Studying the effects of genetic factors on the female reproductive lifespan

Submitted by

Saleh Shekari

to the University of Exeter as a thesis for the degree of Doctoral of Philosophy in Medical Studies at the University of Queensland and the University of Exeter

September 2023

Abstract

The objective of my research was to investigate the rare and very-rare genetic factors influencing female reproductive ageing in humans using large-scale population exome-sequencing data. Over the past decade, most studies have relied on non-sequencing genomic data, which only allowed analysis of common genomic variants. However, these genome-wide array studies have limitations in capturing the complete range of genetic variation. Consequently, our understanding of the role of rare genomic variants, which may have a significant impact on menopause timing, has been limited. Furthermore, comprehensive studies exploring genetic factors associated with menopause age, particularly early and very early menopause, have been limited by the lack of large-scale sequencing genomic data, such as population-based datasets. Most of the previously published research has been derived from clinical and family studies, and there has been a dearth of population-based studies that can validate and identify novel genomic factors using a cohort of healthy individuals. Consequently, my aim was to utilise population whole-exome sequencing data for the first time to advance our understanding of genomic factors that impact female reproductive lifespan.

In Chapter 1, I provide an introduction to the biology of menopause. I emphasise the importance of studying menopause timing and the revolutionary impact of using population sequencing genomic data to improve our understanding of the underlying genomic causes of menopause timing.

Chapter 2 comprises analysis focusing on the correlation between bone morphogenetic protein 15 (*BMP15*) and its previously reported variants in relation to menopause timing. The *BMP15* gene and its missense variants have been identified as a potential candidate for premature ovarian insufficiency (POI) based on prior investigations. However, our study revealed no evidence of the previously reported variants being causative factors for POI. Furthermore, when conducting a gene burden association test, we found no significant association between various types of *BMP15* variants and early menopause.

Chapter 3 builds based on the previous chapter, which presents an in-depth analysis aimed at assessing the penetrance of over 100 genes associated with premature ovarian insufficiency (POI). The findings of this investigation provide limited evidence supporting the existence of autosomal dominant effects in the

reported POI genes. Surprisingly, the vast majority of heterozygous effects on these genes were ruled out, with 99.9% of all protein-truncating variants being observed in women with normal reproductive health. However, we did observe evidence of haploinsufficiency effects in certain genes, including *TWNK* and *SOHLH2*.

Chapter 4 is an exome-wide association study to identify rare genetic variants associated with menopause timing. We identified effects ~5 times larger than previously discovered in analyses of common variants, highlighting protein-coding variants in *ETAA1*, *ZNF518A*, *PNPLA8*, *PALB2* and *SAMHD1*. We found rare loss-of-function variants in the *ZNF518A* gene, which reduced menopause age by approximately six years.

Chapter 5 culminates by assessing the significant contributions made by this study in advancing our comprehension of the variation in genetic risk factors associated with female reproductive lifespan. Additionally, it outlines potential directions for future research in this field, highlighting areas that warrant further exploration and investigation.

List of contents

List of Table

List of Figure

Acknowledgements

I extend my heartfelt gratitude to my esteemed supervisors, Professor Anna Murray, Professor Gita Mishra, and Dr Katherine Ruth, for their invaluable guidance, unwavering support, and inspiration. Their mentorship has been instrumental in shaping my research journey and providing me with opportunities to engage in numerous projects and collaborations.

I am also deeply thankful to Professor Mike Weedon and Professor Caroline Wright for their supervision and invaluable advice. This research would not have been possible without the collective efforts and dedication of all the mentioned individuals. I am truly grateful for their support and mentorship throughout my academic journey.

I would also like to express my sincere appreciation to my parents and sisters for their unwavering belief in me and their constant support and encouragement over the years. To my friends, especially Charlie and Kathryn, thank you for always being there with your encouragement and for providing a fun distraction from the work. Special thanks go to my best friend, Ace, for the unlimited love and joy during our walks together.

Lastly, I am thankful to all those who provided me with the opportunity to enter a field that I find interesting, challenging, and, most of all, enjoyable. I look forward to continuing my work in this field in the future.

Author's Declaration

I was involved in the study design, analysis, and manuscript preparation for all the studies encompassed within this thesis. Some of the studies include analyses performed by other authors; however, in each case, I had a major role and was the first or joint first author on the paper. My specific contribution to each chapter is listed below:

Chapter 1: Introduction

I conducted a literature review and wrote the content.

Chapter 2: Heterozygous genetic variants in *BMP15* are not a common cause of Premature Ovarian Insufficiency.

I was a lead analyst on the project. I carried out exome sequencing quality control, ran the exome-wide association analyses, contributed to writing the introduction, methods, and results, producing Figures, Tables, and supplementary materials. This chapter was submitted in Human Reproduction in 2022.

Chapter 3: Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency.

I was a joint lead analyst on the project. I carried out exome sequencing quality control, analysed the exome-wide association results, and prepared the manuscript. I produce Figures, Tables, and supplementary materials. This chapter was published in Nature Medicine in 2023.

Chapter 4: Genetic susceptibility to earlier ovarian ageing increases de novo mutation rate in offspring.

I analysed the exome-wide association results for age at natural menopause phenotype, sensitivity analysis and quality control of *ZNF518A* genetic variants. This chapter was submitted to Nature in 2023 and subsequently resubmitted in 2024.

Chapter 5: Discussion

I wrote the content.

Chapter 1:

Introduction

Introductory notes

Chapter 1 provides an overview of menopause biology within the context of the female reproductive lifespan. I describe the importance of age at menopause in relation to female fertility. I present current knowledge regarding genetic risk factors that influence age at menopause. Additionally, the importance of population sequencing data for analysing female reproductive traits is discussed, emphasising the role of comprehensive genomic studies in advancing our knowledge of reproductive ageing.

Age at menopause as a marker of female reproductive lifespan

The age at which menopause occurs is an important indicator of a woman's reproductive lifespan. It signifies the end of her natural ability to conceive. Menopause is a natural biological process in a woman's life when menstruation stops permanently. As women age, they will generally go through this natural process, typically taking place between the ages of 50 and 52 years[1]. This excludes those who undergo non-natural menopause due to factors like surgery, chemotherapy, and other medical interventions. However, it is imperative to note that our understanding of the age at natural menopause (ANM) is hindered by a paucity of information from developing and economically challenged nations. Contradictory perspectives exist regarding the uniformity of ANM across populations. While some argue that ANM exhibits no significant differences, others contend that substantial variations persist. Notably, there is discernible diversity in ANM across racial, ethnic, and regional groups, a phenomenon that persists even after adjusting for influential factors. Recent studies have shed light on this variability, indicating that South Asian, Middle Eastern, Southeast Asian, and African American/Black women experience the youngest mean ages at natural menopause, ranging between 48.8 and 49.8 years. These differences remained evident even after further adjustments for menarche and parity. In contrast, Japanese (USA/UK) and Japanese (Japan)

women demonstrated the highest mean ages at natural menopause, with figures reaching 51.4 and 51.9 years[1]. This intricate interplay of factors underscores the need for a more comprehensive and inclusive approach to menopausal research, considering the diverse contexts in which women experience this crucial life transition.

Many women assume that once they reach a certain age, their natural fertility declines rapidly until it completely disappears at menopause. However, this is only partially accurate because while menopause marks the end of a woman's reproductive lifespan, natural infertility can occur ten years prior to menopause[2,3].

Human ovarian reserve from 6 months gestation to menopause

The human ovarian reserve refers to the number and quality of oocytes present in the ovaries. It is a critical factor in female reproductive health, as it determines a woman's ability to conceive naturally. The ovarian reserve is established during fetal development and gradually declines with age until menopause[4].

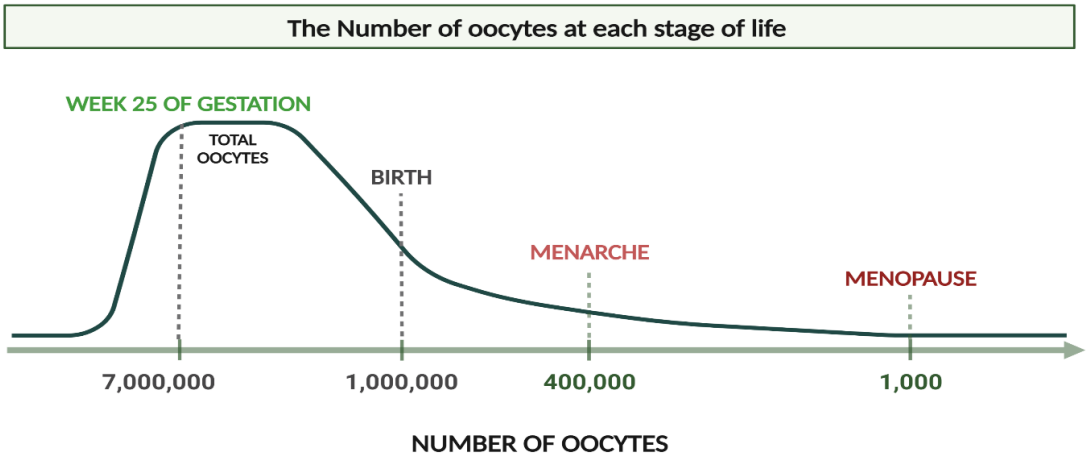
During fetal development, the ovaries contain approximately 6-7 million oogonia, the precursor cells that give rise to oocytes. However, this number decreases rapidly, and by the time of birth, the ovaries contain approximately 1-2 million oocytes[5,6]. Throughout a woman's life, spanning from before birth through adolescence and beyond, there is a gradual decrease in the number of oocytes due to a process called atresia. Atresia refers to the degeneration and reabsorption of immature oocytes[7].

Throughout puberty, a woman's reproductive potential reaches its full development, and a regular monthly process of ovulation takes place, wherein mature oocytes are released from the ovaries. However, the ovarian reserve still decreases even at this stage due to follicular atresia[8]. During each menstrual cycle, multiple follicles are initially recruited for development. However, it is noteworthy to mention that recruitment does not occur on a monthly basis; rather, it spans a period of approximately 12 months[9]. Ultimately, only one follicle will undergo maturation and ovulate, while the rest of the recruited follicles will undergo atresia, a process of degeneration and reabsorption by the body[10]. It is essential to expand on this aspect, as it

highlights the temporal dynamics and complexity involved in the selection and maturation of the dominant follicle leading to monthly ovulation[3].

As women age, the rate of follicular atresia, the natural degeneration of ovarian follicles, increases. Notably, after week 25 of gestation, the decline in the ovarian reserve becomes more pronounced [11,12]. This decline accelerates further in the years leading up to menopause, which is marked by the cessation of ovulation and a significant decrease in the number of remaining follicles[3]. Upon reaching menopause, the oocyte count in the ovary is bound to dip below the critical threshold level of 1000[13] (Figure 1).

Figure 0–1:Dynamic Changes in the Ovarian Reserve Throughout a Woman's Life.



Importance of studying age at menopause in population

According to the World Health Organization (WHO), approximately 1 in 6 adults worldwide experience infertility[14]. Infertility can be influenced by a multitude of factors, and among them, a significant contributor at the population level, particularly in high- and middle-income countries, is the increasing trend of women postponing childbirth until their 30s and beyond[15-17]. This decision can be influenced by a variety of personal, social, and economic factors, such as carrier and education, financial security, social norms, expectations and etc[18-20]. This shift in childbirth patterns can be attributed to several societal changes, including advancements in women's education, career opportunities, and access to effective contraception[15,21]. As women prioritise their personal and professional aspirations, they often choose to postpone starting a family.

In contemporary society, the issue of infertility associated with delayed childbirth has become a prominent concern in numerous developed countries[22,23].

Simultaneously, young women in their 20s often experience apprehension regarding the limited duration of their reproductive lifespan as they embark on their professional journeys[24,25]. The awareness of a finite fertility window, societal expectations, and the fear of potential infertility create a unique psychological burden. The pressure to balance personal and professional aspirations with the desire for motherhood can significantly impact the mental well-being and decision-making processes of young women[26].

At the population level, investigating the age of natural menopause and its associated factors presents a promising strategy to address concerns regarding reproductive timing. Viewing menopause as an indicator of the conclusion of the female reproductive lifespan and examining factors linked to early and premature menopause provide an opportunity to identify optimal reproductive timelines for women, potentially mitigating challenges related to delayed childbirth and infertility.

Distinguishing early and premature menopause as distinct phenotypes offers valuable insights into the contributing factors that lead to variations in menopausal onset. These variations significantly impact women's reproductive health and fertility prospects. With this research, we can establish correlations between specific factors and early or premature menopause, yielding crucial knowledge for identifying risk factors and potential interventions.

Consequently, this research could offer women and healthcare practitioners important guidance regarding reproductive choices and family planning to optimize fertility potential and address fertility challenges associated with delayed childbirth. Additionally, the investigation sheds light on the factors influencing menopause timing, contributing to a deeper understanding of the biology of female reproductive ageing and its relevance to the broader field of reproductive medicine and women's health.

Factors influencing age at menopause.

The age at which women experience menopause can vary widely and is influenced by a complex interplay of genetic, lifestyle, and reproductive factors. Various environmental factors play a crucial role in shaping the age at menopause. Notably, early and late-life events, encompassing aspects like multiple births, an earlier age at menarche, and a lower educational level, have

been identified as influential contributors[27]. Lifestyle choices, including smoking and alcohol consumption, have also demonstrated associations with the timing of menopause[28]. Additionally, non-genomic factors such as social influences, body mass index (BMI), and smoking duration have emerged as significant determinants[29-31]. Intriguingly, even early-life experiences, such as parental divorce during childhood and weight at 2 years, have been shown to exert discernible impacts on the age at menopause, emphasizing the multifaceted interplay between environmental circumstances and reproductive aging.[27,32]. Other factors that have been associated with earlier menopause include exposure to certain environmental toxins, such as pesticides and Polychlorinated biphenyls (PCBs), and certain medical treatments, such as chemotherapy and radiation therapy[33].

One of the most well-established predictors of age at menopause is genomics[34-38]. Empirical investigations have provided evidence supporting the familial nature of the age at menopause, indicating a heritable component in this biological phenomenon[39]. Furthermore, numerous specific genetic variants have been discovered through genomic research, showing significant associations with either earlier or later onset of menopause [38,40]. A noteworthy aspect of studying genomic factors that influence menopause timing is the ability to approximate a woman's reproductive timeline from birth. This predictive capability enables young females to gain an overview of their fertility timeline and potentially manage it in conjunction with other lifestyle factors. Such insights can facilitate informed decision-making regarding family planning and reproductive health strategies.

Genomic factors

Recent studies have elucidated the impact of genomic factors on menopause timing, offering compelling evidence and highlighting a substantial heritability rate for menopause. Researchers have estimated the heritability rate to be approximately 40%, signifying a significant genetic influence on the timing of menopause[38,41]. Several genomic factors have been firmly established as potential contributors to early menopause. Among these factors, chromosomal abnormalities account for approximately 13% of early menopause cases, while the fragile X premutation is associated with about 2% of early menopause occurrences[42-47].

Chromosomal abnormalities and menopause timing

Chromosome abnormalities have been associated with alterations in the timing of menopause. Specifically, women with certain X chromosomal abnormalities, namely aneuploidies and rearrangements, experience earlier menopause[42]. One of the most well-known chromosomal abnormalities associated with menopause timing is Turner syndrome, which occurs when a female is born with only one X chromosome instead of two. Women with Turner syndrome often experience early menopause[42].Conversely, women with trisomy X, or the presence of an extra X chromosome, have been observed to undergo menopause either earlier or later than women without this chromosomal abnormality[43,44].

Fragile X premutation and menopause

The fragile X premutation is a genetic condition that is caused by a repeat expansion in the Fragile X Mental Retardation 1 (*FMR1*) gene[45]. Normally, the *FMR1* gene contains up to 44 CGG repeats, but in individuals with the premutation, the gene contains between 55 and 200 CGG repeats[46]. While individuals with the premutation do not typically exhibit the full-blown symptoms of Fragile X syndrome, which is caused by over 200 CGG repeats in the *FMR1* gene, they are still at risk for developing certain health problems[48-51].

One of the most well-known health risks associated with the Fragile X premutation is an increased risk of premature ovarian insufficiency (POI), which can lead to early menopause[2,47]. Women with the premutation have been found to have an increased risk of POI, with some studies estimating that up to 20% of women with the premutation experience early menopause[2].

Premature ovarian insufficiency (POI)

Premature ovarian insufficiency (POI) is a condition characterized by the loss of normal ovarian function before the age of 40[52]. This condition affects approximately 1% of the population and it is a significant cause of female infertility[52-55]. POI can be either syndromic, occurring alongside other phenotypic features like in Turner's syndrome, or non-syndromic with various aetiologies[56-58]. While genetic causes are reported in 1-10% of cases, other causes include autoimmune and iatrogenic factors. Idiopathic cases account for 50-90% of POI[59], with 10-30% of those being familial, suggesting a genetic basis[60]. Moreover, estimates show that menopausal age has a heritability

ranging from 44% to 65%, and daughters of affected mothers have a six-fold increased risk of early menopause[34,38,61,62]. This genetic component adds to our understanding of the mechanisms underlying POI, providing valuable insights into potential genetic contributors to this condition.

More than 100 monogenic causes of POI have been documented, where a single genetic variant is sufficient to cause the phenotype (Table 1). To compile a comprehensive list of monogenic causes of POI, we primarily relied on the Genomics England Panel App. The Genomics England POI panel represents a cutting-edge initiative in advancing our understanding of the genetic determinants of POI. This panel is derived through a meticulous process that begins with an exhaustive literature review, identifying genes previously associated with POI. Expert consultation ensures the inclusion of genes with well-established connections to ovarian function, and rigorous validation processes confirm their relevance to the condition. The panel's dynamic nature allows for continuous updates, aligning with the evolving landscape of genomic research. By leveraging state-of-the-art sequencing technologies and methodologies, the Genomics England POI panel contributes significantly to the precision and efficiency of genetic screening efforts, facilitating a more nuanced comprehension of the genetic factors influencing POI. This resource employs a traffic light system, categorising genes as "GREEN" (high evidence), "AMBER" (moderate evidence), or "RED" (insufficient evidence).

Our analysis identified 67 genes, supplemented with 38 manually curated POI genes from literature up to June 2022. Selection criteria included a focus on POI phenotypes, discovery through family segregation studies, consanguineous pedigree analysis, and validation in animal models or cell-based assays. Our focus was on non-syndromic POI genes (details provided in Chapter 3). The gene list categorised by their mode of inheritance. Approximately half of these cases demonstrate an autosomal dominant (AD) inheritance pattern. Other genes show an autosomal recessive (AR) inheritance pattern, necessitating disruption of both gene copies to manifest the phenotype. Additionally, X chromosome genes have long been implicated in maintaining ovarian development and function. Structural variants on the X chromosome account for about 13% of POI cases in certain published series[59,60].

Table 0-1: Genes reported as associated with premature ovarian insufficiency.

Gene	Gene Name	Reported gene inheritance	Reference
AARS2	alanyl-tRNA synthetase 2, mitochondrial	AR	[63,64]
AIRE	autoimmune regulator	AD	[65]
ALOX12B	arachidonate 12-lipoxygenase, 12R type	AD	[66]
AMH	anti-Mullerian hormone	AD	[67]
ANKRD31	ankyrin repeat domain 31	AD	[68]
AR	androgen receptor	X-linked	[69]
ATG7	autophagy related 7	AD	[70]
ATG9A	autophagy related 9A	AD	[71]
ATM	ATM serine/threonine kinase	AR	[72]
BLM	BLM RecQ like helicase	AR	[73,74]
BMP15	bone morphogenetic protein 15	X-linked	[75-89]
BMPR1A	bone morphogenetic protein receptor type 1A	AD	[90]
BMPR1B	bone morphogenetic protein receptor type 1B	AD	[70,90]
BNC1	basonuclin 1	AD	[91]
BRCA2	BRCA2 DNA repair associated	AR	[92]
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	AD	[93]

C14orf39	chromosome 14 open reading frame 39	AR	94,95
CDKN1B	cyclin dependent kinase inhibitor 1B	AD	96
CLPP	caseinolytic mitochondrial matrix peptidase proteolytic subunit	AR	97-99
CPEB1	cytoplasmic polyadenylation element binding protein 1	AD	100-102
CYP17A1	cytochrome P450 family 17 subfamily A member 1	AR	103
CYP19A1	cytochrome P450 family 19 subfamily A member 1	AR	103
DACH2	dachshund family transcription factor 2	X-linked	104
DIAPH2	diaphanous related formin 2	X-linked	105,106
DMC1	DNA meiotic recombinase 1	AR	107,108
EIF2B2	eukaryotic translation initiation factor 2B subunit beta	AR	72,109
EIF2B4	eukaryotic translation initiation factor 2B subunit delta	AR	109
EIF2B5	eukaryotic translation initiation factor 2B subunit epsilon	AR	109
EIF4ENIF1	eukaryotic translation initiation factor 4E nuclear import factor 1	AD	110
ERAL1	Era like 12S mitochondrial rRNA chaperone 1	AR	111
ERCC6	ERCC excision repair 6, chromatin remodeling factor	AD	112
FANCA	FA complementation group A	AR	72,113
FANCC	FA complementation group C	AR	114
FANCG	FA complementation group G	AR	115

FANCL	FA complementation group L	AD	116
FANCM	FA complementation group M	AR	117,118
FIGLA	folliculogenesis specific bHLH transcription factor	AD	80,119-121
FMN2	formin 2	AD	122
FMR1	fragile X messenger ribonucleoprotein 1	X-linked	123
FOXL2	forkhead box L2	AD	72,80,124-128
FOXO4	forkhead box O4	X-linked	129
FSHB	follicle stimulating hormone subunit beta	AR	130
FSHR	follicle stimulating hormone receptor	AR	102,131,132
GALT	galactose-1-phosphate uridylyltransferase	AR	133
GDF9	growth differentiation factor 9	AD	85,134-137
GGPS1	geranylgeranyl diphosphate synthase 1	AR	138
HARS2	histidyl-tRNA synthetase 2, mitochondrial	AR	139
HFM1	helicase for meiosis 1	AR	72,140-142
HROB	homologous recombination factor with OB-fold	AR	142
HSD17B4	hydroxysteroid 17-beta dehydrogenase 4	AR	143,144
HSF2BP	heat shock transcription factor 2 binding protein	AR	145,146
IGSF10	immunoglobulin superfamily member 10	AR	147

KASH5	KASH domain containing 5	AR	148
KHDRBS1	KH RNA binding domain containing, signal transduction associated 1	AD	149
LARS2	leucyl-tRNA synthetase 2, mitochondrial	AR	150
LHCGR	luteinizing hormone/choriogonadotropin receptor	AR	151,152
LHX8	LIM homeobox 8	AD	80
LMNA	lamin A/C	AD	153
LRPPRC	leucine rich pentatricopeptide repeat containing	AR	72
MCM8	minichromosome maintenance 8 homologous recombination repair factor	AR	142,154-162
MCM9	minichromosome maintenance 9 homologous recombination repair factor	AR	72,157,162-169
MEIOB	meiosis specific with OB-fold	AR	170
MRPS22	mitochondrial ribosomal protein S22	AR	147,171
MSH4	mutS homolog 4	AR	172,173
MSH5	mutS homolog 5	AR	102,174
NANOS3	nanos C2HC-type zinc finger 3	AD	175,176
NBN	nibrin	AR	177
NHEJ1	non-homologous end joining factor 1	AD	178
NOBOX	NOBOX oogenesis homeobox	AD	80,168,179-184
NOG	noggin	AD	185

NOTCH2	notch receptor 2	AD	70,186,187
NR5A1	nuclear receptor subfamily 5 group A member 1	AD	102,167,188-200
NUP107	nucleoporin 107	AR	102,142,201,202
PGRMC1	progesterone receptor membrane component 1	X-linked	203
PMM2	phosphomannomutase 2	AR	165,204,205
POF1B	POF1B actin binding protein	X-linked	102,142,196,206
POLG	DNA polymerase gamma, catalytic subunit	AD	169,207-213
POLR2C	RNA polymerase II subunit C	AD	169
POLR3B	RNA polymerase III subunit B	AR	214
POLR3H	RNA polymerase III subunit H	AR	215
POU5F1	POU class 5 homeobox 1	AD	216
PRDM9	PR/SET domain 9	AD	68
PSMC3IP	PSMC3 interacting protein	AR	99,165,217-219
RAD51	RAD51 recombinase	AD	220
RCBTB1	RCC1 and BTB domain containing protein 1	AR	221
REC8	REC8 meiotic recombination protein	AD	80,102,142
RECQL4	RecQ like helicase 4	AR	222
SALL4	spalt like transcription factor 4	AD	223,224

SGO2	shugoshin 2	AR	225
SMC1B	structural maintenance of chromosomes 1B	AD	80
SOHLH1	spermatogenesis and oogenesis specific basic helix-loop-helix 1	AR	226
SOHLH2	spermatogenesis and oogenesis specific basic helix-loop-helix 2	AD	227
SPATA22	spermatogenesis associated 22	AR	228
SPIDR	scaffold protein involved in DNA repair	AR	102,229,230
STAG3	stromal antigen 3	AR	102,142,231-240
SYCE1	synaptonemal complex central element protein 1	AR	95,102,241
SYCP2L	synaptonemal complex protein 2 like	AR	242
TP63	tumor protein p63	AD	243,244
TRIM37	tripartite motif containing 37	AR	245
TWNK	twinkle mtDNA helicase	AR	246-251
WDR62	WD repeat domain 62	AD	252
WRN	WRN RecQ like helicase	AR	253
WT1	WT1 Transcription Factor	AD	254,255
YTHDC2	YTH domain containing 2	AR	256
ZSWIM7	zinc finger SWIM-type containing 7	AR	257,258

Decoding the complexity of genomic factors involved in female reproductive ageing.

