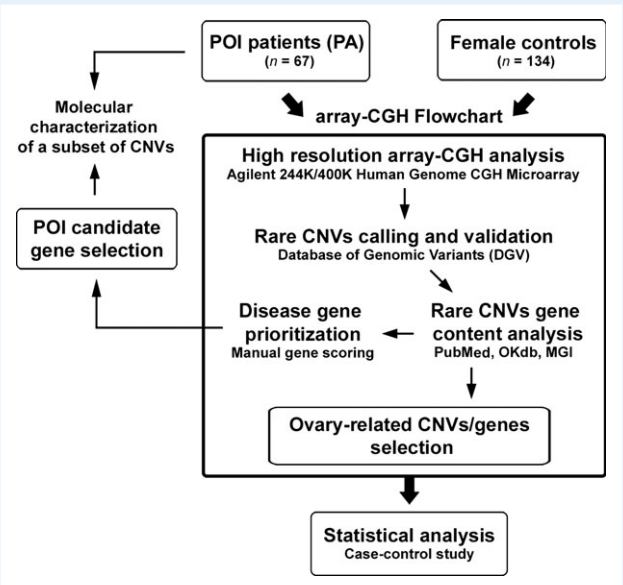

Figure 1 Flowchart of the study design.

Ethical approval

Written informed consent to the research investigation, approved by the Ethical Clinical Research Committees of IRCCS Istituto Auxologico Italiano and of San Raffaele Research Institute, was obtained from either the adult patients or parents of underage patients.

Patients and controls

There were 67 Caucasian patients affected by POI, mean age 19.2 ± 7.3 years (median: 16 years), were enrolled from July 2011 to December 2014. The patients were referred and gave their informed consent to the study at the Division of Endocrine and Metabolic Diseases of Istituto Auxologico Italiano and at the Division of Genetics and Cell Biology of San Raffaele Research Institute in Milan. The inclusion criteria were 46,XX karyotype, primary amenorrhea (61 patients) or very early secondary amenorrhea (6 patients) occurring immediately after menarche. Mean age at diagnosis was 15.1 ± 1.7 years and the hormone levels in the analyzed patients were in the postmenopausal range (FSH > 40 UI/l and estradiol <30 pg/ml). All included patients were phenotypically normal and considered idiopathic because they did not have any POI-related conditions such as ovarian surgery or previous chemo- or radio-therapy (Supplementary Table SI). Patients cohort included 35 sporadic (52.23%), and 25 familial (37.31%) POI cases; clinical details are reported in Supplementary Table SI.

The control cohort, selected from our in-house array-CGH data of healthy individuals, consisted of 134 healthy Caucasian women with a normal karyotype, who had given birth to at least one child, and with a mean age of 46.5 ± 10.4 years.

Array-CGH analysis, CNV classification and validation

The genome scan was performed using the Human Genome CGH Microarray Kits 1 × 244 K (9 patients) and 2 × 400 K (58 patients), according to the manufacturer's instructions (Agilent Technologies, Santa Clara, CA, USA). Data were extracted and analyzed for copy number changes using Agilent CytoGenomics v.3.0 (Agilent Technologies). CNV classification was performed according to the Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/, release March 2016) to exclude common polymorphic CNVs with a frequency >1% in healthy controls. To confirm the actual rarity of the CNVs identified in POI patients, in-house array-CGH results were drawn for the female controls previously analyzed with 244 K ($n = 71$) and 400 K ($n = 63$) platforms. CNVs detected by ≤5 probes or with an unknown inheritance were validated by means of real-time quantitative PCR (qPCR) or Long-Range PCR (LR-PCR). Specifically, qPCR was performed on gDNA using SYBR Green methodology (Thermo Fisher Scientific, Waltham, MA, USA) and data were analyzed through the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). LR-PCR with specific primers able to amplify the aberrations breakpoints was carried out using TaKaRa LA Taq™ kit (TaKaRa, JP). PCR products were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). Electropherograms obtained were analyzed using ChromasPro 1.5 software (Technelysium Pty Ltd., Tewantin QLD, Australia). A full list of primers is available upon request.

Ovarian CNVs selection and classification

In order to pick out CNVs with a putative role in POI etiology, the rare CNVs identified in patients and controls were assessed for gene content and compared with similar already reported CNVs using public databases such as ISCA (https://www.clinicalgenome.org/, update November 2012), and DECIPHER v9.10 (http://www.sanger.ac.uk/PostGenomics/decipher/, last accessed March 2017). Genes involved in rare CNVs were analyzed in both patients and controls according to the following features: (i) expression and function in ovary, (ii) association with POI in humans and (iii) presence of animal models. A positive feedback in at least one out of three evaluated features classified the genes and relative CNVs as ovary-related. Ovarian Kaleidoscope Database (OKdb, http://okdb.appliedbioinfo.net, last accessed July 2017) was consulted for expression, function, and reported variants, and Mouse Genome Informatics database (MGI, http://www.informatics.jax.org/, last accessed July 2017) was investigated to evaluate the presence of infertile or subfertile animal models.

Prioritization strategy

To identify the top POI candidates, we applied a prioritization strategy by scoring the genes as reported in Table I. Based on final ranking [1–8], the genes were classified as 'low score' [1–3], or 'high score' [4–8] genes. This approach was applied to both patients and controls.

Statistical analyses

Statistical analyses for the case-control study were performed by online 2 × 2 contingency table (www.vassarstats.net/odds2x2.html) for Pearson's Chi-square test (Chi-q) and R.3.1.0 software for Wilcoxon rank sum test. Chi-q was used to investigate significant differences between patients and controls in the number of individuals having either rare CNVs or ovary-related rare CNVs. Wilcoxon rank sum test was used to evaluate a relation between CNVs and POI disease. In detail, the total number of: (i) rare CNVs, (ii) ovary-related rare CNVs (total number, duplications and deletions) and (iii) common susceptibility variants were compared between patients and controls. Wilcoxon rank sum test was used to match the number of: (i) ovary-related genes [score 1–8], (ii) high score genes [4–8] and (iii) low score genes [1–3]. Data used for statistical analyses are reported in Supplementary Table SII. Identical rearrangements found in

Table I Criteria used for gene POI association score assessment.

