Causal and Candidate Gene Variants in a Large Cohort of Women with Primary Ovarian Insufficiency

Gorsi B[1], Hernandez EJ[1], Moore B[1], Moriwaki M[2], Chow CY[3], Coelho E[3], Taylor E[4], Lu C[4], Walker A[4], Touraine P[5], Nelson LM[6], Cooper AR[7], Mardis ER[8], Rajkovic A[9], Yandell M[1], Welt CK[2]

[1]Utah Center for Genetic Discovery, Department of Human Genetics, University of Utah, Salt Lake City, UT, USA

[2]Division of Endocrinology, Metabolism and Diabetes, University of Utah, Salt Lake City, UT, USA

[3]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA

[4]University of Utah School of Medicine, Salt Lake City, UT, USA

[5]Sorbonne Universite, Hôpital Universitaire Pitié Salpêtrière-Charles Foix, Service d'Endocrinologie et Médecine de la Reproduction, Centre de Maladies Endocriniennes Rares de la Croissance et du Développement, Centre de Pathologies Gynécologiques Rares, Paris, France

[6]Conover Foundation, McLean, VA, USA

[7]Vios Fertility Institute, St. Louis, MO, USA

[8]Institute for Genomic Medicine, Nationwide Children's Hospital, Ohio State University College of Medicine, Columbus, OH 43215

[9]Department of Pathology, University of California San Francisco School of Medicine, San Francisco, CA, 94143 USA

Correspondence to

Corrine Welt

Division of Endocrinology, Metabolism and Diabetes

University of Utah

Eccles Institute of Human Genetics

15 N 2030 E

Salt Lake City, UT 84112

USA

Phone: 801-585-1875

Email: cwelt@genetics.utah.edu

Short Title:

Disclosure: MY is a stock holder or has received stock option awards from Fabric Genomics Inc. BM and MY have received consulting fees from Fabric Genomics Inc. ARC is a consultant for Ferring and a scientific advisory board member for INVObioscience and Celmatix.

Author Contributions:

Conception of study CKW, BG

Design of work BG, EJH, MM, CYC

Acquisition of Data LMN, ARC, ERM, AR, PT, AW, MM, EC, CYC

Analysis of data ET, CL, AW, CKW, EJH, BM, BG, MJ, EC, CYC

Drafted and revised AW, MY, CKW, MM, EJH, BG, CYC, BM, MY

All authors approved the final version.

Abstract

A genetic etiology accounts for unexplained primary ovarian insufficiency (POI; amenorrhea with an elevated FSH level). Subjects with POI (n=291) and controls recruited for health in old age or 1000 Genomes (n=233) underwent whole exome or whole genome sequencing. Data were analyzed using a rare variant scoring method and a Bayes factor-based framework for identifying genes harboring pathogenic variants. Candidate heterozygous variants were identified in known genes and genes with functional evidence. Gene sets with increased burden of deleterious alleles included the categories transcription and translation, DNA damage and repair, meiosis and cell division. Variants were found in novel genes from the enhanced categories. Functional evidence supported 7 new risk genes for POI (*USP36, VCP, WDR33, PIWIL3*, *NPM2*, *LLGL1* and *BOD1L1*). Aggregating clinical data and genetic risk with a categorical approach may expand the genetic architecture of heterozygous rare gene variants causing risk for POI.

Introduction

Primary ovarian insufficiency (POI) encompasses a continuum from infertility in women with ovarian dysfunction to early menopause[1]. The cause of POI remains unknown in the majority of women, making intervention impossible to initiate until it is too late[1].

Data overwhelmingly support a genetic cause in women with POI[2-4]. Twin studies estimate heritability from 53-71%[2-4]. There is a strong relationship between age at menopause in mothers and daughters, with an odds ratio of 6 (95% confidence intervals 3.4, 10.7) for early menopause in daughters whose mothers had early menopause[5]. In small studies it has been estimated that up to 30% of POI cases are familial[6]. The most common known genetic causes include X chromosome defects, *FMR1* premutations and autoimmune causes[1]. Nevertheless, the additive effect of these and known iatrogenic causes explain less than 30% of POI. A remarkable number of new POI-associated genes have been discovered, facilitated by whole exome sequencing (WES) in consanguineous and large families[7-15]. The women in these families typically develop POI before puberty, also termed ovarian dysgenesis. Mutations in the DNA of these women have been identified in genes important for mitochondrial function, meiosis, homologous recombination and DNA damage repair[7,8,10,12,16].

The inheritance pattern for POI is not recessive in all cases. Heterozygous mutations in genes such as *eIF4ENIF1* cause POI in women in the mid reproductive years[9]. Recessive gene mutations found to cause POI and primary amenorrhea also cause POI or earlier menopause in heterozygous mothers, demonstrating that dominant and semi-dominant mutations may be causal[17,18], with heterozygous damaging variants in known genes or in two or more candidate genes causing POI[19,20]. One study of POI suggested an additive effect from common variants contributing to age at menopause, with a recent study suggesting that common variation may explain a portion of earlier age at menopause, as low as age 34 years[21,22].

Most previous WES analyses in large numbers of women with sporadic POI used a candidate gene approach to identify gene mutations most likely to cause POI. However, a

variant-centric approach has identified novel POI candidate genes[19,23]. We used an unbiased approach and a new prioritization algorithm (GEM)[24-26] to identify damaging gene variants in known POI genes. We then used a category-wide association approach to test the hypothesis that additional candidate mutations could be found in clustered gene sets created using known genes and gene candidates from model organisms[27,28]. We demonstrated a significant enhancement in identified gene sets in women with POI compared to controls. These gene sets revealed additional candidate genes in POI, with seven genes confirmed by functional studies to play a role in oocyte or ovary development. These findings improve our understanding of the genetic architecture of POI, an extremely heterogeneous disorder.

Subjects

All subjects were diagnosed with POI as defined by at least 4 months of amenorrhea and an FSH level in the menopausal range. All women were 18 years or older, and had a 46XX karyotype, and normal *FMR1* repeat number. Subjects (n=35) were recruited in Boston. Additional subjects were recruited from the Partners Biobank (n=63). A second cohort (n=98) was recruited at the National Institute of Health (NIH) for a study of non-syndromic POI[29]. These subjects were re-consented to have their DNA undergo WES at Washington University (LMN, ARC and ERM). A third cohort was recruited from Pittsburgh (n=20), Italy (n=43) and France (n=32)(AR and PT). All Boston subjects underwent a medical history and physical exam and family history. Subjects from the NIH, Pittsburgh, Italy and France had limited phenotypic data.

Control subjects for category-wide association using GEM[24] included 96 unrelated, unaffected subjects recruited for health in old age and 137 CEU, FIN and GBR samples from the 1000 Genomes Project (total n=233 controls)[30,31]. The majority of CEU samples are from Utah families recruited for large family size (n=47 of 61)[32]. The control subjects underwent whole genome sequencing, as previously described[33]. All subjects provided written, informed consent from the University of Utah, Washington University, University of Pittsburgh or the Sorbonne Universite IRB.

Methods

DNA samples were extracted (Qiagen) and subjected to WES. The Boston cohort was sequenced using the Illumina HiSeq 2000 (Illumina). All candidate susceptibility variants in the Boston cohort were Sanger sequenced for verification. Sequencing of the NIH/Washington University cohort was performed using the Roche NimbleGen VCRome 2.1 (HGSC design) exome capture and the Illumina HiSeq 2500 for sequencing at the McDonnell Genome Institute at Washington University. The dataset was accessed through dbGAP (NIH approved request #47895-1, Project #11971). The Pittsburgh, French and Italian cohorts were sequenced at the

Pittsburgh Clinical Genomics Laboratory using the Haloplex Exome Target Enrichment System or the Agilent SureSelect V5 Capture Kit (Agilent Technologies, Santa Clara, CA), and 2× 100 bp paired-end WES was performed on an Illumina HiSeq 2500 (San Diego, CA, USA).

The control subjects' DNA underwent whole genome sequencing (WGS) using the Illumina X Ten sequencing platform (Nantomics, Culver City, CA). The comparison of variants in cases using WES versus controls using WGS would result in a conservative estimate of variants in cases based on the higher coverage expected from WGS.

Alignment and Variant Calling

Alignment and variant calling were performed by the Utah Center for Genetic Discovery (UCGD) core services. Fastq files were downloaded from the Pittsburgh Clinical Genomics Laboratory and dbGAP. Variants were called through the UCGD pipeline using the Sentieon software package (https://www.sentieon.com) [34]. Reads were aligned to the human reference build GRCh37 using BWA-MEM (Burrows-Wheeler Aligner). SAMBLASTER was used to mark duplicate reads and de-duplicate aligned BAM files. Aligned BAM files underwent INDEL realignment and base recalibration using Realigner and QualCal algorithms from the Sentieon software package3 to produce polished BAM files. Each polished BAM file was processed using the Sentieon's Haplotyper algorithm to produce gVCF files[35]. Sample gVCF files were combined and jointly genotyped with 728 samples comprised of the 1000 genomes project (CEU) samples and samples unrelated to reproduction or cancer phenotypes to produce a multi-sample VCF file. To produce the final VCF variant quality scores, VCF files were recalibrated using Sentieon's VarCal algorithm to estimate the accuracy of variant calls and reduce potential false positive calls.

Quality Control

Quality control algorithms were applied to sequence reads (Fastq files), aligned reads (BAM files) and variants (VCF files)[36]. Fastp was used to evaluate read quality, read duplication rate, presence of adapter and overrepresented sequences in Fastq files[37]. Indexcov was used to estimate depth and coverage of aligned sequence data using BAM indexes. Further alignment quality metrics were calculated on BAM files with samtools stats. The 291 cases were sequenced using different exome capture kits, we therefore standardized QC analysis regions with a bed file made up of exonic regions from coding gene models from RefSeq and Ensemble gene sets. These regions were used to obtain the total number of reads, percentage aligned reads and mean and median coverage for all samples.

Variant quality metrics were calculated by running bcftools stats[38-40]. The overall quality of VCF callsets were evaluated using Peddy to confirm sex, relatedness, heterozygosity and ancestry of each individual and identifying potential sample-level data quality issues[41].

Identification of Damaging Gene Variants

For each case, the uploaded VCF file was scored with VAAST Variant Prioritizer (VVP) and Variant Annotation Analysis and Search Tool (VAAST) to prioritize potentially deleterious variants and damaged genes[25,42]. VVP and VAAST use a likelihood ratio test (LRT) to score each variant and the aggregate burden of variants for each gene in affected individuals relative to a set of 2,492 control genomes of healthy individuals from the 1000 Genomes Project[43]. The LRT incorporates three components of each variant; the severity of amino acid substitution, phylogenetic conservation of the variant, and the frequency of the variant relative to the control population. The sum of the top scoring variant(s) based (one variant for dominant inheritance and two variants for recessive inheritance) represents the cumulative likelihood ratio (CLR) for a given gene. The significance of each gene's VAAST CLRT score is evaluated by a permutation test that randomizes the case/control status of individuals in each of $1e^6$ permutations and

generates a permutation *p* value for the gene. The output from VAAST is an ordered gene list ranked for the probability of being damaged relative to the control genomes.

Variants identified in the VAAST analysis above were further refined by selecting only variants found at a minor allele frequency (MAF) <0.001 and with no homozygotes found in gnomAD[44,45]. The choice of a MAF <0.001 cutoff was based on the frequency of the fragile X premutation. The prevalence of the premutation in the population is 0.004 and it is the most common single gene cause of POI identified to date. A fragile X premutation accounts for only 6% of sporadic POI cases[46], making 0.001 a conservative upper bound for the risk attributable to any one gene. We also removed variants in genes known to tolerate a large burden of genetic variation such as olfactory receptors, snoRNAs, mucins and T cell receptors[47]. Finally, we required an Omicia score of >0.7; a meta-classifier that combines scores from SIFT, PolyPhen, MutationTaster and PhyloP to predict pathogenicity[48-53]. A range of 4-25 damaging gene variants were found per person.

GEM Analysis

We also used GEM to identify gene variants in each subject that were most likely to be pathogenic[24]. GEM is an Electronic Clinical Decision Support System (eCDSS) framework that aggregates and adjudicates data from multiple algorithms and clinical datasets to provide rapid and accurate diagnosis of individual genomes[24]. GEM generates a Bayes Factor-based score that calculates the degree of support for and against a given model (a gene allele is pathogenic vs. benign) considering multiple lines of evidence from the following variant analysis tools and data sources: VVP, VAAST, Phevor, mode of inheritance for disease genes from Online Mendelian Inheritance in Man (OMIM), pathogenicity of variants (ClinVar), population specific allele frequencies (gnomAD), quality of variants and the overall genome (data) and quality of the genomic location (gnomAD)[25,26,42,44,54-56]. Using this data, GEM identifies potentially pathogenic genotypes and evaluates support for their association with disease. Gene variants for each

subject were considered candidates if they had a GEM score ≥ 1 (strong support for the model of pathogenicity)[57], together with genes having a GEM score ≥ 0.69 (substantial support for the model of pathogenicity), and a Phevor Bayes factor ≥ 0.9 (genes with a strong association with POI)[54]. One to twenty-four gene variants were identified for each subject.

Creating Categories for Enrichment Analysis

We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) to functionally annotate known POI genes and candidate genes identified in model organisms [14,58]. The analysis yielded 47 clusters with one removed for too few genes (<20; Cluster 37)(Supplementary Table 1). Cluster 37 contained mismatch repair genes, but the genes were also found in other clusters and was therefore redundant.

Calculating Gene Burden on Resampled Gene Lists and Housekeeping Genes

We randomly selected 146 genes from a list of housekeeping genes that are constitutively expressed over many developmental time points in 16 tissue types (Supplementary Table 2). We ensured that none of the housekeeping genes were found in our gene sets identified in the GEM analysis. In addition, we created a burden-matched set of genes. For this, we created a burden ratio for every gene by summing the number of rare variants (MAF ≤ 0.005)[59] in the longest coding transcript of each gene and dividing by transcript length. For each decile in the distribution of this burden ratio, we determined the mean and standard deviation of the burden ratio. We then used these mean and standard deviations to generate randomly sampled gene sets for each decile that had matched mean and standard deviation for burden ratio. These burden matched gene sets were used to test the significance of gene set enrichment in the subjects.

Permutation Tests and Case/Control Comparison

For the permutation tests and analysis, we used the GEM results with a GEM score ≥ 1 generated from the POI cases and control individuals to test for enrichment in individuals with POI compared to controls. Both sets of data generated GEM results using two different HPO terms: POI (HP:0008209) and phenotypic abnormality (HP:0000118) to create 4 sets of data: Cases POI, Cases Phenotypic Abnormality (root), Controls POI, Controls Phenotypic Abnormality (root). The reason we ran GEM using the root of the HPO ontology (Phenotypic Abnormality) was to control for overly connected genes that might have inflated Phevor scores, thus reducing biases due to the nature of the ontology. We then determined the number of damaged genes found in GEM results (number of successes) from genes listed in the individual pathways, a burden-matched gene list and a housekeeping gene list.

Permutation analyses were performed using a random sampling strategy to evaluate enrichment of the POI dataset against gene lists related to functional aspects of the disease. A gene list containing 18,876 RefSeq genes was first created, excluding mucins and olfactory receptor genes. For each functional gene list of size N, random samples of equal size were drawn from the 18,876 genes. This process was repeated 100,000 times, each time with an independently generated random gene list, to create an empirical distribution of the number of damaged genes (GEM score ≥ 1) for each functional gene list. To test whether the probands show enrichment in the functional gene lists, the actual number of damaged genes found for each list was compared to the distribution of damaged genes found using burden-matched gene lists and the housekeeping genes list. To test for statistical significance, we used Fisher's exact test to calculate a *p* value using the 2x2 contingency table testing the hypothesis that the number of damaging genes that matched the functional list was significantly larger in the POI GEM runs than in the root Phenotypic Abnormality GEM runs. All *p* values were adjusted for multiple testing (False Discovery Rate, FDR). To generate a final score depicting the most significantly enriched pathways, adjusted POI *p* values were divided by Phenotypic Abnormality *p* values to generate a normalized score that represents enrichment. The higher the ratio, the

more enriched the pathway. Pathways with a corrected *p* value <0.05 and a $\log_2$ ratio of greater than 2 for the POI *p* value/Phenotypic Abnormality *p* value were considered significant pathways.

Oocyte Expression

To determine whether candidate genes are expressed in mammalian oocytes, 35 day-old female mice were treated with an intraperitoneal injection of 5 IU PMSG to initiate follicular development and 5 IU hCG 48 hours later to induce ovulation[60]. Eighteen hours later, mice were sacrificed, oviducts dissected to remove oocytes and cumulus cells manually removed. RNA was isolated from oocytes using RNeasy (Qiagen, Valencia, CA)[9]. Reverse transcription was performed with SuperScript Master Mix (Life Technologies, Carlsbad, CA) using SuperScript III RT and random primers. Quantitative real-time polymerase chain reaction was performed for the expression of candidate genes and glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) as an endogenous control using PowerUp SYBR Green Master Mix (Applied Biosystems, Foster City, CA). Primers were designed to span two exons to avoid amplifying genomic DNA. Primer sequences are provided (Supplementary Table 3). Samples were examined in triplicate and at three dilutions. mRNA levels were determined using the $2^{-\Delta\Delta CT}$ method to calculate relative quantification and to correct for expression of endogenous controls.