Understanding the genetic architecture involved in female reproductive ageing has been a crucial area of research in recent years. Scientists have undertaken extensive studies to unravel the intricate genetic mechanisms underlying the process of age-related decline in female fertility and menopause. One extensively studied approach is the genome-wide association study (GWAS), which focuses on low frequencies and common genetic variations and has successfully identified numerous genetic variants associated with age at menopause (Figure 2).

In a recent publication in Nature by Ruth et al. (2021), GWAS involving a large cohort of more than 200,000 women revealed the presence of 290 common genetic variants that collectively account for 10% to 12% of the variation in age at natural menopause (ANM)[38]. These variants were found to have effects ranging from a difference of approximately 3.5 weeks to 1.3 years in ANM [38]. The researchers investigated the genetic factors underlying menopause timing and identified several biological mechanisms associated with this process. Through their study, they found that genetic variants associated with DNA repair mechanisms, estrogen signalling, and follicle development played significant roles in determining the timing of menopause[38]. These findings provide valuable insights into the intricate biological pathways involved in menopause timing and deepen our understanding of the underlying genetic mechanisms in this complex process.

Notwithstanding the significant advancements achieved through GWAS, it is imperative to recognise their primary focus on the assessment of common genetic variants (with an allele frequency exceeding 0.1%) utilising genome-wide arrays. This approach, however, bears limitations in its ability to encapsulate the entirety of genetic diversity. Array-based investigations are intrinsically constrained by their predetermined design, thereby restricting their capacity to only detect previously identified variants that are intentionally included on the array[259-261]. Conversely, the utilisation of sequencing data introduces a markedly broader spectrum of genetic exploration. This technology facilitates the identification of an extensive repertoire of genetic variants, spanning single nucleotide polymorphisms (SNPs), insertions, deletions, and

even structural variations that can exert profound influences on the genome[262,263]. This heightened versatility in variant detection underscores the potential for a more holistic comprehension of the genetic landscape in contrast to the constrained scope of array-based studies.

For a more comprehensive elucidation of the loci associated with diseases, an enhanced comprehension of disease pathogenicity, and a deeper insight into disease heritability, the integration of sequencing data emerges as an indispensable facet. Sequencing data provides a pivotal avenue for the exploration of rare and very rare variants, thus augmenting the breadth and depth of our understanding of the genetic underpinnings of complex diseases[264,265]. By studying rare variants in large cohorts, there is potential to identify novel genes and variants associated with the trait of interest[266]. Therefore, a complementary approach that combines exome sequencing with array-based GWAS holds promise for a biological explanation of genetic architecture involved in female reproductive lifespan[267] (Figure 2). The figure depicting the genetic architecture of female reproductive ageing and the effect sizes of variants on early menopause illustrates a compelling opportunity to uncover novel genes and variants associated with this trait.

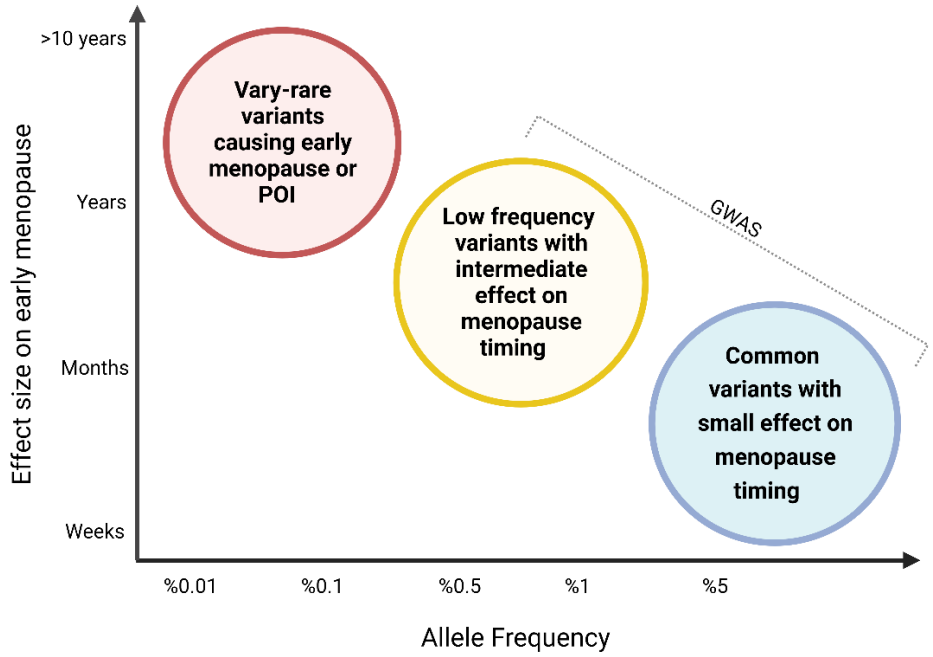


***Figure 2-1*: Illustrates the genetic architecture of early menopause.**

Population sequencing is the "gold standard" for genomic analysis.
Population sequencing is considered the gold standard for genomic analysis because it provides a comprehensive view of genetic variation in a given population[268,269]. This approach involves sequencing the protein-coding region of it in a population of individuals, allowing for a complete assessment of genetic variation and its impact on human health and disease.

Population sequencing data offers several advantages over targeted sequencing or array data. Firstly, it provides a comprehensive picture of genetic variation by enabling the identification of both common and rare variants across the entire genome[269]. This allows researchers to capture a more complete spectrum of genetic diversity and understand the full range of genetic contributions to complex traits and diseases[270,271]. In contrast, targeted sequencing focuses on specific genomic regions of interest, limiting the scope of genetic variation analysis.

Furthermore, population sequencing data enhances our understanding of population genetics and evolutionary patterns[272]. Researchers can investigate population structure, migration patterns, and genetic adaptations by studying genetic variation across diverse populations. This broader perspective provides insights into human populations' demographic history and evolutionary dynamics, shedding light on the genetic factors underlying population-specific traits and disease susceptibilities[273].

Recent advances in sequencing technologies, such as whole exome sequencing (WES), have made population sequencing more accessible and affordable. For example, the 1000 Genomes Project was a large-scale effort to sequence the genomes of over 2,500 individuals from diverse populations worldwide, providing a comprehensive reference dataset for human genetic variation[269]. The UK Biobank is another large-scale population sequencing effort, involving the sequencing of over 500,000 individuals to study the genetic basis of common diseases[274].

Comparing clinical and population sequencing data: Advantages of studying healthy individuals
Population sequencing data from healthy individuals provides a unique and valuable resource for the study of human genetics. Unlike traditional

approaches such as clinical and family studies, population sequencing data can reveal the full spectrum of genetic variation in a diverse and representative sample of individuals[275]. This can provide insights into the genetic basis of complex traits and diseases that may not be apparent in smaller, more homogeneous study populations.

Population sequencing studies involve the analysis of genetic data from large population-based cohorts, typically without a known family history of disease. This approach makes it possible to detect new genomic variations that may be linked to disease susceptibility among the general population. Furthermore, unlike clinical and family studies, which may be affected by selection bias, population genomic studies aim to be more unbiased in their representation of the overall population[264,276]. This reduces potential confounding factors and improves the reliability of results. Additionally, it enables the validation of previously reported variants for their actual impact on disease development. Furthermore, the use of healthy individuals as controls in these studies can increase the statistical power to detect genetic associations with disease.

UK Biobank is the largest population sequencing databases.

The UK Biobank was established in 2006, and it is a large-scale population sequencing effort that has become one of the largest databases of genetic and health data in the world. This project involves the collection of biological samples, health information, and lifestyle data from over 500,000 individuals aged between 40 and 69 years in the United Kingdom[274]. These participants have consented to the use of their data for research purposes, including genetic analysis, which has resulted in the creation of a comprehensive resource for studying the genetic basis of common diseases.

The UK Biobank has been instrumental in identifying genetic variants associated with a wide range of diseases, including cardiovascular disease, cancer, endometriosis, and polycystic ovary syndrome (PCOS). This resource has enabled researchers to investigate the genetic basis of these diseases in large, well-characterised populations, which can help to identify novel targets for drug development and personalized treatment strategies. Furthermore, the UK Biobank is a publicly accessible resource, and data access is granted to researchers worldwide through a formal application process. This has facilitated collaboration and enabled researchers to undertake large-scale genomic

studies across diverse populations and diseases. As such, the UK Biobank serves as a powerful tool for accelerating medical research and improving human health.

Summary

The importance of age at menopause as an indicator of female reproductive lifespan is emphasized, along with an examination of various factors that impact the timing of menopause. Menopause signifies the conclusion of a woman's natural fertility, yet natural infertility may arise a decade before menopause sets in. The timing of menopause is influenced by genetics, lifestyle choices, and reproductive factors. Genetic factors, such as chromosomal abnormalities and the fragile X premutation, have been linked to alterations in menopause timing. GWAS have identified common genetic variations associated with age at menopause and reproductive ageing.

To further advance our understanding of the genetic basis of complex traits like female reproductive lifespan, comprehensive genomic analysis is vital, with a specific focus on rare and very rare genomic variants. Examination of population sequencing data, particularly from healthy individuals, allows the identification and capture of variants related to menopause timing, offering valuable insights into the underlying genetic mechanisms. Additionally, validating previously reported variants as monogenic causal genes for POI is essential to elucidate the causality of this condition, primarily studied through family and clinical data from limited cases.

Combining findings from both common and rare genomic variations derived from population sequencing data enables a more comprehensive understanding of how genomic factors influence female reproductive ageing. Ultimately, this knowledge facilitates more accurate predictions of young women's reproductive lifespan from birth. My PhD projects aimed to contribute to a deeper comprehension of the genetic determinants of female reproductive ageing, benefiting future healthcare and personalised reproductive planning.

References

1 Variations in reproductive events across life: a pooled analysis of data from 505 147 women across 10 countries. *Hum Reprod* **34**, 881-893 (2019). https://doi.org:10.1093/humrep/dez015

2 Murray, A., Ennis, S., MacSwiney, F., Webb, J. & Morton, N. E. Reproductive and menstrual history of females with fragile X expansions. *Eur J Hum Genet* **8**, 247-252 (2000). https://doi.org:10.1038/sj.ejhg.5200451

3 Broekmans, F. J., Soules, M. R. & Fauser, B. C. Ovarian aging: mechanisms and clinical consequences. *Endocr Rev* **30**, 465-493 (2009). https://doi.org:10.1210/er.2009-0006

4 Wallace, W. H. & Kelsey, T. W. Human ovarian reserve from conception to the menopause. *PLoS One* **5**, e8772 (2010). https://doi.org:10.1371/journal.pone.0008772

5 Mandl, A. M. & Zuckerman, S. The relation of age to numbers of oocytes. *J Endocrinol* **7**, 190-193 (1951). https://doi.org:10.1677/joe.0.0070190

6 Allen, J. (Oxford, NY: Oxford University Press, 2010).

7 Gruhn, J. R. *et al.* Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science* **365**, 1466-1469 (2019). https://doi.org:10.1126/science.aav7321

8 Kerr, J. B. *et al.* Cisplatin-induced primordial follicle oocyte killing and loss of fertility are not prevented by imatinib. *Nat Med* **18**, 1170-1172; author reply 1172-1174 (2012). https://doi.org:10.1038/nm.2889

9 Reed, B. G. & Carr, B. R. in *Endotext* (eds K. R. Feingold *et al.*) (MDText.com, Inc.Copyright © 2000-2023, MDText.com, Inc., 2000).

10 Hirshfield, A. N. Development of follicles in the mammalian ovary. *Int Rev Cytol* **124**, 43-101 (1991). https://doi.org:10.1016/s0074-7696(08)61524-7

11 Faddy, M. J., Gosden, R. G., Gougeon, A., Richardson, S. J. & Nelson, J. F. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod* **7**, 1342-1346 (1992). https://doi.org:10.1093/oxfordjournals.humrep.a137570

12 Hansen, K. R. *et al.* A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod* **23**, 699-708 (2008). https://doi.org:10.1093/humrep/dem408

13 Perheentupa, A. & Huhtaniemi, I. Aging of the human ovary and testis. *Mol Cell Endocrinol* **299**, 2-13 (2009). https://doi.org:10.1016/j.mce.2008.11.004

14 Organization, W. H. (2023).

15 Sobotka, T., Skirbekk, V. & Philipov, D. Economic recession and fertility in the developed world. *Population and development review* **37**, 267-306 (2011).

16 Bunting, L. & Boivin, J. Decision-making about seeking medical advice in an internet sample of women trying to get pregnant. *Hum Reprod* **22**, 1662-1668 (2007). https://doi.org:10.1093/humrep/dem057

17 Schmidt, L., Sobotka, T., Bentzen, J. G. & Nyboe Andersen, A. Demographic and medical consequences of the postponement of parenthood. *Hum Reprod Update* **18**, 29-43 (2012). https://doi.org:10.1093/humupd/dmr040

18 Lundberg, S. & Rose, E. Child gender and the transition to marriage. *Demography* **40**, 333-349 (2003). https://doi.org:10.1353/dem.2003.0015

19 Wu, L. & Wolfe, B. *Out of wedlock: Causes and consequences of nonmarital fertility*. (Russell Sage Foundation, 2001).

20 Booth, A. & Crouter, A. C. *The new population problem: Why families in developed countries are shrinking and what it means*. (Psychology Press, 2005).

21 Kushnir, V. A., Barad, D. H., Albertini, D. F., Darmon, S. K. & Gleicher, N. Systematic review of worldwide trends in assisted reproductive technology 2004-2013. *Reprod Biol Endocrinol* **15**, 6 (2017). https://doi.org:10.1186/s12958-016-0225-2

22 Nargund, G. Declining birth rate in Developed Countries: A radical policy re-think is required. *Facts Views Vis Obgyn* **1**, 191-193 (2009).

23 Ombelet, W. Global access to infertility care in developing countries: a case of human rights, equity and social justice. *Facts Views Vis Obgyn* **3**, 257-266 (2011).

24 Miller, W. B., Severy, L. J. & Pasta, D. J. A framework for modelling fertility motivation in couples. *Popul Stud (Camb)* **58**, 193-205 (2004). https://doi.org:10.1080/0032472042000213712

25 Song, M. *et al.* Differences in Family Planning and Fertility Among Female and Male Gynecologic Oncologists. *Womens Health Rep (New Rochelle)* **2**, 78-84 (2021). https://doi.org:10.1089/whr.2020.0046

26 Singh, V., Rajpal, G., Majumdar, S., Mandal, S. & Sahu, R. Awareness, knowledge, and psycho-emotional aspect of infertility among women seeking fertility assistance at a tertiary care hospital in Chhattisgarh, a cross-sectional study. *J Educ Health Promot* **12**, 3 (2023). https://doi.org:10.4103/jehp.jehp_581_22

27 Mishra, G. D. *et al.* Early menarche, nulliparity and the risk for premature and early natural menopause. *Hum Reprod* **32**, 679-686 (2017). https://doi.org:10.1093/humrep/dew350

28 Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci Rep* **6**, 24710 (2016). https://doi.org:10.1038/srep24710

29 Christensen, K., Carlson, L. M. & Lehmann, G. M. The role of epidemiology studies in human health risk assessment of polychlorinated biphenyls. *Environ Res* **194**, 110662 (2021). https://doi.org:10.1016/j.envres.2020.110662

30 Perry, J. R. *et al.* DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. *Hum Mol Genet* **23**, 2490-2497 (2014). https://doi.org:10.1093/hmg/ddt620

31 Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* **44**, 260-268 (2012). https://doi.org:10.1038/ng.1051

32 He, C. *et al.* A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum Genet* **128**, 515-527 (2010). https://doi.org:10.1007/s00439-010-0878-4

33 Murabito, J. M., Yang, Q., Fox, C., Wilson, P. W. & Cupples, L. A. Heritability of age at natural menopause in the Framingham Heart Study. *J Clin Endocrinol Metab* **90**, 3427-3430 (2005). https://doi.org:10.1210/jc.2005-0181

34 Ruth, K. S. *et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393-397 (2021). https://doi.org:10.1038/s41586-021-03779-7

35 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. *J Clin Endocrinol Metab* **83**, 1875-1880 (1998). https://doi.org:10.1210/jcem.83.6.4890

36 Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97 (2014). https://doi.org:10.1038/nature13545

37 Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* **47**, 1294-1303 (2015). https://doi.org:10.1038/ng.3412

38 Stochholm, K., Juul, S., Juel, K., Naeraa, R. W. & Gravholt, C. H. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. *J Clin Endocrinol Metab* **91**, 3897-3902 (2006). https://doi.org:10.1210/jc.2006-0558

39 Jacobs, P. A. *et al.* Evidence for the existence of the human "super female". *Lancet* **2**, 423-425 (1959). https://doi.org:10.1016/s0140-6736(59)90415-5

40 Otter, M., Schrander-Stumpel, C. T. & Curfs, L. M. Triple X syndrome: a review of the literature. *Eur J Hum Genet* **18**, 265-271 (2010). https://doi.org:10.1038/ejhg.2009.109

41 Tassone, F. *et al.* Clinical involvement and protein expression in individuals with the FMR1 premutation. *Am J Med Genet* **91**, 144-152 (2000). https://doi.org:10.1002/(sici)1096-8628(20000313)91:2<144::aid-ajmg14>3.0.co;2-v

42 Pembrey, M., Winter, R., Davies, K., Opitz, J. M. & Reynolds, J. F. A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. *American journal of medical genetics* **21**, 709-717 (1985).

43 Ennis, S., Ward, D. & Murray, A. Nonlinear association between CGG repeat number and age of menopause in FMR1 premutation carriers. *Eur J Hum Genet* **14**, 253-255 (2006). https://doi.org:10.1038/sj.ejhg.5201510

44 Verkerk, A. J. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905-914 (1991).

45 Hagerman, P. J. & Hagerman, R. J. The Fragile-X Premutation: A Maturing Perspective. *The American Journal of Human Genetics* **74**, 805-816 (2004). https://doi.org:https://doi.org/10.1086/386296

46 Jacquemont, S. *et al.* Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet* **72**, 869-878 (2003). https://doi.org:10.1086/374321

47 Sherman, S., Pletcher, B. A. & Driscoll, D. A. Fragile X syndrome: diagnostic and carrier testing. *Genet Med* **7**, 584-587 (2005). https://doi.org:10.1097/01.gim.0000182468.22666.dd

48 Wesevich, V., Kellen, A. N. & Pal, L. Recent advances in understanding primary ovarian insufficiency. *F1000Res* **9** (2020). https://doi.org:10.12688/f1000research.26423.1

49 Golezar, S., Ramezani Tehrani, F., Khazaei, S., Ebadi, A. & Keshavarz, Z. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric* **22**, 403-411 (2019). https://doi.org:10.1080/13697137.2019.1574738

50 Rudnicka, E. *et al.* Premature ovarian insufficiency - aetiopathology, epidemiology, and diagnostic evaluation. *Prz Menopauzalny* **17**, 105-108 (2018). https://doi.org:10.5114/pm.2018.78550

51 Coulam, C. B., Adamson, S. C. & Annegers, J. F. Incidence of premature ovarian failure. *Obstet Gynecol* **67**, 604-606 (1986).

52 Harlow, B. L. & Signorello, L. B. Factors associated with early menopause. *Maturitas* **35**, 3-9 (2000). https://doi.org:10.1016/s0378-5122(00)00092-x

53 Goswami, D. & Conway, G. S. Premature ovarian failure. *Hum Reprod Update* **11**, 391-410 (2005). https://doi.org:10.1093/humupd/dmi012

54 Szeliga, A. *et al.* Autoimmune Diseases in Patients with Premature Ovarian Insufficiency-Our Current State of Knowledge. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms22052594

55 Fortuño, C. & Labarta, E. Genetics of primary ovarian insufficiency: a review. *J Assist Reprod Genet* **31**, 1573-1585 (2014). https://doi.org:10.1007/s10815-014-0342-9

56 Venturella, R. *et al.* The Genetics of Non-Syndromic Primary Ovarian Insufficiency: A Systematic Review. *Int J Fertil Steril* **13**, 161-168 (2019). https://doi.org:10.22074/ijfs.2019.5599

57 van Kasteren, Y. M. *et al.* Familial idiopathic premature ovarian failure: an overrated and underestimated genetic disease? *Hum Reprod* **14**, 2455-2459 (1999). https://doi.org:10.1093/humrep/14.10.2455

58 Qin, Y. *et al.* ESR1, HK3 and BRSK1 gene variants are associated with both age at natural menopause and premature ovarian failure. *Orphanet J Rare Dis* **7**, 5 (2012). https://doi.org:10.1186/1750-1172-7-5

59 Zhou, Y. *et al.* Novel alanyl-tRNA synthetase 2 (AARS2) homozygous mutation in a consanguineous Chinese family with premature ovarian insufficiency. *Fertil Steril* **112**, 569-576.e562 (2019). https://doi.org:10.1016/j.fertnstert.2019.05.005

60 Dallabona, C. *et al.* Novel (ovario) leukodystrophy related to AARS2 mutations. *Neurology* **82**, 2063-2071 (2014). https://doi.org:10.1212/wnl.0000000000000497

61 Zheng, W. B. *et al.* A novel variant in AIRE causing a rare, non-classical autoimmune polyendocrine syndrome type 1. *Mol Med Rep* **22**, 1285-1294 (2020). https://doi.org:10.3892/mmr.2020.11227

62 Alavi, A. *et al.* Mutation in ALOX12B likely cause of POI and also ichthyosis in a large Iranian pedigree. *Mol Genet Genomics* **295**, 1039-1053 (2020). https://doi.org:10.1007/s00438-020-01663-z

63 Alvaro Mercadal, B. *et al.* AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Hum Reprod* **30**, 1196-1202 (2015). https://doi.org:10.1093/humrep/dev042

64 Wang, Y. *et al.* Pathogenic variants of meiotic double strand break (DSB) formation genes PRDM9 and ANKRD31 in premature ovarian insufficiency. *Genet Med* **23**, 2309-2315 (2021). https://doi.org:10.1038/s41436-021-01266-y

65 Chatterjee, S. *et al.* Longer CAG repeat length in the androgen receptor gene is associated with premature ovarian failure. *Hum Reprod* **24**, 3230-3235 (2009). https://doi.org:10.1093/humrep/dep296

66 Patiño, L. C. *et al.* New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Hum Reprod* **32**, 1512-1520 (2017). https://doi.org:10.1093/humrep/dex089

67 Delcour, C. *et al.* ATG7 and ATG9A loss-of-function variants trigger autophagy impairment and ovarian failure. *Genet Med* **21**, 930-938 (2019). https://doi.org:10.1038/s41436-018-0287-y

68 Liu, H. *et al.* Whole-exome sequencing in patients with premature ovarian insufficiency: early detection and early intervention. *J Ovarian Res* **13**, 114 (2020). https://doi.org:10.1186/s13048-020-00716-6

69 Ding, X. *et al.* DNA double-strand break genetic variants in patients with premature ovarian insufficiency. *Journal of Ovarian Research* **16**, 135 (2023). https://doi.org:10.1186/s13048-023-01221-2

70 Arora, H. *et al.* Bloom syndrome. *Int J Dermatol* **53**, 798-802 (2014). https://doi.org:10.1111/ijd.12408

71 Di Pasquale, E., Beck-Peccoz, P. & Persani, L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. *Am J Hum Genet* **75**, 106-111 (2004). https://doi.org:10.1086/422103

72 Dixit, H. *et al.* Missense mutations in the BMP15 gene are associated with ovarian failure. *Hum Genet* **119**, 408-415 (2006). https://doi.org:10.1007/s00439-006-0150-0

73 Gónzalez, A. *et al.* Association of genetic markers within the BMP15 gene with anovulation and infertility in women with polycystic ovary syndrome. *Fertil Steril* **90**, 447-449 (2008). https://doi.org:10.1016/j.fertnstert.2007.06.083

74 Fonseca, D. J. *et al.* BMP15 c.-9C>G promoter sequence variant may contribute to the cause of non-syndromic premature ovarian failure. *Reprod Biomed Online* **29**, 627-633 (2014). https://doi.org:10.1016/j.rbmo.2014.07.018

75 Rossetti, R. *et al.* BMP15 mutations associated with primary ovarian insufficiency cause a defective production of bioactive protein. *Hum Mutat* **30**, 804-810 (2009). https://doi.org:10.1002/humu.20961

76 Bouilly, J. *et al.* Identification of Multiple Gene Mutations Accounts for a new Genetic Architecture of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **101**, 4541-4550 (2016). https://doi.org:10.1210/jc.2016-2152

77 Di Pasquale, E. *et al.* Identification of new variants of human BMP15 gene in a large cohort of women with premature ovarian failure. *J Clin Endocrinol Metab* **91**, 1976-1979 (2006). https://doi.org:10.1210/jc.2005-2650

78 Tiotiu, D. *et al.* Variants of the BMP15 gene in a cohort of patients with premature ovarian failure. *Hum Reprod* **25**, 1581-1587 (2010). https://doi.org:10.1093/humrep/deq073

79 Laissue, P. *et al.* Mutations and sequence variants in GDF9 and BMP15 in patients with premature ovarian failure. *Eur J Endocrinol* **154**, 739-744 (2006). https://doi.org:10.1530/eje.1.02135