Total score [0–8]
Expression and Function in ovary [score 0–3]
0 nothing reported related to ovary
1 gene expressed during oocyte/follicle development in animals[a]
2 gene with a known ovarian function
3 gene with more clarified ovarian function
Association to POI in humans [0–3]
0 never associated to POI
1 Reported in 1 or 2 patients with fertility defects or POI
2 Reported in more than 2 patients with POI
3 OMIM disease gene for POI
Animal model [0–1]
0 No infertile or subfertile animal models
1 Presence of infertile or subfertile animal models
Literature ovary-related [0–1]
0 Less than 5 articles related to ovarian function/POI
1 More than 5 articles related to ovarian function/POI

[a]Expression evaluated based on published articles or expression data from Ovarian Kaleidoscope Database (http://okdb.appliedbioinfo.net).

familial cases were counted as one. A *P*-value <0.05 assigned a statistically significant test.

Gene expression analysis, CNVs molecular characterizations, and luciferase assay

Transcriptional expression of ovary-related genes was evaluated on RNA from (i) adult ovary (Human Ovary Poly A+ RNA, Clontech), (ii) fetus (Human Fetus, Whole Poly A+ RNA, Clontech) and (iii) peripheral blood. cDNAs from each tissue were obtained with the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) using 500 ng of RNAs. PCR reactions were performed using the AmpliTaq Gold® kit (Thermo Fisher Scientific).

For genomic characterizations of CNVs involving *TP63* and *VLDLR*, specific primers to amplify duplication/deletion breakpoints were designed for LR-PCR reactions (TaKaRa LA Taq™) and sequencing; qPCR on gDNA was carried out with SYBR-Green technology (Thermo Fisher Scientific) for *SYCE1* rearrangements. RT-qPCR reactions for *TP63*, *VLDLR* and *DIAPH2* were performed using the TaqMan method with three housekeeping genes as normalizers (*TBP*, *GUSB*, and *RPLP0*) (Supplementary Table SIII). The $2^{-\Delta\Delta Ct}$ method was applied.

For CNVs position effect assessment using luciferase assays, we selected both CNVs adjacent to genes with a high POI association score and CNVs present in patients whose mRNA was available for further validations. The only variant that met these conditions was the 9p24.2 CNV (*VLDLR*) on POI-44. To characterize the regulatory activity of the predicted ENCODE elements (Ernst *et al.*, 2011), several firefly luciferase reporter plasmids were constructed by cloning 11 DNA fragments (BglII restriction sites) into the pGL3 promoter plasmid (Promega, USA). To characterize the activity of the fragment INS-BglII on the *VLDLR* promoter, a firefly luciferase reporter plasmid was constructed by subcloning the promoter from positions −1000 to +100 (HindIII restriction sites) into the promoterless pGL3-Basic (Promega). The fragment INS-BglII was then subcloned upstream of the *VLDLR* promoter through BglII restriction sites. Constructs were transfected on both COV434 granulosa cells and HeLa using Fugene (Invitrogen) according to the manufacturer's instructions. After treating for 16 h, cell lysates obtained with Passive Lysis Buffer 1X (Promega) were centrifuged and the supernatants were then assayed for luciferase activity using the Dual Luciferase reporter Assay kit (Promega). Luminescence in relative light units (RLU) was measured for 10 s in a Fluoroskan Ascent instrument (Labsystems). Results were then calculated as fold induction relative to pGL3/pGL3b wild-type expression vector.

Results

Array-CGH analysis and ovary-related CNVs

The array-CGH analysis identified a total of 1402 CNVs with an average of about 21 aberrations per patient. Based on the DGV, common variants were excluded and 72 rare validated CNVs in 49 patients (Supplementary Table SIV) were selected. Similarly, according to the DGV, 118 rare validated CNVs were found in 78 controls (Supplementary Table SIV). The analysis of rare CNVs gene content in patients and controls selected 44 and 48 unique ovary-related genes, respectively (Supplementary Table SV). The selected genes are included or adjacent to 37 (51.4%) CNVs in 32 patients and to 42 (35.6%) CNVs in 38 controls (Supplementary Table SIV).

To support the pathogenic effect of the selected CNVs, all detected ovary-related variants were exclusively identified in patients, except for the Xp22.31 duplication involving *STS* also found in CTR_F097 (Supplementary Table SIV). Three CNVs are adjacent to or involve two OMIM POI genes, and five CNVs overlap with variants previously reported in POI patients (Supplementary Fig. S1).

Among common rearrangements counted as susceptibility loci for POI (i.e. *NAIP* at 5q13.2, *DUSP22* at 6p25.2, and *SYCE1* at 10q26.3), 26 aberrations were found in 25 patients and 41 common aberrations were found in 39 controls (Supplementary Table SIV). All rare and selected common CNVs identified in patients were submitted to ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/, July 24 2018, date last accessed; accession numbers SCV000787656 to SCV000787743).

Gene content analysis of CNVs and disease gene prioritization indicate potentially relevant regions for POI

Among the 44 ovary-related genes, we identified two OMIM POI genes (*DIAPH2*, POF2A, OMIM#300511; *BMP15*, POF4, OMIM#300510), seven genes already associated with POI in human (*BNC1, CPEB1, PSMB1, SH3GL3, STS, TBP* and *ZFR2*) and 35 genes either not well or never associated with POI (Supplementary Table SV). The expression in the ovary was confirmed by RT-PCR on all selected ovary-related genes (data not shown) and detailed information about their function are reported in Supplementary Table SVI. The disease prioritization strategy attributing to each gene a POI association score [1–8] (Table I, Supplementary Table SVI) identified the most promising candidates (Fig. 2). Similarly, 48 ovary-related genes were identified and prioritized in controls (Supplementary Tables SV and SVI).