Functional Analysis

Flies were raised at 25º C on standard diet based on the Bloomington *Drosophila* Stock Center standard medium with malt. We obtained 20 RNAi lines from the Bloomington *Drosophila* Stock Center. Ovary/germline specific RNAi knockdowns were performed using Gal4 DNA-binding protein and Upstream Activator Sequence (*GAL4/UAS*) technology, as previously described[61]. We crossed flies carrying the *Maternal Triple Driver-GAL4* (*MTD-GAL4*; BDSC 31777) transgene to flies carrying each respective UAS-RNAi transgene to generate female flies

with ovary specific knockdown of each gene. Control flies were generated by crossing flies carrying the *MTD-GAL4* transgene to the appropriate *AttP* RNAi background strain (does not carry UAS-RNAi transgene). Virgin female RNAi knockdown (and control) flies were collected on CO_2 anesthesia and aged 3-5 days on standard media supplemented with dry yeast. Female knockdown flies were singly mated with a 3-5 day old Canton S male. Individual mating pairs were observed to ensure successful mating. Males were removed after mating. We measured four female reproductive phenotypes: 1) egg number: number of eggs laid in first 8 hrs post mating; 2) hatchability: number of adults that hatched from those eggs; 3) total fertility over 10 days post mating; 4) overall ovary appearance and morphology[62]. For egg number, newly mated females were place in vials for 8 hrs and egg number was counted. For hatchability, all the progeny that eclosed from the egg number vial were counted (progeny #/egg #). To measure total fertility, mated females were transferred to new vials every two days for ten days and all the progeny were summed over the entire period. For ovary images, adult females were collected under CO_2 anesthesia, dissected and immediately imaged. Ovaries were imaged at 3X magnification using a Leica EC3 camera. We assayed at 8-10 females per RNAi knockdown. Statistical analysis was performed using R software. P-values were determined using ANOVA. A p value < 0.05 was used for significance.

Results

Whole exome sequencing produced a mean of 96 million reads per individual (range 31-186 million) and an average of 99.6% mapped/aligned reads to the GRCh37 human reference genome with an average duplication rate of 8.4% (Supplementary Figure 1A-C) for the 280 samples that passed QC metric cutoffs. Fastp identified read quality and insert sizes within the normal ranges. Variant calling produced an average of 21,923 SNVs and 576 indels per sample, with an average depth of 55x per sample (Supplementary Figure 1G-I). Peddy was used to infer the sex, heterozygosity, ancestry, and relatedness of subjects, and compared to known metadata about samples (Supplementary Figure 2A-D)[41]. From these quality control metrics we identified and removed four samples (Supplementary Table 4) that had very low heterozygosity and low coverage. Three samples were removed due to high duplication rates. One sample was removed due to excess heterozygosity. We also discovered a previously unidentified deletion of the long arm of the X chromosome (93.7% homozygous X:130678467-X:155171537) in sample IPOF32. In total, we removed 9 samples from analysis for the quality issues described above leaving 282 samples for the analysis. Peddy confirmed the sex of the POI subjects (Supplementary Figure 2A), and the known relatedness of a few individuals (Supplementary Table 5, Supplementary Figure 2D-E). An additional four sib pairs were identified in the cohort and one family with dominant inheritance was included (Supplementary Figure 2E). For these related individuals, only one subject was included in the joint analyses. PCA projection of the samples together with data from the 1000 genomes identified individuals of European descent (n=235), admixed American (n=18), African (n=10), South Asian (n=4), East Asian (n=3), and unknown (n=12) ancestry (Supplementary Table 6, Supplementary Figure 2C).

In 19 subjects, we identified variants in genes previously determined to cause POI, including confirmation of previously identified variants in 12 subjects with primary amenorrhea (6.7%; Table 1)[19,63-66]. Five of these variants were found as heterozygous genotypes in the genes *NR5A1, PTPN22* and *eIF4ENIF1*[9,17,67].

Sixty-four subjects (23%) carried at least one variant in a previously identified POI gene that was determined to cause ovarian dysgenesis or primary amenorrhea with autosomal recessive inheritance (Table 2). Twenty-seven subjects (10%) carried a heterozygous variant in a gene for which there was a previously identified functional model (Table 3)[68,69]. Variants at genomic loci that were not conserved across species were not considered for analysis, although variants impacting conserved amino acids that were found only in mammalian species were included. One subject carried two variants in *FANCM* and one subject carried two variants in *RECQL4,* however, it was not possible to confirm whether these variants were in *cis* or *trans*. Fourteen subjects carried 2, one subject carried 3 and one subject carried 4 candidate POI risk variants in different genes.

We next determined gene clusters for known genes for POI in women and candidates from animal models. DAVID analysis identified 47 gene list clusters with enrichment >2 (Supplementary Table 1). We then examined enrichment of these 47 gene sets in women with POI compared to controls and found 13 significant gene sets. These gene sets encompassed GO term biological processes including transcription/translation, DNA damage and repair, oogenesis, cell proliferation, hormone regulation, growth factors, regulation of gene expression, embryogenesis, cytoplasmic signaling, male gonad development, chromatin binding, cell division and protein phosphorylation (Table 4, Figure 1). Further, there was significant enhancement compared to housekeeping genes and burden-matched gene lists in POI cases compared to controls (Figure 1A-C and Supplementary Table 2). The majority of the causal or candidate genes were found in the enriched gene sets (Tables 1-3). The two genes that were not found in the gene lists are important for meiosis (*MARF1* and *ANKRD31*).

We examined the remaining candidate genes that had not yet been implicated in a woman with POI or in an animal model. We identified several deleterious variants in genes found in the implicated gene sets (Table 5)[70-77]. None of these gene variants was identified in the control groups that we assessed. Additional candidates were identified in these and other

gene sets, although their pathogenicity was not as strong based on conservation, allele frequency or gene constraint (Supplementary Table 7).

Functional Studies

The potential pathogenicity of the variants not previously identified is outlined in the Supplementary Data (Supplementary Information). For genes with variants not previously identified in POI or in animal models of POI or not previously examined in oocytes (Table 5), RTPCR in mouse oocytes that have resumed and/or completed meiosis I was performed to ensure that the candidate gene was expressed. Of the 24 genes tested, four were present but not highly expressed in the oocyte (Table 5 and Supplementary Tables 3 and 8).

D. melanogaster orthologues were identified for 20 of 35 candidate genes queried (Supplemental Table 9). Thirteen candidates could not be obtained based on availability or could not be tested based on lack of orthology or absence of ovarian expression. Two of the candidates had multiple weak orthologues and were not pursued (*POLK* and *ANKRD31*).

Five knockdowns (*USP36, VCP, WDR33, PIWIL3* and *NPM2*) were completely infertile with atrophic ovaries (Table 6 and Supplementary Figures 18 and 19). Two gene knockdowns demonstrated decreased hatchability and fertility, with abnormal (*LLGL1*) or normal ovaries (*BOD1L1*). Two gene knockdowns had variable or mild ovarian defects that were not statistically significant (*CDK7* and *BRIP1*). One gene knockdown was lethal (*RUVBL2*).

Discussion

We performed WES in 291 subjects with POI from three cohorts. Using two methods, a broad and unbiased discovery method and a more robust prioritization algorithm (GEM), we identified the most likely pathogenic variants in these women with POI. Our data suggest that the candidate genes for POI in individual women are highly heterogeneous. However, when the most likely candidate genes were categorized into functionally related groups, the genes aligned into 13 clusters that were enriched in cases compared to controls after correcting for multiple testing, gene size and pathogenic specificity for POI. New candidate genes were found in enhanced gene sets that included genes important for transcription/translation, DNA damage and repair, meiosis and cell division. Functional analysis in *D. melanogaster* supported a role in oocyte or ovary development for seven genes not previously associated with POI. Taken together, the data support a categorical approach to understanding the genetic architecture for POI.

After an initial broad search for damaging variants, we used an AI-based eCDSS tool, GEM, that employs variant impact (VAAST and VVP), patient phenotypes (Phevor), known Mendelian and pathogenic variants (OMIM, ClinVar) and ancestry to identify disease-causing genotypes[15,24,25,42,54]. Using GEM, we supplemented our data with previously analyzed WESs and have replicated genetic findings in 11 out of 12 subjects with primary amenorrhea, demonstrating the utility of the new software. The only gene variant that was not identified was in *MARF1*, which has not yet been associated with POI in OMIM. GEM identified additional homozygous or compound heterozygous mutations (*HFM1, DCAF17*) or heterozygous mutations (*NR5A1*) in previously identified genes[78-80].

GEM also identified heterozygous deleterious variants in known genes, particularly in women with POI and secondary amenorrhea. GEM uncovered variants in 28 genes previously demonstrated to cause POI with recessive inheritance, out of 42 gene variants total (67%). GEM also identified 13 of 20 candidate variants in genes with evidence for ovarian insufficiency

in an animal or other experimental species model (65%). The results are not surprising based on the use of HPO POI terms in the algorithm, which emphasizes phenotype in GEM[54]. Further tool development will encompass gene pathways for discovery.

With the exception of subjects with POI and primary amenorrhea, the majority of candidate variants identified in the current study are heterozygous, arguing for a dominant, semi-dominant, or complex inheritance pattern for POI that occurs later in the reproductive years. Genome-wide association studies (GWASs) of age at natural menopause support the concept that menopause has a complex inheritance pattern[68]. Further, the largest GWAS of early menopause, defined as menopause before the age of 45 years, replicated 4 common variants associated with age at natural menopause and demonstrated that menopause risk alleles have an additive contribution to age at menopause[21]. Recent data also suggest that common variants contribute to menopause occurring as early as 34 years[22]. However, the contribution of common variation explains only a small portion of the genetic risk for menopause under age 40 years[22]. The current study using WES was not able to assess common variants, but did demonstrate overlap between common GWA variants and rare, deleterious variants in the same candidate genes. For example, nonsynonymous variants in *BRCA1* are associated with earlier menopause by approximately 6 months[68]. In the current study, we identified a frameshift mutation, expected to result in early protein termination, possibly causative for POI. Additional genes with deleterious variants (*MSH6, CHD7)* and some with rare missense variants (*RAD54L, HELQ, POLG)* also overlap with candidate genes associated with age at natural menopause (Table 2). The apparent overlap of common variants associated with menopause age and deleterious variants in the same candidate genes is consistent with the hypothesis that mutations in these genes play a causative role in POI.

Further support for the causative role of heterozygous gene variants in POI comes from the reproductive history of the mothers of girls with primary amenorrhea. A heterozygous *MND1* gene mutation in a mother resulted in POI at age 35 years[45]. Similarly, a heterozygous mutation

in *MCM8* caused POI in a mother at age 29 years[18]. In both families, the daughters with homozygous mutations presented with primary amenorrhea. With the exception of a few reports, age at menopause is rarely mentioned or may not yet have occurred for mothers of girls with POI. However, age at menopause is heritable supporting the segregation of ovarian damaging genes with an effect on age at menopause through the mother[3]. It is also not surprising that heterozygous variants that relatively decrease fertility would be removed from the population through decreased progeny[81], and might therefore be inherited from the father since reproductive lifespan is not limited in men. Taken together, these cases also support the hypothesis that heterozygous mutations can result in earlier age at menopause.

Although the number of subjects in the current study is not sufficient to replicate the genes individually, we were able to demonstrate significantly enriched gene clusters controlled for multiple testing. Previous work in autism and congenital heart disease has used a similar category-wide association study approach[27,28]. Our approach was unbiased; first examining the most deleterious variants in women with POI to identify known genes and candidates with previous functional models, and subsequently determining whether additional genes were found in the clustered gene sets. Interestingly, a MAGENTA analysis of age at natural menopause variants identified similar enhanced categories for candidate genes inferred from genome-wide associated variants[68]. In addition, known genes causing male azoospermia were enriched in comparable pathways[82]. Taken together, a category enhanced approach identifies consistent gene sets across reproductive studies. Genes falling into gene sets including oogenesis, spermatogenesis, meiosis, DNA damage and repair, transcription and translation, chromatin binding, regulation of gene expression, growth factors, embryo development, cell division, extracellular to cytoplasmic signaling, protein kinase phophorylation, and vasoactivity and hormone regulation were enriched compared to controls in our unbiased candidate gene search for damaging mutations across the genome (Figure 1)[83]. New candidate genes were identified

within these gene sets demonstrating that the category approach provides a mechanism for new candidate gene discovery.

Our *D. melanogaster* knockdown model affords a mechanism to determine an oocyte and ovarian phenotype at scale for genes in enhanced pathways. The genes and developmental processes involved in oogenesis in *D. melanogaster* overlap with those in the mouse[84]. We chose quantifiable fertility assays including egg laying rates, hatchability and ovarian morphology[62]. The use of RNAi technology also presumes that the gene is not fully deleted and serves as an excellent model for heterozygous gene variants. Using our *D. melanogaster* model, we identified five genes that are critical for ovarian or oocyte development and that fall into the enriched pathways we defined: transcription/translation, meiosis, DNA repair and DNA damage. RNAi knockdown resulted in atrophic ovaries with no eggs or progeny (Table 6 and Supplementary Figures 18 and 19).

USP36 is a deubiquitinase demonstrated to promote RNA polymerase I stability for the ribosomal RNA processing and translation[85]. Previous studies found that the scny *D. melanogaster* homologue also acts as a histone H2B ubiquitin protease[86]. The atrophic ovary in the knockdown shows that the ribosomal RNA translation and/or the chromatin modification function may affect oocyte or ovarian development in addition to its role in embryogenesis[85].

WDR33 plays a role as one of 4 proteins that recognize the polyadenylation signal in the 3'-end processing of mRNA precursors[87]. The gene is highly expressed in testes and we have now demonstrated that it is also highly expressed in mouse oocytes (Supplementary Table 8). RNAi knockdown results in an atrophic ovary. Thus, *WDR33* may also play critical role in ovarian or oocyte development.

PIWIL3 is a P-element induced wimpy testis protein short RNA found in human, nonhuman primate and bovine oocytes. It is specifically expressed in maturing human oocytes during oogenesis[88] and in bovine oocytes from the GV stage onward[89]. It is critical for germline integrity from DNA transposable element activity[90]. The affected subject carries two *PIWIL3*

variants; a frameshift mutation and a stop gain mutation that both remove the PIWI domain from the protein[90]. We also identified a stop gain mutation in *PIWIL2,* a family member that is expressed in fetal human germ cells[89]. Although knockouts of the mouse *PIWIL2* homologue *Mili* were described as fertile, there were no details provided across the reproductive lifespan[91]. These data demonstrate the importance of the *PIWIL* genes in the ovary in addition to the testes.

NPM2 is found in oocytes before germinal vesicle breakdown[92]. The *Npm2* knockout females are infertile, with normal sized pronuclei that lack nucleoli[92,93]. Although previous studies suggest that infertility is caused by failure of zygote development, our data suggest that *NPM2* is critical for oocyte and ovary development.

VCP, or valosin-containing protein, is an ATPase associated with a variety of activities[94]. It is expressed in GV oocytes and preimplantation embryos in the mouse and controls germinal vesicle breakdown. *Vcp* knockout mice demonstrate no homozygotes because they have a defect in early embryonic development. Our model demonstrates atrophic ovaries. A missense variant in a highly conserved threonine in the N terminal was found in two sisters and a mother with POI (Supplementary Figure 2E). The N terminal is the portion that interacts with other proteins. Therefore, VCP should be considered a new candidate for POI.

In contrast to genes described above, there were only subtle phenotypes identified using our model in genes associated with DNA damage and repair pathways. Two genes in the pathway, *LLGL1* and *BOD1L1,* were highly expressed in the oocyte and demonstrated decreased hatchability and decreased fertility. The *LLGL1* cytoskeletal network is involved in maintaining cell polarity and epithelial integrity[95]. Mutations including the gene region on chr 17 cause Smith Magenis syndrome, a disorder of developmental delay, behavioral abnormalities, sleep disturbance and abdominal obesity. An indel upstream of *LLGL1* in Shaanbei White Cashmere goats is associated with change in litter size[96]. *BOD1L1* stabilizes *RAD51* at the site of DNA replication forks[97]. The frameshift variant in our subject would remove all ATM

phosphorylation sites, along with the majority of the protein. Other genes in the DNA damage and repair pathway with no previous functional models or human mutations had no phenotype in our *D. melanogaster* model and more sensitive functional models may be needed. Nevertheless, a number of previously well validated variants involved in the homologous recombination steps in meiosis were discovered in our cohort (Figure 2)[83]. Many of these gene mutations may result in meiotic failure and oocyte loss. Given the large number of gene mutations falling into the DNA damage and repair pathway, intervening to rescue meiosis for development of normal gametes may be a treatment opportunity in POI.