80 Patiño, L. C. *et al.* BMP15 Mutations Associated With Primary Ovarian Insufficiency Reduce Expression, Activity, or Synergy With GDF9. *J Clin Endocrinol Metab* **102**, 1009-1019 (2017). https://doi.org:10.1210/jc.2016-3503

81 Wang, B. *et al.* Analyses of growth differentiation factor 9 (GDF9) and bone morphogenetic protein 15 (BMP15) mutation in Chinese women with premature ovarian failure. *Clin Endocrinol (Oxf)* **72**, 135-136 (2010). https://doi.org:10.1111/j.1365-2265.2009.03613.x

82 Santos, M. *et al.* Association of BMP15 and GDF9 variants to premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 2163-2169 (2019). https://doi.org:10.1007/s10815-019-01548-0

83 Zhang, W. *et al.* A novel homozygous mutation of bone morphogenetic protein 15 identified in a consanguineous marriage family with primary ovarian insufficiency. *Reprod Biomed Online* **36**, 206-209 (2018). https://doi.org:10.1016/j.rbmo.2017.10.104

84 Mayer, A., Fouquet, B., Pugeat, M. & Misrahi, M. BMP15 "knockout-like" effect in familial premature ovarian insufficiency with persistent ovarian reserve. *Clin Genet* **92**, 208-212 (2017). https://doi.org:10.1111/cge.12970

85 Rossetti, R. *et al.* Fundamental role of BMP15 in human ovarian folliculogenesis revealed by null and missense mutations associated with primary ovarian insufficiency. *Hum Mutat* **41**, 983-997 (2020). https://doi.org:10.1002/humu.23988

86 Renault, L. *et al.* BMPR1A and BMPR1B Missense Mutations Cause Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgz226

87 Zhang, D. *et al.* Basonuclin 1 deficiency is a cause of primary ovarian insufficiency. *Hum Mol Genet* **27**, 3787-3800 (2018). https://doi.org:10.1093/hmg/ddy261

88 Turchetti, D., Zuntini, R., Tricarico, R. & Bellacosa, A. BRCA2 in Ovarian Development and Function. *N Engl J Med* **380**, 1086-1087 (2019). https://doi.org:10.1056/NEJMc1813800

89 Chen, Q. *et al.* Rare deleterious BUB1B variants induce premature ovarian insufficiency and early menopause. *Hum Mol Genet* **29**, 2698-2707 (2020). https://doi.org:10.1093/hmg/ddaa153

90 Fan, S. *et al.* Homozygous mutations in C14orf39/SIX6OS1 cause non-obstructive azoospermia and premature ovarian insufficiency in humans. *Am J Hum Genet* **108**, 324-336 (2021). https://doi.org:10.1016/j.ajhg.2021.01.010

91 Hou, D. *et al.* Variations of C14ORF39 and SYCE1 Identified in Idiopathic Premature Ovarian Insufficiency and Nonobstructive Azoospermia. *J Clin Endocrinol Metab* **107**, 724-734 (2022). https://doi.org:10.1210/clinem/dgab777

92 Chasseloup, F. *et al.* Germline CDKN1B Loss-of-Function Variants Cause Pediatric Cushing's Disease With or Without an MEN4 Phenotype. *J Clin Endocrinol Metab* **105**, 1983-2005 (2020). https://doi.org:10.1210/clinem/dgaa160

93 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 (2015). https://doi.org:10.1016/j.jns.2015.04.038

94 Dursun, F. *et al.* A Novel Missense Mutation in the CLPP Gene Causing Perrault Syndrome Type 3 in a Turkish Family. *J Clin Res Pediatr Endocrinol* **8**, 472-477 (2016). https://doi.org:10.4274/jcrpe.2717

95 Al-Agha, A. E. *et al.* Primary Ovarian Insufficiency and Azoospermia in Carriers of a Homozygous PSMC3IP Stop Gain Mutation. *J Clin Endocrinol Metab* **103**, 555-563 (2018). https://doi.org:10.1210/jc.2017-01966

96 Tšuiko, O. *et al.* Copy number variation analysis detects novel candidate genes involved in follicular growth and oocyte maturation in a cohort of premature ovarian failure cases. *Hum Reprod* **31**, 1913-1925 (2016). https://doi.org:10.1093/humrep/dew142

97 Jiao, W., Zhao, S., Liu, R., Guo, T. & Qin, Y. CPEB1 deletion is not a common explanation for premature ovarian insufficiency in a Chinese cohort. *J Ovarian Res* **13**, 49 (2020). https://doi.org:10.1186/s13048-020-00630-x

98 Rouen, A. *et al.* Whole exome sequencing in a cohort of familial premature ovarian insufficiency cases reveals a broad array of pathogenic or likely pathogenic variants in 50% of families. *Fertil Steril* **117**, 843-853 (2022). https://doi.org:10.1016/j.fertnstert.2021.12.023

99 Carvalho, L. C. *et al.* Clinical, hormonal, ovarian, and genetic aspects of 46,XX patients with congenital adrenal hyperplasia due to CYP17A1 defects. *Fertil Steril* **105**, 1612-1619 (2016). https://doi.org:10.1016/j.fertnstert.2016.02.008

100 Bione, S. *et al.* Mutation analysis of two candidate genes for premature ovarian failure, DACH2 and POF1B. *Hum Reprod* **19**, 2759-2766 (2004). https://doi.org:10.1093/humrep/deh502

101 Genesio, R. *et al.* Short stature and primary ovarian insufficiency possibly due to chromosomal position effect in a balanced X;1 translocation. *Mol Cytogenet* **8**, 50 (2015). https://doi.org:10.1186/s13039-015-0154-3

102 Rizzolio, F. *et al.* Epigenetic control of the critical region for premature ovarian failure on autosomal genes translocated to the X chromosome: a hypothesis. *Hum Genet* **121**, 441-450 (2007). https://doi.org:10.1007/s00439-007-0329-z

103 Mandon-Pépin, B. *et al.* Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur J Endocrinol* **158**, 107-115 (2008). https://doi.org:10.1530/eje-07-0400

104 He, W. B. *et al.* DMC1 mutation that causes human non-obstructive azoospermia and premature ovarian insufficiency identified by whole-exome sequencing. *J Med Genet* **55**, 198-204 (2018). https://doi.org:10.1136/jmedgenet-2017-104992

105 Fogli, A. *et al.* Ovarian failure related to eukaryotic initiation factor 2B mutations. *Am J Hum Genet* **72**, 1544-1550 (2003). https://doi.org:10.1086/375404

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

107 Chatzispyrou, I. A. *et al.* A homozygous missense mutation in ERAL1, encoding a mitochondrial rRNA chaperone, causes Perrault syndrome. *Hum Mol Genet* **26**, 2541-2550 (2017). https://doi.org:10.1093/hmg/ddx152

108 Qin, Y. *et al.* CSB-PGBD3 Mutations Cause Premature Ovarian Failure. *PLoS Genet* **11**, e1005419 (2015). https://doi.org:10.1371/journal.pgen.1005419

109 Yang, X. *et al.* Rare variants in FANCA induce premature ovarian insufficiency. *Hum Genet* **138**, 1227-1236 (2019). https://doi.org:10.1007/s00439-019-02059-9

110 França, M. M. & Mendonca, B. B. Genetics of ovarian insufficiency and defects of folliculogenesis. *Best Pract Res Clin Endocrinol Metab* **36**, 101594 (2022). https://doi.org:10.1016/j.beem.2021.101594

111 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* **37**, 609-635 (2016). https://doi.org:10.1210/er.2016-1047

112 Yang, Y. *et al.* FANCL gene mutations in premature ovarian insufficiency. *Hum Mutat* **41**, 1033-1041 (2020). https://doi.org:10.1002/humu.23997

113 Fouquet, B. *et al.* A homozygous FANCM mutation underlies a familial case of non-syndromic primary ovarian insufficiency. *Elife* **6** (2017). https://doi.org:10.7554/eLife.30490

114 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 (2020). https://doi.org:10.1016/j.maturitas.2020.06.004

115 Zhao, H. *et al.* Transcription factor FIGLA is mutated in patients with premature ovarian failure. *Am J Hum Genet* **82**, 1342-1348 (2008). https://doi.org:10.1016/j.ajhg.2008.04.018

116 Chen, B. *et al.* Consanguineous familial study revealed biallelic FIGLA mutation associated with premature ovarian insufficiency. *J Ovarian Res* **11**, 48 (2018). https://doi.org:10.1186/s13048-018-0413-0

117 Mei, L. *et al.* Mutations in FIGLA Associated With Premature Ovarian Insufficiency in a Chinese Population. *Front Med (Lausanne)* **8**, 714306 (2021). https://doi.org:10.3389/fmed.2021.714306

118 Li, J. *et al.* Heterozygous FMN2 missense variant found in a family case of premature ovarian insufficiency. *J Ovarian Res* **15**, 31 (2022). https://doi.org:10.1186/s13048-022-00960-y

119 Yoo, S., Yoon, J. Y., Keum, C. & Cheon, C. K. The first case of novel variants of the FSHR mutation causing primary amenorrhea in 2 siblings in Korea. *Ann Pediatr Endocrinol Metab* **28**, 54-60 (2023). https://doi.org:10.6065/apem.2142116.058

120 Crisponi, L. *et al.* The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* **27**, 159-166 (2001). https://doi.org:10.1038/84781

121 Laissue, P. *et al.* Functional evidence implicating FOXL2 in non-syndromic premature ovarian failure and in the regulation of the transcription factor OSR2. *J Med Genet* **46**, 455-457 (2009). https://doi.org:10.1136/jmg.2008.065086

122 Yang, X. W. *et al.* Novel FOXL2 mutations cause blepharophimosis-ptosis-epicanthus inversus syndrome with premature ovarian insufficiency. *Mol Genet Genomic Med* **6**, 261-267 (2018). https://doi.org:10.1002/mgg3.366

123 Nomura, R. *et al.* Nr5a1 suppression during the murine fetal period optimizes ovarian development by fine-tuning Notch signaling. *J Cell Sci* **132** (2019). https://doi.org:10.1242/jcs.223768

124 Grzechocińska, B., Warzecha, D., Wypchło, M., Ploski, R. & Wielgoś, M. Premature ovarian insufficiency as a variable feature of blepharophimosis, ptosis, and epicanthus inversus syndrome associated with c.223C > T p.(Leu75Phe) FOXL2 mutation: a case report. *BMC Med Genet* **20**, 132 (2019). https://doi.org:10.1186/s12881-019-0865-0

125 Fonseca, D. J. *et al.* Screening for mutations of the FOXO4 gene in premature ovarian failure patients. *Reprod Biomed Online* **24**, 339-341 (2012). https://doi.org:10.1016/j.rbmo.2011.11.017

126 Eskenazi, S. *et al.* Next Generation Sequencing Should Be Proposed to Every Woman With "Idiopathic" Primary Ovarian Insufficiency. *J Endocr Soc* **5**, bvab032 (2021). https://doi.org:10.1210/jendso/bvab032

127 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. *J Clin Endocrinol Metab* **88**, 3491-3498 (2003). https://doi.org:10.1210/jc.2003-030217

128 Liu, H. *et al.* A novel homozygous mutation in the FSHR gene is causative for primary ovarian insufficiency. *Fertil Steril* **108**, 1050-1055.e1052 (2017). https://doi.org:10.1016/j.fertnstert.2017.09.010

129 Fridovich-Keil, J. L. *et al.* Ovarian function in girls and women with GALT-deficiency galactosemia. *J Inherit Metab Dis* **34**, 357-366 (2011). https://doi.org:10.1007/s10545-010-9221-4

130 Dixit, H. *et al.* Mutational screening of the coding region of growth differentiation factor 9 gene in Indian women with ovarian failure. *Menopause* **12**, 749-754 (2005). https://doi.org:10.1097/01.gme.0000184424.96437.7a

131 Wang, T. T. *et al.* Identification of a mutation in GDF9 as a novel cause of diminished ovarian reserve in young women. *Hum Reprod* **28**, 2473-2481 (2013). https://doi.org:10.1093/humrep/det291

132 Verma, K. P. *et al.* A homozygous truncating variant in GDF9 in siblings with primary ovarian insufficiency. *J Assist Reprod Genet* **38**, 1539-1543 (2021). https://doi.org:10.1007/s10815-021-02144-x

133 Marinakis, N. M. *et al.* Ovarian insufficiency and secondary amenorrhea in a patient with a novel variant within GDF9 gene. *Menopause* **29**, 491-495 (2022). https://doi.org:10.1097/gme.0000000000001928

134 Foley, A. R. *et al.* GGPS1 Mutations Cause Muscular Dystrophy/Hearing Loss/Ovarian Insufficiency Syndrome. *Ann Neurol* **88**, 332-347 (2020). https://doi.org:10.1002/ana.25772

135 Pierce, S. B. *et al.* Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. *Proc Natl Acad Sci U S A* **108**, 6543-6548 (2011). https://doi.org:10.1073/pnas.1103471108

136 Wang, J., Zhang, W., Jiang, H. & Wu, B. L. Mutations in HFM1 in recessive primary ovarian insufficiency. *N Engl J Med* **370**, 972-974 (2014). https://doi.org:10.1056/NEJMc1310150

137 Pu, D. *et al.* Association analysis between HFM1 variation and primary ovarian insufficiency in Chinese women. *Clin Genet* **89**, 597-602 (2016). https://doi.org:10.1111/cge.12718

138 Tucker, E. J. *et al.* Meiotic genes in premature ovarian insufficiency: variants in HROB and REC8 as likely genetic causes. *Eur J Hum Genet* **30**, 219-228 (2022). https://doi.org:10.1038/s41431-021-00977-9

139 Pierce, S. B. *et al.* Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet* **87**, 282-288 (2010). https://doi.org:10.1016/j.ajhg.2010.07.007

140 Chen, K. *et al.* A homozygous missense variant in HSD17B4 identified in a consanguineous Chinese Han family with type II Perrault syndrome. *BMC Med Genet* **18**, 91 (2017). https://doi.org:10.1186/s12881-017-0453-0

141 Felipe-Medina, N. *et al.* A missense in HSF2BP causing primary ovarian insufficiency affects meiotic recombination by its novel interactor C19ORF57/BRME1. *Elife* **9** (2020). https://doi.org:10.7554/eLife.56996

142 Li, S. *et al.* Pathogenic Variations of Homologous Recombination Gene HSF2BP Identified in Sporadic Patients With Premature Ovarian Insufficiency. *Front Cell Dev Biol* **9**, 768123 (2021). https://doi.org:10.3389/fcell.2021.768123

143 Jolly, A. *et al.* Exome Sequencing of a Primary Ovarian Insufficiency Cohort Reveals Common Molecular Etiologies for a Spectrum of Disease. *J Clin Endocrinol Metab* **104**, 3049-3067 (2019). https://doi.org:10.1210/jc.2019-00248

144 Wu, H. *et al.* Homozygous missense mutation in CCDC155 disrupts the transmembrane distribution of CCDC155 and SUN1, resulting in non-obstructive azoospermia and premature ovarian insufficiency in humans. *Hum Genet* **141**, 1795-1809 (2022). https://doi.org:10.1007/s00439-022-02459-4

145 Wang, B. *et al.* Sequence variants of KHDRBS1 as high penetrance susceptibility risks for primary ovarian insufficiency by mis-regulating mRNA

alternative splicing. *Hum Reprod* **32**, 2138-2146 (2017). https://doi.org:10.1093/humrep/dex263

146 Pierce, S. B. *et al.* Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. *Am J Hum Genet* **92**, 614-620 (2013). https://doi.org:10.1016/j.ajhg.2013.03.007

147 Latronico, A. C. *et al.* Brief report: testicular and ovarian resistance to luteinizing hormone caused by inactivating mutations of the luteinizing hormone-receptor gene. *N Engl J Med* **334**, 507-512 (1996). https://doi.org:10.1056/nejm199602223340805

148 Zielen, A. C. *et al.* A novel homozygous frame-shift variant in the LHCGR gene is associated with primary ovarian insufficiency in a Pakistani family. *Clin Genet* **94**, 396-397 (2018). https://doi.org:10.1111/cge.13406

149 McPherson, E. *et al.* Ovarian failure and dilated cardiomyopathy due to a novel lamin mutation. *Am J Med Genet A* **149a**, 567-572 (2009). https://doi.org:10.1002/ajmg.a.32627

150 AlAsiri, S. *et al.* Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J Clin Invest* **125**, 258-262 (2015). https://doi.org:10.1172/jci78473

151 Tenenbaum-Rakover, Y. *et al.* Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. *J Med Genet* **52**, 391-399 (2015). https://doi.org:10.1136/jmedgenet-2014-102921

152 Dou, X. *et al.* Minichromosome maintenance complex component 8 mutations cause primary ovarian insufficiency. *Fertil Steril* **106**, 1485-1489.e1482 (2016). https://doi.org:10.1016/j.fertnstert.2016.08.018

153 Desai, S. *et al.* MCM8 and MCM9 Nucleotide Variants in Women With Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **102**, 576-582 (2017). https://doi.org:10.1210/jc.2016-2565

154 Bouali, N. *et al.* New MCM8 mutation associated with premature ovarian insufficiency and chromosomal instability in a highly consanguineous Tunisian family. *Fertil Steril* **108**, 694-702 (2017). https://doi.org:10.1016/j.fertnstert.2017.07.015

155 Zhang, Y. X. *et al.* Novel loss-of-function mutation in MCM8 causes premature ovarian insufficiency. *Mol Genet Genomic Med* **8**, e1165 (2020). https://doi.org:10.1002/mgg3.1165

156 Heddar, A., Beckers, D., Fouquet, B., Roland, D. & Misrahi, M. A Novel Phenotype Combining Primary Ovarian Insufficiency Growth Retardation and

Pilomatricomas With MCM8 Mutation. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgaa155

157 Wang, F., Guo, S. & Li, P. Two novel mutations in the MCM8 gene shared by two Chinese siblings with primary ovarian insufficiency and short stature. *Mol Genet Genomic Med* **8**, e1396 (2020). https://doi.org:10.1002/mgg3.1396

158 Jin, H. *et al.* Identification of potential causal variants for premature ovarian failure by whole exome sequencing. *BMC Med Genomics* **13**, 159 (2020). https://doi.org:10.1186/s12920-020-00813-x

159 Wood-Trageser, M. A. *et al.* MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am J Hum Genet* **95**, 754-762 (2014). https://doi.org:10.1016/j.ajhg.2014.11.002

160 Fauchereau, F. *et al.* A non-sense MCM9 mutation in a familial case of primary ovarian insufficiency. *Clin Genet* **89**, 603-607 (2016). https://doi.org:10.1111/cge.12736

161 Yang, X. *et al.* Gene variants identified by whole-exome sequencing in 33 French women with premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 39-45 (2019). https://doi.org:10.1007/s10815-018-1349-4

162 Guo, T. *et al.* Novel pathogenic mutations in minichromosome maintenance complex component 9 (MCM9) responsible for premature ovarian insufficiency. *Fertil Steril* **113**, 845-852 (2020). https://doi.org:10.1016/j.fertnstert.2019.11.015

163 Alvarez-Mora, M. I. *et al.* An exome-wide exploration of cases of primary ovarian insufficiency uncovers novel sequence variants and candidate genes. *Clin Genet* **98**, 293-298 (2020). https://doi.org:10.1111/cge.13803

164 França, M. M. *et al.* Screening of targeted panel genes in Brazilian patients with primary ovarian insufficiency. *PLoS One* **15**, e0240795 (2020). https://doi.org:10.1371/journal.pone.0240795

165 Shen, J. *et al.* Genetic etiologic analysis in 74 Chinese Han women with idiopathic premature ovarian insufficiency by combined molecular genetic testing. *J Assist Reprod Genet* **38**, 965-978 (2021). https://doi.org:10.1007/s10815-021-02083-7

166 Caburet, S. *et al.* A truncating MEIOB mutation responsible for familial primary ovarian insufficiency abolishes its interaction with its partner SPATA22 and their recruitment to DNA double-strand breaks. *EBioMedicine* **42**, 524-531 (2019). https://doi.org:10.1016/j.ebiom.2019.03.075

167 Chen, A. *et al.* Mutations in the mitochondrial ribosomal protein MRPS22 lead to primary ovarian insufficiency. *Hum Mol Genet* **27**, 1913-1926 (2018). https://doi.org:10.1093/hmg/ddy098

168 Carlosama, C. *et al.* A homozygous donor splice-site mutation in the meiotic gene MSH4 causes primary ovarian insufficiency. *Hum Mol Genet* **26**, 3161-3166 (2017). https://doi.org:10.1093/hmg/ddx199

169 Akbari, A. *et al.* Rare missense variant in MSH4 associated with primary gonadal failure in both 46, XX and 46, XY individuals. *Hum Reprod* **36**, 1134-1145 (2021). https://doi.org:10.1093/humrep/deaa362

170 Guo, T. *et al.* Mutations in MSH5 in primary ovarian insufficiency. *Hum Mol Genet* **26**, 1452-1457 (2017). https://doi.org:10.1093/hmg/ddx044

171 Wu, X. *et al.* A NANOS3 mutation linked to protein degradation causes premature ovarian insufficiency. *Cell Death Dis* **4**, e825 (2013). https://doi.org:10.1038/cddis.2013.368

172 Santos, M. G. *et al.* Homozygous inactivating mutation in NANOS3 in two sisters with primary ovarian insufficiency. *Biomed Res Int* **2014**, 787465 (2014). https://doi.org:10.1155/2014/787465

173 Tucker, E. J. *et al.* Identification of variants in pleiotropic genes causing "isolated" premature ovarian insufficiency: implications for medical practice. *Eur J Hum Genet* **26**, 1319-1328 (2018). https://doi.org:10.1038/s41431-018-0140-4

174 Li, G. *et al.* Haploinsufficiency in non-homologous end joining factor 1 induces ovarian dysfunction in humans and mice. *J Med Genet* **59**, 579-588 (2022). https://doi.org:10.1136/jmedgenet-2020-107398

175 Qin, Y. *et al.* NOBOX homeobox mutation causes premature ovarian failure. *Am J Hum Genet* **81**, 576-581 (2007). https://doi.org:10.1086/519496

176 Bouilly, J., Bachelot, A., Broutin, I., Touraine, P. & Binart, N. Novel NOBOX loss-of-function mutations account for 6.2% of cases in a large primary ovarian insufficiency cohort. *Hum Mutat* **32**, 1108-1113 (2011). https://doi.org:10.1002/humu.21543

177 Bouali, N. *et al.* NOBOX is a strong autosomal candidate gene in Tunisian patients with primary ovarian insufficiency. *Clin Genet* **89**, 608-613 (2016). https://doi.org:10.1111/cge.12750

178 Ferrari, I. *et al.* Impaired protein stability and nuclear localization of NOBOX variants associated with premature ovarian insufficiency. *Hum Mol Genet* **25**, 5223-5233 (2016). https://doi.org:10.1093/hmg/ddw342

179 Li, L. *et al.* A homozygous NOBOX truncating variant causes defective transcriptional activation and leads to primary ovarian insufficiency. *Hum Reprod* **32**, 248-255 (2017). https://doi.org:10.1093/humrep/dew271

180 Sassi, A. *et al.* Compound heterozygous null mutations of NOBOX in sisters with delayed puberty and primary amenorrhea. *Mol Genet Genomic Med* **9**, e1776 (2021). https://doi.org:10.1002/mgg3.1776

181 Kosaki, K. *et al.* Premature ovarian failure in a female with proximal symphalangism and Noggin mutation. *Fertil Steril* **81**, 1137-1139 (2004). https://doi.org:10.1016/j.fertnstert.2003.08.054

182 Patiño, L. C. *et al.* Functional evidence implicating NOTCH2 missense mutations in primary ovarian insufficiency etiology. *Hum Mutat* **40**, 25-30 (2019). https://doi.org:10.1002/humu.23667

183 Li, L. *et al.* NOTCH2 variant D1853H is mutated in two non-syndromic premature ovarian insufficiency patients from a Chinese pedigree. *J Ovarian Res* **13**, 41 (2020). https://doi.org:10.1186/s13048-020-00645-4

184 Lourenço, D. *et al.* Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med* **360**, 1200-1210 (2009). https://doi.org:10.1056/NEJMoa0806228

185 Warman, D. M. *et al.* Three new SF-1 (NR5A1) gene mutations in two unrelated families with multiple affected members: within-family variability in 46,XY subjects and low ovarian reserve in fertile 46,XX subjects. *Horm Res Paediatr* **75**, 70-77 (2011). https://doi.org:10.1159/000320029

186 Janse, F. *et al.* Limited contribution of NR5A1 (SF-1) mutations in women with primary ovarian insufficiency (POI). *Fertil Steril* **97**, 141-146.e142 (2012). https://doi.org:10.1016/j.fertnstert.2011.10.032

187 Lakhal, B. *et al.* Mutational screening of SF1 and WNT4 in Tunisian women with premature ovarian failure. *Gene* **509**, 298-301 (2012). https://doi.org:10.1016/j.gene.2012.08.007

188 Philibert, P. *et al.* NR5A1 (SF-1) gene variants in a group of 26 young women with XX primary ovarian insufficiency. *Fertil Steril* **99**, 484-489 (2013). https://doi.org:10.1016/j.fertnstert.2012.10.026

189 Safari, S. *et al.* An Iranian family with azoospermia and premature ovarian insufficiency segregating NR5A1 mutation. *Climacteric* **17**, 301-303 (2014). https://doi.org:10.3109/13697137.2013.847079