Statistical analysis revealed in patients a significant enrichment in ovary-related CNVs and genes

In order to validate our methodological approach and results obtained with prioritization, a case-control analysis was performed. Chi-q test showed that the number of patients carrying either rare CNVs (n = 49) or ovary-related CNVs (n = 32) is statistically significantly higher compared to the number of controls ($P = 0.0386$, $P = 0.0065$, respectively) (Fig. 3A, C). Wilcoxon rank sum test did not reveal any differences in the number of rare CNVs between patients and controls (P = 0.1852); conversely the number of rare ovary-related CNVs was statistically significant (P = 0.0132), in particular for duplications (P = 0.0128). Similar results were obtained for ovary-related genes (P = 0.0126), particularly for the genes classified with the highest POI association score ($P = 0.0192$) (Fig. 3B and D). Even though the Wilcoxon test did not reveal any difference regarding the number of previously associated common CNVs (P = 0.2494; Fig. 3B and D), no *SYCE1* deletions were detected in controls.

Pathogenic effect evaluation of particular CNVs

The identified 161 kb paternal intragenic duplication at 3q28 in POI-8 and POI-9 (Fig. 4A) involves exons 2–9 of the longest *TP63* transcript variant (NM_003722) (Fig. 4B and C). LR-PCR and sequencing of duplication breakpoints confirmed a direct tandem orientation and the inclusion of exon 10 (Fig. 4D), refining the duplication at chr3:189,424,338-189,592,155 (hg19). RT-PCR and sequencing

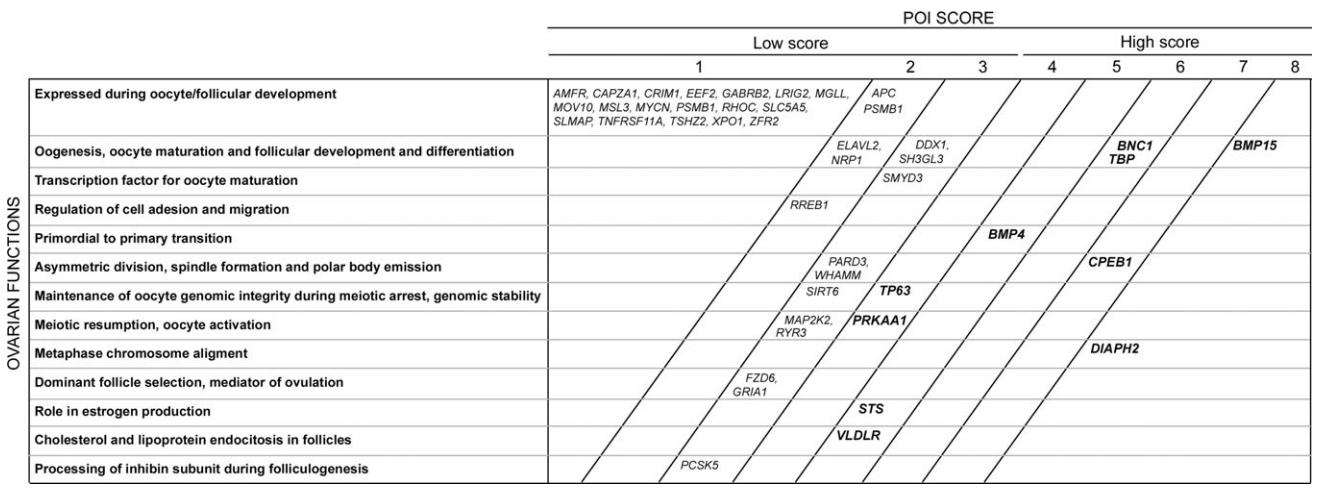

Figure 2 **Schematic view of genes according to their distribution among POI association scores from 1 (low) to 8 (high) and related functional categories**.

A

	Percentage				
	Cases	Controls	*P*	OR	CI 95%
Individuals with rare CNVs	73.13	58.12	**0.0386**	1.95	[1.0305–3.7067]
Individuals with ovary-related CNVs	47.76	28.36	**0.0065**	2.31	[1.2562–4.2469]

B

	Mean		SD		*P*
	Cases	Controls	Cases	Controls	
Rare CNVs	1.0746	0.8806	0.8699	0.9180	0.1852
Rare ovary-related CNVs	0.5075	0.3134	0.5871	0.5263	**0.0132**
Rare ovary-related duplications	0.2836	0.1493	0.4541	0.3975	**0.0128**
Rare ovary-related deletions	0.2239	0.1642	0.4200	0.3915	0.2559
Common CNVs	0.3881	0.3060	0.5209	0.4940	0.2494
Ovary-related genes	0.7015	0.3657	1.1012	0.6551	**0.0126**
Ovary-related high score genes [4–8]	0.1791	0.0746	0.4237	0.3157	**0.0192**
Ovary-related low score genes [1–3]	0.5224	0.2910	0.9432	0.5596	0.1130