An association between autoimmune oophoritis, with POI as the end-stage, has been demonstrated only with adrenal autoimmunity[98]. We identified a novel *PTPN22* variant, a gene associated with adrenal insufficiency, and in*TARBP1,* a gene associated with autoimmune syndromes[99,100]. A final subject carried a variant in IL1B, which has been associated with ovarian inflammation. Further delineation of the associated autoimmune risk genes and diseases will clarify the relationship between autoimmunity, genetics and POI.

Our study is limited by whole exome sequencing. We were not able to evaluate common variation, some promoter regions and could not evaluate copy number variants. We did not have trios for the majority of subjects and did not recruit family members to clarify segregation or *de novo* mutations. Future studies will also be needed to more carefully analyze the mitochondrial genome.

The current cohort forms one of the largest WES datasets analyzed for POI. We used an unbiased approach and a new AI-based algorithm to identify the most likely pathogenic variants. We also demonstrated new genes important for oogenesis and ovarian development using a model *D. melanogaster* system. Our more global approach contrasts to previous studies that examined individual consanguineous families and/or were restricted to candidate gene lists. Collectively, our results identify not only disease-causing variants, but also gene categories involved in POI. These results should prove useful for precision medicine efforts aimed at early

identification of gene variants increasing a woman's likelihood to experience infertility or a shortened reproductive lifespan. The early identification of women at risk for POI may enable fertility preserving measures. More broadly, better understanding of the genetic architecture of POI might also aid in identifying additional comorbid risks in a subset of the subjects.

Acknowledgements

The work in this publication was supported by the Center for Genomic Medicine and its Functional Analysis Service. The work was also supported by R56HD090159 and R01HD099487 (CKW) and R35GM124780 (CYC and EC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Table 1. Candidate variants in previously identified genes causing POI.

Gene	Chr: Location	Coding Change	Protein Change	Zygosity	Consequence	gnomAD Allele Frequency (Female/Male)	Conservation[1]	POI ID	Primary-1 or Secondary-2 (Age yrs)	Reference[2]	Z score or Obs/exp[3]	Gene Set[4]
HFM1	1:91816375	c.2123A>T	p.Ile709Asn	Homo[5]	Missense	Novel	M(not C)XZL	NIH897		[78]	0.74	16,40
PTPN22	1:114380534	c.1488A>C	p.Glu496Asp	Het	Missense	0.000072 / NA[6]	M	11896X16	2 (34)	[67]	0.87	7,9,11,32,40
FSHR	2:49190159	c.1801G>C	p.Leu601Val	Compound Het	Missense	Novel	MCXZL	FPOF28	1 (19)	[65 2]	-0.8	1,5,7,11,25,32,40
	2:49190405	c.1555G>T	p.Pro519Thr	Homo	Missense	Novel	MCXZL	FPOF21	1 (17)	[64 2]	-0.8	
	2:49190686	c.1274C>T	p.Thr425Ile	Compound Het	Missense	Novel	MCXZ(not L)	FPOF12	1 (18)	[63 2]	-0.8	
	2:49190751	c.1334>T	p.Asn403Tyr	Compound Het	Missense	Novel	MCX(not Z)L	FPOF12	1 (18)	[63 2]	-0.8	
	2:49196020	c.671T>A	p.Asp224Val	Compound Het	Missense	NA / 0.00005721	MCXZL	FPOF28	1 (19)	[65 2]	-0.8	
DCAF17	2:172306465	c.535C>T	p.Gln179Ter	Compound Het	Stop Gain	0.0000719 / NA	NA	PPOF142	1	[66 2]	0.53 (0.35-0.82)	7,16,25
	2:172325465	c.906G>A	p.Trp302Ter	Compound Het	Stop Gain	NA / 0.00000147	NA	PPOF142	1	[66 2]	0.53 (0.35-0.82	
GDF9	5:132197489	c.1161C>	p.Cys386Phe	Homo	Missense	Novel	MCXZ	FPOF1	1 (17)	[63 2]	-0.5	7,11,25,32,40

		A						0			7	
HARS2	5:140076804	c.1010A>G	p.Tyr337Cys	Homo	Missense	0.00000866 / NA	MCXZL	FpoF32	1 (27)	63 2	0.04	7,16,40
MCM9	6:119147976	c.1785G>C	p.Thr595Ser	Compound Het	Missense	Novel	M(not elephant or Z)CX	FPOF38	1 (20)	63,1012	1.45	7,16,40
	6:119149171	c.1649A>C	p.Gln551Ter	Homo	Stop-gain	0.000017 / 0.0000148	NA	FPOF24	1 (20)	63,1012	0.64 (0.47-0.87)	
	6:119234586	c.905-1G>T	NA	Compound Het	Splice Acceptor	0.00000088 / 0.00000075	MCXZL	FPOF38	1 (20)	63,1012	0.64 (0.47-0.87)	
NR5A1	chr9:127253405	c.1095G>A	p.Arg365Trp	Het	Missense	Novel	MCX (not ZL)	FPOF8	1 (14)	63,1022	2.34	5,7,13,25
	chr9:127253414	c.1084C>T	p.Gln362Ter	Het	Stop Gain	Novel	NA	IPOF17			0.006 (0.02-0.26)	
	9:127262988	c.253C>A	p.Arg84Leu	Het	Missense	Novel	MCXZL	NIH930		79	2.34	
PMM2	16:8898701	c.255+1G>A	NA	Compound Het	Splice Donor	0.00000866 / 0.0000147	MXL	FPOF16	1 (15)	63 2		7,40
	16:8900240	c.323C>T	p.Ala108Val	Compound Het	Missense	0.00000866 / 0.0000147	MCX (not ZL)	NIH811		103	-1.39	
	16:8900255	c.338C>T	p.Pro113Leu	Compound Het	Missense	0.00000173 / 0.0000244	MCXZL	NIH811		103	-1.39	
	16:8905531	c.484C>T	p.Arg162Trp	Compound Het	Missense	0.00000866 / 0.0000221	MCL (not XZ)	FPOF16	1 (15)	63 2		

MARF1	16:15719495	c.1688G>A	p.Arg563Cys	Homo	Missense	Novel	MCX	FPOF11	1 (15)	[63 2]	2.73	NA Meiosis
PSMC3IP	17:40725368-40725370	c.490CCT>C	p.Arg166AlafsTer5	Compound Het	Frameshift	0.0000087 / 0.00000074	NA	FPOF41	1 (28)	[63 2]	0.62 (0.37-1.08)	7,16
	17:40725549-40725549	c.431A>ATC	p.Leu144AspfsTer2	Compound Het	Frameshift	0.0000087 / 0.00000074	NA	FPOF41	1 (28)	[63 2]	0.62 (0.37-1.08)	
EIF4ENIF1	22:31836842	c.2564G>A	p.Pro855Leu	Homo	Missense	0.00047 / 0.0004	MCX (not L)	NIH790		[104]	1.99	7,16,25,34
[7]	chr22:31859102	c.603T>G	p.Ser201Arg	Het	Missense	0.000070 / 0.000098	MCXZ	13429X45	2 (40)		1.99	

[1]Conservation – M-mammals including rhesus, mouse, dog, elephant, C-chicken, X-Xenopus tropicalis, Z-zebrafish, L-lamprey. Exceptions are noted in parentheses.
[2]Previously published
[3]Z score for missense variants and observed/expected (90% confidence intervals) for loss of function variantsgnomAD.[59] A Z score >3 indicates a gene constrained for missense variants and an upper bound 90% confidence interval <1 indicates a gene constrained for loss of function variants.
[4] DAVID Gene Sets: 1-oogenesis/spermatogenesis, 5-male gonad development, 7-transcription/translation/DNA binding, 9-regulation of gene expression, 11-growth factor/cytokine/TGFβ, 13-chromatin binding, 16-meiosis/DNA repair/homologous recombination, 25-extracellular to cytoplasmic signaling, 32-protein kinase/phophorylation, 34-cell division/meiosis, 40-cell proliferation/DNA damage. Also see Table 4. Other known biologic functions have been added for genes that were not found in DAVID.
[5]Homo = homozygous, Het = heterozygous
[6]NA = not applicable, i.e. not identified
[7]Confirmed by Sanger Sequencing

Table 2. Candidate variants found as heterozygotes in previously identified genes causing POI or associated with age at natural menopause.

Gene	Chr Location	Coding Change	Protein Change	Zygosity	Consequence	gnomAD Allele Frequency (Female/Male)	Conservation[1]	POI ID	Primar-1 or Secondary-2 (Age yrs)	Reference[2]	Z score or Obs/exp[3]	Gene Set[4]
HFM1	1:91859764	c.382T>C	p.Lys127Arg	Het[5]	Missense	Novel	M	13429X55	2 (26)		0.74	16,40
MSH6[7]	2:48033757	c.3969_4001+52dupTGAGA...	NA	Het	Splice Site	Novel	NA[6]	11896X6	2 (40)	[105]	0.34 (0.23-0.5)	7,11,13,16,25,34,40
FSHR	2:49190111	c.660delG	p.Pro616ProfsTer21	Het	Frameshift	Novel	NA	FPOF30	1 (16)	[63]	0.83 (0.58-1.22	1,5,7,11,25,32,40
FIGLA	2:71014888	c.277G>T	p.Pro93Thr	Het	Missense	0.00031 / 0.00037	MX	NIH607		[106]	0.08	1,7,11
	2:71017770	c.1T>C	NA	Het	Start Lost	Novel	M	11896X9	2 (39)	[106]	0.63 (0.33-1.32)	
IL1B	2:113593794	c.13G>T	p.Pro5Thr	Het	Missense	0.00004630 / 0.00009130	M	13429X53	2 (38)	[107]	1.15	7,9,11,25,32,40
LARS2	3:45500250	c.622G>A	p.Asp208Asn	Het	Missense	Novel	MCXZ	13429X70	2 (23)	[108]	1.33	7,16,32,40
	3:	c.2675C>T	p.Pro89	Het	Miss	Novel	MCX	IP		[108]	1.3	

	45588985		2Leu		ense		Z	OF1			3	
MRPS22	3:139069121	c.605G>A	p.Arg202His	Het	Missense	0.000 0232 / 0.000 0130	MCXZ	FPOF34	1 (17)	[109]	0.13	7,25,40
	3:139071635	c.878+2dupT	NA	Het	Splice Donor	0.000 00869 / NA	MCXZL	NIH356		[109]	0.54 (0.34-0.89)	
HELQ	4:84348761	c.2631A>T	p.Asp877Glu	Het	Missense	0.000 177 / 0.000 094	MCXZL	NIH923		[68,69,110]	0.6	7,16,40
	4:84358150	c.1909C>T	p.Pro637Ser	Het	Missense	NA / 0.000 00738	MCXZL	IPOF11		[110]	0.6	
BMPR1B	4:96036840	c.341G>T	p.Thr114Ile	Het	Missense	0.000 0697 / 0.000 144	MCXZ	13429X17	2 (39)	[20,111]	0.27	7,11,16,32,40
	chr4:96073907	c.1456C>T	p.Arg486Trp	Het	Missense	0.000 149 / 0.000 127	MCXZL	NIH933		[20,111]	0.27	
	4:96075786	c.1561C>G	p.Leu521Val	Het	Missense	0.000 00867 / 0.000 00738	MCXZ	NIH762		[20,111]	0.27	
MND1	4:154315483	c.349delA	p.Thr117Argfs Ter7	Het	Frameshift	Novel	NA	NIH899		[45]	0.89 (0.59-1.4)	7,16
HSD17B4	5:118837758	c.1308dupA	p.Glu437Argfs Ter4	Het	Missense	Novel	NA	13429X43	2 (40)	[16]	0.53 (0.38 - 0.74)	7,11,25,32,40
GDF9	5:132197397	c.1251G>A	p.Pro417Ser	Het	Missense	Novel	MCXZ	13429X50	2 (21)	[112]	0.7 (0.43-1.19)	7,11,25,32,40
MC	6:	c.1477T>A	p.Lys49	Het	Stop	NA /	NA	PP		[10]	0.6	7,16,4

M9	11915026 2		3Ter			0.000012 42		OF105			4 (0.47-0.87)	0
	6:119232914	c.1051C>A	p.Val351Phe	Het	Missense	0.000039 / 0.000026	MCXZL	NIH933		[10]	1.45	
	6:119234571	c.919A>C	p.Leu307Val	Het	Missense	0.0000174 / 0.0000148	MCXZ	11896X4	2 (39)	[10]	1.45	
	6:119243238	c.635G>C	p.Ser212Cys	Het	Missense	0.000062 / 0.000066	MCXZ	NIH926		[10]	1.45	
STAG3	7:99795404	c.1069C>T	p.Arg357Ter	Het	Stop Gain	0.00000087 / NA	NA	IPOF36		[8]	0.45 (0.34-0.61)	7,11,16,25,34,40
NOBOX	7:144098166	c.815TC>T	p.Arg272fsTer14	Het	Frameshift	Novel	NA	PPOF14		[113]	0.47 (0.3-0.76)	1,7,11
SPIDR	8:48308973	c.563G>A	p.Ser188Asn	Het	Missense	0.000368 / 0.000367	M	13429X13	2 (37)	[114]	-0.22	7,16
	8:48511607	c.1393C>T	p.Leu465Phe	Het	Missense	0.0007180 / NA	M	13429X29	2 (35)	[114]	-0.22	
	8:48625275	c.2029G>A	p.Asp677Asn	Het	Missense	0.000258 / 0.000262	MC	13429X40	2 (39)	[114]	-0.22	
CHD7	8:61773535	c.7681G>T	p.Gly2561Ter	Het	Stop Gain	novel	NA	PPOF105		[68]	0.04 (0.02-0.08)	7,9,11,13,16,18,40
NBN	8:90983470	c.633T>C	p.Asp211Glu	Het	Missense	0.00000232 / 0.000	MCX	13429X7	2 (23)	[115]	0.61	7,11,16,34,40

						0392		0				
	8:90983520	c.585-2A>G	NA	Het	Splice	NA / 0.0000007402	MCX	NIH848		115	0.74 (0.56-1.01)	
RECQL4	8:145737353	c.3334C>T	p.Glu1112Lys	Het	Missense	0.00000179 / 0.0000298	M	IPOF28		116	-4.22	7,16,25,40
	8:145738522	c.2464-1G>C	NA	Het	Splice Site	0.000120 / 0.0000922	MXZL	13429X54	2 (40)	117	0.96 (0.77-1.21)	
	8:145738678	c.2386C>T	p.Glu796Lys	Het	Missense	0.00000278 / 0.0000261	MCXZL	IPOF28		116	-4.22	
	8:145741182	c.1223C>CA	p.Gln408fs12	Het	Frameshift	Novel	NA	NIH584			0.96 (0.77-1.21)	
SETX	9:135206809	c.865C>T	p.Ala289Thr	Het	Missense	0.00000433 / 0.0000736	MCXZ	PPOF21		118	-0.11	1,7,13,16,32,34,40
SYCE1	10:135369340	c.658_662delGAGGG	p.Glu220ProfsTer5	Het	Frameshift	Novel	NA	NIH735		7	0.63 (0.42-0.96)	1,7,34
	10:135370567	c.464+4	NA	Het	splice	Novel	M	FPOF30	1(16)	7	0.63 (0.42-0.96)	
FANCF	11:22647025	c.332T>C	p.Leu111Pro	Het	Missense	0.00000261 / 0.0000294	MC	NIH830		119	-1.83	1,7,16
	11:2264	c.67G>T	p.Val23Phe	Het	Missense	0.0000390 /	MC	NIH4			-1.8	

	7290					0.0000196		25			3	
ZP1	11:60640970	c.1364dupA	p.His455GlnfsTer93	Het	Frameshift	0.00002	NA	NIH822		[120]	0.99 (0.7501.33)	11,25
ATM	11:108099934	c.216_217CAG:C	p.Glu73MetfsTer26	Het	Frameshift	NA / 0.0000131	NA	13429X20	2 (35)	[121]	0.06 (0.51-0.71)	7,11,16,32,34,40
	11:108122683	c.1727T>C	p.Ile576Thr	Het	Missense	Novel	MCX	NIH796		[121]	1.1	
	11:108181006	c.5882A>G	p.Tyr1961Cys	Het	Missense	Novel	MCXZ	13429X5	2 (40)	[121]	1.1	
	11:1082011108	c.7475T>G	p.Leu2492Arg	Het	Missense	Novel	MCXZL	NIH356		[121]	1.1	
BRCA2	13:32910420	c.1929delG	p.Arg645GlufsTer15	Het	Frameshift	Novel	NA	FPOF17	1 (16)	[122]	0.51 (0.42-0.64)	1,7,11,16,25,34,40
[7]	13:32913836	c.5344C>T	p.Gln1782Ter	Het	Stop Gain	Novel	NA	11896X30	2 (38)	[122]	0.51 (0.42-0.64)	
	13:32945190	c.8585T>C	p.Leu2862Pro	Het	Missense	Novel	MCXZL	IPOF14		[122]	-1.29	
REC8	14:24642106	c.126-2A>G	-2 splice	Het	Splice	Novel	M	NIH727		[123,124]	0.42 (0.28-0.65)	1,7,16,34,40
	14:24642546	c.388G>A	p.Glu130Lys	Het	Missense	0.0000350 / 0.0000369	M	11896X3	2 (40)	[123,124]	1.47	
	14:24647	c.884C>G	p.Pro294Arg	Het	Missense	0.000723 /	M	13429	2 (40)	[123,124]	1.47	