190 Jiao, X. *et al.* Novel NR5A1 missense mutation in premature ovarian failure: detection in han chinese indicates causation in different ethnic groups. *PLoS One* **8**, e74759 (2013). https://doi.org:10.1371/journal.pone.0074759

191 Eggers, S. *et al.* Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1. *Eur J Hum Genet* **23**, 486-493 (2015). https://doi.org:10.1038/ejhg.2014.130

192 Oral, E. *et al.* Clinical and Genetic Investigation of Premature Ovarian Insufficiency Cases from Turkey. *J Gynecol Obstet Hum Reprod* **48**, 817-823 (2019). https://doi.org:10.1016/j.jogoh.2019.04.007

193 Jaillard, S. *et al.* Analysis of NR5A1 in 142 patients with premature ovarian insufficiency, diminished ovarian reserve, or unexplained infertility. *Maturitas* **131**, 78-86 (2020). https://doi.org:10.1016/j.maturitas.2019.10.011

194 Bertrand-Delepine, J. *et al.* In cases of familial primary ovarian insufficiency and disorders of gonadal development, consider NR5A1/SF-1 sequence variants. *Reprod Biomed Online* **40**, 151-159 (2020). https://doi.org:10.1016/j.rbmo.2019.10.002

195 Cattoni, A. *et al.* The Potential Synergic Effect of a Complex Pattern of Multiple Inherited Genetic Variants as a Pathogenic Factor for Ovarian Dysgenesis: A Case Report. *Front Endocrinol (Lausanne)* **11**, 540683 (2020). https://doi.org:10.3389/fendo.2020.540683

196 Brauner, R., Pierrepont, S., Bignon-Topalovic, J., McElreavey, K. & Bashamboo, A. Etiology of primary ovarian insufficiency in a series young girls presenting at a pediatric endocrinology center. *Eur J Pediatr* **174**, 767-773 (2015). https://doi.org:10.1007/s00431-014-2457-5

197 Weinberg-Shukron, A. *et al.* A mutation in the nucleoporin-107 gene causes XX gonadal dysgenesis. *J Clin Invest* **125**, 4295-4304 (2015). https://doi.org:10.1172/jci83553

198 Ren, Y. *et al.* Functional study of a novel missense single-nucleotide variant of NUP107 in two daughters of Mexican origin with premature ovarian insufficiency. *Mol Genet Genomic Med* **6**, 276-281 (2018). https://doi.org:10.1002/mgg3.345

199 Qin, Y., Jiao, X., Simpson, J. L. & Chen, Z. J. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update* **21**, 787-808 (2015). https://doi.org:10.1093/humupd/dmv036

200 Peng, T. *et al.* Novel PMM2 missense mutation in a Chinese family with non-syndromic premature ovarian insufficiency. *J Assist Reprod Genet* **37**, 443-450 (2020). https://doi.org:10.1007/s10815-019-01675-8

201 Masunaga, Y. *et al.* Primary ovarian insufficiency in a female with phosphomannomutase-2 gene (PMM2) mutations for congenital disorder of glycosylation. *Endocr J* **68**, 605-611 (2021). https://doi.org:10.1507/endocrj.EJ20-0706

202 Lacombe, A. *et al.* Disruption of POF1B binding to nonmuscle actin filaments is associated with premature ovarian failure. *Am J Hum Genet* **79**, 113-119 (2006). https://doi.org:10.1086/505406

203 Luoma, P. *et al.* Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. *Lancet* **364**, 875-882 (2004). https://doi.org:10.1016/s0140-6736(04)16983-3

204 Mancuso, M., Filosto, M., Oh, S. J. & DiMauro, S. A novel polymerase gamma mutation in a family with ophthalmoplegia, neuropathy, and Parkinsonism. *Arch Neurol* **61**, 1777-1779 (2004). https://doi.org:10.1001/archneur.61.11.1777

205 Hakonen, A. H. *et al.* Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *Am J Hum Genet* **77**, 430-441 (2005). https://doi.org:10.1086/444548

206 Pagnamenta, A. T. *et al.* Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. *Hum Reprod* **21**, 2467-2473 (2006). https://doi.org:10.1093/humrep/del076

207 Blok, M. J. *et al.* The unfolding clinical spectrum of POLG mutations. *J Med Genet* **46**, 776-785 (2009). https://doi.org:10.1136/jmg.2009.067686

208 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 (2012). https://doi.org:10.1016/j.gene.2012.02.034

209 Chen, B. *et al.* Identification of the first homozygous POLG mutation causing non-syndromic ovarian dysfunction. *Climacteric* **21**, 467-471 (2018). https://doi.org:10.1080/13697137.2018.1467891

210 Yang, H. J., Park, G., Nam-Goong, I. S., Ahn, J. W. & Weon, Y. C. The First Korean Siblings With Adult-Onset 4H Leukodystrophy Related to Nonsynonymous POLR3B Mutations. *Neurol Genet* **8**, e667 (2022). https://doi.org:10.1212/nxg.0000000000000667

211 Franca, M. M. *et al.* Exome Sequencing Reveals the POLR3H Gene as a Novel Cause of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **104**, 2827-2841 (2019). https://doi.org:10.1210/jc.2018-02485

212 Wang, J. *et al.* New candidate gene POU5F1 associated with premature ovarian failure in Chinese patients. *Reprod Biomed Online* **22**, 312-316 (2011). https://doi.org:10.1016/j.rbmo.2010.11.008

213 Zangen, D. *et al.* XX ovarian dysgenesis is caused by a PSMC3IP/HOP2 mutation that abolishes coactivation of estrogen-driven transcription. *Am J Hum Genet* **89**, 572-579 (2011). https://doi.org:10.1016/j.ajhg.2011.09.006

214 Mei, L. *et al.* Two novel biallelic mutations in PSMC3IP in a patient affected by premature ovarian insufficiency. *Mol Med Rep* **25** (2022). https://doi.org:10.3892/mmr.2021.12561

215 Sirchia, F., Giorgio, E., Cucinella, L., Valente, E. M. & Nappi, R. E. Biallelic mutations in PSMC3IP are associated with secondary amenorrhea: expanding the spectrum of premature ovarian insufficiency. *J Assist Reprod Genet* **39**, 1177-1181 (2022). https://doi.org:10.1007/s10815-022-02471-7

216 Luo, W. *et al.* Variants in Homologous Recombination Genes EXO1 and RAD51 Related with Premature Ovarian Insufficiency. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgaa505

217 Coppieters, F. *et al.* Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. *Am J Hum Genet* **99**, 470-480 (2016). https://doi.org:10.1016/j.ajhg.2016.06.017

218 Huang, C., Guo, T. & Qin, Y. Meiotic Recombination Defects and Premature Ovarian Insufficiency. *Front Cell Dev Biol* **9**, 652407 (2021). https://doi.org:10.3389/fcell.2021.652407

219 Wang, B. *et al.* Mutational analysis of SAL-Like 4 (SALL4) in Han Chinese women with premature ovarian failure. *Mol Hum Reprod* **15**, 557-562 (2009). https://doi.org:10.1093/molehr/gap046

220 Wang, Q. *et al.* Whole-exome sequencing reveals SALL4 variants in premature ovarian insufficiency: an update on genotype-phenotype correlations. *Hum Genet* **138**, 83-92 (2019). https://doi.org:10.1007/s00439-018-1962-4

221 Faridi, R. *et al.* Mutations of SGO2 and CLDN14 collectively cause coincidental Perrault syndrome. *Clin Genet* **91**, 328-332 (2017). https://doi.org:10.1111/cge.12867

222 Zhao, S. *et al.* Transcription factor SOHLH1 potentially associated with primary ovarian insufficiency. *Fertil Steril* **103**, 548-553.e545 (2015). https://doi.org:10.1016/j.fertnstert.2014.11.011

223 Qin, Y. *et al.* Novel variants in the SOHLH2 gene are implicated in human premature ovarian failure. *Fertil Steril* **101**, 1104-1109.e1106 (2014). https://doi.org:10.1016/j.fertnstert.2014.01.001

224 Yao, C. *et al.* Bi-allelic SPATA22 variants cause premature ovarian insufficiency and nonobstructive azoospermia due to meiotic arrest. *Clin Genet* **101**, 507-516 (2022). https://doi.org:10.1111/cge.14129

225 Smirin-Yosef, P. *et al.* A Biallelic Mutation in the Homologous Recombination Repair Gene SPIDR Is Associated With Human Gonadal Dysgenesis. *J Clin Endocrinol Metab* **102**, 681-688 (2017). https://doi.org:10.1210/jc.2016-2714

226 Heddar, A., Guichoux, N., Auger, N. & Misrahi, M. A SPIDR homozygous nonsense pathogenic variant in isolated primary ovarian insufficiency with chromosomal instability. *Clin Genet* **101**, 242-246 (2022). https://doi.org:10.1111/cge.14080

227 Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *N Engl J Med* **370**, 943-949 (2014). https://doi.org:10.1056/NEJMoa1309635

228 Le Quesne Stabej, P. *et al.* STAG3 truncating variant as the cause of primary ovarian insufficiency. *Eur J Hum Genet* **24**, 135-138 (2016). https://doi.org:10.1038/ejhg.2015.107

229 He, W. B. *et al.* Whole-exome sequencing identifies a homozygous donor splice-site mutation in STAG3 that causes primary ovarian insufficiency. *Clin Genet* **93**, 340-344 (2018). https://doi.org:10.1111/cge.13034

230 Colombo, R., Pontoglio, A. & Bini, M. A STAG3 missense mutation in two sisters with primary ovarian insufficiency. *Eur J Obstet Gynecol Reprod Biol* **216**, 269-271 (2017). https://doi.org:10.1016/j.ejogrb.2017.08.005

231 França, M. M. *et al.* Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency. *Eur J Med Genet* **62**, 186-189 (2019). https://doi.org:10.1016/j.ejmg.2018.07.008

232 Heddar, A., Dessen, P., Flatters, D. & Misrahi, M. Novel STAG3 mutations in a Caucasian family with primary ovarian insufficiency. *Mol Genet Genomics* **294**, 1527-1534 (2019). https://doi.org:10.1007/s00438-019-01594-4

233 Xiao, W. J. *et al.* In-Frame Variants in STAG3 Gene Cause Premature Ovarian Insufficiency. *Front Genet* **10**, 1016 (2019). https://doi.org:10.3389/fgene.2019.01016

234 Mellone, S. *et al.* A Long Contiguous Stretch of Homozygosity Disclosed a Novel STAG3 Biallelic Pathogenic Variant Causing Primary Ovarian Insufficiency: A Case Report and Review of the Literature. *Genes (Basel)* **12** (2021). https://doi.org:10.3390/genes12111709

235 Akbari, A. *et al.* Novel STAG3 variant associated with primary ovarian insufficiency and non-obstructive azoospermia in an Iranian consanguineous family. *Gene* **821**, 146281 (2022). https://doi.org:10.1016/j.gene.2022.146281

236 Gómez-Rojas, S., Aristizábal-Duque, J. E., Muñoz-Fernández, L. F., Sarmiento-Ramón, M. P. & Pereira-Gómez, M. D. P. New STAG3 gene variant as a cause of premature ovarian insufficiency. *Rev Colomb Obstet Ginecol* **73**, 142-148 (2022). https://doi.org:10.18597/rcog.3806

237 de Vries, L. *et al.* Exome sequencing reveals SYCE1 mutation associated with autosomal recessive primary ovarian insufficiency. *J Clin Endocrinol Metab* **99**, E2129-2132 (2014). https://doi.org:10.1210/jc.2014-1268

238 He, W. B. *et al.* Homozygous variants in SYCP2L cause premature ovarian insufficiency. *J Med Genet* **58**, 168-172 (2021). https://doi.org:10.1136/jmedgenet-2019-106789

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

240 Mathorne, S. W. *et al.* Novel phenotype of syndromic premature ovarian insufficiency associated with TP63 molecular defect. *Clin Genet* **97**, 779-784 (2020). https://doi.org:10.1111/cge.13725

241 Karlberg, S., Tiitinen, A., Alfthan, H. & Lipsanen-Nyman, M. Premature ovarian insufficiency and early depletion of the ovarian reserve in the monogenic Mulibrey nanism disorder. *Hum Reprod* **33**, 1254-1261 (2018). https://doi.org:10.1093/humrep/dey103

242 Morino, H. *et al.* Mutations in Twinkle primase-helicase cause Perrault syndrome with neurologic features. *Neurology* **83**, 2054-2061 (2014). https://doi.org:10.1212/wnl.0000000000001036

243 Demain, L. A. *et al.* Expanding the genotypic spectrum of Perrault syndrome. *Clin Genet* **91**, 302-312 (2017). https://doi.org:10.1111/cge.12776

244 Lerat, J. *et al.* An Application of NGS for Molecular Investigations in Perrault Syndrome: Study of 14 Families and Review of the Literature. *Hum Mutat* **37**, 1354-1362 (2016). https://doi.org:10.1002/humu.23120

245 Ołdak, M. *et al.* Novel neuro-audiological findings and further evidence for TWNK involvement in Perrault syndrome. *J Transl Med* **15**, 25 (2017). https://doi.org:10.1186/s12967-017-1129-4

246 Domínguez-Ruiz, M. *et al.* Perrault syndrome with neurological features in a compound heterozygote for two TWNK mutations: overlap of TWNK-related recessive disorders. *J Transl Med* **17**, 290 (2019). https://doi.org:10.1186/s12967-019-2041-x

247 Gotta, F. *et al.* A novel mutation of Twinkle in Perrault syndrome: A not rare diagnosis? *Ann Hum Genet* **84**, 417-422 (2020). https://doi.org:10.1111/ahg.12384

248 Zhou, Y. *et al.* Wdr62 is involved in female meiotic initiation via activating JNK signaling and associated with POI in humans. *PLoS Genet* **14**, e1007463 (2018). https://doi.org:10.1371/journal.pgen.1007463

249 Rossi, M. L., Ghosh, A. K. & Bohr, V. A. Roles of Werner syndrome protein in protection of genome integrity. *DNA Repair (Amst)* **9**, 331-344 (2010). https://doi.org:10.1016/j.dnarep.2009.12.011

250 Wang, H. *et al.* Novel WT1 Missense Mutations in Han Chinese Women with Premature Ovarian Failure. *Sci Rep* **5**, 13983 (2015). https://doi.org:10.1038/srep13983

251 Wang, Y. *et al.* Whole exome sequencing identified a rare WT1 loss-of-function variant in a non-syndromic POI patient. *Mol Genet Genomic Med* **10**, e1820 (2022). https://doi.org:10.1002/mgg3.1820

252 McGlacken-Byrne, S. M. *et al.* Pathogenic variants in the human m6A reader YTHDC2 are associated with primary ovarian insufficiency. *JCI Insight* **7** (2022). https://doi.org:10.1172/jci.insight.154671

253 McGlacken-Byrne, S. M. *et al.* ZSWIM7 Is Associated With Human Female Meiosis and Familial Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **107**, e254-e263 (2022). https://doi.org:10.1210/clinem/dgab597

254 Yatsenko, S. A. *et al.* Pathogenic Variants in ZSWIM7 Cause Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **107**, e2359-e2364 (2022). https://doi.org:10.1210/clinem/dgac090

255 Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-69 (2012). https://doi.org:10.1126/science.1219240

256 Zuk, O. *et al.* Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455-E464 (2014). https://doi.org:doi:10.1073/pnas.1322563111

257 McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356-369 (2008). https://doi.org:10.1038/nrg2344

258 Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31-46 (2010). https://doi.org:10.1038/nrg2626

259 Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**, 363-376 (2011). https://doi.org:10.1038/nrg2958

260 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009). https://doi.org:10.1038/nature08494

261 Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**, 415-425 (2010). https://doi.org:10.1038/nrg2779

262 Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet* **48**, 314-317 (2016). https://doi.org:10.1038/ng.3507

263 Stasa, S. *et al.* Genetic susceptibility to earlier ovarian ageing increases de novo mutation rate in offspring. *medRxiv*, 2022.2006.2023.22276698 (2022). https://doi.org:10.1101/2022.06.23.22276698

264 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001). https://doi.org:10.1038/35057062

265 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015). https://doi.org:10.1038/nature15393

266 Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* **44**, 243-246 (2012). https://doi.org:10.1038/ng.1074

267 Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A* **108**, 11983-11988 (2011). https://doi.org:10.1073/pnas.1019276108

268 Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* **49**, 1421-1427 (2017). https://doi.org:10.1038/ng.3954

269 Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet* **100**, 635-649 (2017). https://doi.org:10.1016/j.ajhg.2017.03.004

270 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018). https://doi.org:10.1038/s41586-018-0579-z

271 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021). https://doi.org:10.1038/s41586-021-03205-y

272 Gravel, S. Population Genetics Models of Local Ancestry. *Genetics* **191**, 607-619 (2012). https://doi.org:10.1534/genetics.112.139808

Chapter 2:

Heterozygous genetic variants in BMP15 are not a common cause of Premature Ovarian Insufficiency.

Saleh Shekari, Stuart Cannon, Caroline F. Wright, Michael N. Weedon, Andrew R. Wood, Robin N. Beaumont, Gareth Hawkes, Julia K. Prague, Timothy M. Frayling, Gita Mishra, Katherine S. Ruth, Anna Murray

Main text

Abstract

Premature Ovarian Insufficiency (POI) has been linked to deleterious genetic variations in more than 100 genes, including *BMP15* an X-linked gene with an important role in female ovarian development and folliculogenesis. More than 20 genomic variants in *BMP15* have previously been associated with POI, but the strength of evidence to support pathogenicity differs between variants. To examine the impact of *BMP15* gene genomic variants on ovarian function, we utilized exome sequence data from approximately 105,000 women in the UK Biobank.

We examined the association between 16 previously reported variants and age at natural menopause (ANM), as well as 143 rare protein-coding variants in BMP15. These rare variants were annotated as either predicted loss-of-function (LOF) or missense, and their implications in POI were previously unknown. We assessed the genomic variations individually and also employed a gene burden test.

Our findings revealed three novel heterozygous LOF variants in BMP15, and interestingly, neither carrier had an ANM less than 40 years, providing additional evidence that haploinsufficiency of *BMP15* does not cause POI. Furthermore, none of the 16 *BMP15* variations previously reported in the literature exhibited an association with early menopause in our cohort. However, after adjusting for multiple testing, rare missense variants in *BMP15* were not found to be related to ANM in the UK Biobank.

Our findings indicate that heterozygous missense variants in the *BMP15* gene should be approached with caution and are unlikely to be causative factors for POI or early menopause. This study significantly enhances our comprehension of the involvement of *BMP15* genomic variants in ovarian function. Moreover, it emphasizes the importance of meticulous evaluation of the pathogenicity of specific variants within the context of POI and early menopause.

This study represents one of the largest population genomic analyses of POI to date, comprising 113 clinical POI cases and 2,231 non-clinical POI cases.

Introduction

Premature ovarian insufficiency (POI), defined as menopause before the age of 40, has been associated with chromosomal abnormalities and variation single gene in over 100 genes, but the evidence to support causality of individual variants varies between studies [1-3]. POI occurs in 1 in 100 women [4], is a major cause of female infertility and is characterised by increased follicle-stimulating hormone (FSH) levels and hypoestrogenism. Single-gene causes of POI can be used within families to predict ovarian insufficiency in other family members, enabling women to make informed reproductive choices. As large-scale, population genomic data becomes more accessible, it is also essential to ascertain the penetrance of genomic variants in the absence of family history [5].

BMP15 (Bone Morphogenetic Protein 15) is an X-linked gene that was first reported to be associated with hypergonadotropic ovarian failure in two sisters who were heterozygous for a paternally inherited missense variant-NP_005439.2: p.(Tyr235Cys)-, with a suspected dominant-negative effect [6]. Subsequently, family and cohort studies have reported the association of *BMP15* variants with female ovarian development and folliculogenesis, with supporting evidence from animal models [6-18]. The majority of the previously reported variants are heterozygous missense variants in the mature *BMP15* protein, although there are three cases where individuals were homozygous for loss-of-function (LOF) variants [19,20]. Haploinsufficiency of *BMP15* is not thought to be pathogenic, as, in all three homozygous cases, the mother carried one of the LOF variants but was obviously fertile and did not have POI. *BMP15* is an oocyte-specific growth factor and member of the transforming growth factor-β superfamily, which regulates granulosa cell proliferation and differentiation during folliculogenesis[21-23]. Animal models have further characterised the biological function of *BMP15* in mammalian reproduction [6,24-29]. However, there are differences between species; for example, mice with *BMP15* inactivating mutations show minor defects in follicular growth[28], while sheep with *BMP15* knock-out are infertile[29].

We sought to explore the pathogenicity of *BMP15* genomic variants reported as associated with POI in females who were not selected through clinical features but from the general population by using age at natural menopause (ANM). We report *BMP15* genomic variation in 104,733 females from the UK Biobank with

whole exome sequencing (WES) data. We evaluated previously reported individual *BMP15* variants from the literature and also all variants that were predicted as LOF, or non-synonymous variants in aggregate, using gene burden tests.

Materials and methods

Exome sequencing in UK Biobank

We tested the effect of individual variants in *BMP15* plus combined variant burden analyses in Whole Exome Sequencing data (WES) from 454,787 individuals in the UK Biobank study [32]. Details of this study, including data collection and processing, are described elsewhere [33]. Informed consent was provided by all participants. Study approval was received from the National Research Ethics Service Committee North West–Haydock and all study procedures were performed according to the World Medical Association Declaration of Helsinki ethical principles for medical research. WES data were generated with the IDT xGen Exome Research Panel v1.0, which targeted 39Mbp of the human genome with mean coverage exceeding 20x on 95.6% of sites. The OQFE protocol was used for mapping and variant calling to the GRCh38 reference. Quality control filters applied by UK Biobank were individual and variant missingness <10% and Hardy Weinberg Equilibrium P-value >10-15. In addition to UK Biobank quality control filters described previously, we excluded variants with <10X coverage in 90% of the samples that were provided by Backman et al. [32]. We selected variants in *BMP15* MANE transcript (ENST00000252677), and variants were annotated using the Ensembl Variant Effect Predictor [34] and LOFTEE plugin (https://github.com/konradjk/loftee). Minor allele frequency (MAF) and allele count analysis was calculated using PLINK [35]. Furthermore, we manually assessed the variants using Integrative Genomics Viewer (IGV) [36,37]. Analyses were performed on the UK Biobank Research Analysis Platform (RAP; https://ukbiobank.dnanexus.com/) using Compressed Reference-oriented Alignment Map (CRAM) and Variant Call Format (VCF) sequence data files.

Phenotype definition

UK Biobank provided phenotypic data (e.g., health status, lifestyle and demographics) for ~500,000 individuals aged 37-73 years at the time of recruitment from across the UK [5]. ANM was derived from self-reported questionnaire data as the age at last naturally occurring menstrual period, excluding those with surgical menopause (field 2824 and 3882) or taking hormone replacement therapy (field 3536), as described previously [30]. There

were 104,733 European female participants with ANM (range 18 to 65 years, mean=50.1, SD=4.5); of those, 2,231 individuals reported ANM<40 years which we refer to as non-clinical POI cases. Participants who reported ANM < 40 years during the questionnaire were asked to confirm their ANM. Controls had ANM ≥ 40, plus those who were pre-menopausal and older than 40 at the time of participation (n=192,438).

We identified primary and secondary amenorrhea cases from UK Biobank primary care data of 230,060 participants (field 42040). We selected severe ovarian failure GP codes (more than 30 codes) and excluded participants with X chromosomal aneuploidy using UK Biobank SNP array data, as described previously[31]. We further excluded women who self-reported ANM ≥40 or had cancer <40, leaving 113 clinical POI cases, including 10 primary and 103 secondary amenorrhea cases. However, among all clinical cases that were identified, a total of 95 cases also were included in our ANM cohort and had undergone menopause before the age of 40 (non-clinical cases). All manipulations were conducted in R (v4.0.3) on the UK Biobank Research Analysis Platform (RAP; https://ukbiobank.dnanexus.com/).

Analysis of previously published *BMP15* variants associated with ANM.
From the literature, we identified 25 variants in *BMP15* (**Supplementary Table 1**) that were reported as risk alleles for POI[6-18]. In addition, there were four LOF variants reported in the literature in women with POI, two in homozygotes, two in a compound heterozygote and one heterozygous carrier. Each variant with a minor allele count (MAC) of more than 5 was tested individually for association with ANM. In addition to testing each variant individually, we investigated the effect of having any of these variants on menopause timing by aggregating them into a single unit by coding an individual carrying any variant as 1 and non-carriers as 0.
We used REGENIE to implement a generalised mixed-model region-based association test that can account for population stratification and sample relatedness in large-scale analyses. We applied an inverse normal rank transformation to ANM and included recruitment centre, sequence batch and 40 genetic principal components as covariates. The decision of whether to adjust for certain variables, such as age and BMI, in a study involves a careful

consideration of the research objectives, potential confounding factors, and statistical implications. Adjusting for variables is often done to isolate the specific effect of the variable of interest. However, in our study, the primary focus was on understanding the pure genetic effect without the influence of other factors.

In cases where age and BMI were not adjusted for, the goal was to avoid overadjustment and maintain a clear focus on the unadulterated impact of genetic factors. Overadjustment, or controlling for variables on the causal pathway, can distort results, particularly when those variables are intricately linked to the genetic effect under investigation. By excluding adjustments for age and BMI, we aimed to simplify the analysis and directly examine the genetic component, aligning with our study's central objective of emphasizing the unadjusted, intrinsic genetic effects on the outcome variable. This intentional choice streamlines the investigation, providing a more straightforward and transparent exploration of the genetic factors in question.