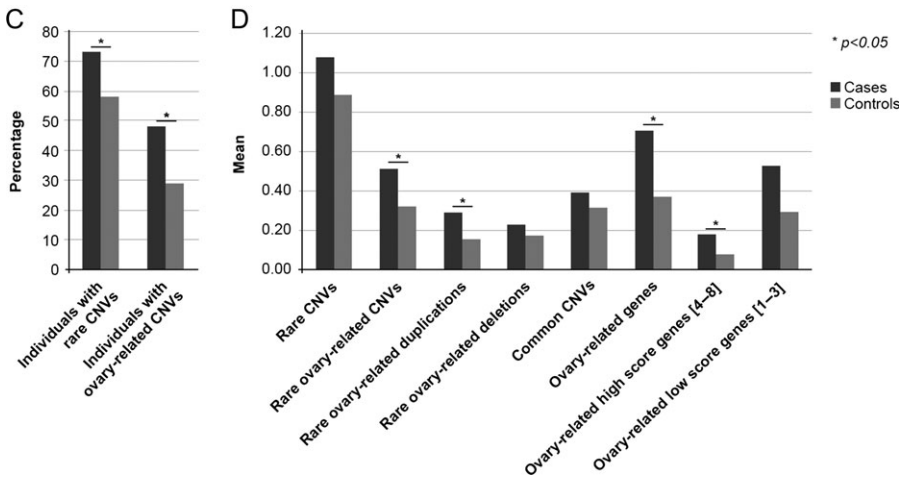

Figure 3 **Case-control statistical analysis**. (**A)** Pearson Chi-q test. Significant *P*-values (<0.05) are in bold. (**B)** Wilcoxon rank sum test without continuity correction. Mean and standard deviation (SD) are represented for both cases and controls. Significant *P*-values (<0.05) are in bold. (**C** and **D)** Ideogram view showing the results of Chi-q and Wilcoxon tests, respectively.

reactions specific for exon junction 10-2 confirmed that 3q28 duplication produces an aberrant transcript without disruption of the reading frame (Fig. 4E). RT-qPCRs carried out on junction 7–8, 11–12 indicated that *TP63* transcript levels in the patient are comparable to that in controls, and only a small amount of the duplicated aberrant mRNA is present (about 30% of the total level) (Fig. 4E). Since the

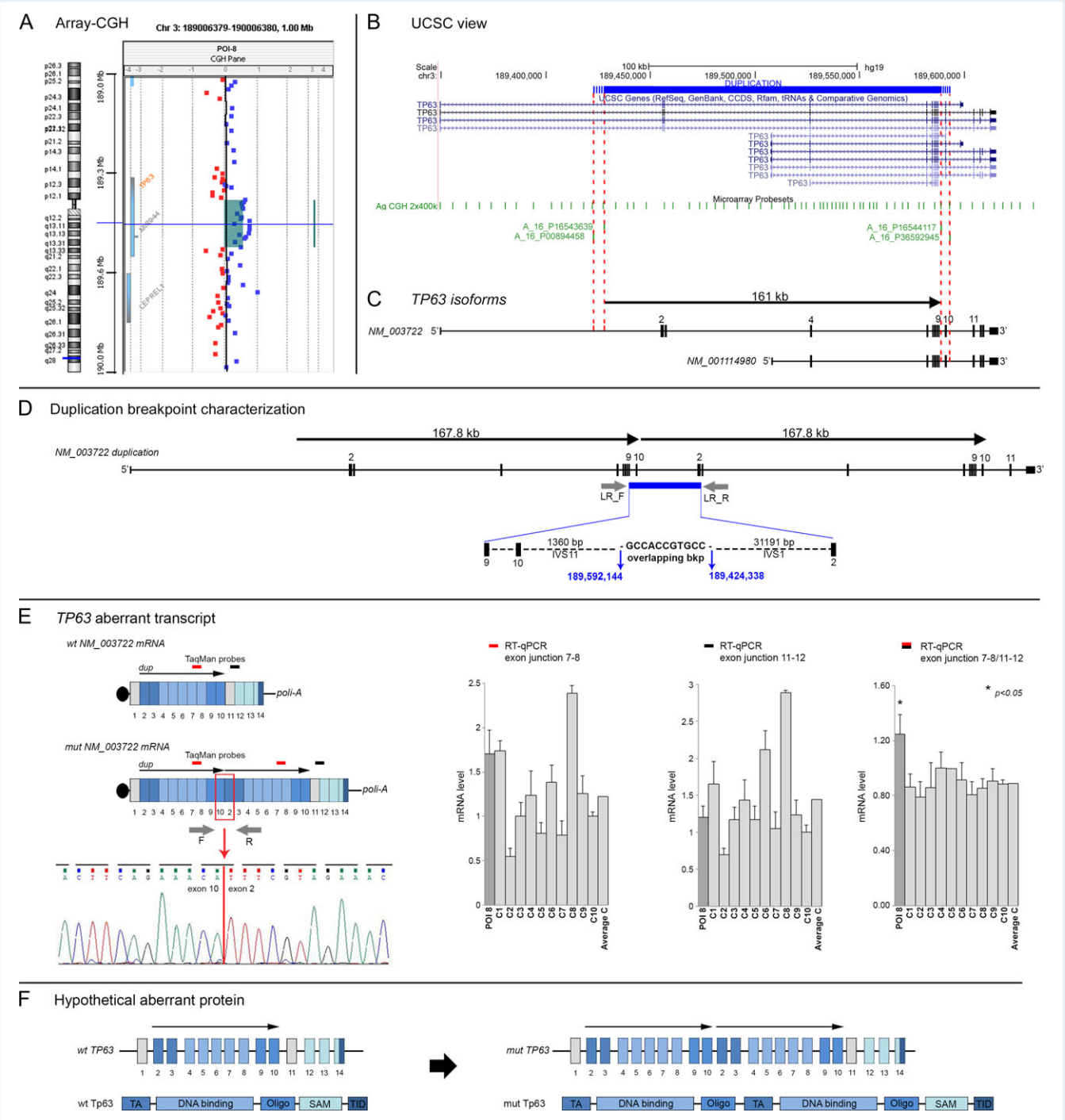

Figure 4 **Molecular characterization of *TP63* gene duplication**. (**A**) array-CGH profile output. (**B**) Duplication view according to UCSC genome browser; intragenic duplication is represented by a blue bar from IVS1 to IVS9. (**C**) *TP63* isoforms are shown with the respective duplicated portion of 161 kb based on array-CGH results. (**D**) Duplication breakpoint characterization on gDNA performed by LR-PCR and sequencing with specific primers for the duplicated allele is shown (direct tandem orientation is considered); an overlapping sequence of 11 bp between the breakpoint junctions in IVS11 and IVS1 is shown. (**E**) Wild-type mRNA and mutated transcript with the relative sequences are shown. RT-qPCR analysis revealed a statistically significant *TP63* overexpression in patients compared to controls which is indicative of the amount of the aberrant transcript. (**F**) Wild-type and hypothetical aberrant Tp63 proteins are shown.

reading frame is unaffected and that wild-type Tp63 protein is composed of 680 aa, the hypothetical aberrant protein might be formed by 1109 aa (Fig. 4F).