	316					0.000617		X43				
	14:24647872	c.1057A>C	p.Thr352Pro	Homo	Missense	novel	M (except elephant)	IPOF2		[124]	1.47	
FANCM	14:45605596	c.362C>T	p.Ala121Val	Het	Missense	Novel	MCXZL	IPOF6		[125]	-0.13	7,16,40
FANCM	14:45633715	c.1735C>T	p.Arg579Cys	Het	Missense	0.00000771 / 0.0000196	MCXL	IPOF6		[125]	-0.13	
FANCM	14:45644539	c.2583_2586delTAAAA:T	p.Lys863IlefsTer12	Het	Frameshift	0.0000407 / 0.000000849	NA	NIH875		[125]	0.46 (0.35-0.59)	
SIX6OS1	14:60950431	c.207_210delTGAG	p.Ser69ArgfsTer53	Het	Frameshift	0.0000181 / 0.0000231	NA	11896X5	2 (19)	[126,127]	0.22 0.45 (0.3-0.69)	40
EIF2B2[7]	chr14:75471447	c.443_446dupTGGA	p.Glu149AspfsTer16	Het	Frameshift	Novel	NA	13429X10	2 (39)	[128]	0.99 (0.66-1.52)	7,16,34,40
PIF1	15:65113400	c.1051dup	p.Thr351AsnfsTer52	Het	Frameshift	0.0000542 / 0.0000457	NA	IPOF29		[129]	0.95 (0.69-1.32)	7,16,40
POLG	15:89865023	c.2542C>T	p.Gly848Ser	Het	Missense	0.000178 / 0.000163	MCXZL	IPOF28		[130]	-0.74	7,16,40
BLM	15:91312735	c.2474C>T	p.Pro825Leu	Het	Missense	0.0000173 / 0.00 0029	MCXZL	NIH735		[131]	0.89	7,16,34,40

						43						
	15:91337408	c.3031G>A	p.Gly1011Arg	Het	Missense	0.00000744 / 0.00000175	MCXZ	PPOF101		[131]	0.89	
	15:91347405	c.3568dupA	p.Met1190AsnfsTer27	Het	Frameshift	Novel	NA	IPOF3		[131]	0.56 (0.43-0.75)	
NR2F2	15:96877380	c.518A>G	p.Asn173Ser	Het	Missense	Novel	MCX	11896X17	2 ()	[132]	3.6	7,32,40
MARF1	16:15719528	c.1654T>C	p.Ile552Val	Het	Missense	0.000002343 / 0.00001047	MCX	11896X19	2 (19)	[63]	2.73	None, Meiosis
	16:15729800	c.544C>T	p.Glu182Lys	Het	Missense	NA / 0.00002955	MCX	NIH830		[63]	2.73	
APC2	19:1468911	c.5616dupC	p.Ser1873LeufsTer280	Het	Frameshift	Novel	NA	IPOF29				7,11,40
PSMC3IP	17:40725043	c.591-1G>A	NA	Het	Splice Acceptor	0.000054 / 0.000039	MCXZL	13429X39	2 (28)	[133]	0.62 (0.37-1.08)	7,16
	17:40729698	c.7T>C	p.Lys3Glu	Het	Missense	0.00000780 / 0.00002614	M	NIH560		[133]	-0.43	
BRCA1	17:41209079	c.5266dupT/TG	p.Gln1777ProfsTer74	Het	Frameshift	0.0002471 / 0.00012339	NA	PPOF39		[68]	Recalc	7,9,11,16,25,34,40
AMH	chr19:2249397	c.67delG	p.Glu23ArgfsTer54	Het	Frameshift	Novel	NA	NIH584		[134]	0.06	1,9,11,25,32,40,41

	19:22494449	c.118C>T	p.Arg40Ter	Het	Stop Gain	0.0000166 / 0.0000208	NA	NIH881		[134]	1.21	
ZFR2	19:38061125	c.2644-2A>G	NA	Het	Splice Acceptor	0.000306 / 0.000141	MZ	11896X23		[112]	0.97 (0.75-1.27)	11,16
	19:3831384	c.755_768delCACCGCTTCC...	p.Pro252GlnfsTer152	Het	Frameshift	novel	NA	IPOF38		[112]	0.97 (0.75-1.27)	
NLRP5	19:56539687	c.2090_2091delAA	p.Lys697ArgfsTer18	Het	Frameshift	0.000195 / 0.000210	NA	PPOF105		[135]	0.58 (0.43-0.8)	7,9,11,16,18,32,40
MCM3AP[8]	21:47666541	c.4544C>T	NA	Het	Splice Donor	Novel	MCXZ	NIH7 77		[136]	0.36 (0.27-0.49)	7,16,25,34,40
	21:47676826	c.3810_3812delA:ACG	p.Val1270AlafsTer23	Het	Frameshift	NA / 0.0000572	NA	IPOF19		[136]	0.36 (0.27-0.49)	
AFF2	X:147743481	c.233A>G	p.Tyr78Cys	Heterozygous	Missense	0.0000309 / NA	MCXZL	NIH622		[137]	1.41	7,16,40

[1] Conservation – M-mammals including rhesus, mouse, dog, elephant, C-chicken, X-Xenopus tropicalis, Z-zebrafish, L-lamprey

[2] Previous evidence from human POI cases or GWAS of age at natural menopause

[3] Z score for missense variants and observed/expected (90% confidence intervals) for loss of function variantsgnomAD.[59] A Z score >3 indicates a gene constrained for missense variants and an upper bound 90% confidence interval <1 indicates a gene constrained for loss of function variants.

[4] DAVID Gene Sets: 1-oogenesis/spermatogenesis, 5-male gonad development, 7-transcription/translation/DNA binding, 9-regulation of gene expression, 11-growth factor/cytokine/TGFβ, 13-chromatin binding, 16-meiosis/DNA repair/homologous recombination, 18-embryo development, 25-extracellular to cytoplasmic signaling, 32-protein kinase/phophorylation, 34-cell division/meiosis, 40-cell proliferation/DNA damage, 41-vasoactivity/hormone regulation. Also see Table 4. Other known biologic functions have been added for genes that were not found in DAVID.

[5]Homo = homozygous, Het = heterozygous
[6]NA = not applicable, i.e. not identified
[7]Confirmed by Sanger Sequencing
[8]Expressed in superovulated mouse oocytes (Supplementary Table 9)

Table 3. Candidate variants in genes from pathways with functional models

Gene	Chr Location	Coding Change	Protein Change	Zygosity	Consequence	gnomAD Allele Frequency	Conservation[1]	POI ID	Primary-1 or Secondary-2 (Age yrs)	Reference[2]	Z Score or Obs/exp[3]	Gene Set[4]
FANCD2	3:1008476 3	c.919dupC	p.Leu307ProfsTer20	Het[5]	Frameshift	0.0000086 6 / NA	NA[6]	IPOF 33		[138]	0.71 (0.58-0.89)	7,16,34,40
KIT	4:55597550	c.2198C>T	p.Pro733Leu	Het	Missense	0.0000867 / 0.0000738	MCX	FPOF 23	1 (19)	[139]	2.58	1,7,9,11,16,25,32,40
CASP3	4:185556495	c.84_130delCTCTGGAATAT...	p.Ser29GlyfsTer7	Het	Frameshift	0.0000174 / 0.0000295	NA	13429X45	2 (40)	[140]	0.3 (0.15-0.69)	7,11,16,25,32,40
SKP2	5:36153055	c.197dupC	p.Arg67ThrfsTer10	Het	Frameshift	Novel	NA	NIH710		[141]	0.13 (0.06-0.33)	7,11,16,32,34,40
	5:36170475	c.701G>A	p.Arg234Gln	Het	Missense	0.0000086 8 / 0.0	MZ	IPOF 9		[141]	1.95	

						0000738						
NIPBL	5:36984879:A:T	c.1597>T	p.Asn533Tyr	Het	Missense	Novel	NA	NIH897		[142]	5.57	7,11,16,34,40
	5:37000554	c.3384T>G	p.His1128Gln	Het	Missense	Novel	NA	13429X66 and 13429X67	2 (29, 30)	[142]	5.57	
ANKRD31	5:74400949	c.4434dup	p.Gln1479ThrfsTer12	Het	Frameshift	0.000176 / 0.000275	NA	IPOF47		[143,144]	0.67 (0.53-0.87)	None Meiosis – Double strand break accessory protein
RAD50	5:131924526	c.1200_1201delT>TGA		Het	Frameshift	Novel	NA	NIH830		[19]	0.7 (0.56-0.87)	7,13,16,25,34,40
	5:131972871	c.3454C>T	p.Arg1152Ter	Het	Stop Gain	NA / 0.000147	NA	IPOF39		[19]	0.7 (0.56-0.87)	
GJA1	6:121768529	c.536T>A	p.Phe179Tyr	Het	Missense	Novel	MCXZL	13429X50	2 (21)	[145]	1.28	7,9,11,16,18,25,32,40,41
MIOS	7:7613260	c.1154C>A	p.Ser385Tyr	Het	Missense	Novel	MCXZ	11896X11	2 (25)	[146]	0.49 (0.35-0.7)	7,34
FZD4	11:8666303 2	c.766A>G	p.Ile256Val	Het	Missense	0.000379 /	MCXZL	11896X	2 (37)	[147]	0.73	7,11,25,32,40

						0.000646		13				
MRE11	11:94180454	c.1714G>A	p.Arg572Ter	Het	Stop gain	0.0000463 / 0.0000718	NA	NIH225		[19,148]	0.66 (0.44-0.84)	7,16,34,40
	11:94204867	c.718C>G	p.Leu240Val	Het	Missense	0.00000869 / NA	MCXZL	PPOF105		[19,148]	0.13	
MLH3	14:75485611	c.4162delC	p.Gln1388SerfsTer16	Het	Frameshift	Novel	NA	NIH768		[149]	0.41 (0.29-0.59)	7,16,40
	14:75515643	c.716T>G	p.Ile239Ser	Het	Missense	0.0000436 / 0.0000591	MCXZL	NIH849		[149]	0.74	
TDRD9	14:104436945	c.833C>G	p.Ser278Ter	Het	Stop Gain	0.0000228 / 0.0000199	NA	NIH225		[150,151]	0.49 (0.38-0.65)	1,7,16,34
NUPR1	16:28549371	c.272C>G	p.Ser91Ter	Het	Stop Gain	Novel	NA	IPOF21		[152,153]	1.29 (0.52-1.91)	5,7,11,40
BCL2	18:60985767	c.132dupC	p.Ala45ArgfsTer108	heterozygous	FRAMESHIFT_VARIANT	Novel	NA	NIH727		[154]	0.15 (0.05-0.71)	5,7,11,16,25,32,34,40
ERCC	19:458557	c.2041C>T	p.Asp681Asn	Het	Missense	0.000	MXZL	1342	2 (40)	[19,155]	0.38	7,9,11,16,18,34,

2	69					260 / 0.0 000 147		9X 18				40
PC NA	20:51 0032 0	c.125G>A	p.Ser4 2Leu	Het	Missense	NA / 0.0 000 074 9	MCX ZL	13 42 9X 37	2 (20)	156	2.4 7	7,11,13, 16,34,4 0
AD AM TS 1	21: 2821 6862	c.412G>T	p.Ala1 38Ser	Het	Missense	0.0 001 01 / 0.0 001 09	MCX ZL	13 42 9X 20	2	157	0.6 7	7,11,16, 25
US P9 X	X: 4099 6058	c.437G>A	p.Arg1 46Lys	Het	Missense	0.0 000 173 / NA	MCX Z	NI H6 34		158	z= 6.4 1	7,11,16, 25,34,4 0
	X:41 0274 59	c.2639T> A	Leu87 5His	Het	Missense	Nov el	MCX ZL	NI H9 32		158	6.4 1	
	X:41 0298 41	c.2996A> G	Asn99 9Ser	Het	Missense	Nov el	MCX ZL	13 42 9X 39	2 (28)	158	6.4 1	
	X: 4104 7363	c.3803A> G	p.Tyr1 268Cy s	Het	Missense	0.0 000 087 5 / NA	MCX Z	NI H8 02		158	6.4 1	
AT RX	X: 7684 9311	c.5965T> C	Thr198 9Ala	Het	Missense	Nov el	MCX	FP OF 35	1 (29)	159,1 60	3.1	1,7,11,1 6,34,40

[1] Conservation – M-mammals including rhesus, mouse, dog, elephant, C-chicken, X-Xenopus tropicalis, Z-zebrafish, L-lamprey

[2] Previous evidence from human or animal models

[3]Z score for missense variants and observed/expected (90% confidence intervals) for loss of function variantsgnomAD.[59] A Z score >3 indicates a gene constrained for missense variants and an upper bound 90% confidence interval <1 indicates a gene constrained for loss of function variants.

[4] DAVID Gene Sets: 1-oogenesis/spermatogenesis, 5-male gonad development, 7-transcription/translation/DNA binding, 9-regulation of gene expression, 11-growth factor/cytokine/TGFβ, 13-chromatin binding, 16-meiosis/DNA repair/homologous recombination, 18- embryo development, 25-extracellular to cytoplasmic signaling, 32-protein kinase/phophorylation, 34-cell division/meiosis, 40-cell proliferation/DNA damage, 41-

vasoactivity/hormone regulation. Also see Table 4. Other known biologic functions have been added for genes that were not found in DAVID.

[5]Homo = homozygous, Het = heterozygous

[6]NA = not applicable, i.e. not identified

[7]Expressed in superovulated mouse oocytes (Supplementary Table 9)

[8]Confirmed by Sanger Sequencing

Table 4. Enhanced biological pathways or clusters in women with POI compared to controls.

Cluster 3. Biological Process	# Damaged Genes Cases POI	# Damaged Genes Cases Phenotypic Abnormality	# Damaged Genes Controls POI	# Damaged Genes Controls Phenotypic Abnormality	Total Number of Genes in Category	Cases: Controls POI *p* value[a]	Cases: Controls Phenotypic Abnormality *p* value[b]	Cases: Controls POI *p* value[c]$_{FDR}$	Cases: Controls Phenotypic Abnormality *p* value[d]$_{FDR}$	Log_2 Ratio *p* values$_{FDR}$ Phenotypic Abnormality/POI
7. Transcription/ Translation/DNA Binding	120	95	46	43	740	0.00000	0.00001	0.00000	0.00013	8.4
40. Cell Proliferation/DNA Damage	178	138	92	77	830	0.00000	0.00002	0.00000	0.00030	6.1
16. Meiosis/DNA Repair/Homologous Recombination	70	56	31	32	351	0.00007	0.0069	0.0089	0.029	5.0
41. Vasoactivity/Hormone Regulation	22	17	2	3	75	0.00002	0.0013	0.00030	0.0082	4.8
11. Growth Factor/Cytokine/TGFβ	87	66	36	29	474	0.00000	0.00009	0.00008	0.0011	3.8
9. Regulation of Gene Expression	49	35	24	23	123	0.0023	0.074	0.011	0.14	3.6
18. Embryo Development	23	16	4	3	61	0.00016	0.0022	0.0014	0.011	3.0
25. Extracellular to Cytoplasmic Signaling	44	33	15	12	284	0.00010	0.0012	0.0011	0.0082	2.9
1. Oogenesis/Spermatogenesis	13	8	3	3	91	0.011	0.11	0.035	0.19	2.4
34. Cell Cycle/Meiosis/ Nuclear Membrane	55	35	28	19	213	0.0020	0.020	0.011	0.057	2.4

5. Male Gonad Development	11	8	2	3	31	0.011	0.11	0.036	0.19	2.4
13. Chromatin Binding	11	13	1	4	58	0.0032	0.025	0.015	0.067	2.2
32. Protein Kinase/Phosphorylation	57	33	28	16	244	0.0011	0.011	0.0081	0.035	2.1

[a] Fisher exact test *p* value for the number of damaged genes in the biological category found using GEM with the term POI in cases compared to controls. [b] Fisher exact test *p* value for the number of damaged genes in the biological category found using GEM with the term phenotypic abnormality in cases compared to controls. [c] Fisher exact test *p* value corrected for the false discovery rate for the number of damaged genes in the biological category found using GEM with the term POI in cases compared to controls. [d] Fisher exact test *p* value corrected for the false discovery rate for the number of damaged genes in the biological category found using GEM with the term phenotypic abnormality in cases compared to controls.