To estimate the approximate effect on ANM per year, we multiplied the estimated effect on inverse normalised ANM by the standard deviation of ANM (SD = 4.53). For gene burden tests, REGENIE runs in two steps [38], which we implemented on the UK Biobank RAP. Firstly, genetic variants are aggregated into gene-specific units and then, secondly, these units are tested for association with ANM.

Testing of all rare variants in *BMP15* in UK Biobank

We tested the association with ANM of all variants in BMP15 with MAF less than 0.1%, both individually (AC ≥5) and by aggregating them into gene-specific units, using REGENIE as described. Gene-specific units included in our analysis were: high-confidence LOF, including stop-gain, frameshift, or abolishing a canonical splice site (-2 or +2 bp from exon, excluding the ones in the last exon); missense variants; damaging variants that included high-confidence LOF variants or/and missense variants with REVEL score >0.7; and synonymous variants.

Results

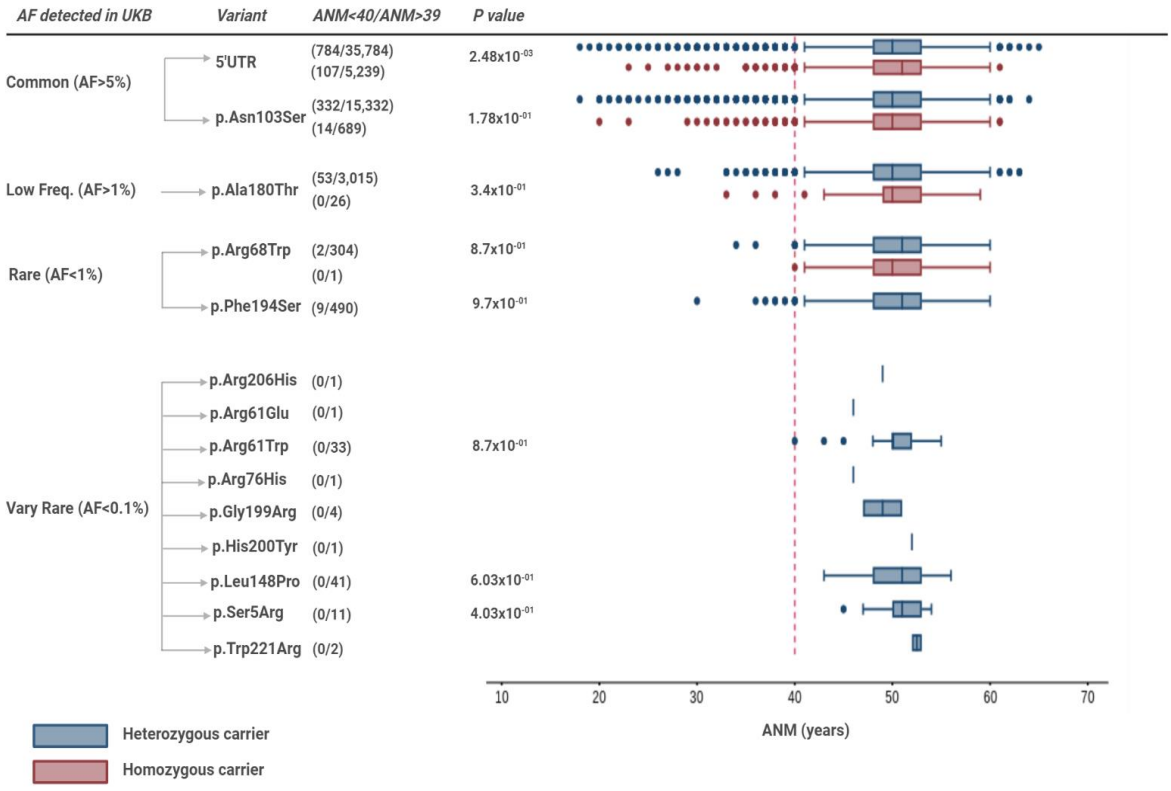
Previously published *BMP15* variants are not associated with ANM in UK Biobank

Of the 25 variants identified in the literature as associated with POI, 24 were exonic, and there was a non-coding 5 prime untranslated (5′ UTR) variant that was previously reported as being possibly or likely pathogenic. Of these, 19 were missense variants reported to be risk alleles in the heterozygous state, of which there was evidence that two were acting via a dominant-negative mechanism (**Supplementary Table I**). For five of the 19 missense variants plus the 5′ UTR variant, there was functional evidence in the literature that the variant had a deleterious effect on the protein through reporter assays and Western blots (**Supplementary Table I**) [10-15]. An additional missense variant was reported in a homozygous state in a POI proband, while her mother was heterozygous and did not have POI [18]. We were able to identify 16 out of the 25 variants in UK Biobank. In total, 104,489 women carried one or more previously reported variants in *BMP15* and found out that the control group has at least one *BMP15* variant that has been reported before (as shown in **Figure 1** and Supplementary Table 1). Figure 1 illustrates the distribution of previously reported variants within ANM cohorts, accompanied by summary statistics. However, out of the 16 detected variants, two were found exclusively in the control cohort and thus were excluded from Figure 1 (Supplementary Table 1). Among the remaining variants, four were identified in both heterozygous (represented by blue colour bars) and homozygous (represented by red colour bars) states.

However, none of the four previously reported homozygous or compound heterozygous LOF variants were present in our data, even in a heterozygous state. Among the 2,231 cases of non-clinical POI, a total of 1,180 females were identified as carriers of five previously reported variants (p.Arg68Trp, p.Asn103Ser, p.Ala180Thr, p.Phe194Ser, and c.-9C>G) in a heterozygous state. Out of these, 56 POI cases identified by health records were found to be carriers for three of these variants (p.Asn103Ser, p.Ala180Thr and c.-9C>G). However, all five variants were also found in controls (103,134 heterozygotes

and 11,303 homozygotes) and were present in the Genome Aggregation Database (gnomAD) [39] with similar allele frequencies. An analysis that considered the combined effects of all five variants found no evidence of earlier menopause (Beta = 0.07 years later menopause per allele, SE = 0.02, P = 0.0015). The remaining 11 previously published variants were only detected in controls in the UK Biobank, with 175 heterozygous carriers.

Figure 0–1: Previously reported BMP15 variants that were detected in UK Biobank.



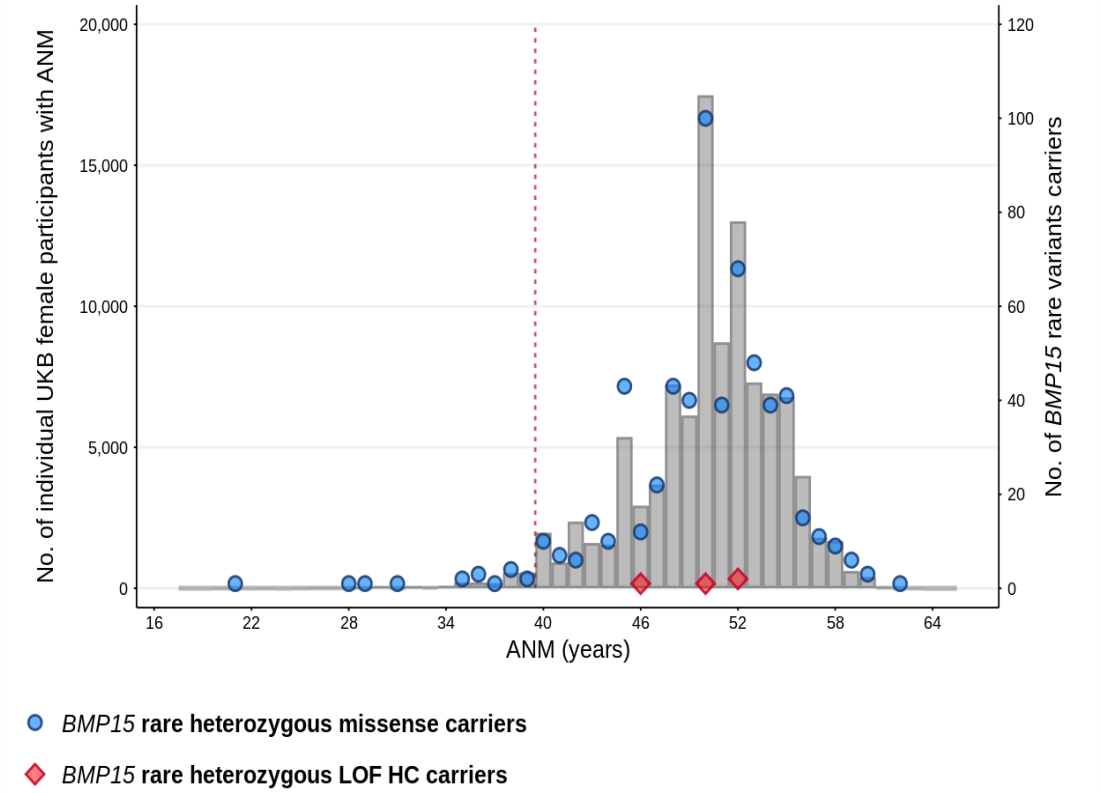
Note: Count and distribution of ANM in individuals carrying 14 previously reported BMP15 missense/5'UTR variants (further details in supplementary table 1). The blue and red bars are for heterozygous and homozygous carriers, respectively. AF = allele frequency; 5'UTR = 5' untranslated region.

Rare missense variants in *BMP15* are not associated with ANM.

In addition to the 16 previously published variants, there were 191 additional rare variants (MAF<0.1%) in the protein-coding region of *BMP15* detected in the 104,733 European ancestry females with ANM. Of the 191 variants, three were deletions which were predicted to result in loss of functional protein; one removing a splice site acceptor sequence (c.329-10_338del) and two frameshift variants (c.131del - p.Leu44ArgfsTer6; c.1138del - p.Glu380SerfsTer5). A further 140 variants were missense (**Supplementary Table 2, figure 2**), and 48 were synonymous. None of the carriers of heterozygous LoF variants in *BMP15*

were a clinical POI case or had ANM<40, nor was there evidence of an effect on menopause timing based on the gene burden tests (P>0.05; Table 1 and **Supplementary Table 3**).

Figure 0–2: Comparison of ANM in BMP15 LOF and missense carriers compared to controls.



Note: Distribution of ANM for 104,733 women in UK Biobank, with number of LOF (N=4) and missense (N=605) variant carriers indicated with red diamonds and blue circles, respectively.

Table 0-1: REGENIE burden testing for rare variants in BMP15.

VARIANT CLASS	VARIANTS (N)	ALLELE COUNT (N)	BURDEN P.VALUE	BURDEN BETA	BURDEN SE
LOF	3	4	0.95	0.13	2.16
MISSENSE	140	605	0.54	-0.11	0.18
DAMAGING	14	92	0.82	-0.1	0.45
SYN.	48	676	0.74	0.05	0.17

Note: LOF= Loss-of-function; Syn= synonymous; Bonferroni corrected p-value for 5 tests; 0.05/5 = 0.01.

There were 605 women with ANM who were heterozygous for the 140 missense variants with MAF <0.01%. There was no evidence that rare missense variants combined in a burden test were associated with ANM as a quantitative trait (Burden P=0.54, Beta=-0.11 years earlier per allele) (**Table 1**).

We further tested a subset of the predicted most damaging variants by combining missense variants with a REVEL score > 0.7 (n=11) with LOF variants (n=3). There was no association of ‘damaging’ variants with ANM (P=0.82).

Discussion

Previous studies have reported more than 20 *BMP15* genetic variants as likely to cause or contribute to POI in either an X-linked dominant negative or recessive inheritance pattern [6-18]. Numerous studies have focused on POI cases identified in clinical or family settings. However, our study examined a population-based cohort of over 2000 women with ANM under 40 and discovered that many of the pathogenic variants previously reported were common in controls. We found little evidence that these variations reduce menopause age. Additionally, all previously reported variants found in women with a clinical diagnosis of POI from health records and women with menopause <40 were also present in controls, indicating that they are not a monogenic cause of POI, at least in the heterozygous state.

Complete biallelic loss of *BMP15* in humans has been reported in several cases. Two sisters inherited compound heterozygous LOF variants; one had primary amenorrhea and the other POI at age 17; however, their (heterozygous) mother was fertile with at least three children [19]. There are two further examples of probands with POI and homozygous LOF variants in *BMP15*, but heterozygous mothers with menopause at 50 and 56 years of age [20]. We did not find any of the previously reported LOF variants in our study, but the three novel heterozygous LOF variants in 4 carriers we did detect were found in controls, supporting previous reports that haploinsufficiency of *BMP15* is not a causal mechanism for POI [19,20].

Most heterozygous missense variants previously reported as pathogenic localise to key functional elements of the N-terminal prodomain of the *BMP15* protein and may affect how *BMP15* interacts with *GDF9* (Growth Differentiation Factor 9) to regulate granulosa cell growth in a dominant negative fashion [15]. In our study, all of the previously reported missense variants were found in controls, suggesting that they are either not pathogenic or incompletely penetrant in the heterozygous state. We categorised the heterozygous variants depending on the strength of previous evidence that the variant affected the function of the protein. The group with functional evidence in the literature included a variant in the 5'UTR (c.-9C>G) of the gene that has been analysed *in vitro* with luciferase assays to determine its potential for changing the expression level of *BMP15* protein [7-9]. The *BMP15* c.-9C>G was identified in

approximately 40% of both clinical and non-clinical POI cases and controls, with 9,938 in control exhibiting homozygosity for the variant. A further four variants with experimental evidence for an effect on the protein function were also detected (p.Arg68Trp; p.Arg138His; p.Leu148Pro; p.Phe194Ser), but none were associated with ANM. The two variants with functional evidence of a dominant negative effect (p.Tyr235Cys; p.Arg329Cys) were not detected in our study. A further five missense variants were not detected in the UK Biobank, and we were, therefore not able to draw any conclusions about the pathogenicity of these variants.

Previous studies have mostly included relatively small cohorts of POI cases (between 50-300 women) and often have had similarly small numbers of controls. Our study included more than double the number of cases, plus tens of thousands of controls, which enabled more robust statistical comparisons to be made. There are some limitations to our study. Other published studies reported families with POI, often consanguineous, where they could assess co-segregation of variants, which we were not able to do in our cohort study. Our cohort of POI cases was mostly based on self-reported menopause age, although participants were asked to confirm that age if it was <40 years. There was no clinical assessment of reproductive status in UK Biobank participants, including appropriate hormone measures, which may have resulted in inclusion of cases in addition to those with ANM <40. Women with primary amenorrhea were unlikely to be included in our study, because they wouldn't have an age at menopause recorded. Our non-clinical POI cohort may therefore not be completely comparable to the previously reported clinically ascertained series of cases. We were able to identify 113 cases of primary and secondary amenorrhea based on UK Biobank GP and hospital record data, but this is likely to be an under-estimate of cases. The UK Biobank is known to have a bias towards healthier individuals [40]. Our POI cohort may therefore not be completely comparable to the previously reported clinically ascertained series of cases. We and others have shown previously that the penetrance of rare Mendelian disease variants can be lower in population-based cohorts compared to clinically ascertained cohorts [5].

As genetic tests have become more accessible it is increasingly important to ascertain the penetrance of genetic variants in the general population in

addition to clinical cases. Robust statistical evidence plus functional supporting data should be used to assess pathogenicity. Finding the causative gene and variant for women with POI is important to enable appropriate risks to be given to other family members who may be able to take steps to preserve fertility. Understanding the underlying causative mechanism also has the potential to drive therapeutic strategies in the future.

In conclusion, our results indicate that many of the previously published variants in the *BMP15* gene are not sufficient to cause POI when heterozygous and that heterozygous *BMP15* missense variants do not cause monogenic POI. Therefore, although BMP15 remains a gene of interest in POI due to its role in ovarian function, clinicians should be cautious when attributing causality to heterozygous variants found in the gene.

Acknowledgements

This research has been conducted using the UK Biobank Resource under application numbers 9072 and 871. The authors acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. This study was supported by the NIHR Exeter Biomedical Research Centre. This work was supported by the QUEX Institute (the University of Exeter, UK and University of Queensland, Australia).

Funding

CW, MNW, TMF & AM funded by MRC project grant (MR/T00200X/1).GH has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 875534. This Joint Undertaking received support from the European Union's Horizon 2020 research and innovation programme and EFPIA and T1D Exchange, JDRF, and Obesity Action Coalition

References

1 Qin, Y., Jiao, X., Simpson, J. L. & Chen, Z. J. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update* **21**, 787-808 (2015). https://doi.org:10.1093/humupd/dmv036

2 Laven, J. S. Primary Ovarian Insufficiency. *Semin Reprod Med* **34**, 230-234 (2016). https://doi.org:10.1055/s-0036-1585402

3 França, M. M. & Mendonca, B. B. Genetics of Primary Ovarian Insufficiency in the Next-Generation Sequencing Era. *J Endocr Soc* **4**, bvz037 (2020). https://doi.org:10.1210/jendso/bvz037

4 Tsiligiannis, S., Panay, N. & Stevenson, J. C. Premature Ovarian Insufficiency and Long-Term Health Consequences. *Curr Vasc Pharmacol* **17**, 604-609 (2019). https://doi.org:10.2174/1570161117666190122101611

5 Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* **104**, 275-286 (2019). https://doi.org:10.1016/j.ajhg.2018.12.015

6 Di Pasquale, E., Beck-Peccoz, P. & Persani, L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. *Am J Hum Genet* **75**, 106-111 (2004). https://doi.org:10.1086/422103

7 Dixit, H. *et al.* Missense mutations in the BMP15 gene are associated with ovarian failure. *Hum Genet* **119**, 408-415 (2006). https://doi.org:10.1007/s00439-006-0150-0

8 Gónzalez, A. *et al.* Association of genetic markers within the BMP15 gene with anovulation and infertility in women with polycystic ovary syndrome. *Fertil Steril* **90**, 447-449 (2008). https://doi.org:10.1016/j.fertnstert.2007.06.083

9 Fonseca, D. J. *et al.* BMP15 c.-9C>G promoter sequence variant may contribute to the cause of non-syndromic premature ovarian failure. *Reprod Biomed Online* **29**, 627-633 (2014). https://doi.org:10.1016/j.rbmo.2014.07.018

10 Rossetti, R. *et al.* BMP15 mutations associated with primary ovarian insufficiency cause a defective production of bioactive protein. *Hum Mutat* **30**, 804-810 (2009). https://doi.org:10.1002/humu.20961

11 Bouilly, J. *et al.* Identification of Multiple Gene Mutations Accounts for a new Genetic Architecture of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **101**, 4541-4550 (2016). https://doi.org:10.1210/jc.2016-2152

12 Di Pasquale, E. *et al.* Identification of new variants of human BMP15 gene in a large cohort of women with premature ovarian failure. *J Clin Endocrinol Metab* **91**, 1976-1979 (2006). https://doi.org:10.1210/jc.2005-2650

13 Tiotiu, D. *et al.* Variants of the BMP15 gene in a cohort of patients with premature ovarian failure. *Hum Reprod* **25**, 1581-1587 (2010). https://doi.org:10.1093/humrep/deq073

14 Laissue, P. *et al.* Mutations and sequence variants in GDF9 and BMP15 in patients with premature ovarian failure. *Eur J Endocrinol* **154**, 739-744 (2006). https://doi.org:10.1530/eje.1.02135

15 Patiño, L. C. *et al.* BMP15 Mutations Associated With Primary Ovarian Insufficiency Reduce Expression, Activity, or Synergy With GDF9. *J Clin Endocrinol Metab* **102**, 1009-1019 (2017). https://doi.org:10.1210/jc.2016-3503

16 Wang, B. *et al.* Analyses of growth differentiation factor 9 (GDF9) and bone morphogenetic protein 15 (BMP15) mutation in Chinese women with premature ovarian failure. *Clin Endocrinol (Oxf)* **72**, 135-136 (2010). https://doi.org:10.1111/j.1365-2265.2009.03613.x

17 Santos, M. *et al.* Association of BMP15 and GDF9 variants to premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 2163-2169 (2019). https://doi.org:10.1007/s10815-019-01548-0

18 Zhang, W. *et al.* A novel homozygous mutation of bone morphogenetic protein 15 identified in a consanguineous marriage family with primary ovarian insufficiency. *Reprod Biomed Online* **36**, 206-209 (2018). https://doi.org:10.1016/j.rbmo.2017.10.104

19 Mayer, A., Fouquet, B., Pugeat, M. & Misrahi, M. BMP15 "knockout-like" effect in familial premature ovarian insufficiency with persistent ovarian reserve. *Clin Genet* **92**, 208-212 (2017). https://doi.org:10.1111/cge.12970

20 Rossetti, R. *et al.* Fundamental role of BMP15 in human ovarian folliculogenesis revealed by null and missense mutations associated with primary ovarian insufficiency. *Hum Mutat* **41**, 983-997 (2020). https://doi.org:10.1002/humu.23988

21 Roh, J. S. *et al.* Growth differentiation factor-9 stimulates inhibin production and activates Smad2 in cultured rat granulosa cells. *Endocrinology* **144**, 172-178 (2003). https://doi.org:10.1210/en.2002-220618

22 Kaivo-Oja, N. *et al.* Growth differentiation factor-9 induces Smad2 activation and inhibin B production in cultured human granulosa-luteal cells. *J Clin Endocrinol Metab* **88**, 755-762 (2003). https://doi.org:10.1210/jc.2002-021317

23 Otsuka, F. *et al.* Bone morphogenetic protein-15. Identification of target cells and biological functions. *J Biol Chem* **275**, 39523-39528 (2000). https://doi.org:10.1074/jbc.M007428200

24 Galloway, S. M. *et al.* Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat Genet* **25**, 279-283 (2000). https://doi.org:10.1038/77033

25 McNatty, K. P. *et al.* Oocyte-derived growth factors and ovulation rate in sheep. *Reprod Suppl* **61**, 339-351 (2003).

26 Shimasaki, S., Moore, R. K., Otsuka, F. & Erickson, G. F. The bone morphogenetic protein system in mammalian reproduction. *Endocr Rev* **25**, 72-101 (2004). https://doi.org:10.1210/er.2003-0007

27 Flossmann, G. *et al.* A nonsense mutation of bone morphogenetic protein-15 (BMP15) causes both infertility and increased litter size in pigs. *BMC Genomics* **22**, 38 (2021). https://doi.org:10.1186/s12864-020-07343-x

28 Dong, J. *et al.* Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383**, 531-535 (1996). https://doi.org:10.1038/383531a0

29 Hanrahan, J. P. *et al.* Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (Ovis aries). *Biol Reprod* **70**, 900-909 (2004). https://doi.org:10.1095/biolreprod.103.023093

30 Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci Rep* **6**, 24710 (2016). https://doi.org:10.1038/srep24710

31 Tuke, M. A. *et al.* Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genet Med* **21**, 877-886 (2019). https://doi.org:10.1038/s41436-018-0271-6

32 Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628-634 (2021). https://doi.org:10.1038/s41586-021-04103-z

33 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015). https://doi.org:10.1371/journal.pmed.1001779

34 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016). https://doi.org:10.1186/s13059-016-0974-4

35 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007). https://doi.org:10.1086/519795

36 Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34 (2017). https://doi.org:10.1158/0008-5472.Can-17-0337

37 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011). https://doi.org:10.1038/nbt.1754

38 Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021). https://doi.org:10.1038/s41588-021-00870-7

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

40 Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017). https://doi.org:10.1093/aje/kwx246

Chapter 3:

Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency.

Saleh Shekari, Stasa Stankovic, Eugene J. Gardner, Gareth Hawkes, Katherine A. Kentistou, Robin N. Beaumont, Alexander Mörseburg, Andrew R. Wood, Gita Mishra, Felix Day, Julia Baptista, Caroline F. Wright, Michael N. Weedon, Eva Hoffmann, Katherine S. Ruth, Ken Ong, John R. B. Perry and Anna Murray

Main text

Abstract

Premature ovarian insufficiency (POI) affects 1% of women and is a leading cause of infertility. It is often considered to be a monogenic disorder, with pathogenic variants in ~100 genes described in the literature. We sought to systematically evaluate the penetrance of variants in these genes using exome sequence data in 104,733 women from the UK Biobank, 2,231 (1.14%) of whom reported natural menopause under the age of 40. We found limited evidence to support any previously reported autosomal dominant effect. For nearly all heterozygous effects on previously reported POI genes we ruled out even modest penetrance, with 99.9% (13,699/13,708) of all protein truncating variants found in reproductively healthy women. We found evidence of haploinsufficiency effects in several genes, including TWNK (1.54 years earlier menopause, P=1.59*10-6) and SOHLH2 (3.48 years earlier menopause, P=1.03*10-4). Collectively our results suggest that for the vast majority of women, POI is not caused by autosomal dominant variants either in genes previously reported or currently evaluated in clinical diagnostic panels. Our findings, plus previous studies, suggest that the majority of POI cases are likely oligogenic or polygenic in nature, which has major implications for future clinical genetic studies, and genetic counselling for families affected by POI.

Introduction

Premature ovarian insufficiency (POI) is the loss of ovarian activity and permanent cessation of menstruation occurring before the age of 40[1]. It represents a major cause of female infertility, affecting 1 in 100 women[1-4]. Some POI cases are syndromic, in which POI accompanies other phenotypic features, such as in Turner's syndrome. Genetic causes of POI have been reported in 1-10% of cases, while other causes include autoimmune and iatrogenic[5-7]. Approximately 50-90% of POI cases are idiopathic[8,9], 10-30% of those being familial, suggesting a genetic basis. Furthermore, heritability estimates of menopausal age from mother-daughter pairs range from 44% to 65%[10,11], and there is six times increased risk of early menopause in daughters of affected mothers[12,13]. A genetic diagnosis can provide important information to families about the risks of POI as well as the aetiology of the condition.