Although 11 CNVs might be perturbed by position effect, we focused our attention on deletion at 9p24.2 in POI-44 because it involves the high score gene *VLDLR* and patient mRNA was available

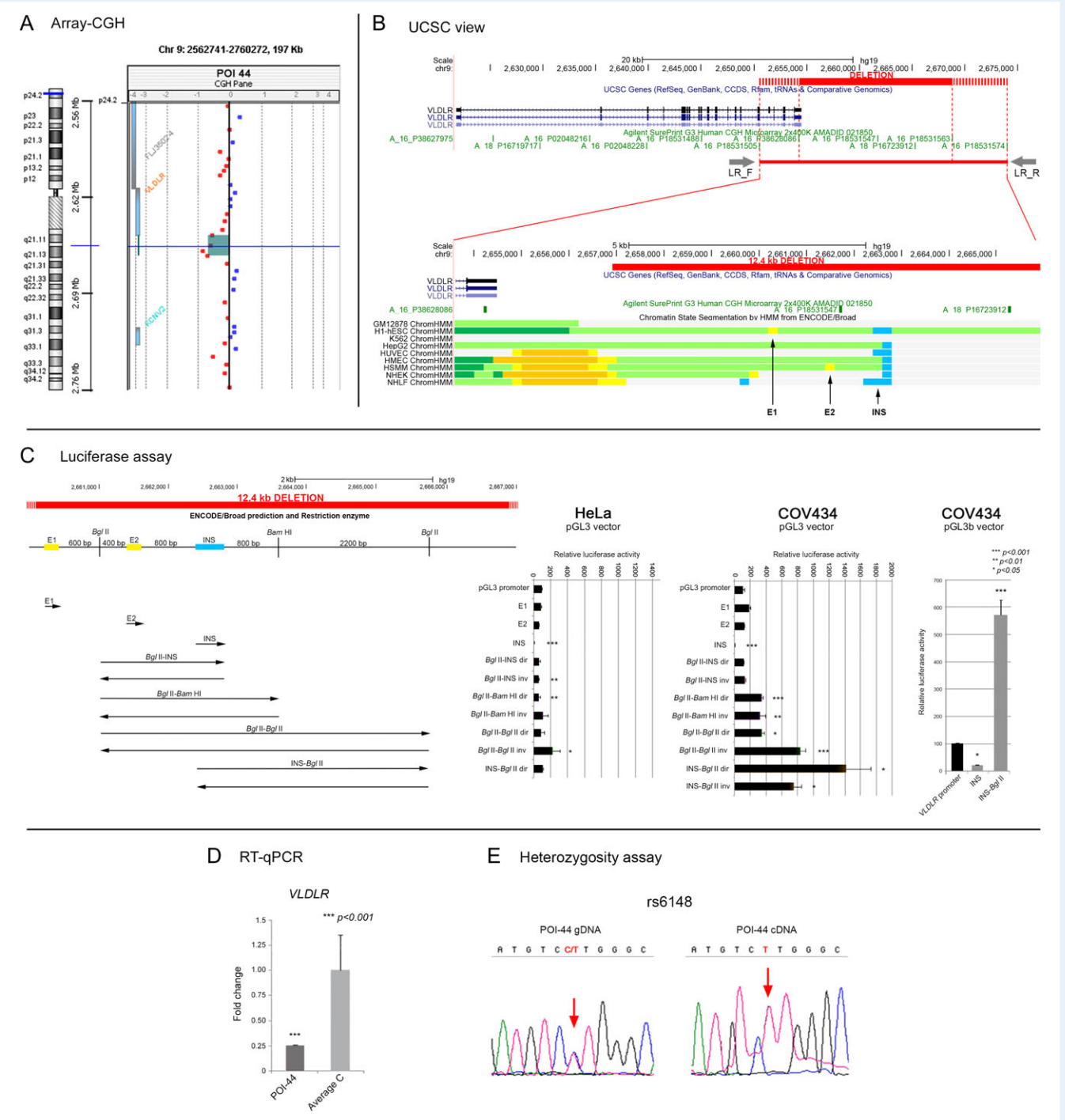

Figure 5 Molecular characterization of 9p24.2 deletion. (**A**) array-CGH profile output. (**B**) Deletion view according to UCSC genome browser; deletion is represented by a red bar from *VLDLR* 3'UTR to downstream. Primers used for LR-PCR analysis are shown and the relative output is enlarged below the red bar. Regulatory elements included are shown (E1, E2, enhancers; INS, insulator). (**C**) Luciferase assay experiments using both HeLa and COV434 cell lines. Ideograms showing the chemiluminescent emission for each construct are depicted. (**D**) RT-qPCR analysis highlighted in patient a dramatic downregulation of *VLDLR* mRNA compared to the average of controls. (**E**) The heterozygosity assay specific for rs6148 in *VLDLR* exon 14 showed a mono-allelic expression.

for further studies. Array-CGH analysis revealed a four-probe deletion of about 14.6 kb (chr9:2654,203-2668,811, hg19), involving *VLDLR* 3'UTR (NM_003383) and predicted regulatory elements (Fig. 5A and B). LR-PCR and sequencing revealed that the *VLDLR* gene was preserved confirming a deletion of 12.4 kb (chr9:2656,920-2669,483, hg19) (Fig. 5B). Through luciferase assays, the region containing the predicted

insulator and the 3000 bp downstream (INS-BglII) showed a strong specific enhancer effect in COV434, and particularly on the *VLDLR* promoter (Fig. 5C). Indeed, RT-qPCR performed on patient peripheral blood cDNA revealed a significant transcript downregulation ($P < 0.001$; Fig. 5D). Accordingly, the identification on the patients' DNA of a heterozygous SNP (rs6148) in the coding region, highlighted mono-allelic mRNA expression (Fig. 5D).