Table 5. Variants in candidate genes and candidate pathways with no previous model for primary ovarian insufficiency.

Gene	Chr Location	Coding Change	Protein Change	Zygosity	Consequence	gnomAD Allele Frequency	Conservation[1]	POI Age	Primary-1 or Secondary-2 (Age yrs)	Ref	Z score or Obs/exp[2]	Gene Set[3]
MUTYH[6]	1:45797230	c.1103-2A>G	NA[4]	Het[5]	Splice acceptor	0.00003112 / 0.00001916	MCXZ	NIH75			0.88 (0.66-1.19)	16
	1:45799108	c.241C>T	p.Arg106Trp	Het	Missense	0.00003886 / 0.00003911	MCXZ	NIH701			-0.21	
RPL5[6]	1:93303032	c.547T>G	p.Tyr183Asp	Het	Missense	Novel	MCXZL	IPOF29			1.9	7,16,25,34,40
PRMT6	1:107600153	c.819dup	p.Arg274AlafsTer146	Homo	Frameshift	NA / 0.00000747 3	NA	IPOF15		161	0.47 (0.25-0.92)	7
CHD1L[6]	1:146740448	c.998delC	p.Pro333ArgfsTer10	Het	Frameshift	Novel	NA	13429X22	2 (37)	162	0.04 (0.83-1.03)	16
	1:146757078	c.1937_1941delAAAGA	p.Lys646ThrfsTer22	Het	Frameshift	0.00003887 / 0.00002611	NA	13429X8	2 (18)	162	0.04 (0.83-1.03)	
TDRKH	1:151747243	c.1574_1575delCT	p.Thr525ArgfsTer27	Het	Frameshift	0.00010 9 / 0.00017 7	NA	13429X38	2	89, 90	0.29 (0.17-0.52)	1,25,34,40

CENPF[6]	1:214830419	c.8629G>C	p.Ala2877Pro	Homo	Missense	Novel	MC(not XL)	IPOF24			-0.03	7,34,4
TARBP1[6]	1:234529540	c.4286_4287insT	p.Val1430ArgfsTer33	Het	Frameshift	Novel	NA	NIH914			0.66 (0.52-0.84)	7
APLF[6]	2:68694942	c.79C>T	p.Arg27Cys	Het	Missense	Novel	MC	13429X52	2 (38)	[163]	0.1	16
	2:68717330	c.106_107delAA	p.Lys36GlufsTer10	Het	Frameshift	0.000008 84 / 0.000007 49	NA	IPOF26			1.1 (0.81-1.53)	
	2:68717350	c.125A>T	p.His42Leu	Het	Missense	NA / 0.000057 3	M	IPOF16			0.1	
	2:68729901	c.210_213dupTCAG	p.Leu72SerfsTer24	Het	Frameshift	0.000085 8 / 0.000078 5	NA	13429X35	2		1.1 (0.81-1.53)	
WDR33[6,7,8]	2:128467392	c.3346_3347insGA	p.Ala1116GlyfsTer166	Het	Frameshift	novel	NA	13429X11	2 (38)	[87]	0.13 (0.08-0.21)	7,16
RNF168[6]	3:196214335	c.493G>A	p.Arg165Ter	Het	Stop Gain	Novel	NA	NIH926		[164]	0.84 (0.6-1.2)	16,34
CPZ	4:86032201	c.493C>T	p.Arg165Trp	Het	Missense	0.000001 52 / 0.000055 1	MCXZ	13429X28 and 13429X71	2 (37 and 39)		-4,43	7,25
BOD1L1[6,8]	4:136615981	c.1009_1012delGAAA	p.Glu337ArgfsTer32	Het	Frameshift	0.000022 / 0.00	NA	NIH633		[97]	0.17 (0.11-0.25	16

						00295					)	
CDK7	5:68531230	c.78_79delTT	p.Tyr27GlnfsTer5	Het	Frame shift	0.000008 66 / NA	NA	11896 X15	2 (31)	[165]	0.64 (0.42-1.02)	7,34,40
BDP1[6]	5:70856038	c.7471delT	p.Ser2491ArgfsTer20	Het	Frame shift	Novel	NA	FPOF19	1			7
POLK[6,7]	5:74892232	c.1714C>T	p.Arg572Ter	Het	Stop Gain	0.000117 / 0.000785	NA	13429 X24	2 (40)	[166]	0.62 (0.46-0.87)	16
DCP2[6,7,8]	5:112343710	c.1018C>T	p.Gln340Ter	Het	Stop Gain	0.000034 7 / NA	NA	13429 X13	2 (36)	[167,168]	0.2 (0.1-0.41)	7
NUP43[6]	6:150067167	c.141_151delGTCTATTGGAG	p.Trp47Ter	Het	Stop Gain	0.000017 3 / 0.000014 7	NA	NIH832		[169,170]	0.51 (0.32-0.84)	7,34
BRAT1[6]	7:2584678	c.294dupA	p.Leu99ThrfsTer92	Het	Frame shift	0.000252 / 0.000234	NA	NIH859 and 13429 X36	2 (X,36)	[171]	0.71 (0.51-0.99)	16,40
CASTOR3 (GATS)[7]	7:99821688	c.218-15_227dupTGTGTCT...	p.Ser77ValfsTer52	Het	Frame shift	0.000017 5 / 0.000007 39	NA	13429 X38	2	[56]	0.55 (0.27-1.25)	None POI
NCAPG2[6]	7:158455043	c.1830_1831delAC	p.Leu611ValfsTer23	Het	Frame shift	Novel	NA	13429 X31	2 (40)	[172]	0.18 (0.11-0.31)	7,34
NPM2[6,7,8]	8:21882984	c.97dupA	p.Arg33LysfsTer90	Het	Frame shift	0.000030 9 /	NA	11896 X1	2 (34)	[92]	0.51 (0.29-	7,11,13,40

						0.000006 52 all african 0.000203		8			0.95)	
PIWIL2[8]	8:22136912	c.13C>T	p.Arg5Ter	Het	Stop gain	0.000023 / 0.000026	NA	NIH790		[88,89]	0.42 (0.3-0.6)	1,16,34,40
ZNF572	8:125989932	c.188dupC	p.Ala64SerfsTer9	Het	Frame shift	NA / 0.000071 7	NA	13429X20	2 (35)		0.36 (0.2-0.72)	7,16
TONSL[6]	8:145662282	c.1747delG	p.Asp583ThrfsTer60	Het	Frame shift	Novel	NA	NIH937		[173,174]	0.51 (0.38-0.69)	7
HAUS6[6]	9:19082982	c.755_758delGAGA	p.Arg252LysfsTer15	Het	Frame shift	Novel	NA	NIH800		[175]	0.72 (0.54-0.96)	34
VCP[6,7,8]	9:35068336	c.41G>A	p>Thr14Ile	Het	Missense	0.000072 /NA	MCXZ	13429X66 and 13429X67	2	[94]	5.41	7,11,16,40
HNRNPK[6]	9:86586818	c.928G>A	p.Pro311Leu	Het	Missense	Novel	MCXZ	13429X61	2	[176]	3.99	7,34,40
CPEB3[6]	10:93999849	c.258delT	p.Pro87LeufsTer33	Het	Frame shift	novel	NA	NIH878		[177]	0.07 (0.03-0.23)	7,16,40
C10orf90[6]	10:128192549	c.1219_1220insAA	p.Ile407LysfsTer28	Het	Frame shift	0.000114 / 0.00		13429X49	2 (38)	NA	0.86 (0.62-1.21	7

						00718					)	
TSG101[6]	11:18503395	c.863_864delTA	p.Ile288ArgfsTer8	Het	Frameshift	0.000008 67 / NA	NA	13429 X38	2		0.23 (0.12-0.48)	7,11,16,25,32,40
NAP1L4[6,7]	11:2999567	c.20C>G	p.Ser7Ter	Het	Stop Gain	0.000478 / 0.000485	NA	13429 X34 and NIH 708	2 (20, X)	[178]	0.19 (0.01-0.39)	7
NANOGNB[7]	12:7923084	c.486_490delCAAGA	p.Arg164AlafsTer2	Het	Frameshift	0.000012 9 / NA	NA	11896 X12	2 (39)	[179]	0.54 (0.2801.12)	7
NIN[6,7]	14:51224657	c.3109T>C	p.Met1031Val	Homo	Missense	0.000124 / 0.000124	Not conserved	11896 X2	1		0.33	7,16
	14:51288759	c.16C>T	p.Gln6Ter	Het	Stop Gain	Novel	NA	11896 X8	2 (32)	[180]	0.32 (0.25-0.43)	
ORC6[6]	16:46724920	c.84dupG	p.Arg29AlafsTer42	Het	Frameshift	Novel	NA	13429 X63	2 (40)	[181]	0.84 (0.53-1.39)	7,16,25,34,40
LLGL1[6]	17:18138789	c.1290G>A	p.Trp430Ter	Het	Stop Gain	0.000008 67 / 0.000007 36	NA	13429 X26	2 (40)	[96]	0.35 (0.24-0.53)	7,25,40
BRIP1[6]	17:59878688	c.1066G>A	p.Arg356Ter	Het	Stop Gain	0.000007 73 / 0.000013 1	NA	11896 X9	2 (39)	[182]	0.59 (0.46-0.79)	7,11,16
USP36[7,8]	17:7679	c.2952_2977delAGAT	p.Asp985LeufsT	Homo	Frameshift	0.000018	NA	13429	2 (39)	[85,18]	0.35 (0.2	7

	8450	GCTG..	er31			6 / NA		X12		[3]	4-0.52)	
EEF2[6]	19:39842 03	c.149C>T	p.Arg50Gln	Het	Missense	Novel	MCXZL	IPOF21			4.88	7,25,40
RUVBL2[6]	19:49514507	c.941G>A	p.Arg314Gln	Het	Missense	0.000017 6 / 0.000051 7	MCX	13429X6	2 (40)	[184]	3.11	7,16
NINL[6]	20:25462625	c.1782_1788delACGGCAG	p.Arg594SerfsTer19	Het	Frameshift	Novel	NA	FPOF19	1 (19)		0.82 (0.79-1.09)	34
SAMHD1[6]	20:35555623	c.658G>A	p.Arg220Ter	Het	Stop Gain	0.000008 65 / 0.000014 7	NA	13429X11	2 (38)		0.67 (0.49-0.94)	7,11,1
	20:35555633	c.647_648delCAT>C	p.Met216fsTer1	Het	Frameshift	Novel	NA	11896X24	2		0.67 (0.49-0.94)	
PIWIL3[7,8]	22:25124021	c.1933-7_1933-2delACTTAA	NA	Het	Splice site	0.000797 / 0.000647	NA	13429X37	2 (20)	[88,185]	0.96 (0.77-1.22)	1,7,16,34,40
	22:25152652	c.376C>T	p.Gln126Ter	Het	Stop Gain	0.000008 69 / 0.000014 8	NA	13429X37	2 (20)	[88,185]	0.96 (0.77-1.22)	
SMARCA1	X:128620995	c.2217G>A	p.Lys739=	Het	Splice site	Novel	NA	11896X10	2 (16)	[186]	0.04 (0.02-0.14)	7,16

[1] Conservation – M-mammals including rhesus, mouse, dog, elephant, C-chicken, X-Xenopus tropicalis, Z-zebrafish, L-lamprey

[2]Z score for missense variants and observed/expected (90% confidence intervals) for loss of function variantsgnomAD.[59] A Z score >3 indicates a gene constrained for missense variants and an upper bound 90% confidence interval <1 indicates a gene constrained for loss of function variants.
[3] DAVID Gene Sets: 1-oogenesis/spermatogenesis, 5-male gonad development, 7-transcription/translation/DNA binding, 9-regulation of gene expression, 11-growth factor/cytokine/TGFβ, 13-chromatin binding, 16-meiosis/DNA repair/homologous recombination, 18-embryo development, 25-extracellular to cytoplasmic signaling, 32-protein kinase/phophorylation, 34-cell division/meiosis, 40-cell proliferation/DNA damage, 41-vasoactivity/hormone regulation. Also see Table 4. Other known biologic functions have been added for genes that were not found in DAVID.
[4]NA = not applicable, i.e. not identified
[5]Homo = homozygous, Het = heterozygous
[6]Expressed in superovulated mouse oocytes (Supplementary Table 9)
[7]Confirmed by Sanger Sequencing
[8]Functional support from *D. melanogaster* model

Table 6. Ovary and fertility phenotypes in *D. melanogaster* RNAi knockdown in ovaries/germline.

Human Gene	Fly Ortholog	Ovary Defect	Egg Defect[1]	Hatchability[2]	Fertility[3]
USP36	*Scny*	atrophic	No eggs	--	infertile
VCP	*TER94*	atrophic	No eggs	--	infertile
WDR33	*wdr33*	atrophic	No eggs	--	infertile
PIWIL3	*piwi*	atrophic	No eggs	--	infertile
NPM2	*Nlp*	atrophic	No eggs	--	infertile
LLGL1	*l(2)gl*	normal	normal	~20%	~1%
BOD1L1	*BOD1*	normal	normal	normal	~50%
DCP2	*DCP2*	normal	normal	normal	normal
TDRKH	*papi*	normal	normal	normal	normal
SMRCA/CHDL1	*Iswi*	normal	normal	normal	normal
TTLL5	*TTLL5*	normal	normal	normal	normal
CPEB3	*orb2*	normal	normal	normal	normal
CDK7	*Cdk7*	Normal[4]	normal	normal	normal
BRIP1	*CG4078*	Normal[4]	normal	normal	normal
NIN	*Bsg25D*	normal	normal	normal	normal
NAP1L4	*Nap1*	normal	normal	normal	normal
BRAT1	*CG7044*	normal	normal	normal	normal

[1]Number of eggs laid in first 8 hrs after mating
[2]Number of progeny hatched from eggs counted, "--" indicates no value because no eggs were laid.
[3]Total progeny count 10 days post mating
[4]*CDK7* and *BRIP1* displayed inconsistent minor ovary defects, but fertility was completely normal

Figure 1. Three examples of enriched pathways in the POI data set as determined by the permutation tests in cases (upper panels) and controls (lower panels). Enriched pathways that encompassed novel POI genes (Table 5) included: A) Transcription/Translation/DNA binding, B) Meiosis/DNA Repair/Homologous Recombination and C) Cell Division/Meiosis, compared to D) Housekeeping Genes. The number of damaged genes from the target gene list in the pathways of interest (red arrow) is compared to the distribution of damaged genes in random gene lists of equal number to the lists of interest (gray bars), burden-matched control genes (pink arrows) and housekeeping genes (green arrows). The burden-matched genes and housekeeping genes are not significantly enriched for any gene set. *p* values are controlled for the false discovery rate.

Figure 2. Candidate genes in women with POI. Variants in a number of genes involved in chromosome pairing and DNA damage and repair are involved in meiosis. The figure depicts candidate genes that are involved in chromosome movement, double strand breaks, end resection, double strand break repair, crossovers and dissociation and resolution of Holliday junctions. Members of the nuclear pore complex (NUP43) play a role in chromosome movement and organization. After DNA replication (*ORC6*), the synaptonemal complex pairs homologous chromosomes (*PSMC3IP*) loaded with condensin and cohesion complex proteins (*STAG3, REC8, NIPBL*) and connects the synaptonemal complex to DNA repair proteins (*SYCE1*). During recombination, double strand breaks form (*ATM*, *ANKRD3, PIF1*), ends are resected (*BRCA1, SAMHD1, BOD1L1*), and crossovers occur (*HFM1*) through strand invasion (*PSMC3IP, MND1, RAD51*). Subsequently, DNA double strand break repair (*CHD1L, POLG, POLK, MSH6, PCNA, NUPR1, APLF, NBN, RAD50, RUVBL2, MRE11*), DNA repair (*CDK7, MLH3, PRMT6, HELQ, TONSL*), strand annealing (*RECQL4*) and repair via homologous recombination (*BRCA2, BRIP1, FANCD2, HELQ, FANCM, FANCF, BLM, MCM9, USP36*) take

place. Kinetochore/chromosome assembly, orientation and segregation (*HAUS6*, *CENPF, NUP43, NCAPG2, LLGL1, NINL, ATRX*) follow recombination.

Supplementary Table 1. Annotation clusters identified using the Database for Annotation, Visualization and Integrated Discovery (DAVID). Data were organized into 47 clusters with enrichment scores of >2.

Supplementary Table 2. Housekeeping genes. Genes that are constantly and uniformly expressed over many developmental and adult time points in 16 tissue types were chosen as housekeeping genes to examine enrichment.

Supplementary Table 3. PCR primers used to analyze gene expression in super ovulated mouse oocytes using RTPCR.

Supplementary Table 4. Samples removed by Peddy for very low heterozygosity and low coverage.

Supplementary Table 5. Related subjects identified by Peddy.

Supplementary Table 6. Ancestry identified by PCA plot and projection onto 1000 genomes data.

Supplementary Table 7. GEM results for all subjects. The GEM results for all genes with a GEM score greater than 0 are presented.