More than 100 monogenic causes of POI have been reported, where a single genetic variant is sufficient to cause the phenotype, with approximately half showing an autosomal dominant (AD) inheritance pattern (e.g., *BNC1*, *FANCA* and *NOBOX*). Variants in other genes are described as being inherited in an autosomal recessive (AR) manner, requiring both copies of the gene to be disrupted in order to cause the phenotype (e.g., *HFM1*, *LARS2* and *MCM8*). In addition to the autosomal genes, X chromosome genes have long been suggested to play an essential role in the maintenance of ovarian development and function, with X chromosome structural variants representing about 13% of POI cases in some published series[8,14,15].

More recently, GWAS have identified ~300 common genetic variants associated with population variation in the timing of menopause[11,16]. These studies have provided evidence that some POI cases may be polygenic in nature[11], where women inherit large numbers of common alleles associated with earlier menopause that, when combined with other risk factors, could push them into the extreme end of the phenotypic distribution.

With decreasing cost and improved analytical pipelines, whole exome sequencing (WES) is increasingly being used in clinical settings as a powerful diagnostic tool, including for POI[17-21]. However, the reported evidence for causal POI genes and variants is inconsistent, often based on small numbers of families or individuals and with variable degrees of functional validation[21]. As

genetic testing becomes more widespread in both clinical and non-clinical settings, there is an increasing need to better understand the phenotypic consequences of finding variants in these genes to help ensure appropriate advice and treatment is offered to women. Therefore, we aimed to assess the penetrance of variants in genes previously reported to cause POI in a general population study. We focused on the POI genes that are part of the Genomics England diagnostic gene panel for POI, an expert-reviewed and publicly available panel database, which we additionally supplemented with literature-reported POI genes. Our results indicate that the reported autosomal dominant (AD) causes of POI are likely to be either only partially penetrant or not pathogenic. Furthermore, we conclude that most cases of menopause under 40 years are likely to be multifactorial.

Methods

Identification of reported POI genes

In order to identify relevant gene candidates reported to cause POI, we initially focused on the POI gene panel available through Genomics England Panel App, a publicly accessible virtual panel database (https://panelapp.genomicsengland.co.uk/panels/155/). This panel was selected as the 'gold standard' resource as it is the most thoroughly curated one, reviewed by 12 professional clinical geneticists. We considered the following evidence as part of our gene evaluation: (1) Selection and categorisation: inheritance and phenotype, and (2) Number of reviews and gene ranking based on their traffic light system. This includes "RED" genes that do not have enough evidence for the association with the condition and should not be used for clinical interpretation, "AMBER" genes with moderate evidence that should not be yet used for the interpretation, and "GREEN" genes with high level of evidence, which demonstrates the confidence that this gene should be used for clinical interpretation (Supplementary Table 1). In total, we identified 67 genes: 28 green, 23 amber and 16 red. We reviewed the evidence provided on the Genomics England Panel App webpage for these genes and identified the specific genetic variants reported as associated with the phenotype (Supplementary Table 2).

This list was additionally supplemented with 38 manually curated POI genes reported in the literature. The search was performed using PubMed and Google Scholar, focusing on original articles published up to June 2022. The keyword combinations included 'premature ovarian failure', 'primary ovarian insufficiency', 'premature ovarian insufficiency', 'early menopause', 'premature menopause', 'POI', 'POF', 'infertility', 'hypergonadotropic hypogonadism', 'ovarian dysgenesis', 'genetic variants', 'sequencing', and 'primary amenorrhea'. Studies were also identified by a manual search of original publications described in review articles. Where appropriate, reference lists of identified articles were also searched for further relevant papers. Identified articles were restricted to English-language full-text papers. Studies were included according to following criteria: (1) the phenotype of interest was described as POI, primary or secondary amenorrhea, (2) one or more affected individuals for particular

causal variant were identified, (3) the focus was on either the autosomes or the X chromosome, (4) genetic variants were discovered by traditional family segregation studies, consanguineous pedigree analysis, unrelated cohort studies on whole exome (WES)/targeted next-generation sequencing data, (5) variant discovery was supported by validation in animal models and/or cell-based assays. We excluded studies that: (1) described hypothalamic pituitary adrenal axis and/or puberty-related phenotypes, (2) genes that were discovered through genome-wide association studies due to the lack of statistical power as a result of small sample sizes and the challenge to locate causative genes, and finally (3) genes that were discovered through array analysis due to the high inconsistency of the results coming from varied resolution of arrays across studies and thus uncommon replications. We recorded and analysed genes described for either non-syndromic or syndromic POI, however, the main focus of our paper was on genes associated with non-syndromic POI. Papers that exclusively reported the role of candidate genes in animal models were only used as supporting evidence when assessing the functional evaluation of the gene and to guide our conclusions. Following information was extracted from each study: (1) Publication info: PMID, (2) Inheritance: autosomal dominant (AD), autosomal recessive (AR) or X-linked, (3) Sample size: number of the genetic variant carriers, cases versus controls, if reported, and (4) Genetic variant info: genomic position, transcript and protein sequence (Supplementary Table 2). If the data were missing from published papers, relevant information was obtained by direct communication with the corresponding authors. In cases where response was not received, the information was recorded as NA. All data were extracted independently by two authors (S. Shekari and S. Stankovic).

Overall, we identified 105 unique POI genes that we classified according to their mode of inheritance. This includes 67 validated genes rated as either 'GREEN' (high level of evidence for disease association), 'AMBER' (moderate evidence) or 'RED' (not enough evidence) on the Genomics England POI Panel App (version 1.67). We also identified a further 38 genes reported as being causal for POI (Supplementary Table 1).

Genes were considered as inherited through the AD pattern if the reported variants in the heterozygous state were sufficient to cause POI, leading to 40 genes in total. Of those, seven were reported to act through the LoF mechanism

only, while in 34 genes, both LoF and missense genetic alterations caused the phenotype. If variants in both copies of the gene were necessary for the phenotype development, the gene was classified as AR (N=57). For two genes (*POLG*, *REC8*), both dominant and recessive causes were identified, and so we investigated them with other AD genes, while seven genes had an X-linked inheritance pattern.

Constraint metric of pathogenicity

We annotated each gene identified with the Genome Aggregation Database (gnomAD) v2.1.1 predicted constraint metric of pathogenicity to identify genes that are subject to strong selection against PTV variation[24]. The metric encompassed observed and expected variant counts per gene, observed/expected ratio (O/E) and probability of loss of function intolerance (pLI) (Supplementary Table 1). In short, the observed count represents the number of unique SNPs in each gene (MAF < 0.1%), while expected count relies on a depth-corrected probability prediction model that takes into account sequence context, coverage and methylation to predict expected variant count. The O/E is a continuous measurement that assesses how tolerant a gene is to a certain class of variation. Low O/E value indicates that the gene is under stronger selection for that class of variation. Finally, the pLI score reflects the constraint or intolerance of a given gene to a PTV variation, with a score closer to 1 indicating that the gene cannot tolerate PTV variation.

UK Biobank Data Processing and Quality Control

To perform rare variant burden analyses described in this study, we accessed Whole Exome Sequencing data (WES) for 454,787 individuals from the UK Biobank study[44]. Details of this study, including data collection and processing, are extensively described elsewhere[45]. Informed consent was provided by all participants. Study approval was received from the National Research Ethics Service Committee North West–Haydock and all study procedures were performed according to the World Medical Association Declaration of Helsinki ethical principles for medical research.

WES data were generated with the IDT xGen Exome Research Panel v1.0, which targeted 39Mbp of the human genome with mean coverage exceeding 20x on 95.6% of sites. The OQFE protocol was used for mapping and variant calling to the GRCh38 reference. Quality control filters applied by UK Biobank were individual and variant missingness <10% and Hardy Weinberg Equilibrium P-value >10^{-15}. In addition, we excluded variants with <10X coverage in 90% of the samples that were provided by Backman *et al.* [44]. We selected variants in the Consensus CDS (CCDS) transcripts and variants were annotated using the Ensembl Variant Effect Predictor[46] and LOFTEE plugin (https://github.com/konradjk/loftee). Minor allele frequency (MAF) was calculated using PLINK[47]. Furthermore, for homozygous variants, we manually assessed the variants using the Integrative Genomics Viewer[48,49]. Analyses were performed on the UK Biobank Research Analysis Platform (RAP; https://ukbiobank.dnanexus.com/).

Phenotype derivation

ANM was derived from self-reported questionnaire data as the age at last naturally occurring menstrual period, excluding those with surgical menopause (field 2824 and 3882) or taking hormone replacement therapy (field 3536), as described previously[11]. There were 104,733 female participants with ANM included in our analyses (range 18 to 65 years, mean=50.1, SD=4.5), of whom 2,231 individuals reported ANM under 40 years. During the data collection process, participants who reported ANM under 40 years on the questionnaire were asked to confirm their ANM. For comparisons of variant counts, we identified a control cohort of women with ANM at ≥40 years, including those who reported still menstruating (n=192,438). Analyses were performed in Stata: Release 16 on the UK Biobank RAP.

Primary exome-wide association analysis

In order to perform rare variant burden tests, we used the REGENIE regression algorithm (REGENIEv2.2.4; https://github.com/rgcgithub/regenie). REGENIE implements a generalised mixed-model region-based association test that can

account for population stratification and sample relatedness in large-scale analyses. REGENIE runs in 2 steps[50], which we implemented on the UKBiobank RAP: In the first step, genetic variants are aggregated into gene-specific units for each class of variant called masks: high confidence protein-truncating variants included stop-gain, frameshift, or abolishing a canonical splice site (-2 or +2 bp from exon, excluding the ones in the last exon); non-synonymous (missense) variants with CADD score > 25; damaging that included high confidence protein-truncating variants or/and non-synonymous variants with CADD score >25. The three masks were tested for association with ANM in the second step. As described previously, in our analyses we included individuals identified as European, excluding participants who had subsequently withdrawn from the study and those for whom self-reported sex did not match genetic sex[51]. We applied an inverse normal rank transformation to ANM and included recruitment centre, sequence batch and 40 principal components as covariates. We transformed the effect estimates from our analyses to approximate values in years by multiplying by the standard deviation of ANM in our study cohort (4.53 years). Analyses were performed on the UK Biobank RAP. To identify significant gene associations we Bonferroni corrected P<0.05 for the number of masks (n=3) and genes tested (n=105) giving a significance threshold of $P<1.6*10^{-4}$ ($P=0.05/(3*105)=1.6*10^{-4}$).

In a similar way, we used REGENIE to test the association of individual genetic variants reported in the literature with ANM. Variants with allele count >5 were tested in an additive model, applying an inverse normal rank transformation to ANM and including recruitment centre, sequence batch and 40 principal components calculated by UK Biobank as covariates. Of 421 uniquely identified variants,182 were present in the UK Biobank.

Replication and sensitivity analyses

A second analysis team (Cambridge) independently performed analyses of WES data in UK Biobank (Supplementary Table 10). The ANM phenotype was derived as described in Stankovic *et al.* (2022)[27]. Briefly, a different approach was used to generate the phenotype by handling data from multiple visits and missing data differently to the main method of generating the phenotype. This resulted in 106,973 female individuals for analyses. All manipulations were conducted in R (v4.1.2) on the UK Biobank RAP. Rare variant burden tests of

functional variant categories (defined as for main analyses) were performed using a custom implementation of BOLT-LMM v2.3.6[52] for the UK Biobank RAP using MANE Select plus Clinical transcripts v0.93 in VEP v104, as described in Stankovic *et al.* (2022)[27]. Analyses used a winsorised ANM phenotype, with everyone reporting ANM at younger than 34 years given a value of 34. Analyses were adjusted for age, age^2, sex, and the first ten genetic principal components as calculated in Bycroft *et al.*[53] and study participant exome sequencing batch as a categorical covariate (either 50k, 200k, or 450k).

We repeated our primary REGENIE analysis, including only variants with MAF<0.01% and secondly excluding HC-PTVs in the last exon or in the last 50bp of the penultimate exon (results presented in Supplementary Table 3). A further sensitivity analysis of HC-PTVs with MAF <0.001 in gnomAD (https://gnomad.broadinstitute.org/) was carried out in BOLT-LMM. In order to test the prevalence of damaging variants in our target genes in a clinically diagnosed cohort, we identified 113 women with POI from primary care and hospital record data, including 12 with primary amenorrhea. Clinical cases were selected with relevant codes, including primary ovarian failure, secondary ovarian failure, other ovarian failure, ovarian agenesis, ovarian hypogonadism, primary amenorrhoea, and a self-reported natural age at menopause <40 years for those without primary amenorrhea.

Gene-set burden analysis

We ran gene-set burden tests by collapsing the genes of interest and their variants into one unit for analysis. The gene-set burden tests were performed by extending an association testing workflow of applets designed for the UK Biobank RAP for single genes to gene sets. The RAP association workflow is described in detail in Gardner *et al.*, 2022[54]. In total, we conducted four gene-set burden tests, collapsing variants and genes into the following categories: (1 AD only genes (N=38), (2) AR genes (N=57), (3) genes with both AD and AR inheritance (N=2), and (4) all 105 genes (Supplementary Table 11).

Briefly, for each of the gene sets, we included variants with MAF < 0.1% that were HC-PTVs, as predicted by the LOFTEE tool[24]. For each gene set, we ran two related approaches. Firstly, we implemented a generalised linear model

(GLM) using the Python package 'stats models' [55]. For the GLM, the number of variant alleles across the gene set was summed up into a single score under a simple additive model. This score was used as a predictor of the ANM phenotype in a three-step regression.

Secondly, we ran the STAAR method (implemented in R package "STAAR")[25]. This method corrects for population stratification by including a genetic relatedness matrix (GRM) in the test framework. The GRM used was based on pre-computed autosomal kinship coefficients from Bycroft *et al.* [53]. For each STAAR test, the genotype information was represented by a single n*p matrix where n was the sample size, and p was the number of included genetic variants across all genes of interest. For all association tests, we corrected for age, age^2, the first ten genetic principal components provided by Bycroft *et al.* [53] and study participants' WES batch as a categorical covariate.

Frequency of homozygous or compound heterozygous LOF individuals

We estimated the frequency of homozygous or compound heterozygous HC-PTV individuals for each gene as *F*^2, where *F* is the frequency of individuals with any high-confidence HC-PTV allele with MAF<0.1% in a gene as estimated from the primary analysis (Supplementary Table 3). To find the total frequency of individuals with homozygous or compound heterozygous HC-PTVs, we then summed *F*^2 for the 105 POI genes reported in the literature.

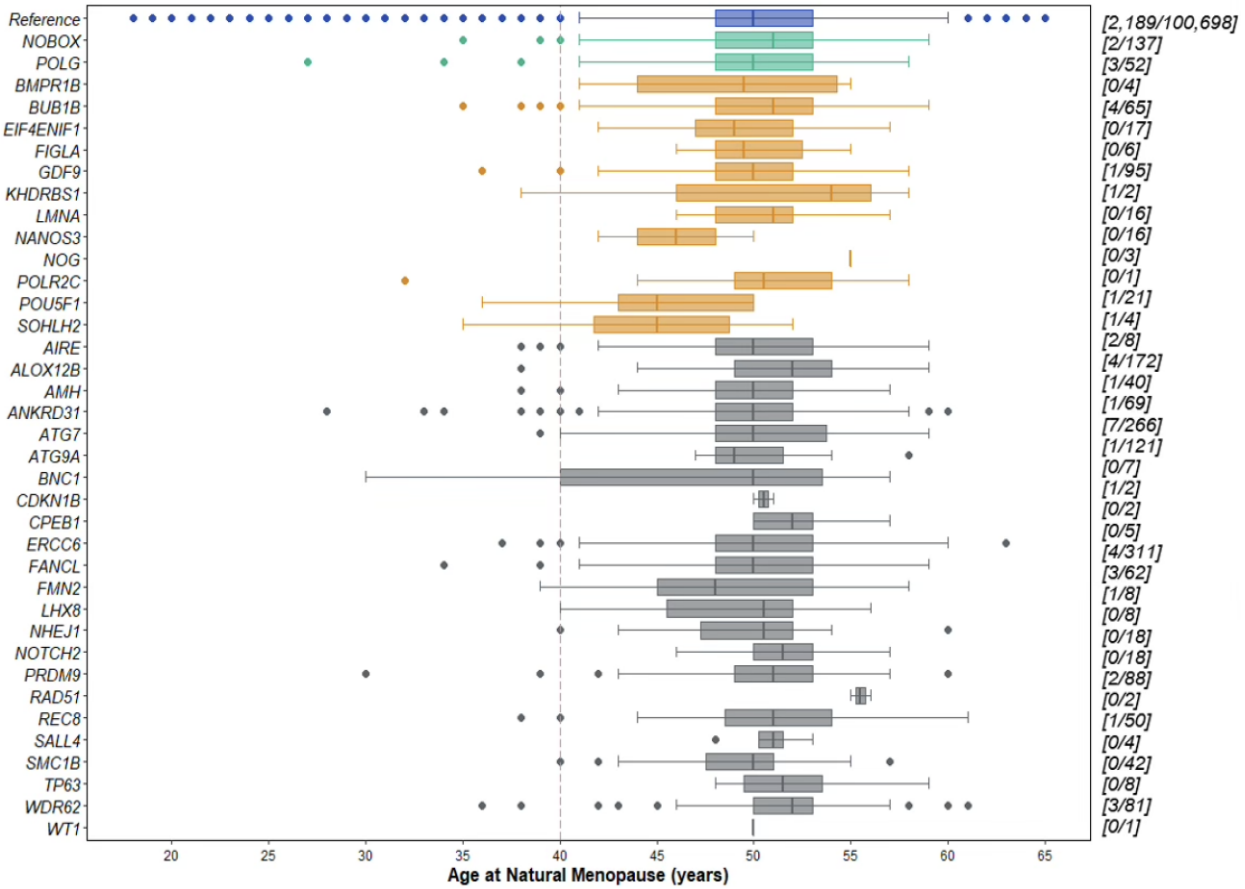
The expected frequency of having a gene with a homozygous or compound heterozygous LOF knockout is 6 per billion individuals, based on the median frequency in gnomAD[26]. From this estimate we would expect 1.2 per 10,000 people to carry a homozygous or compound heterozygous LOF knockout in any of the ~20,000 genes in the genome ($20000*6*10^{-9}=1.2*10^{-4}$). Assuming 100% penetrance, the number of genes with a homozygous or compound heterozygous LOF knockout that would be needed to reach the observed 1% frequency of POI in the population (1 per 100 individuals) would be $0.01/6*10^{-9}=1.7*10^6$ genes.

Results

Heterozygous damaging variants do not often cause POI.

The Genomics England POI Panel App (version 1.67) includes 67 validated genes rated as either 'GREEN' (high level of evidence for disease association), 'AMBER' (moderate evidence) or 'RED' (not enough evidence). We also identified a further 38 genes reported as being causal for POI. We classified these 105 genes according to the reported mode of inheritance (Supplementary Table 1). We then identified genetic variants in these 105 putative POI genes using WES data available in 104,733 UK Biobank post-menopausal female participants of European genetic ancestry[22], of which 2,231 reported age at natural menopause (ANM) below the age of 40. High confidence protein truncating variants (HC-PTVs) were found in 100 genes, but never only in the cases: there were 41 women with menopause under 40 years (ANM range: 27-39, mean ANM: 36.4, SD=3.2) who had an HC-PTV in at least one of the 40 genes reported to be autosomal dominant, but these variants were also detected in 1,817 women with ANM over 40 years (ANM range: 40-63, mean ANM: 50.4, SD=3.9). For three of the 40 POI genes (*BMPR1A*, *FOXL2* and *NR5A1)* there were no HC-PTVs carriers in either cases or controls, but for all 37 genes with HC-PTVs, the median ANM for those with heterozygous loss of function (LOF) alleles was between 45 and 56 years (Figure 1, Supplementary Table 3).

Figure 0–1 Age at natural menopause in women with HC-PTVs in POI genes reported to have an autosomal dominant pattern of inheritance.



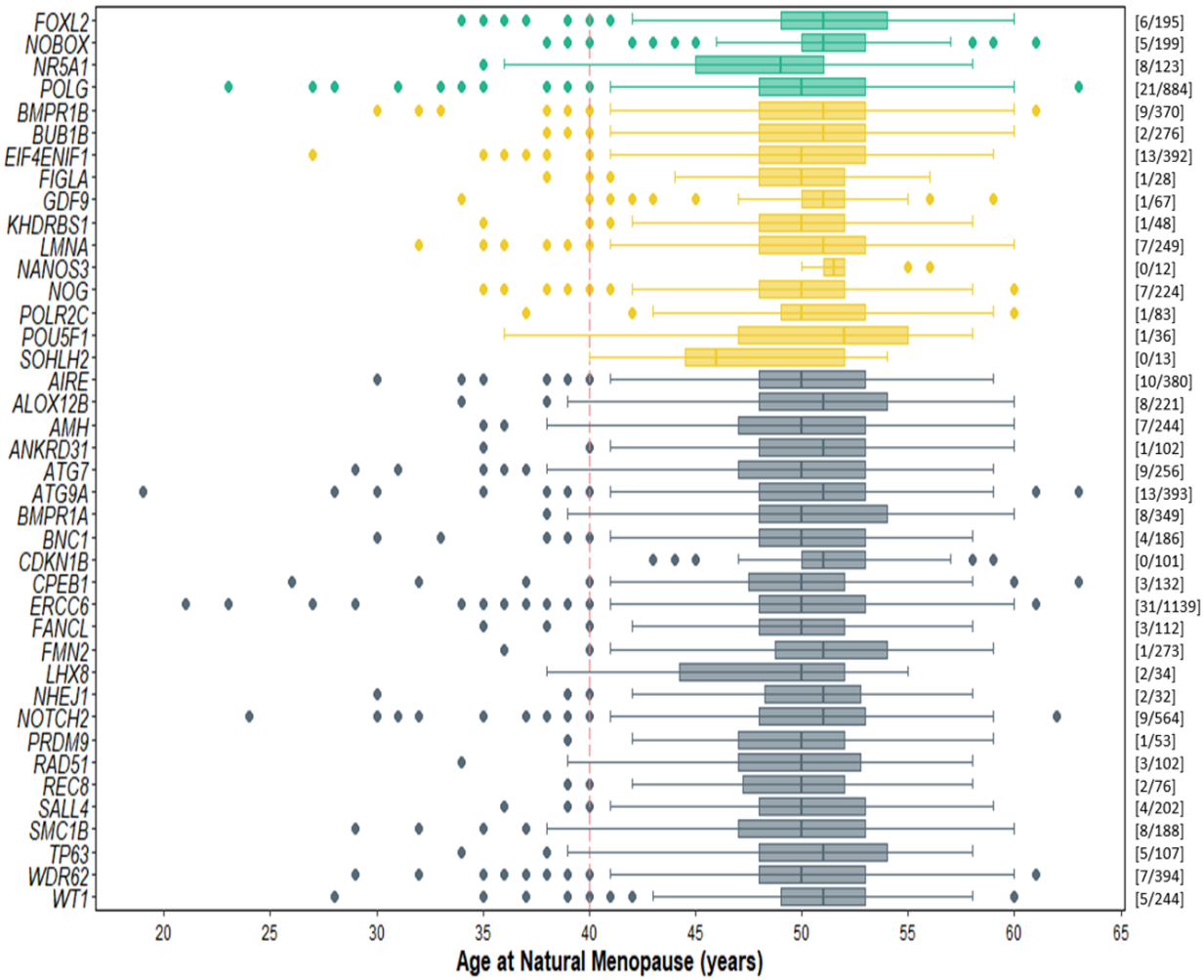
Note: Genes are coloured by the strength of evidence for POI in either the Genomics England Panel App (Green or Amber; no HC-PTV were detected in red genes) or our own manual curation of the POI literature (grey; N = 23). The number of women with ANM <40 (cases) compared to >40 years (controls) is shown in brackets on the right Y axis [cases/controls].

Women who did not carry an HC-PTV are indicated in blue (a histogram of the distribution of ANM for these women is included in supplementary figure 1). In the plot, the boxes show the values of the lower quartile, median and upper quartile; the whiskers show the most extreme value within a distance of 1.5 times of the interquartile range from the lower and upper quartiles, respectively; outliers are shown as individual points.

The intolerance for individual genes to harbour protein truncating variation, also known as genic 'constraint', has previously been linked to reproductive success[23]. Our results demonstrate that the majority of AD POI genes (26/40, 67.5%) have limited evidence of being under strong selective constraint (pLI ≤ 0.9) as assessed by gnomAD[24], which further supports that these genes are unlikely to play an important role for reproductive success.

Next, we tested individual variants in the 40 AD genes that have been previously reported to be pathogenic for POI (Supplementary Table 4). There were 153 variants reported, of which 126 were predicted to be missense and, of these, 58 (46%) were detected in our study with 37 only found in controls. A further 20 missense variants were found both in women with ANM under 40 years and controls, and only one missense variant was found only in cases (NM_002693.3:c.2828G>A [p.Arg943His] in *POLG*); however, the burden tests of all HC-PTVs or deleterious missense variants in *POLG* were not associated with menopause timing (P=0.7 and P=0.05, respectively; Supplementary Table 5). Therefore, while the variant in cases alone could have a gain of function or dominant negative effect, the finding is also consistent with chance. Having tested reported 'pathogenic' missense variants in the 40 AD genes, we tested all missense variants with MAF<0.1% in UK Biobank. We next collated a broader set of 17,374 rare missense variants in the 40 AD genes, including 2,740 with CADD score >25 (Figure 2, Supplementary Table 8) and 1,120 with REVEL score >0.7 (Figure 3, Supplementary Table 9). We identified no robust associations with ANM for any of these individual variants (all were $P>3.11*10^{-4}$ and so above our threshold for multiple testing of all missense variants with AC>5; $0.05/4,737=1.06*10^{-5}$). These results support our previous observation that POI genes are generally not pathogenic in the heterozygous state.