Regarding the intronic duplication at Xq21.33 involving *DIAPH2* IVS26 and the entire long non-coding RNA (*DIAPH2*-antisense), RT-qPCR was performed on POI-6 blood mRNA to evaluate a possible *DIAPH2* alteration, but no mRNA level changes were detected in the analyzed tissue (data not shown).

CNVs at 10q26.3 involving *CYP2E1* and *SYCE1* were identified in patients and controls. Namely, a deletion in POI-60 and a duplication in POI-9, POI-11, POI-17, POI-62, CTR_F003, CTR_F008, CTR_F021, CTR_F053, CTR_F084, and CTR_F086. To evaluate any differences between patients and controls *SYCE1* duplication CNVs, which might explain a different variant effect reflecting a diverse phenotype, qPCR experiments were carried out to clarify *SYCE1* 5'UTR complete or partial involvement. The analyses revealed no differences between POI patients and controls rearrangements (Supplementary Fig. S2).

Discussion

We report an array-CGH study, carried out on a cohort of 67 46,XX young patients with PA or OD, which identified both old and novel (i.e. *TP63* and *VLDLR*) potentially deleterious CNVs for ovarian function with a detection rate of 47.8%. The study carried out on a control cohort confirmed the rarity of identified variants and their possible involvement in POI pathogenesis: it revealed a significant enrichment in the number of patients presenting ovary-related variants compared to controls, and similarly in the number of ovary-related CNVs, genes and high score genes (Fig. 3). Despite ovary-related rare CNVs being identified also in controls, most of them likely do not exert any effect on the gene content expression, thus being less deleterious rearrangements compared to those of patients. In contrast, as previously reported (Tsuiko *et al.*, 2016), no differences in the number of selected common variants were found between patients and controls suggesting that common CNVs might enhance the POI burden in combination with other variants (Fig. 3).

Besides the identification of CNVs involving POI genes, such as *DIAPH2* and *BMP15*, we detected variants previously reported on POI cohorts, supporting their contribution in disease pathogenesis (Supplementary Fig. S1). Among them, the 15q25.2 deletion (POI 54), involving the ovary-related genes *CPEB1*, *BNC1, WHAMM, SH3GL3*, and *BTBD1*, was previously identified in five POI unrelated patients (McGuire *et al.*, 2011; Hyon *et al.*, 2016; Tsuiko *et al.*, 2016). Based on the severe phenotype of most patients (PA or early SA), it can be hypothesized that such a variant might contribute to an early onset of POI and that *CPEB1* can be considered a leading candidate for ovarian failure because it is involved in synaptonemal complex formation during oocyte maturation (Tay and Richter, 2001). The additional identification of a homozygous microdeletion involving *CPEB1* IVS1 (POI-46) further support its role in POI pathogenesis. The 15q25.2 region seems a POI-specific locus as other ovary-related genes are included, such as the high prioritized *BNC1* which plays a role in mouse oocytes differentiation and causes its knock-out mouse to be sub-fertile (Tian *et al.*, 2001).

The duplications at 6q27 (POI-53) and at 19p13.3 (POI-63) involving *TBP* and *ZFR2*, respectively, had been previously reported in single cases (Norling *et al.*, 2014; Tsuiko *et al.*, 2016). *TBP* is essential for mouse oocyte maturation (Gazdag *et al.*, 2007; Sun *et al.*, 2013), whereas only expression in follicles is known for *ZFR2*, whose role in the ovary is still unknown (Fig. 2).

Another reported CNV, of uncertain significance because it was shared between patients and controls (Supplementary Fig. S1), was the duplication encompassing the *STS* gene which is involved in the steroidogenesis (Otsuka *et al.*, 2005). Besides its previous identification in two SA patients (Quilter *et al.*, 2010; Tsuiko *et al.*, 2016), the evidence of the same CNV in controls highlights its unclear role in POI pathogenesis.

Deletions or duplications at 10q26.3 encompassing *SYCE1* gene (Supplementary Fig. S2) had been previously reported in several POI cohorts (McGuire *et al.*, 2011; Zhen *et al.*, 2013; de Vries *et al.*, 2014; Jaillard *et al.*, 2016; Tsuiko *et al.*, 2016). *SYCE1* plays a role in the synaptonemal complex stability during meiosis (Costa *et al.*, 2005) and its haploinsufficiency in mouse causes oocytes apoptosis and infertility (Bolcun-Filas *et al.*, 2009). In humans, a nonsense homozygous mutation in the *SYCE1* gene, observed in two POI sisters might have a similar effect (de Vries *et al.*, 2014). Supporting its pathogenic role, we found the *SYCE1* deletion only in patients (POI-60) but not among controls, confirming a frequency (1.49%) similar to that previously reported (Tsuiko *et al.*, 2016). Conversely, *SYCE1* duplication was found with a similar frequency among patients and controls (5.97% and 4.48%, respectively) and should not have any POI susceptibility role.