Supplementary Table 8. Oocyte expression. RTPCR was performed in superovulated mouse oocytes for gene targets with no previous functional studies.

Supplementary Table 9. Genes chosen for RNAi knockdown in a *D. melanogaster* model with Bloomington *Drosophila* Stock Center Number (BDSC#).

Supplementary Figure 1. Quality control metrics for 283 POI cases.

Box whisker plots of the alignment statistics and vcf statistics for 283 cram and vcf files that passed QC metrics: A) Total number of reads per sample, B) Percentage of Aligned reads, C) Percentage of duplicate reads, D) Mean coverage per sample, E) Median coverage per sample, F) Percentage of Coverage over 20 bases, G) Number of SNPs per sample extracted from the Bcftools statistics, H) Number of Indels found per sample, and I) Average Depth per sample.

Supplementary Figure 2. Peddy analysis of 283 samples from the final VCF files of women with POI**.** A) The predicted sex was female for all cases. B) The proportion of heterozygous calls ranged from 0.12 to 0.18 at a median depth of 30 to 65. C) PCA projection of the 283 cases onto ancestry of 1000 Genomes data. The majority of subjects were of European ancestry as expected. D) Coefficient of relatedness between two samples plotted by sampling 25K sites in the genome and comparing the relatedness reported in the ped file to the relatedness inferred from the genotypes. Thus, five sib pairs were confirmed, along with grandparent-parent and parent-child relationships. E) Five pedigrees of relationships confirmed by Peddy and investigators.

Supplementary Figure 3-16. Enriched pathways in the POI data set as determined by the permutation tests in cases (upper left panels) and controls (lower left panels) and compared to data from the same pathways for the root phenotypic abnormality (upper right and lower right panels). The number of damaged genes from the pathways of interest (red arrow) is compared to the distribution of damaged genes in random gene lists of equal number to the lists of interest (gray bars), burden-matched control genes (pink arrows), and housekeeping genes (green arrows). The burden-matched genes and 16A) housekeeping genes are not significantly enriched for any gene set.

Supplementary Figure 17. PA-1 cells were transfected using PolyJet transfection reagent (SignaGen Laboratories, Rockville, MD) with WT *eIF4ENIF1* or *eIF4ENIF1* containing the c.603T>G variant created using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) into a pcDNA3.1(-) expression vector (Invitrogen, Carlsbad, CA) using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswich, MA). Stable cell lines were generated by selection of colonies resistant to 750 µg/ml G418 (Life Technologies, Carlsbad, CA). Cells were seeded at 3×10^4 cells/well in an 8-well chamber slide. After 48 hours, cells were fixed with ice-cold 100% methanol for 5 minutes at room temperature (RT), followed by washing with PBST comprising 0.1% Tween-20 in 1× PBS (Fisher Scientific, Waltham, MA). Cells were blocked with 1% BSA and 22.52 mg/ml glycine (Fisher Scientific, Waltham, MA) in PBST for 30 minutes at RT and then incubated with an N-terminal antibody (Novus Biologicals, Centennial, CO) diluted in 1% BSA in PBST at 4℃ overnight. After another wash with PBST, cells were labeled with an anti-rabbit Alexa Fluor 594 (Invitrogen, Carlsbad, CA) for 1 hour at RT, washed with PBST as before, counterstained with DAPI (Southern Biotech, Birmingham, AL), and mounted with glycerol mounting medium with DABCO (Electron Microscopy Sciences, Hatfield, PA). The Nikon fluorescent microscope was used for image acquisition. The c.603T>G variant, p.S201R, is located in the nuclear import signal of *eIF4ENIF1*. The top panels show the N terminal eIF4ENIF1 images with DAPI staining of the nucleus, while the bottom panels show the N terminal eIF4ENIF1 images. Compared to the A) wild type *eIF4ENIF1*, the B) S201R variant transfected cells demonstrated disorganized localization of *eIF4ENIF1* with increased intranuclear protein.

Supplementary Figure 18. *Drosophila melanogaster* phenotypes.
Hatchability and total fertility values are plotted for all genes tested by RNAi in *Drosophila*. P values for genes with significantly different phenotypic values are highlighted in red. N= 8-10 for all measurements. C= control; KD= RNAi knockdown

Supplementary Figure 19. *Drosophila melanogaster* ovarian phenotype.

Representative images of ovaries from RNAi knockdowns that produced atrophic ovaries and a control. All other RNAi knockdowns that produced normal ovaries appear identical to the control and are not shown.

References

1 Welt, C. K. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clinical endocrinology* **68**, 499-509, doi:10.1111/j.1365-2265.2007.03073.x (2008).

2 de Bruin, J. P. *et al.* The role of genetic factors in age at natural menopause. *Human reproduction* **16**, 2014-2018 (2001).

3 Treloar, S. A., Do, K. A. & Martin, N. G. Genetic influences on the age at menopause. *Lancet* **352**, 1084-1085, doi:10.1016/S0140-6736(05)79753-1 (1998).

4 Snieder, H., MacGregor, A. J. & Spector, T. D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *The Journal of clinical endocrinology and metabolism* **83**, 1875-1880, doi:10.1210/jcem.83.6.4890 (1998).

5 Torgerson, D. J., Thomas, R. E. & Reid, D. M. Mothers and daughters menopausal ages: is there a link? *European journal of obstetrics, gynecology, and reproductive biology* **74**, 63-66 (1997).

6 Vegetti, W. *et al.* Inheritance in idiopathic premature ovarian failure: analysis of 71 cases. *Human reproduction* **13**, 1796-1800 (1998).

7 Vries, L. D. *et al.* Exome sequencing reveals SYCE1 mutation associated with autosomal recessive primary ovarian insufficiency. *The Journal of clinical endocrinology and metabolism*, jc20141268, doi:10.1210/jc.2014-1268 (2014).

8 Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *The New England journal of medicine* **370**, 943-949, doi:10.1056/NEJMoa1309635 (2014).

9 Kasippillai, T. *et al.* Mutations in eIF4ENIF1 are associated with primary ovarian insufficiency. *The Journal of clinical endocrinology and metabolism* **98**, E1534-1539, doi:10.1210/jc.2013-1102 (2013).

10 Wood-Trageser, M. A. *et al.* MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *American journal of human genetics* **95**, 754-762, doi:10.1016/j.ajhg.2014.11.002 (2014).

11 Katari, S. *et al.* Novel Inactivating Mutation of the FSH Receptor in Two Siblings of Indian Origin With Premature Ovarian Failure. *The Journal of clinical endocrinology and metabolism* **100**, 2154-2157, doi:10.1210/jc.2015-1401 (2015).

12 AlAsiri, S. *et al.* Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *The Journal of clinical investigation* **125**, 258-262, doi:10.1172/JCI78473 (2015).

13 Qin, Y., Jiao, X., Simpson, J. L. & Chen, Z. J. Genetics of primary ovarian insufficiency: new developments and opportunities. *Human reproduction update* **21**, 787-808, doi:10.1093/humupd/dmv036 (2015).

14 Tucker, E. J., Grover, S. R., Bachelot, A., Touraine, P. & Sinclair, A. H. Premature Ovarian Insufficiency: New perspectives on genetic cause and phenotypic spectrum. *Endocrine reviews*, er20161047, doi:10.1210/er.2016-1047 (2016).

15 Al-Agha, A. E. *et al.* Primary Ovarian Insufficiency and Azoospermia in Carriers of a Homozygous PSMC3IP Stop Gain Mutation. *The Journal of clinical endocrinology and metabolism* **103**, 555-563, doi:10.1210/jc.2017-01966 (2018).

16 Ahmed, S. *et al.* Exome analysis identified a novel missense mutation in the CLPP gene in a consanguineous Saudi family expanding the clinical spectrum of Perrault Syndrome type-3. *J Neurol Sci* **353**, 149-154, doi:10.1016/j.jns.2015.04.038 (2015).

17 Lourenco, D. *et al.* Mutations in NR5A1 associated with ovarian insufficiency. *The New England journal of medicine* **360**, 1200-1210, doi:10.1056/NEJMoa0806228 (2009).

18 Heddar, A., Beckers, D., Fouquet, B., Roland, D. & Misrahi, M. A Novel Phenotype Combining Primary Ovarian Insufficiency Growth Retardation and Pilomatricomas With MCM8 Mutation. *The Journal of clinical endocrinology and metabolism* **105**, doi:10.1210/clinem/dgaa155 (2020).

19 Tucker, E. J. *et al.* TP63-truncating variants cause isolated premature ovarian insufficiency. *Hum Mutat* **40**, 886-892, doi:10.1002/humu.23744 (2019).

20 Patino, L. C. *et al.* New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Human reproduction* **32**, 1512-1520, doi:10.1093/humrep/dex089 (2017).
21 Murray, A. *et al.* Common genetic variants are significant risk factors for early menopause: results from the Breakthrough Generations Study. *Human molecular genetics* **20**, 186-192, doi:10.1093/hmg/ddq417 (2011).
22 *https://doi.org/10.1101/2021.01.11.20248322*.
23 Jaillard, S. *et al.* New insights into the genetic basis of premature ovarian insufficiency: Novel causative variants and candidate genes revealed by genomic sequencing. *Maturitas* **141**, 9-19, doi:10.1016/j.maturitas.2020.06.004 (2020).
24 De La Vega, F. M. *et al.* Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *medRxiv*, 2021.2002.2009.21251456, doi:10.1101/2021.02.09.21251456 (2021).
25 Hu, H. *et al.* VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genetic epidemiology* **37**, 622-634, doi:10.1002/gepi.21743 (2013).
26 Kennedy, B. *et al.* Using VAAST to Identify Disease-Associated Variants in Next-Generation Sequencing Data. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* **81**, 6 14 11-16 14 25, doi:10.1002/0471142905.hg0614s81 (2014).
27 Watkins, W. S. *et al.* De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nat Commun* **10**, 4722, doi:10.1038/s41467-019-12582-y (2019).
28 Werling, D. M. *et al.* An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nature genetics* **50**, 727-736, doi:10.1038/s41588-018-0107-y (2018).
29 Popat, V. B., Vanderhoof, V. H., Calis, K. A., Troendle, J. F. & Nelson, L. M. Normalization of serum luteinizing hormone levels in women with 46,XX spontaneous primary ovarian insufficiency. *Fertility and sterility* **89**, 429-433, doi:10.1016/j.fertnstert.2007.02.032 (2008).
30 Kerber, R. A., O'Brien, E., Boucher, K. M., Smith, K. R. & Cawthon, R. M. A genome-wide study replicates linkage of 3p22-24 to extreme longevity in humans and identifies possible additional loci. *PLoS One* **7**, e34746, doi:10.1371/journal.pone.0034746 (2012).
31 Smith, K. R. *et al.* Familial aggregation of survival and late female reproduction. *J Gerontol A Biol Sci Med Sci* **64**, 740-744, doi:10.1093/gerona/glp055 (2009).
32 Malhotra, A., Cromer, K., Leppert, M. F. & Hasstedt, S. J. The power to detect genetic linkage for quantitative traits in the Utah CEPH pedigrees. *J. Hum. Genet.* **50**, 69-75, doi:10.1007/s10038-004-0222-8 (2005).
33 Moriwaki, M. M., B.; Mosbruger,T.; Neklason,D.W.; Yandell,M.; Jorde,L.B.; Welt,C.K. POLR2C Mutations are Associated with Primary Ovarian Insufficiency in Women. *J Endocr Soc* **1**, 162-173 (2017).
34 Freed, D., Aldana, R. Weber, J.A., Edwards, J.S. The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv* (2017).
35 Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505, doi:10.1093/bioinformatics/btu314 (2014).
36 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
37 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).
38 Pedersen, B. S., Collins, R. L., Talkowski, M. E. & Quinlan, A. R. Indexcov: fast coverage quality control for whole-genome sequencing. *Gigascience* **6**, 1-6, doi:10.1093/gigascience/gix090 (2017).

39 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

40 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

41 Pedersen, B. S. & Quinlan, A. R. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *American journal of human genetics* **100**, 406-413, doi:10.1016/j.ajhg.2017.01.017 (2017).

42 Flygare, S. *et al.* The VAAST Variant Prioritizer (VVP): ultrafast, easy to use whole genome variant prioritization tool. *BMC Bioinformatics* **19**, 57, doi:10.1186/s12859-018-2056-y (2018).

43 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

44 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).

45 Jolly, A. *et al.* Exome Sequencing of a Primary Ovarian Insufficiency Cohort Reveals Common Molecular Etiologies for a Spectrum of Disease. *The Journal of clinical endocrinology and metabolism* **104**, 3049-3067, doi:10.1210/jc.2019-00248 (2019).

46 Hantash, F. M. *et al.* FMR1 premutation carrier frequency in patients undergoing routine population-based carrier screening: insights into the prevalence of fragile X syndrome, fragile X-associated tremor/ataxia syndrome, and fragile X-associated primary ovarian insufficiency in the United States. *Genet. Med.* **13**, 39-45, doi:10.1097/GIM.0b013e3181fa9fad (2011).

47 Fuentes Fajardo, K. V. *et al.* Detecting false-positive signals in exome sequencing. *Hum Mutat* **33**, 609-613, doi:10.1002/humu.22033 (2012).

48 Coonrod, E. M., Margraf, R. L., Russell, A., Voelkerding, K. V. & Reese, M. G. Clinical analysis of genome next-generation sequencing data using the Omicia platform. *Expert Rev Mol Diagn* **13**, 529-540, doi:10.1586/14737159.2013.811907 (2013).

49 Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome research* **11**, 863-874, doi:10.1101/gr.176601 (2001).

50 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

51 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**, 575-576, doi:10.1038/nmeth0810-575 (2010).

52 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* **20**, 110-121, doi:10.1101/gr.097857.109 (2010).

53 Breiman, l. Random forests. *Machine Learning.* **45**, 5-32 (2001).

54 Singleton, M. V. *et al.* Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *American journal of human genetics* **94**, 599-610, doi:10.1016/j.ajhg.2014.03.010 (2014).

55 Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic acids research* **48**, D835-D844, doi:10.1093/nar/gkz972 (2020).

56 *Online Mendelian Inheritance in Man, OMIM®*. (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University).

57 *https://en.wikipedia.org/wiki/Bayes_factor*.

58 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

59 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

60 Luo, C. *et al.* Superovulation strategies for 6 commonly used mouse strains. *J Am Assoc Lab Anim Sci* **50**, 471-478 (2011).
61 Palu, R. A. S. & Chow, C. Y. Baldspot/ELOVL6 is a conserved modifier of disease and the ER stress response. *PLoS genetics* **14**, e1007557, doi:10.1371/journal.pgen.1007557 (2018).
62 Wolfner, M. F. Battle and ballet: molecular interactions between the sexes in Drosophila. *J. Hered.* **100**, 399-410, doi:10.1093/jhered/esp013 (2009).
63 Yang, X. *et al.* Gene variants identified by whole-exome sequencing in 33 French women with premature ovarian insufficiency. *Journal of assisted reproduction and genetics* **36**, 39-45, doi:10.1007/s10815-018-1349-4 (2019).
64 Meduri, G. *et al.* Delayed puberty and primary amenorrhea associated with a novel mutation of the human follicle-stimulating hormone receptor: clinical, histological, and molecular studies. *The Journal of clinical endocrinology and metabolism* **88**, 3491-3498, doi:10.1210/jc.2003-030217 (2003).
65 Touraine, P. *et al.* New natural inactivating mutations of the follicle-stimulating hormone receptor: correlations between receptor function and phenotype. *Molecular endocrinology* **13**, 1844-1854, doi:10.1210/mend.13.11.0370 (1999).
66 Gurbuz, F. *et al.* Novel inactivating mutations of the DCAF17 gene in American and Turkish families cause male infertility and female subfertility in the mouse model. *Clin. Genet.* **93**, 853-859, doi:10.1111/cge.13183 (2018).
67 Zhernakova, A., Withoff, S. & Wijmenga, C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat Rev Endocrinol* **9**, 646-659, doi:10.1038/nrendo.2013.161 (2013).
68 Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature genetics* **47**, 1294-1303, doi:10.1038/ng.3412 (2015).
69 Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nature genetics* **44**, 260-268, doi:10.1038/ng.1051 (2012).
70 Wang, H. & Hoog, C. Structural damage to meiotic chromosomes impairs DNA recombination and checkpoint control in mammalian oocytes. *J Cell Biol* **173**, 485-495, doi:10.1083/jcb.200512077 (2006).
71 Qin, Y. *et al.* CSB-PGBD3 Mutations Cause Premature Ovarian Failure. *PLoS genetics* **11**, e1005419, doi:10.1371/journal.pgen.1005419 (2015).
72 Guzeloglu-Kayisli, O. *et al.* Embryonic poly(A)-binding protein (EPAB) is required for oocyte maturation and female fertility in mice. *Biochem J* **446**, 47-58, doi:10.1042/BJ20120467 (2012).
73 Song, Z. H. *et al.* Germ cell-specific Atg7 knockout results in primary ovarian insufficiency in female mice. *Cell Death Dis.* **6**, e1589, doi:10.1038/cddis.2014.559 (2015).
74 Pangas, S. A. *et al.* Intraovarian activins are required for female fertility. *Molecular endocrinology* **21**, 2458-2471, doi:10.1210/me.2007-0146 (2007).
75 Tomic, D. *et al.* Ovarian follicle development requires Smad3. *Molecular endocrinology* **18**, 2224-2240, doi:10.1210/me.2003-0414 (2004).
76 Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Molecular endocrinology* **14**, 1053-1063, doi:10.1210/mend.14.7.0479 (2000).
77 Myllymaa, S. *et al.* Inhibition of oocyte growth factors in vivo modulates ovarian folliculogenesis in neonatal and immature mice. *Reproduction* **139**, 587-598, doi:10.1530/REP-09-0391 (2010).
78 Wang, J., Zhang, W., Jiang, H., Wu, B. L. & Primary Ovarian Insufficiency, C. Mutations in HFM1 in recessive primary ovarian insufficiency. *The New England journal of medicine* **370**, 972-974, doi:10.1056/NEJMc1310150 (2014).
79 Robevska, G. *et al.* Functional characterization of novel NR5A1 variants reveals multiple complex roles in disorders of sex development. *Hum Mutat* **39**, 124-139, doi:10.1002/humu.23354 (2018).