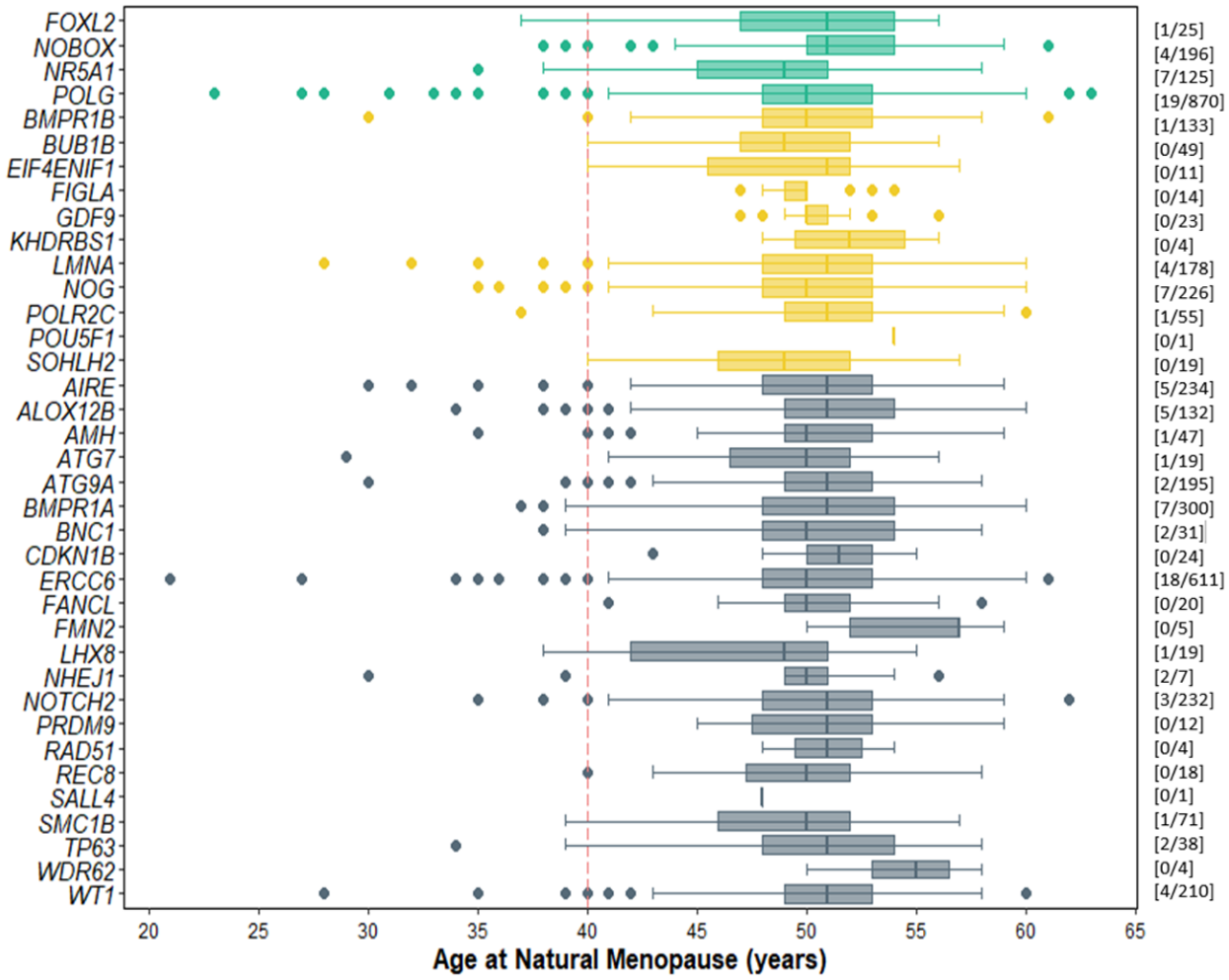
Figure 0–2: Range of age at natural menopause in carriers of missense variants with CADD score greater than 25 in genes reported to have an autosomal dominant pattern of inheritance.



Notes: 17 genes were identified as 'monoallelic' in Genomics England (GeL) Panel App and are coloured according to the strength of evidence categories: "GREEN", and "AMBER" (supplementary table 2). In addition,24 genes were reported in the literature to be a likely monogenic cause of POI in the heterozygous state but were not included on the Panel App (coloured grey). The numbers in brackets in the right corner reported as part of each panel represent [N POI cases/N controls] of women carrying HC PTVs in each gene.

Note: In the plot, the boxes show the values of the lower quartile, median and upper quartile; the whiskers show the most extreme value within a distance of 1.5 times of the interquartile range from the lower and upper quartiles, respectively; outliers are shown as individual points.

Figure 0–3: Range of age at natural menopause in carriers of missense variants with REVEL score greater than 0.7 in genes reported to have an autosomal dominant pattern of inheritance.



Notes: 17 genes were identified as 'monoallelic' in Genomics England (GeL) Panel App and are coloured according to the strength of evidence categories: "GREEN", and "AMBER" (supplementary table 2). In addition,24 genes were reported in the literature to be a likely monogenic cause of POI in the heterozygous state but were not included on the Panel App (coloured grey). The numbers in brackets in the right corner reported as part of each panel represent [N POI cases/N controls] of women carrying HC PTVs in each gene.

Note: In the plot, the boxes show the values of the lower quartile, median and upper quartile; the whiskers show the most extreme value within a distance of 1.5 times of the interquartile range from the lower and upper quartiles, respectively; outliers are shown as individual points.

Due to the relatively small number of protein-truncating variants (PTVs) found within individual genes, to try to increase our statistical power to find any association with POI, we considered the aggregated effect of all PTVs with similar proposed genetic architecture across all putative POI genes. This included a test for (1) AD-only genes (N=38), (2) autosomal recessive (AR) only genes (N=57), (3) genes with both AD and AR inheritance (N=2), and (4) all 105

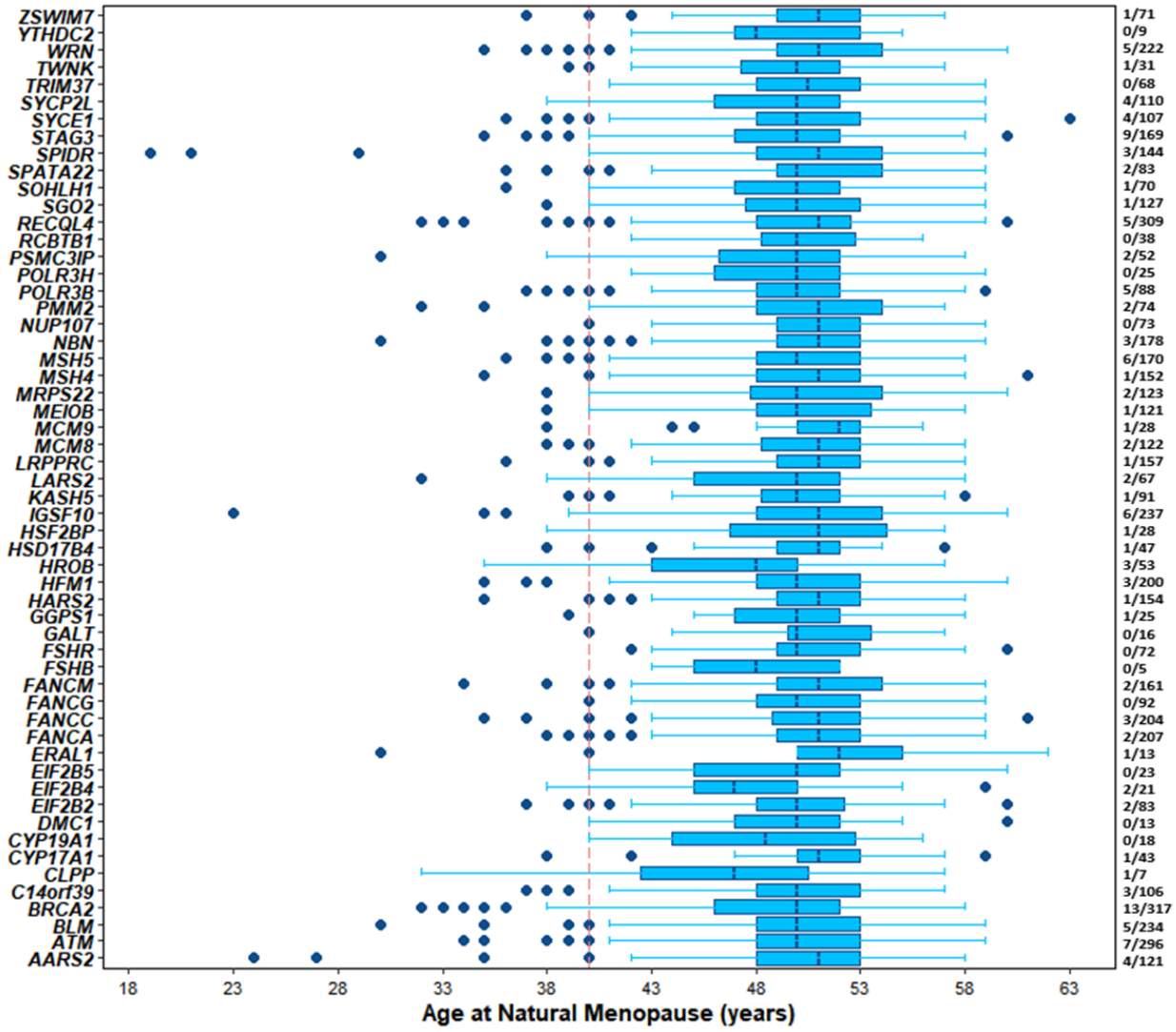
POI genes. None of the tests were associated with ANM at P<0.05, in either a generalised linear model or STAAR Omnibus statistical models[25] (Methods; Supplementary Table 11).

No evidence of haploinsufficiency as a cause of POI

Of the 105 reported monogenic POI genes assessed in our study, 57 were reported to show AR inheritance and a further eight were X-linked. We were unable to evaluate recessive effects as we identified only two women with homozygous HC-PTVs: one with a PTV in *SOHLH1* (NM_001101677.2:c.346-1G>A) with menopause at 45 years and one in *AIRE* (NM_000383.4:c.967_979del [p.Leu323SerfsTer51]) who reported menopause in her 20s. Furthermore, we were unable to identify compound heterozygotes. Instead, by considering HC-PTV allele frequencies in our analyses, we would expect 0.003% of individuals (~4 in the current study) to be homozygous or compound heterozygous for a high-confidence LOF variant in any of the 105 POI genes. This is likely a conservative estimate given we might expect POI genes to be less tolerant than other genes to deleterious alleles as these would impact reproductive fitness. Based on frequencies of gene knockout carriers in gnoMAD[26], we estimate that even if all genes in the genome were true recessive causes of POI (and thus not detected by our study), the population prevalence of carrying a gene knockout would be 100 times smaller than the observed prevalence of POI.

We next hypothesised that there may be an effect on ANM in heterozygous carriers of deleterious variants in these POI recessive genes. In total we identified 122 carriers of HC-PTVs in the 65 recessive or X-linked genes among cases with ANM < 40 years, but also 5,585 carriers among controls (Supplementary Table 3). There was no evidence that haploinsufficiency of any recessive POI gene is sufficient to cause POI (Figure 4).

Figure 0–4: Age at natural menopause in carriers of HC-PTVs in POI genes reported to have an autosomal recessive pattern of inheritance.



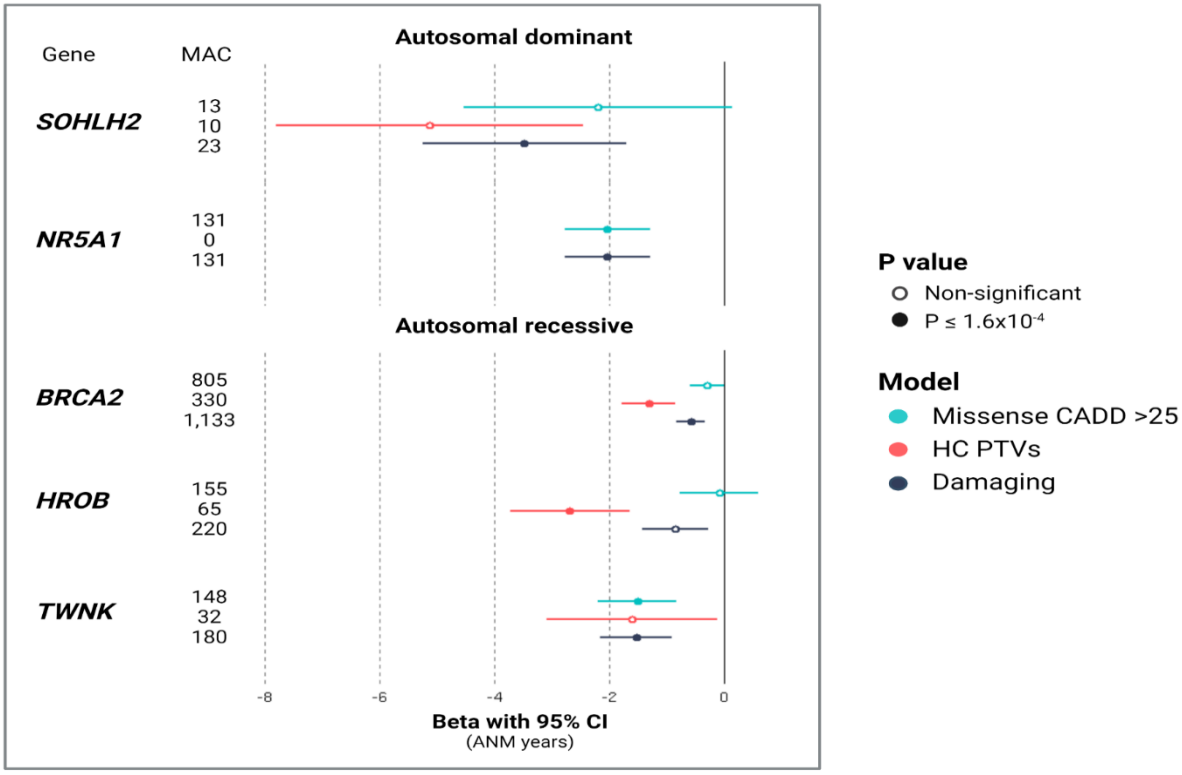
Notes: 65 genes were identified as 'biallelic' in Genomics England (GeL) Panel App (supplementary table 2). [N POI cases/N controls] of women carrying HC PTVs in each gene.

Note: In the plot, the boxes show the values of the lower quartile, median and upper quartile; the whiskers show the most extreme value within a distance of 1.5 times of the interquartile range from the lower and upper quartiles, respectively; outliers are shown as individual points.

Finally, we assessed whether protein-coding variation in any of the 105 monogenic POI genes altered ANM within the normal range. In gene burden tests we grouped genetic variants with MAF < 0.1% into three functional categories: (1) HC-PTVs, (2) missense variants with CADD score ≥ 25, and (3) a combination of 1 and 2, termed 'damaging' variants. For 100 of the 105 POI genes, we did not find an association with ANM ($P<1.6*10^{-4}$; P=0.05/(3 tests ×

105 genes) (Supplementary Table 5). For two AR genes, we have previously reported an effect on ANM: *BRCA2* (P=2.6*10^{-8}; beta:1.32 years earlier ANM [95% CI: -1.79, -0.85]) and *HROB* (P=4.7*10^{-7}; beta: 2.69 years earlier ANM [95% CI: -3.73, -1.65])[27]. There were associations with earlier ANM for a further two AD and one AR genes, with at least one of the variant categories passing our threshold for multiple testing (P<1.6*10^{-4}, Figure 5). These were for damaging variants in *TWNK*, a mitochondrial helicase involved in mtDNA replication and repair (P=1.59*10^{-6}; beta: 1.54 years earlier ANM [95% CI: -2.17,-0.91]; N = 180)[28,29], *NR5A1*, a key gene for gonadal function (P=5.8*10^{-8}; beta: 2.04 years earlier ANM [95% CI: -2.79, -1.30]; N = 131)[30], and *SOHLH2*, a transcription factor involved in both male and female germ cell development and differentiation (P=1.03*10^{-4}; beta: 3.48 years earlier ANM [95% CI: -5.24, -1.72]; N = 23)[31,32].

Figure 0–5: Gene burden associations with age at natural menopause.



Notes: Results are plotted for genes that passed the Bonferroni corrected threshold for 105 genes, each with 3 masks (P<1.6x10^{-4}). There were no HC-PTV carriers for NR5A1.

Discussion

Many genes in the literature have been reported as monogenic causes of POI and are included in diagnostic panels for clinical use[33]. Our literature review identified 105 putative monogenic POI genes, 67 that were included in the Genomics England open access Panel App resource[34] and 38 additional reported genes. Of these 105 genes, 40 are reported to be inherited in an AD fashion. Using UK Biobank exome sequence data in 104,733 post-menopausal women, we found no evidence to support heterozygous HC-PTV of any of these genes as a highly penetrant cause of POI; for each gene the average menopause age for carriers of LOF variants was over 40, with the ANM distribution broadly similar to that of non-carriers. This includes two green Panel App genes - *NOBOX* and *POLG* - where 137/139 and 52/55 of the identified PTV alleles, respectively, were found in controls. Our previous work demonstrated that heterozygous LOF of *ZNF518A* has the largest effect in the protein-coding genome on menopause timing [27], yet carriers report menopause only 6 years earlier than non-carriers, with only 12% experiencing POI. Taken together, our observations suggest that fully, or even largely, penetrant autosomal dominant effects are likely to cause very few cases of POI.

Although heterozygous LOF variants were not penetrant causes of POI, our study suggests that carrying rare coding variants in five of the POI genes can substantially lower an individual's menopause age. Besides previously reported *BRCA2* and *HROB*, three genes, including *NR5A1*, *SOHLH2* and *TWNK,* have not been described as associated with menopause timing in the general population. The effect ranged from 5.13 years earlier menopause for carriers of rare LOF variants in *SOHLH2* to 1.54 years earlier for damaging variants in *TWNK*. We did not identify any heterozygous PTV alleles in *NR5A1*, which is a highly constrained gene (Supplementary Table 1). Therefore, the observation that rare missense variants are associated with a ~2 year reduction in menopause timing suggests that dominant LOF may well be a penetrant, albeit very rare, cause of POI. As for *NR5A1*, *TWNK* is also on the green gene on the Panel App list, but has a reported recessive pattern that causes syndromic POI (Perrault syndrome) and presents in association with other neurologic symptoms[35].

Our findings should be interpreted in the context that the published evidence to support causality of genes and variants for POI is highly variable. Guidelines are available for genomic variant interpretation[36], but many of these genes were reported before such guidelines, making it difficult for non-specialists to interpret the findings. Many studies were based on candidate gene approaches with small numbers of cases or families[18] and in the absence of large-scale reference data or ancestry match controls. More recent POI studies have used exome sequencing, but often revert to candidate gene approaches with relaxed statistical thresholds when no exome-wide association is identified[17,20,37,38]. Furthermore, when studying individual genomes it is also inherently challenging distinguishing between pathogenic variants and private non-functional variants. Functional studies can be informative, but the design and rationale of such studies can be circular. For example, a DNA damage response (DDR) gene that harbours a private variant may be selected as a reasonable candidate, but the downstream functional work is limited to DDR measures, rather than reproductive or ovarian phenotypes. Future studies that aim to investigate novel genetic causes of POI should focus on approaches that more specifically mimic human biology and physiology. Patient-specific induced pluripotent stem cells (iPSCs) lines might offer an individually targeted genetic model for identification, manipulation and better understanding of reproductive biological pathways.

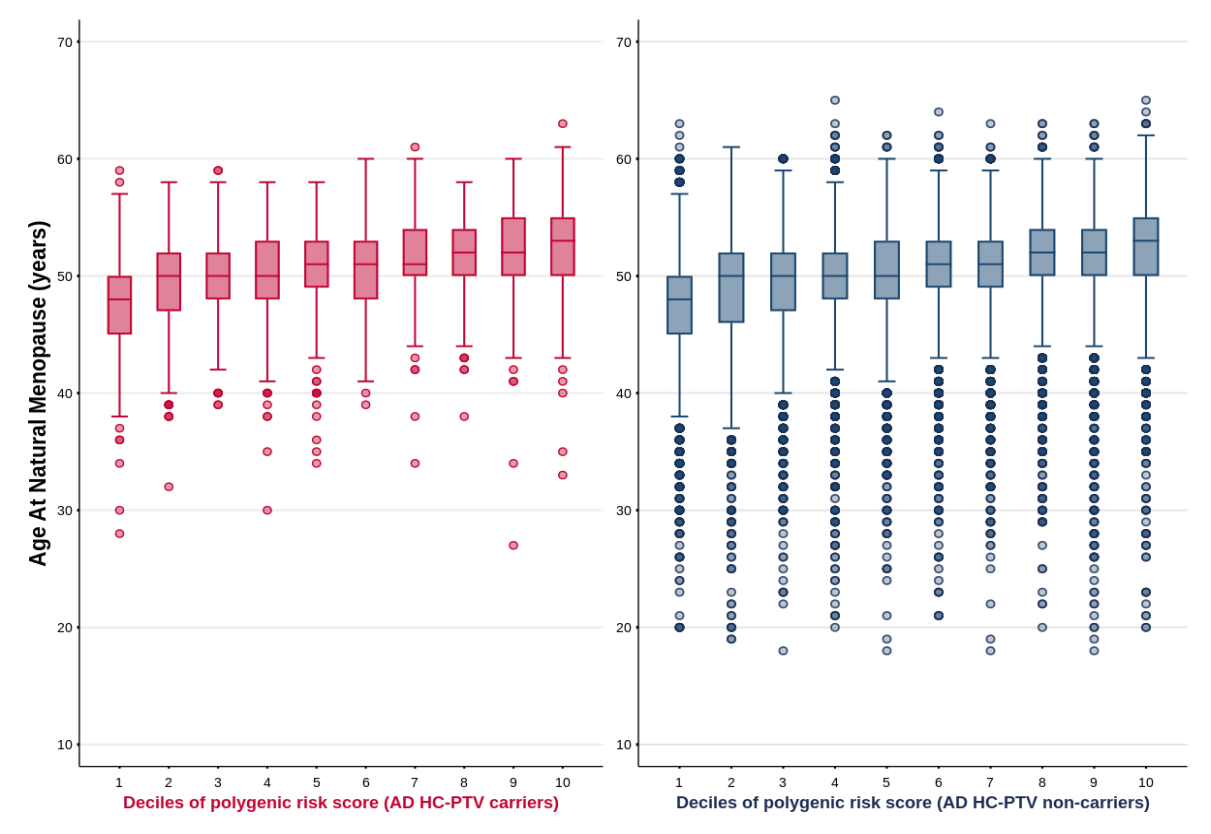
Our study has assessed one of the largest samples to date of women with menopause before 40. A major strength is the analyses of exome sequence data in over 100,000 women with normal ANM, which provides invaluable data on normal genetic variation in a control population. Identification of alleles at high frequency in these samples provides confidence that they are unlikely to be a penetrant cause of POI. Our study does however have a number of limitations. Firstly, we have not investigated a clinically defined cohort of POI cases and not all women with menopause under 40 would be diagnosed as POI. A recent study of clinically defined POI cases reported a high proportion of predicted pathogenic variants in known genes[39], although it did not assess the penetrance of these variants and the rare variant burden association tests did not support the pathogenicity of many of the genes tested. We were not able to validate our penetrance findings in an independent cohort, but all genes tested were previously published as being associated with POI. Furthermore, the UK

Biobank is known to disproportionately include healthier participants, which has been shown to be the case for other conditions[40,41], although this tends to have a greater impact on men[42] and it is not obvious that having POI would influence participation in the study. While these issues will likely lead to underestimates of any potential effect sizes, they do not explain why previously reported pathogenic variants are overwhelmingly found in women with menopause over 40 years. Secondly, we were able to assess the penetrance only of heterozygous variants but not homozygous or compound heterozygous carriers. We also have not considered complex structural variants or cytogenetic abnormalities, so we make no statement on the penetrance of those. The most common variant previously associated with POI is the CGG triplet repeat expansion in *FMR1*. While this variant was not captured in our study, other studies have investigated the penetrance in population-based cohorts[43]. For five genes we did not identify any heterozygous LOF variants in our data so were unable to assess these, although they are unlikely causes of POI given they were not present in over 2000 women with ANM <40 years. Third, we predominantly focused on predicted LOF alleles as the mechanism implied or demonstrated in most studies. It is however possible that some of the literature reported missense variants may act in a gain of function or dominant negative manner such that they have more severe effects than LOF variants. Whilst potentially true of a small number, this is unlikely to be widespread given no highly penetrant effects were seen in the 58 individual literature-reported missense variants that we assessed, or in our burden tests for predicted damaging missense variants by CADD and REVEL. Finally, our study is specific to individuals of European ancestry. While the frequencies of many variants vary between populations, the functional impacts of LOF variants should be widely applicable.