Our array-CGH analysis and the prioritization method uncovered 'high score genes' such as *TP63* (3q28) and *VLDLR* (9p24.2). *TP63* encodes a protein belonging to the p53 family and point mutations are associated in autosomal dominant manner with a wide phenotypic spectrum (OMIM#603273) including ADULT and Rapp-Hodgkin syndromes. In sisters, POI-8 and POI-9 no *TP63*-phenotypes were observed probably due to a different genetic defect. What is consistent with their POI disorder, is that *TP63* turns out to be a main regulator of female germ cells genomic integrity during meiotic arrest (Suh *et al.*, 2006) and its knockout mice are infertile because oocytes do not undergo apoptosis after DNA damage (Deutsch *et al.*, 2011; Kim and Suh, 2014). Indeed, *TP63* intragenic duplications, affecting the main functional domain of the protein (p-53 DNA binding domain), give rise to a stable aberrant mRNA in patients, and consequently a reduction of Tp63 activity might have deleterious effects on the oogenesis through an impaired function on genome integrity maintenance. Interestingly POI-8 and POI-9 are also carriers of a predicted probably damaging *BMP15* heterozygous mutation (p.Y235C) (Di Pasquale *et al.*, 2004), and we propose that the combination of *TP63* and *BMP15* alterations contributes to the ovarian dysgenesis seen in both sisters, consistent with the recently reported oligogenic architecture of early onset POI (Bouilly *et al.*, 2016).

VLDLR in humans is relevant for steroidogenesis as its expression in granulosa cells of pre-ovulatory follicles retains a relevant role in lipoprotein endocytosis during follicular growth (Murata *et al.*, 1998). Moreover, in hens with *VLDLR* point mutations, the oocytes fail to grow and therefore undergo atresia causing anovulation and sterility

(Bujo *et al.*, 1995). In POI-44, based on our results, the observed deletion far (2.4 kb) from *VLDLR* 3'UTR mediates a position effect on the gene, and consistent with its ovarian function, we propose that *VLDLR* haploinsufficiency could alter cholesterol endocytosis in the ovary leading to an impaired folliculogenesis.

Although further functional studies are needed to clarify and prove *TP63* and *VLDLR* involvement on POI phenotype, the whole of these array-CGH data identified novel pathogenic mechanisms for POI and supported the extremely wide genetic heterogeneity of this disease. The combination of extremely and moderately rare variants in an individual patients confirms that the oligogenic architecture can frequently underlie POI with PA or an early age at onset. In such a context, the combination of array-CGH with next generation sequencing represents the most suitable approach to obtain a comprehensive genetic characterization of POI patients.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Acknowledgments

The authors thank the patients and their families for cooperation.

Authors' roles

IB: study design, analysis and interpretation of clinical and experimental data, CNV characterizations, statistical analyses, manuscript preparation; CCast: study design, analysis and interpretation of control array-CGH data; AS: qPCR analyses for CNVs characterization; CCasl: Luciferase analysis, RT-qPCR; CS: clinical data collection, statistical analysis; RR, IF: clinical data collection; MC: interpretation of experimental data; AP: Luciferase analysis; DT, LP: clinical data collection, critical revision of the manuscript; AM: clinical data collection, Luciferase analysis, critical revision of the manuscript; PF: study design, analysis and interpretation of clinical and experimental data, critical revision of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by Italian Ministry of Health grants 'Ricerca Corrente' to IRCCS Istituto Auxologico Italiano (08C203_2012) and 'Ricerca Finalizzata' funds for Young Researchers (GR-2011-02351636, BIOEFFECT) to RR.

Conflict of interest

The authors declare that they have no conflict of interests.

References

Aboura A, Dupas C, Tachdjian G, Portnoi MF, Bourcigaux N, Dewailly D, Frydman R, Fauser B, Ronci-Chaix N, Donadille B *et al*. Array comparative genomic hybridization profiling analysis reveals deoxyribonucleic acid copy number variations associated with premature ovarian failure. *J Clin Endocrinol Metab* 2009;**94**:4540–4546.

Beck-Peccoz P, Persani L. Premature ovarian failure. *Orphanet J Rare Dis* 2006;**1**:9.

Bolcun-Filas E, Hall E, Speed R, Taggart M, Grey C, de Massy B, Benavente R, Cooke HJ. Mutation of the mouse Syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and DNA repair. *PLoS Genet* 2009;**5**:e1000393.

Bouilly J, Beau I, Barraud S, Bernard V, Azibi K, Fagart J, Fèvre A, Todeschini AL, Veitia RA, Beldjord C *et al*. Identification of multiple gene mutations accounts for a new genetic architecture of primary ovarian insufficiency. *J Clin Endocrinol Metab* 2016;**101**:4541–4550.

Bujo H, Yamamoto T, Hayashi K, Hermann M, Nimpf J, Schneider WJ. Mutant oocytic low density lipoprotein receptor gene family member causes atherosclerosis and female sterility. *Proc Natl Acad Sci USA* 1995;**92**:9905–9909.

Castronovo C, Rossetti R, Rusconi D, Recalcati MP, Cacciatore C, Beccaria E, Calcaterra V, Invernizzi P, Larizza D, Finelli P *et al*. Gene dosage as a relevant mechanism contributing to the determination of ovarian function in Turner syndrome. *Hum Reprod* 2014;**29**:368–379.

Chapman C, Cree L, Shelling AN. The genetics of premature ovarian failure: current perspectives. *Int J Womens Health* 2015;**7**:799–810.

Costa Y, Speed R, Ollinger R, Alsheimer M, Semple CA, Gautier P, Maratou K, Novak I, Hoog C, Benavente R *et al*. Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis. *J Cell Sci* 2005;**118**:2755–2762.

Cox L, Liu JH. Primary ovarian insufficiency: an update. *Int J Womens Health* 2014;**6**:235–243.

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**:E2129–E2132.

Deutsch GB, Zielonka EM, Coutandin D, Weber TA, Schafer B, Hannewald J, Luh LM, Durst FG, Ibrahim M, Hoffmann J *et al*. DNA damage in oocytes induces a switch of the quality control factor TAp63alpha from dimer to tetramer. *Cell* 2011;**144**:566–576.