80 Camats, N. *et al.* Ten novel mutations in the NR5A1 gene cause disordered sex development in 46,XY and ovarian insufficiency in 46,XX individuals. *The Journal of clinical endocrinology and metabolism* **97**, E1294-1306, doi:10.1210/jc.2011-3169 (2012).

81 Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nature genetics* **50**, 746-753, doi:10.1038/s41588-018-0101-4 (2018).

82 Kasak, L. & Laan, M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. *Human genetics*, doi:10.1007/s00439-020-02112-y (2020).

83 Lascarez-Lagunas, L., Martinez-Garcia, M. & Colaiacovo, M. SnapShot: Meiosis - Prophase I. *Cell* **181**, 1442-1442 e1441, doi:10.1016/j.cell.2020.04.038 (2020).

84 Lei, L. & Spradling, A. C. Mouse oocytes differentiate through organelle enrichment from sister cyst germ cells. *Science* **352**, 95-99, doi:10.1126/science.aad2156 (2016).

85 Fraile, J. M. *et al.* Loss of the deubiquitinase USP36 destabilizes the RNA helicase DHX33 and causes preimplantation lethality in mice. *The Journal of biological chemistry* **293**, 2183-2194, doi:10.1074/jbc.M117.788430 (2018).

86 Lee, M. C., Skora, A. D. & Spradling, A. C. Identification of Genes Mediating Drosophila Follicle Cell Progenitor Differentiation by Screening for Modifiers of GAL4::UAS Variegation. *G3 (Bethesda)* **7**, 309-318, doi:10.1534/g3.116.036038 (2017).

87 Sun, Y. *et al.* Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E1419-E1428, doi:10.1073/pnas.1718723115 (2018).

88 Gomes Fernandes, M. *et al.* Human-specific subcellular compartmentalization of P-element induced wimpy testis-like (PIWIL) granules during germ cell development and spermatogenesis. *Human reproduction* **33**, 258-269, doi:10.1093/humrep/dex365 (2018).

89 Roovers, E. F. *et al.* Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep* **10**, 2069-2082, doi:10.1016/j.celrep.2015.02.062 (2015).

90 Tan, M. *et al.* PIWIL3 Forms a Complex with TDRKH in Mammalian Oocytes. *Cells* **9**, doi:10.3390/cells9061356 (2020).

91 Kuramochi-Miyagawa, S. *et al.* Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839-849, doi:10.1242/dev.00973 (2004).

92 Burns, K. H. *et al.* Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos. *Science* **300**, 633-636, doi:10.1126/science.1081813 (2003).

93 Inoue, A., Ogushi, S., Saitou, M., Suzuki, M. G. & Aoki, F. Involvement of mouse nucleoplasmin 2 in the decondensation of sperm chromatin after fertilization. *Biol Reprod* **85**, 70-77, doi:10.1095/biolreprod.110.089342 (2011).

94 Peng, H. *et al.* Valosin-containing protein is associated with maintenance of meiotic arrest in mouse oocytesdagger. *Biol Reprod* **100**, 963-970, doi:10.1093/biolre/ioy244 (2019).

95 Beattie, R. *et al.* Mosaic Analysis with Double Markers Reveals Distinct Sequential Functions of Lgl1 in Neural Stem Cells. *Neuron* **94**, 517-533 e513, doi:10.1016/j.neuron.2017.04.012 (2017).

96 Liu, N. *et al.* A 21-bp indel within the LLGL1 gene is significantly associated with litter size in goat. *Anim. Biotechnol.*, 1-6, doi:10.1080/10495398.2019.1677682 (2019).

97 Higgs, M. R. *et al.* BOD1L Is Required to Suppress Deleterious Resection of Stressed Replication Forks. *Mol Cell* **59**, 462-477, doi:10.1016/j.molcel.2015.06.007 (2015).

98 Hoek, A., Schoemaker, J. & Drexhage, H. A. Premature ovarian failure and ovarian autoimmunity. *Endocrine reviews* **18**, 107-134, doi:10.1210/edrv.18.1.0291 (1997).

99 Frommer, L. & Kahaly, G. J. Autoimmune Polyendocrinopathy. *The Journal of clinical endocrinology and metabolism* **104**, 4769-4782, doi:10.1210/jc.2019-00602 (2019).

100 Mandel, M., Gurevich, M., Pauzner, R., Kaminski, N. & Achiron, A. Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. *Clin. Exp. Immunol.* **138**, 164-170, doi:10.1111/j.1365-2249.2004.02587.x (2004).

101 Desai, S. *et al.* MCM8 and MCM9 Nucleotide Variants in Women with Primary Ovarian Insufficiency. *The Journal of clinical endocrinology and metabolism*, jc20162565, doi:10.1210/jc.2016-2565 (2016).

102 Eskenazi, S. B., A.; Hugon-Rodin, J.; Plu-Bureau, G.; Gompel, A.; Catteau-Jonard, S.; Molina-Gomes, D.; Dewailly, D.; Dode, C.; Christin-Maitre, S.; Touraine, P. Next generation sequencing should be proposed to every woman with "idiopathic" premature ovarian insufficiency. *J Endocr Soc* (2021).

103 Vuillaumier-Barrot, S. *et al.* Expanding the Spectrum of PMM2-CDG Phenotype. *JIMD Rep* **5**, 123-125, doi:10.1007/8904_2011_114 (2012).

104 Zhao, M., Feng, F., Chu, C., Yue, W. & Li, L. A novel EIF4ENIF1 mutation associated with a diminished ovarian reserve and premature ovarian insufficiency identified by whole-exome sequencing. *Journal of ovarian research* **12**, 119, doi:10.1186/s13048-019-0595-0 (2019).

105 Perry, J. R. *et al.* DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. *Human molecular genetics* **23**, 2490-2497, doi:10.1093/hmg/ddt620 (2014).

106 Chen, B. *et al.* Consanguineous familial study revealed biallelic FIGLA mutation associated with premature ovarian insufficiency. *Journal of ovarian research* **11**, 48, doi:10.1186/s13048-018-0413-0 (2018).

107 Morita, Y. *et al.* Caspase-2 deficiency prevents programmed germ cell death resulting from cytokine insufficiency but not meiotic defects caused by loss of ataxia telangiectasia-mutated (Atm) gene function. *Cell Death Differ.* **8**, 614-620, doi:10.1038/sj.cdd.4400845 (2001).

108 Pierce, S. B. *et al.* Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. *American journal of human genetics* **92**, 614-620, doi:10.1016/j.ajhg.2013.03.007 (2013).

109 Chen, A. *et al.* Mutations in the mitochondrial ribosomal protein MRPS22 lead to primary ovarian insufficiency. *Human molecular genetics* **27**, 1913-1926, doi:10.1093/hmg/ddy098 (2018).

110 Adelman, C. A. *et al.* HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis. *Nature* **502**, 381-384, doi:10.1038/nature12565 (2013).

111 Renault, L. *et al.* BMPR1A and BMPR1B Missense Mutations Cause Primary Ovarian Insufficiency. *The Journal of clinical endocrinology and metabolism* **105**, doi:10.1210/clinem/dgz226 (2020).

112 Norling, A. *et al.* 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. *Human reproduction* **29**, 1818-1827, doi:10.1093/humrep/deu149 (2014).

113 Qin, Y. *et al.* NOBOX homeobox mutation causes premature ovarian failure. *American journal of human genetics* **81**, 576-581, doi:10.1086/519496 (2007).

114 Smirin-Yosef, P. *et al.* A Biallelic Mutation in the Homologous Recombination Repair Gene SPIDR Is Associated With Human Gonadal Dysgenesis. *The Journal of clinical endocrinology and metabolism* **102**, 681-688, doi:10.1210/jc.2016-2714 (2017).

115 Chrzanowska, K. H. *et al.* High prevalence of primary ovarian insufficiency in girls and young women with Nijmegen breakage syndrome: evidence from a longitudinal study. *The Journal of clinical endocrinology and metabolism* **95**, 3133-3140, doi:10.1210/jc.2009-2628 (2010).

116 Wang, L. L. *et al.* Association between osteosarcoma and deleterious mutations in the RECQL4 gene in Rothmund-Thomson syndrome. *Journal of the National Cancer Institute* **95**, 669-674, doi:10.1093/jnci/95.9.669 (2003).

117 Siitonen, H. A. *et al.* The mutation spectrum in RECQL4 diseases. *Eur J Hum Genet* **17**, 151-158, doi:10.1038/ejhg.2008.154 (2009).

118 Subramanian, G. N., Lavin, M. & Homer, H. A. Premature ovarian ageing following heterozygous loss of Senataxin. *Molecular human reproduction* **27**, doi:10.1093/molehr/gaaa080 (2021).

119 Bakker, S. T. *et al.* Fancf-deficient mice are prone to develop ovarian tumours. *J. Pathol.* **226**, 28-39, doi:10.1002/path.2992 (2012).
120 Huang, H. L. *et al.* Mutant ZP1 in familial infertility. *The New England journal of medicine* **370**, 1220-1226, doi:10.1056/NEJMoa1308851 (2014).
121 Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749-1753, doi:10.1126/science.7792600 (1995).
122 Weinberg-Shukron, A. *et al.* Essential Role of BRCA2 in Ovarian Development and Function. *The New England journal of medicine* **379**, 1042-1049, doi:10.1056/NEJMoa1800024 (2018).
123 Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. & McKay, M. J. Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Dev. Cell* **8**, 949-961, doi:10.1016/j.devcel.2005.03.018 (2005).
124 Bouilly, J. *et al.* Identification of Multiple Gene Mutations Accounts for a new Genetic Architecture of Primary Ovarian Insufficiency. *The Journal of clinical endocrinology and metabolism* **101**, 4541-4550, doi:10.1210/jc.2016-2152 (2016).
125 Fouquet, B. *et al.* A homozygous FANCM mutation underlies a familial case of non-syndromic primary ovarian insufficiency. *Elife* **6**, doi:10.7554/eLife.30490 (2017).
126 Fan, S. *et al.* Homozygous mutations in C14orf39/SIX6OS1 cause non-obstructive azoospermia and premature ovarian insufficiency in humans. *American journal of human genetics* **108**, 324-336, doi:10.1016/j.ajhg.2021.01.010 (2021).
127 Gomez, H. L. *et al.* C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nat Commun* **7**, 13298, doi:10.1038/ncomms13298 (2016).
128 Fogli, A. *et al.* Ovarian failure related to eukaryotic initiation factor 2B mutations. *American journal of human genetics* **72**, 1544-1550, doi:10.1086/375404 (2003).
129 Maddirevula, S. *et al.* A genomics approach to females with infertility and recurrent pregnancy loss. *Human genetics* **139**, 605-613, doi:10.1007/s00439-020-02143-5 (2020).
130 Bekheirnia, M. R. *et al.* POLG mutation in a patient with cataracts, early-onset distal muscle weakness and atrophy, ovarian dysgenesis and 3-methylglutaconic aciduria. *Gene* **499**, 209-212, doi:10.1016/j.gene.2012.02.034 (2012).
131 Arora, H. *et al.* Bloom syndrome. *Int. J. Dermatol.* **53**, 798-802, doi:10.1111/ijd.12408 (2014).
132 Bashamboo, A. *et al.* Loss of Function of the Nuclear Receptor NR2F2, Encoding COUP-TF2, Causes Testis Development and Cardiac Defects in 46,XX Children. *American journal of human genetics* **102**, 487-493, doi:10.1016/j.ajhg.2018.01.021 (2018).
133 Zangen, D. *et al.* XX ovarian dysgenesis is caused by a PSMC3IP/HOP2 mutation that abolishes coactivation of estrogen-driven transcription. *American journal of human genetics* **89**, 572-579, doi:10.1016/j.ajhg.2011.09.006 (2011).
134 Alvaro Mercadal, B. *et al.* AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Human reproduction* **30**, 1196-1202, doi:10.1093/humrep/dev042 (2015).
135 Baronchelli, S. *et al.* Investigating the role of X chromosome breakpoints in premature ovarian failure. *Mol. Cytogenet.* **5**, 32, doi:10.1186/1755-8166-5-32 (2012).
136 Ylikallio, E. *et al.* MCM3AP in recessive Charcot-Marie-Tooth neuropathy and mild intellectual disability. *Brain* **140**, 2093-2103, doi:10.1093/brain/awx138 (2017).
137 Murray, A., Webb, J., Dennis, N., Conway, G. & Morton, N. Microdeletions in FMR2 may be a significant cause of premature ovarian failure. *Journal of medical genetics* **36**, 767-770, doi:10.1136/jmg.36.10.767 (1999).
138 Pitman, J. L. *et al.* The fate of granulosa cells following premature oocyte loss and the development of ovarian cancers. *Int. J. Dev. Biol.* **56**, 949-958, doi:10.1387/ijdb.120144jp (2012).
139 Moniruzzaman, M. & Miyano, T. KIT-KIT ligand in the growth of porcine oocytes in primordial follicles. *J Reprod Dev* **53**, 1273-1281, doi:10.1262/jrd.19107 (2007).

140 Matikainen, T. *et al.* Caspase-3 gene knockout defines cell lineage specificity for programmed cell death signaling in the ovary. *Endocrinology* **142**, 2468-2480, doi:10.1210/endo.142.6.8078 (2001).

141 Fotovati, A., Abu-Ali, S., Nakayama, K. & Nakayama, K. I. Impaired ovarian development and reduced fertility in female mice deficient in Skp2. *J. Anat.* **218**, 668-677, doi:10.1111/j.1469-7580.2011.01370.x (2011).

142 Weng, K. A., Jeffreys, C. A. & Bickel, S. E. Rejuvenation of meiotic cohesion in oocytes during prophase I is required for chiasma maintenance and accurate chromosome segregation. *PLoS genetics* **10**, e1004607, doi:10.1371/journal.pgen.1004607 (2014).

143 Boekhout, M. *et al.* REC114 Partner ANKRD31 Controls Number, Timing, and Location of Meiotic DNA Breaks. *Mol Cell* **74**, 1053-1068 e1058, doi:10.1016/j.molcel.2019.03.023 (2019).

144 Papanikos, F. *et al.* Mouse ANKRD31 Regulates Spatiotemporal Patterning of Meiotic Recombination Initiation and Ensures Recombination between X and Y Sex Chromosomes. *Mol Cell* **74**, 1069-1085 e1011, doi:10.1016/j.molcel.2019.03.022 (2019).

145 Juneja, S. C., Barr, K. J., Enders, G. C. & Kidder, G. M. Defects in the germ line and gonads of mice lacking connexin43. *Biol Reprod* **60**, 1263-1270, doi:10.1095/biolreprod60.5.1263 (1999).

146 Iida, T. & Lilly, M. A. missing oocyte encodes a highly conserved nuclear protein required for the maintenance of the meiotic cycle and oocyte identity in Drosophila. *Development* **131**, 1029-1039, doi:10.1242/dev.01001 (2004).

147 Hsieh, M. *et al.* Mice null for Frizzled4 (Fzd4-/-) are infertile and exhibit impaired corpora lutea formation and function. *Biol Reprod* **73**, 1135-1146, doi:10.1095/biolreprod.105.042739 (2005).

148 Inagaki, A., Roset, R. & Petrini, J. H. Functions of the MRE11 complex in the development and maintenance of oocytes. *Chromosoma* **125**, 151-162, doi:10.1007/s00412-015-0535-8 (2016).

149 Lipkin, S. M. *et al.* Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nature genetics* **31**, 385-390, doi:10.1038/ng931 (2002).

150 Lim, A. K. *et al.* The nuage mediates retrotransposon silencing in mouse primordial ovarian follicles. *Development* **140**, 3819-3825, doi:10.1242/dev.099184 (2013).

151 Arafat, M. *et al.* Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *Journal of medical genetics* **54**, 633-639, doi:10.1136/jmedgenet-2017-104514 (2017).

152 Aboura, A. *et al.* Array comparative genomic hybridization profiling analysis reveals deoxyribonucleic acid copy number variations associated with premature ovarian failure. *The Journal of clinical endocrinology and metabolism* **94**, 4540-4546, doi:10.1210/jc.2009-0186 (2009).

153 Million Passe, C. M. *et al.* Loss of the protein NUPR1 (p8) leads to delayed LHB expression, delayed ovarian maturation, and testicular development of a sertoli-cell-only syndrome-like phenotype in mice. *Biol Reprod* **79**, 598-607, doi:10.1095/biolreprod.108.068304 (2008).

154 Ratts, V. S., Flaws, J. A., Kolp, R., Sorenson, C. M. & Tilly, J. L. Ablation of bcl-2 gene expression decreases the numbers of oocytes and primordial follicles established in the post-natal female mouse gonad. *Endocrinology* **136**, 3665-3668, doi:10.1210/endo.136.8.7628407 (1995).

155 de Boer, J. *et al.* Premature aging in mice deficient in DNA repair and transcription. *Science* **296**, 1276-1279, doi:10.1126/science.1070174 (2002).

156 Xu, B. *et al.* Proliferating cell nuclear antigen (PCNA) regulates primordial follicle assembly by promoting apoptosis of oocytes in fetal and neonatal mouse ovaries. *PLoS One* **6**, e16046, doi:10.1371/journal.pone.0016046 (2011).

157 Shindo, T. *et al.* ADAMTS-1: a metalloproteinase-disintegrin essential for normal growth, fertility, and organ morphology and function. *The Journal of clinical investigation* **105**, 1345-1352, doi:10.1172/JCI8635 (2000).

158 Fischer-Vize, J. A., Rubin, G. M. & Lehmann, R. The fat facets gene is required for Drosophila eye and embryo development. *Development* **116**, 985-1000 (1992).

159 Baumann, C., Viveiros, M. M. & De La Fuente, R. Loss of maternal ATRX results in centromere instability and aneuploidy in the mammalian oocyte and pre-implantation embryo. *PLoS genetics* **6**, e1001137, doi:10.1371/journal.pgen.1001137 (2010).

160 Messiaen, S. *et al.* Rad54 is required for the normal development of male and female germ cells and contributes to the maintainance of their genome integrity after genotoxic stress. *Cell Death Dis.* **4**, e774, doi:10.1038/cddis.2013.281 (2013).

161 Luo, M. *et al.* Protein Arginine Methyltransferase 6 Involved in Germ Cell Viability during Spermatogenesis and Down-Regulated by the Androgen Receptor. *Int. J. Mol. Sci.* **16**, 29467-29481, doi:10.3390/ijms161226186 (2015).

162 Liu, S. S. *et al.* A Novel Regulatory Axis, CHD1L-MicroRNA 486-Matrix Metalloproteinase 2, Controls Spermatogonial Stem Cell Properties. *Molecular and cellular biology* **39**, doi:10.1128/MCB.00357-18 (2019).

163 Iles, N., Rulten, S., El-Khamisy, S. F. & Caldecott, K. W. APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Molecular and cellular biology* **27**, 3793-3803, doi:10.1128/MCB.02269-06 (2007).

164 Bohgaki, T. *et al.* Genomic instability, defective spermatogenesis, immunodeficiency, and cancer in a mouse model of the RIDDLE syndrome. *PLoS genetics* **7**, e1001381, doi:10.1371/journal.pgen.1001381 (2011).

165 Wang, H. *et al.* Inhibition of CDK7 bypasses spindle assembly checkpoint via premature cyclin B degradation during oocyte meiosis. *Biochim. Biophys. Acta* **1863**, 2993-3000, doi:10.1016/j.bbamcr.2016.09.020 (2016).

166 Stern, H. R., Sefcikova, J., Chaparro, V. E. & Beuning, P. J. Mammalian DNA Polymerase Kappa Activity and Specificity. *Molecules* **24**, doi:10.3390/molecules24152805 (2019).

167 Shapouri, F. *et al.* Tob1 is expressed in developing and adult gonads and is associated with the P-body marker, Dcp2. *Cell Tissue Res.* **364**, 443-451, doi:10.1007/s00441-015-2328-z (2016).

168 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).

169 Xu, C. *et al.* Crystal structure of human nuclear pore complex component NUP43. *FEBS Lett.* **589**, 3247-3253, doi:10.1016/j.febslet.2015.09.008 (2015).

170 Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233-1243, doi:10.1016/j.cell.2013.10.055 (2013).

171 Srivastava, S. *et al.* BRAT1 mutations present with a spectrum of clinical severity. *Am J Med Genet A* **170**, 2265-2273, doi:10.1002/ajmg.a.37783 (2016).

172 Khan, T. N. *et al.* Mutations in NCAPG2 Cause a Severe Neurodevelopmental Syndrome that Expands the Phenotypic Spectrum of Condensinopathies. *American journal of human genetics* **104**, 94-111, doi:10.1016/j.ajhg.2018.11.017 (2019).

173 Chang, H. R. *et al.* Hypomorphic Mutations in TONSL Cause SPONASTRIME Dysplasia. *American journal of human genetics* **104**, 439-453, doi:10.1016/j.ajhg.2019.01.009 (2019).

174 Burrage, L. C. *et al.* Bi-allelic Variants in TONSL Cause SPONASTRIME Dysplasia and a Spectrum of Skeletal Dysplasia Phenotypes. *American journal of human genetics* **104**, 422-438, doi:10.1016/j.ajhg.2019.01.007 (2019).

175 Larsson, V. J., Jafferali, M. H., Vijayaraghavan, B., Figueroa, R. A. & Hallberg, E. Mitotic spindle assembly and gamma-tubulin localisation depend on the integral nuclear membrane protein Samp1. *J Cell Sci* **131**, doi:10.1242/jcs.211664 (2018).

176 Wang, N., Zhang, P., Guo, X., Zhou, Z. & Sha, J. Hnrnpk, a protein differentially expressed in immature rat ovarian development, is required for normal primordial follicle assembly and development. *Endocrinology* **152**, 1024-1035, doi:10.1210/en.2010-0797 (2011).

177 Mendez, R. & Richter, J. D. Translational control by CPEB: a means to the end. *Nat. Rev. Mol. Cell Biol.* **2**, 521-529, doi:10.1038/35080081 (2001).

178 Catchpoole, D. *et al.* Mutation analysis of H19 and NAP1L4 (hNAP2) candidate genes and IGF2 DMR2 in Beckwith-Wiedemann syndrome. *Journal of medical genetics* **37**, 212-215, doi:10.1136/jmg.37.3.212 (2000).

179 Dunwell, T. L. & Holland, P. W. H. A sister of NANOG regulates genes expressed in pre-implantation human development. *Open Biol* **7**, doi:10.1098/rsob.170027 (2017).

180 Zheng, Y. *et al.* The Seckel syndrome and centrosomal protein Ninein localizes asymmetrically to stem cell centrosomes but is not required for normal development, behavior, or DNA damage response in Drosophila. *Mol Biol Cell* **27**, 1740-1752, doi:10.1091/mbc.E15-09-0655 (2016).

181 de Munnik, S. A. *et al.* Meier-Gorlin syndrome: growth and secondary sexual development of a microcephalic primordial dwarfism disorder. *Am J Med Genet A* **158A**, 2733-2742, doi:10.1002/ajmg.a.35681 (2012).

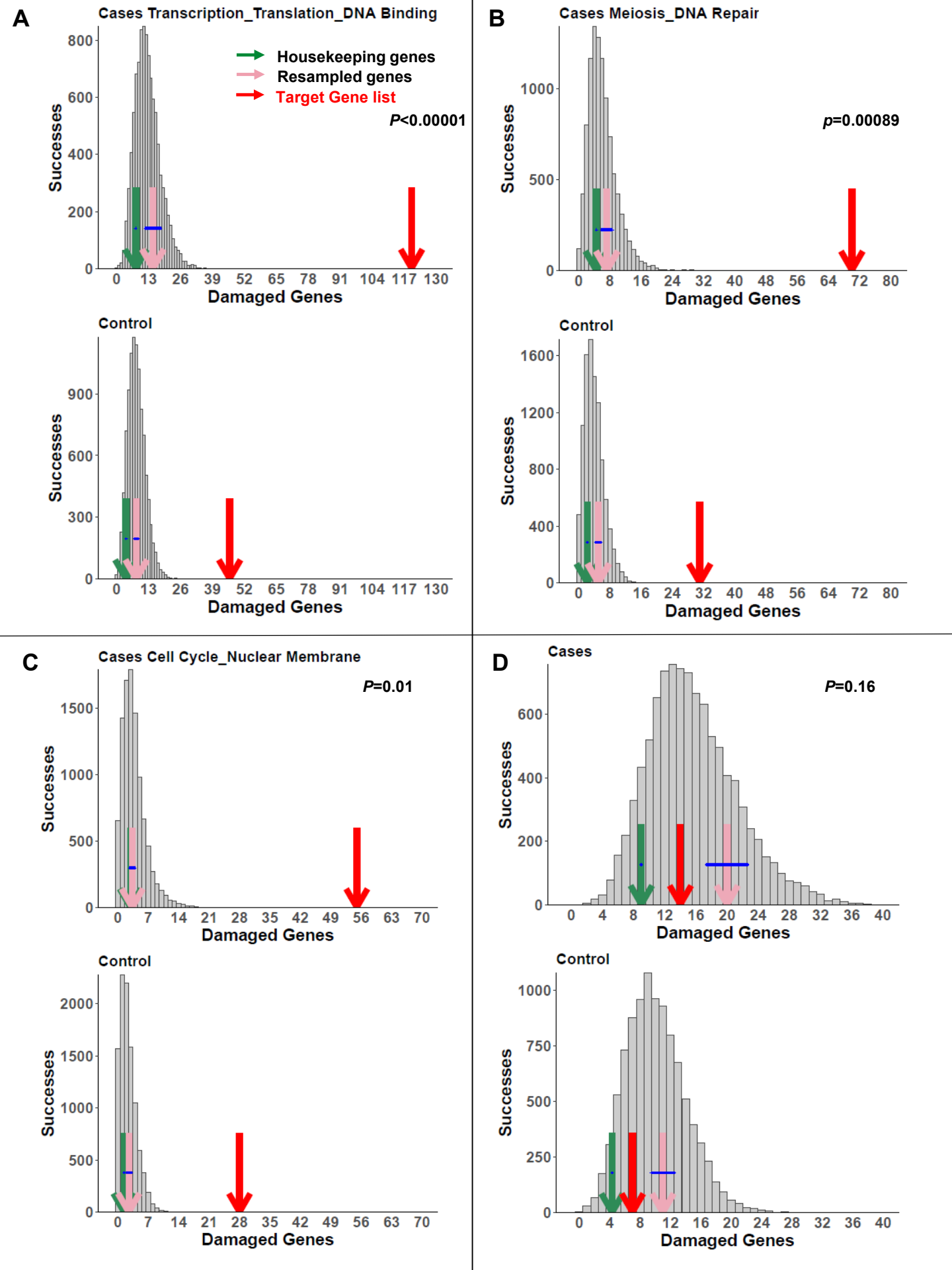
182 Sun, X. *et al.* FancJ (Brip1) loss-of-function allele results in spermatogonial cell depletion during embryogenesis and altered processing of crossover sites during meiotic prophase I in mice. *Chromosoma* **125**, 237-252, doi:10.1007/s00412-015-0549-2 (2016).

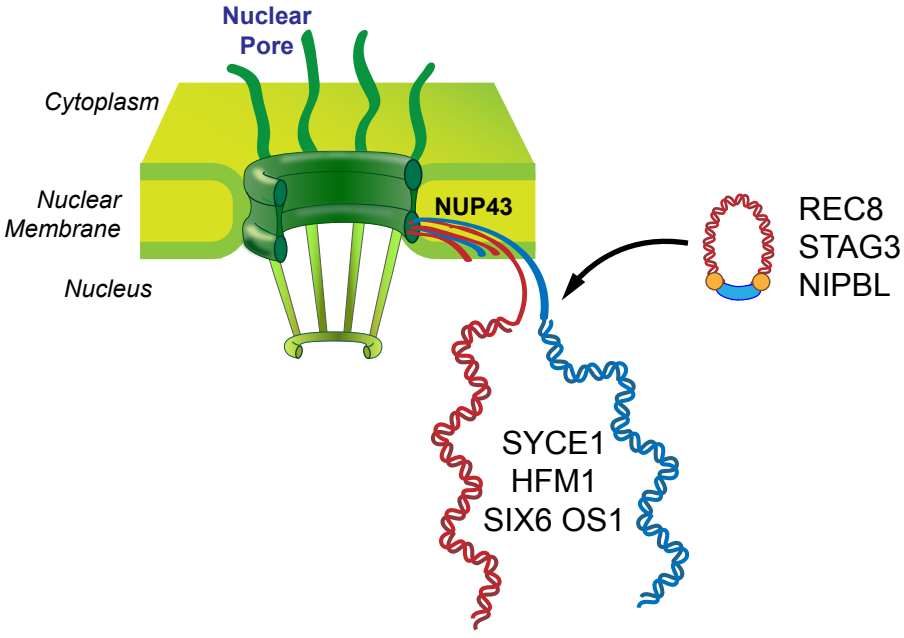
183 Li, J. *et al.* Differential display identifies overexpression of the USP36 gene, encoding a deubiquitinating enzyme, in ovarian cancer. *Int. J. Med. Sci.* **5**, 133-142, doi:10.7150/ijms.5.133 (2008).

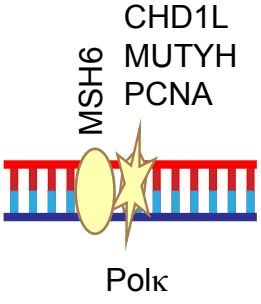
184 Chapman, K. M. *et al.* Linking spermatid ribonucleic acid (RNA) binding protein and retrogene diversity to reproductive success. *Mol. Cell. Proteomics* **12**, 3221-3236, doi:10.1074/mcp.M113.030585 (2013).

185 Virant-Klun, I., Leicht, S., Hughes, C. & Krijgsveld, J. Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. *Mol. Cell. Proteomics* **15**, 2616-2627, doi:10.1074/mcp.M115.056887 (2016).

186 Celichowski, P. *et al.* "Positive Regulation of RNA Metabolic Process" Ontology Group Highly Regulated in Porcine Oocytes Matured In Vitro: A Microarray Approach. *Biomed Res Int* **2018**, 2863068, doi:10.1155/2018/2863068 (2018).


A
Cases Transcription_Translation_DNA Binding
Housekeeping genes
Resampled genes
Target Gene list
P<0.00001
Successes
Damaged Genes
Control
B
Cases Meiosis_DNA Repair
p=0.00089
C
Cases Cell Cycle_Nuclear Membrane
P=0.01
D
Cases
P=0.16

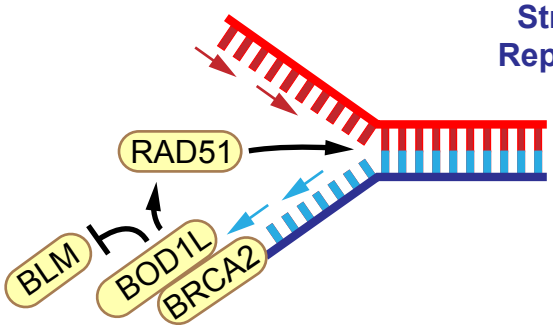




DNA Damage Repair

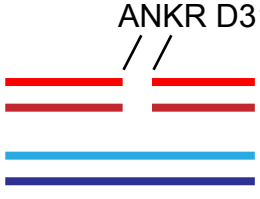
DNA-Base Excision Repair

PRMT6
POLβ
ERCC2

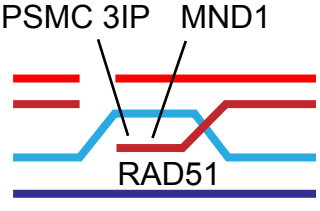


Stress Stalled Replication Fork

SAMHD1
FANCM
SPIDR



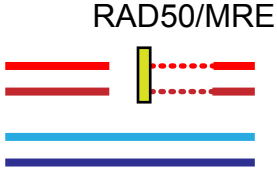
Double Strand Break



Strand Invasion

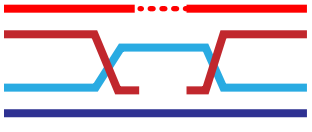
RAD51
MRE11

Homologue Pairing



End Resection

BRCA1
RECQL4



Double Strand Break Repair
Crossovers
Dissociation of Holliday Junctions & Resolution

BRCA1	BLM	BRIP1	RUVBL2
BRCA2	FANCA	MCM9	MIOS
RAD50	FANCD2	HELQ	APLF
RAD51	MLH3	RECQL4	RNF168
ATM	P1F1	FANCF	TONSL