Di Pasquale E, Beck-Peccoz P, Persani L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. *Am J Hum Genet* 2004;**75**:106–111.

Dudding TE, Lawrence O, Winship I, Froyen G, Vandewalle J, Scott R, Shelling AN. Array comparative genomic hybridization for the detection of submicroscopic copy number variations of the X chromosome in women with premature ovarian failure. *Hum Reprod* 2010;**25**:3159–3160.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011;**473**:43–49.

Fonseca DJ, Patino LC, Suarez YC, de Jesus Rodriguez A, Mateus HE, Jimenez KM, Ortega-Recalde O, Diaz-Yamal I, Laissue P. Next generation sequencing in women affected by nonsyndromic premature ovarian failure displays new potential causative genes and mutations. *Fertil Steril* 2015;**104**:154–162.

Gazdag E, Rajkovic A, Torres-Padilla ME, Tora L. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* 2007;**134**:51–62.

Goswami D, Conway GS. Premature ovarian failure. *Hum Reprod Update* 2005;**11**:391–410.

Huhtaniemi I, Hovatta O, La Marca A, Livera G, Monniaux D, Persani L, Heddar A, Jarzabek K, Laisk-Podar T, Salumets A *et al*. Advances in the molecular pathophysiology, genetics, and treatment of primary ovarian insufficiency. *Trends Endocrinol Metab* 2018;**29**:400–419.

Hyon C, Mansour-Hendili L, Chantot-Bastaraud S, Donadille B, Kerlan V, Dode C, Jonard S, Delemer B, Gompel A, Reznik Y *et al*. Deletion of

CPEB1 gene: a rare but recurrent cause of premature ovarian insufficiency. *J Clin Endocrinol Metab* 2016;**101**:2099–2104.

Jaillard S, Akloul L, Beaumont M, Hamdi-Roze H, Dubourg C, Odent S, Duros S, Dejucq-Rainsford N, Belaud-Rotureau MA, Ravel C. Array-CGH diagnosis in ovarian failure: identification of new molecular actors for ovarian physiology. *J Ovarian Res* 2016;**9**:63.

Kim DA, Suh EK. Defying DNA double-strand break-induced death during prophase I meiosis by temporal TAp63alpha phosphorylation regulation in developing mouse oocytes. *Mol Cell Biol* 2014;**34**:1460–1473.

Knauff EA, Blauw HM, Pearson PL, Kok K, Wijmenga C, Veldink JH, van den Berg LH, Bouchard P, Fauser BC, Franke L. Copy number variants on the X chromosome in women with primary ovarian insufficiency. *Fertil Steril* 2011;**95**:1584–1588.

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**:243–257.

Ledig S, Ropke A, Wieacker P. Copy number variants in premature ovarian failure and ovarian dysgenesis. *Sex Dev* 2010;**4**:225–232.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**:402–408.

McGuire MM, Bowden W, Engel NJ, Ahn HW, Kovanci E, Rajkovic A. Genomic analysis using high-resolution single-nucleotide polymorphism arrays reveals novel microdeletions associated with premature ovarian failure. *Fertil Steril* 2011;**95**:1595–1600.

Murata M, Tamura A, Kodama H, Hirano H, Takahashi O, Tanaka T. Possible involvement of very low density lipoproteins in steroidogenesis in the human ovary. *Mol Hum Reprod* 1998;**4**:797–801.

Norling A, Hirschberg AL, Rodriguez-Wallberg KA, Iwarsson E, Wedell A, Barbaro M. Identification of a duplication within the GDF9 gene and novel candidate genes for primary ovarian insufficiency (POI) by a customized high-resolution array comparative genomic hybridization platform. *Hum Reprod* 2014;**29**:1818–1827.

Otsuka Y, Yanaihara A, Iwasaki S, Hasegawa J, Yanaihara T, Okai T. Localization and gene expression of steroid sulfatase by RT-PCR in cumulus cells and relationship to serum FSH levels observed during in vitro fertilization. *J Exp Clin Assist Reprod* 2005;**2**:6.

Quilter CR, Karcanias AC, Bagga MR, Duncan S, Murray A, Conway GS, Sargent CA, Affara NA. Analysis of X chromosome genomic DNA sequence copy number variation associated with premature ovarian failure (POF). *Hum Reprod* 2010;**25**:2139–2150.

Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. *Clin Genet* 2017;**91**:183–198.

Suh EK, Yang A, Kettenbach A, Bamberger C, Michaelis AH, Zhu Z, Elvin JA, Bronson RT, Crum CP, McKeon F. p63 protects the female germ line during meiotic arrest. *Nature* 2006;**444**:624–628.

Sun SC, Wang XG, Ma XS, Huang XJ, Li J, Liu HL. TBP dynamics during mouse oocyte meiotic maturation and early embryo development. *PLoS One* 2013;**8**:e55425.

Tay J, Richter JD. Germ cell differentiation and synaptonemal complex formation are disrupted in CPEB knockout mice. *Dev Cell* 2001;**1**:201–213.

Tian Q, Kopf GS, Brown RS, Tseng H. Function of basonuclin in increasing transcription of the ribosomal RNA genes during mouse oogenesis. *Development* 2001;**128**:407–416.

Tsuiko O, Noukas M, Zilina O, Hensen K, Tapanainen JS, Magi R, Kals M, Kivistik PA, Haller-Kikkatalo K, Salumets A *et al*. Copy number variation analysis detects novel candidate genes involved in follicular growth and oocyte maturation in a cohort of premature ovarian failure cases. *Hum Reprod* 2016;**31**:1913–1925.

Zhen XM, Sun YM, Qiao J, Li R, Wang LN, Liu P. [Genome-wide copy number scan in Chinese patients with premature ovarian failure]. *Beijing Da Xue Xue Bao* 2013;**45**:841–847.