In conclusion, our findings imply that monogenic causes of POI are unlikely for the vast majority of cases. Given our observed results for genes with a dominant mode of inheritance, we advise caution in interpreting reported recessive effects, although we predict this will be by far the most common cause of monogenic POI. Rather than representing a biologically distinct condition, we suggest that POI is part of a continuous distribution of ovarian ageing. Where women are in this distribution is likely determined by a

continuum of multiple risk factors, where the sum of many independent genetic and non-genetic risk factors place women into the tail of the phenotypic distribution. In this study we found that the effect of the polygenic score on menopause timing in women carrying PTVs in the dominant POI genes (beta 1.12; 95% CI 1.04-1.18) was highly similar to the effect in non-carriers (beta 1.11; 95% CI 1.09-1.13) (Figure 6). In a previous study we reported that women with the top 1% of a polygenic risk score, comprising common ANM-reducing alleles, have a five-fold increased risk of POI compared to the median[11]. Collectively, our findings suggest that POI should be considered a genetically complex trait for which genetic testing for monogenic causes is unlikely to be fruitful. Future efforts should address this genetic complexity in the development of new diagnostic approaches for POI to minimise potential mis-diagnoses and inappropriate genetic counselling.

Figure 0–6: Age at natural menopause in carriers of HC-PTVs in POI genes by decile of polygenic risk score.



Notes: Within each variant group (HC-PTV carriers and non-carriers, for autosomal dominants genes) individuals were grouped by decile of polygenic score for 290 ANM-associated common variants[11].
Note: In the plot, the boxes show the values of the lower quartile, median and upper quartile; the whiskers show the most extreme value within a distance of 1.5 times of the interquartile range from the lower and upper quartiles, respectively; outliers are shown as individual points (density of points reflects number of individuals)

Acknowledgements

This work was funded by the Medical Research Council (Unit programs: MC_UU_12015/2, MC_UU_00006/2, MC_UU_12015/1, and MC_UU_00006/1). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The authors acknowledge the use of the University of Exeter High-Performance Computing facility in carrying out this work, funded by a MRC Clinical Research Infrastructure award (MRC Grant: MR/M008924/1). This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. This research was conducted using the UK Biobank Resource under application 9905 (University of Cambridge) and 9072 and 871 (University of Exeter).

Saleh Shekari was supported by the QUEX Institute (University of Exeter, UK and the University of Queensland, Australia). Stasa Stankovic is supported by the Clare Hall Ivan D. Jankovic PhD scholarship from the University of Cambridge. Anna Murray, Caroline Wright and Michael Weedon are supported by the Medical Research Council (MR/T00200X/1). Katherine Ruth is supported by Cancer Research UK [grant number C18281/A29019]. Gareth Hawkes has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 875534. Gita Mishra is supported by National Health and Medical Research Council Investigator grant (APP2009577). Eva Hoffmann was supported by the ERC (724718-ReCAP), Novo Nordisk Foundation (NNF15COC0016662), the Independent Research Foundation Denmark (0134-00299B), and a grant from the Danish National Research Foundation Centre (6110-00344B).

Conflicts of Interest

John Perry and Eugene Gardner hold shares in and are employees of Adrestia Therapeutics. All remaining authors declare no competing interests.

Data Availability Statement

We used publicly available individual-level genotype and phenotype data from the UK Biobank (https://biobank.ndph.ox.ac.uk/showcase/). Access to these data needs to be requested from the UK Biobank.

References

1. Wesevich, V., Kellen, A.N. & Pal, L. Recent advances in understanding primary ovarian insufficiency. *F1000Res* **9**, F1000 Faculty Rev-1101 (2020).

2. Rudnicka, E.*, et al.* Premature ovarian insufficiency - aetiopathology, epidemiology, and diagnostic evaluation. *Prz Menopauzalny* **17**, 105-108 (2018).

3. Coulam, C.B., Adamson, S.C. & Annegers, J.F. Incidence of premature ovarian failure. *Obstet Gynecol* **67**, 604-606 (1986).

4. Golezar, S., Ramezani Tehrani, F., Khazaei, S., Ebadi, A. & Keshavarz, Z. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric* **22**, 403-411 (2019).

5. Harlow, B.L. & Signorello, L.B. Factors associated with early menopause. *Maturitas* **35**, 3-9 (2000).

6. Goswami, D. & Conway, G.S. Premature ovarian failure. *Hum Reprod Update* **11**, 391-410 (2005).

7. Szeliga, A.*, et al.* Autoimmune Diseases in Patients with Premature Ovarian Insufficiency-Our Current State of Knowledge. *Int J Mol Sci* **22**(2021).

8. Fortuño, C. & Labarta, E. Genetics of primary ovarian insufficiency: a review. *J Assist Reprod Genet* **31**, 1573-1585 (2014).

9. Venturella, R.*, et al.* The Genetics of Non-Syndromic Primary Ovarian Insufficiency: A Systematic Review. *Int J Fertil Steril* **13**, 161-168 (2019).

10. Perry, J.R.*, et al.* A genome-wide association study of early menopause and the combined impact of identified variants. *Hum Mol Genet* **22**, 1465-1472 (2013).

11. Ruth, K.S.*, et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393-397 (2021).

12. Qin, Y.*, et al.* ESR1, HK3 and BRSK1 gene variants are associated with both age at natural menopause and premature ovarian failure. *Orphanet J Rare Dis* **7**, 5 (2012).

13. van Kasteren, Y.M.*, et al.* Familial idiopathic premature ovarian failure: an overrated and underestimated genetic disease? *Hum Reprod* **14**, 2455-2459 (1999).

14. Pu, D., Xing, Y., Gao, Y., Gu, L. & Wu, J. Gene variation and premature ovarian failure: a meta-analysis. *Eur J Obstet Gynecol Reprod Biol* **182**, 226-237 (2014).

15. Persani, L., Rossetti, R., Cacciatore, C. & Bonomi, M. Primary Ovarian Insufficiency: X chromosome defects and autoimmunity. *J Autoimmun* **33**, 35-41 (2009).

16. Day, F.R.*, et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* **47**, 1294-1303 (2015).

17. Liu, H.*, et al.* Whole-exome sequencing in patients with premature ovarian insufficiency: early detection and early intervention. *J Ovarian Res* **13**, 114 (2020).

18. França, M.M. & Mendonca, B.B. Genetics of Primary Ovarian Insufficiency in the Next-Generation Sequencing Era. *J Endocr Soc* **4**, bvz037 (2020).

19. Jin, H.*, et al.* Identification of potential causal variants for premature ovarian failure by whole exome sequencing. *BMC Med Genomics* **13**, 159 (2020).

20. Patiño, L.C.*, et al.* New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Hum Reprod* **32**, 1512-1520 (2017).

21. Chapman, C., Cree, L. & Shelling, A.N. The genetics of premature ovarian failure: current perspectives. *Int J Womens Health* **7**, 799-810 (2015).

22. Szustakowski, J.D.*, et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet* **53**, 942-948 (2021).

23. Gardner, E.J.*, et al.* Reduced reproductive success is associated with selective constraint on human genes. *Nature* **603**, 858-863 (2022).

24. Karczewski, K.J.*, et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

25. Li, X.*, et al.* Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet* **52**, 969-983 (2020).

26. Minikel, E.V.*, et al.* Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459-464 (2020).

27. Stankovic, S.*, et al.* Genetic susceptibility to earlier ovarian ageing increases *de novo* mutation rate in offspring. *medRxiv*, 2022.2006.2023.22276698 (2022).

28. Korhonen, J.A., Gaspari, M. & Falkenberg, M. TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem* **278**, 48627-48632 (2003).

29. Khan, I.*, et al.* Biochemical Characterization of the Human Mitochondrial Replicative Twinkle Helicase: Substrate specificity, DNA branch migration and ability to overcome blockades to DNA unwinding. *J Biol Chem* **291**, 14324-14339 (2016).

30. Bashamboo, A. & McElreavey, K. NR5A1/SF-1 and development and function of the ovary. *Ann Endocrinol (Paris)* **71**, 177-182 (2010).

31. Choi, Y., Yuan, D. & Rajkovic, A. Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. *Biol Reprod* **79**, 1176-1182 (2008).

32. Cerván-Martín, M.*, et al.* Intronic variation of the SOHLH2 gene confers risk to male reproductive impairment. *Fertil Steril* **114**, 398-406 (2020).

33. Barros, F., Carvalho, F., Barros, A. & Dória, S. Premature ovarian insufficiency: clinical orientations for genetic testing and genetic counseling. *Porto Biomed J* **5**, e62 (2020).

34. Stark, Z.*, et al.* Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am J Hum Genet* **108**, 1551-1557 (2021).

35. Lerat, J.*, et al.* An Application of NGS for Molecular Investigations in Perrault Syndrome: Study of 14 Families and Review of the Literature. *Hum Mutat* **37**, 1354-1362 (2016).

36. Richards, S.*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015).

37. Rouen, A.*, et al.* Whole exome sequencing in a cohort of familial premature ovarian insufficiency cases reveals a broad array of pathogenic or likely pathogenic variants in 50% of families. *Fertil Steril* **117**, 843-853 (2022).

38. Yang, X.*, et al.* Gene variants identified by whole-exome sequencing in 33 French women with premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 39-45 (2019).

39. Ke, H., *et al*. Landscape of pathogenic mutations in premature ovarian insufficiency.
Nature Medicine, **29**, 483-492, (2023).

40. Li, Z.*, et al.* Validation of UK Biobank data for mental health outcomes: A pilot study using secondary care electronic health records. *Int J Med Inform* **160**, 104704 (2022).

41. Fry, A.*, et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017).

42. Pirastu, N.*, et al.* Genetic analyses identify widespread sex-differential participation bias. *Nat Genet* **53**, 663-671 (2021).

43. Murray, A., *et al.* Population-based estimates of the prevalence of *FMR1* expansion mutations in women with early menopause and primary ovarian insufficiency. *Genetics in Medicine,* **16,** 19-24, (2014).

44. Backman, J.D.*, et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628-634 (2021).

45. Sudlow, C.*, et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).

46. McLaren, W.*, et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

47. Purcell, S.*, et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).

48. Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A. & Mesirov, J.P. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34 (2017).

49. Robinson, J.T.*, et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).

50. Mbatchou, J.*, et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).

51. Tyrrell, J.*, et al.* Using genetics to understand the causal influence of higher BMI on depression. *Int J Epidemiol* **48**, 834-848 (2019).

52. Loh, P.R.*, et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290 (2015).

53. Bycroft, C.*, et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

54. Gardner, E.J.*, et al.* Damaging missense variants in *IGF1R* implicate a role for IGF-1 resistance in the aetiology of type 2 diabetes. *medRxiv*, 2022.2003.2026.22272972 (2022).

55. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. (2010).

Chapter 4:

Genetic susceptibility to earlier ovarian ageing increases *de novo* mutation rate in offspring.

Stasa Stankovic*, Saleh Shekari*, Qin Qin Huang, Eugene J. Gardner, Erna V. Ivarsdottir, Nick D. L. Owens, Ajuna Azad, Gareth Hawkes, Katherine A. Kentistou, Robin N. Beaumont, Felix R. Day, Yajie Zhao, Hakon Jonsson, Thorunn Rafnar, Vinicius Tragante, Gardar Sveinbjornsson, Asmundur Oddsson, Unnur Styrkarsdottir, Julius Gudmundsson, Simon N. Stacey, Daniel F. Gudbjartsson, The Genomics England Research Consortium, Breast Cancer Association Consortium, Kitale Kennedy, Andrew R. Wood, Michael N. Weedon, Ken K. Ong, Caroline F. Wright, Eva R. Hoffmann, Patrick Sulem, Matthew E. Hurles, Katherine S. Ruth, Hilary C. Martin, Kari Stefansson, John R. B. Perry and Anna Murray

Main text

Abstract

Human offspring acquire on average 60-70 single nucleotide *de novo* mutations each generation [2,3], the majority of which are paternal in origin. Population variability in mutation rate is strongly associated with paternal age [2,4], reflecting the increased number of mitoses during continued spermatogenesis. Although a weaker association with maternal age has also been observed [3,7], the biological mechanisms underpinning this remain unclear. To explore this, we tested maternal germline genetic determinants of reproductive ageing[11] on offspring mutation rate using 14,034 sequenced trios. These analyses demonstrated that genetic susceptibility to earlier ovarian ageing in women increases *de novo* mutation rate in their offspring, consistent with the expected role of these implicated DNA damage response genes in maintaining germ cell genetic integrity[11]. We extended these analyses to identify novel genes influencing ovarian ageing by assessing rare variants in 106,973 exome-sequenced women from the UK Biobank study. These analyses identified effects ~5 times larger than previously discovered in analyses of common variants[11], highlighting protein-coding variants in *ETAA1*, *ZNF518A*, *PNPLA8*, *PALB2* and *SAMHD1.* The *SAMHD1* association further reinforces the link between ovarian ageing and cancer susceptibility[11], with damaging germline variants extending reproductive life but increasing all-cause cancer risk in both males (OR=2.1 [1.7-2.6], P=4.7*10^{-13}) and females (OR=1.61 [1.31-1.96], P=4*10^{-6}). Collectively our findings provide direct evidence that female mutation rate is heritable and highlight an example of a mechanism for the maternal genome influencing child health.

Introduction

De novo mutations (DNMs) in the germline are the source of all segregating genetic variation in humans, as well as a major cause of severe paediatric disorders[1]. The DNM rate in humans is ~1.20×10^{-8} per nucleotide per generation, which translates to an average of ~70 *de novo* single nucleotide variants (SNVs) and small indels per individual (typical range ~40-90)[2,3]. Around 80% of these DNMs originate in the father[2,4]. This has long been thought to occur due to the greater number of cell divisions in spermatogenesis than in oogenesis[5], but recent work points to a substantial role for DNA damage-induced mutations rather than primarily replication-coupled mutations[6]. Parental age is the major factor determining DNM rate; the number of *de novo* SNVs increases by ~2 per year of paternal age and ~0.5 per year of maternal age[2,3,7]. Other factors affecting the DNM rate include exposure to certain chemotherapies[3], rare variants in certain DNA repair genes[3] and, at least in mice, exposure to carcinogens in tobacco smoke[8]. One study on ~1,700 families estimated that family-specific factors (which include genetic and environmental effects) explain only ~5% of the variation in DNM rate[9]. Recent work using ~8,000 trios in the Genomics England '100,000 Genomes' project estimated that ~10-20% of the variance in paternal DNM rate is attributable to genome-wide common variants[10]. The equivalent estimate in mothers was not significantly different from zero, although the analysis was less well powered than in fathers due to the lower rate of DNMs.

It has been hypothesised that DNM rates in the female germline may be linked to mechanisms regulating the timing of menopause. Population variation in timing of menopause reflects a complex mix of genetic and environmental factors that population-based studies have begun to unravel. Previous genome-wide association studies (GWAS) have successfully identified ~300 distinct common genomic loci associated with the timing of menopause[11]. These reported variants cumulatively explain 10% - 12% of the variance in age at natural menopause (ANM) and 31-38% of the overall estimated Single Nucleotide Polymorphism (SNP) heritability[11-20] . Furthermore, initial whole exome sequencing (WES) analyses in UK Biobank identified gene-based associations with ANM for *CHEK2*, *DCLRE1A*, *HELB*, *TOP3A*, *BRCA2* and *CLPB*[11,21]. The majority of these loci implicate genes that regulate DNA damage

response and repair (DDR), highlighting the particular sensitivity of oocytes to DNA damage due to the prolonged state of cell cycle arrest across the life-course[11-24]. We hypothesised that variation in these same processes would influence mutation rate in oocytes that might be transmitted to the next generation. Our current study provides evidence to support this notion, demonstrating that women at increased genetic risk of earlier menopause (based on their common variant-based polygenic risk scores) have increased rates of DNMs in their offspring. We then explored the role of rare coding variants in ovarian ageing using WES data from UK Biobank, improving on previous work through a combination of enhanced phenotype curation, better-powered statistical tests and assessment of different types of variant classes at lower allele frequency thresholds (Supplementary Note). Using these approaches, we identify five genes harbouring variants of large effect that have not previously been implicated, highlighting *ETAA1*, *ZNF518A*, *PNPLA8*, *PALB2* and *SAMHD1* as novel regulators of ovarian ageing.

Methods

UK Biobank Data Processing and Quality Control

To conduct rare variant burden analyses described in this study, we obtained whole exome sequencing data (WES) for 454,787 individuals from the UK Biobank study[90]. Participants were excluded based on excess heterozygosity, autosomal variant missingness on genotyping arrays ≥ 5%, or inclusion in the subset of phased samples as defined in Bycroft *et al*[91]. Analysis was restricted to participants with European genetic ancestry, leaving a total of 421,065 individuals. Variant quality control (QC) and annotation were performed using the UK Biobank Research Analysis Platform (RAP; https://ukbiobank.dnanexus.com/), a cloud-based central data repository for UK Biobank WES and phenotypic data. Besides the QC described by Backman *et al.*[90], we performed additional steps using custom applets designed for the RAP. Firstly, we processed provided population-level Variant Call Format (VCF) files by splitting and left-correcting multi-allelic variants into separate alleles using 'bcftools norm'[92]. Secondly, we performed genotype-level filtering applying 'bcftools filter' separately for Single Nucleotide Variants (SNVs) and Insertions/Deletions (InDels) using a missingness-based approach. Using this approach, we set to missing (i.e. ./.) all SNV genotypes with depth < 7 and genotype quality < 20 or InDel genotypes with a depth < 10 and genotype quality < 20. Next, we applied a binomial test to assess an expected alternate allele contribution of 50% for heterozygous SNVs; we set to missing all SNV genotypes with a binomial test *P* value ≤ $1x10^{-3}$. Following genotype-level filtering we recalculated the proportion of individuals with a missing genotype for each variant and filtered all variants with a missingness value > 50%. The variant annotation was performed using the ENSEMBL Variant Effect Predictor (VEP) v104[93] with the '--everything' flag and plugins for CADD[94] and LOFTEE[95] enabled. For each variant we prioritised the highest impact individual consequence as defined by VEP and one ENSEMBL transcript as determined by whether or not the annotated transcript was protein-coding, MANE select v0.97, or the VEP Canonical transcript. Following annotation, variants were categorised based on their predicted impact on the annotated transcript. Protein Truncating Variants (PTVs) were defined as all variants annotated as stop gained, frameshift, splice acceptor, and splice donor. Missense variant

consequences are identical to those defined by VEP. Only autosomal or chrX variants within ENSEMBL protein-coding transcripts and within transcripts included on the UKBB ES assay[90] were retained for subsequent burden testing.

Exome-wide association analyses in the UK Biobank

In order to perform rare variant burden tests, we used a custom implementation of BOLT-LMM v2.3.6[96] for the RAP. Two primary inputs are required by BOLT-LMM: i) a set of genotypes with minor allele count > 100 derived from genotyping arrays to construct a null linear mixed effects model and ii) a larger set of variants collapsed on ENSEMBL transcript to perform association tests. For the former, we queried genotyping data available on the RAP and restricted to an identical set of individuals included for rare variant association tests. For the latter, and as BOLT-LMM expects imputed genotyping data as input rather than per-gene carrier status, we created dummy genotype files where each variant represents one gene and individuals with a qualifying variant within that gene are coded as heterozygous, regardless of the number of variants that individual has in that gene.

To test a range of variant annotation categories for MAF < 0.1%, we created dummy genotype files for high confidence PTVs as defined by LOFTEE, missense variants with CADD ≥ 25, and damaging variants that included both high confidence PTVs and missense variants with CADD ≥ 25. For each phenotype tested, BOLT-LMM was then run with default parameters other than the inclusion of the 'lmmInfOnly' flag. To derive association statistics for individual markers, we also provided all 26,657,229 individual markers regardless of filtering status as input to BOLT-LMM. All tested phenotypes were run as continuous traits corrected by age, age^2, sex, the first ten genetic principal components as calculated in Bycroft *et al*[91] and study participant ES batch as a categorical covariate (either 50k, 200k, or 450k).

For discovery analysis in the primary trait of interest, age at natural menopause (ANM), we analysed 17,475 protein-coding genes with the minimum of 10 rare allele carriers in at least one of the masks tested using BOLT-LMM (Supplementary Table 3). The significant gene-level associations for ANM were identified applying Bonferroni correction for the number of masks with MAC≥10 (N=46,251 masks) in 17,475 protein-coding genes (*P*: 0.05/46,251 = $1.08*10^{-6}$) (Supplementary Table 4).The ANM results obtained via BOLT-LMM are

available in Supplementary Table 3. Furthermore, in order to compare and explain potential differences between our WES results and the previously published one[21], we ran the above described approach using MAF < 1%, a cutoff applied by Ward *et al*. (Supplementary Table 7, Supplementary Note).

To generate accurate odds ratio and standard error estimates for binary traits, we also implemented a generalised linear model using the statsmodels package[97] for python in a three step process. First, a null model was run with the phenotype as a continuous trait, corrected for control covariates as described above. Second, we regressed carrier status for individual genes on the residuals of the null model to obtain a preliminary *P* value. Thirdly, all genes were again tested using a full model to obtain odds ratios and standard errors with the family set to 'binomial'. Generalised linear models utilised identical input to BOLT-LMM converted to a sparse matrix.

Phenotype derivation

Age at natural menopause (ANM) was derived for individuals within the UK Biobank, who were deemed to have undergone natural menopause, i.e. not affected by surgical or pharmaceutical interventions, as follows:

Firstly, European female participants (n=245,820) who indicated during any of the attended visits having had a hysterectomy were collated (fields 3591 and 2724) and their reported hysterectomy ages were extracted (field 2824) and the median age was kept (n=47,218 and 46,260 with reported ages). The same procedure was followed for participants indicating having undergone a bilateral oophorectomy (surgery field 2834 and age field 3882, n=20,495 and 20,001 with reported ages).

For individuals having indicated the use of hormone replacement therapy (HRT; field 2814), HRT start and end ages were collated (fields 3536 and 3546, accordingly) across the different attended visits (n=98,104). In cases where the reported chronological HRT age at later attended visits was greater than that at previous visits, the later instances were prioritised, i.e. as they would potentially indicate an updated use of HRT. In cases where different HRT ages were reported, but not in chronologically increasing order, the median age was kept.

Menopausal status was determined using data across instances (field 2724) and prioritising the latest reported data, to account for changes in menopause status. For participants indicating having undergone menopause, their reported ages at menopause were collated (field 3581) using the same procedure as for HRT ages (n=158,264).

Exclusions were then applied to this age at menopause, as follows:

- Participants reporting undergoing a hysterectomy and/or oophorectomy, but not the age at which this happened (n=958 and 494, accordingly)
- Participants reporting multiple hysterectomy and/or oophorectomy ages, which were more than 10 years apart (n=38 and 23, accordingly)
- Participants reporting multiple HRT start and/or end ages, which were not in chronologically ascending order and were more than 10 years apart (n=124 and 137, accordingly)
- Participants reporting multiple ages at menopause, which were not in chronologically ascending order and were more than 10 years apart (n=73) and participants who reported both having and not having been through menopause and no other interventions (n=98)
- Participants having undergone a hysterectomy/oophorectomy before or during the year they report undergoing menopause
- Participants starting HRT prior to undergoing menopause and participants reporting HRT use, with no accompanying dates

The resulting trait was representative of an ANM (N=115,051) and was used in downstream analyses. Two additional ANM traits were also calculated, windsorized one by coding everyone reporting an ANM younger than 34, as 34 used in the discovery analysis as the primary phenotype (N=115,051 total, reduced to 106,973 after covariate-resulting exclusions), and one by only including participants reporting ANM between 40 and 60, inclusive (N=104,506), treated as a sensitivity analysis.

All manipulations were conducted in R (v4.1.2) on the UKB Research Analysis Platform (RAP; https://ukbiobank.dnanexus.com/).

Replication analyses

Replication was performed using two study populations - the Icelandic deCODE study and the BRIDES study.

The BRIDGES study included women from studies participating in the Breast Cancer Association Consortium (BCAC; v14) (http://bcac.ccge.medschl.cam.ac.uk/; Table). The subset of population or hospital-based studies sampled independently of family history, together with population-matched controls (25 studies) were included in the analyses. ANM (years) was obtained from baseline questionnaire data. Women were considered as having experienced natural menopausal if they indicated that the reason for menopause was reported as 'natural' or 'unknown'. Women were excluded from the analysis if the reason was indicated as either oophorectomy, hysterectomy, chemotherapy, stopping oral contraception or 'any other reason'. Only studies with information on year of birth and age at menopause, and only women with reported age at menopause between ages twenty-five years and sixty years were included. All studies were approved by the relevant ethical review boards and used appropriate consent procedures.

Targeted sequencing of germline DNA from participants for 35 known or suspected breast cancer genes was performed, including the coding sequence and splice sites. Details of library preparation, sequencing, variant calling and quality control procedures are described in Dorling *et al*.[98] Carriers of protein truncating variants (PTVs) in more than one of five main breast cancer susceptibility genes (*BRCA1*, *BRCA2, ATM*, *CHEK2*, *PALB2*) were excluded. Carriers of pathogenic MSVs (as defined by Dorling *et al*[98]) in *BRCA1* or *BRCA2* were also excluded.

We carried out burden analyses, assessing the associations between rare variants in aggregate and age at natural menopause using linear regression, adjusting for country of origin, breast cancer case-control status and year of birth (categorized as <=1935, 1936-1945, 1946-1955, >=1956), and for some analyses BMI. For each gene we considered PTVs in aggregate. The primary analyses included covariates to adjust for population, which was defined by country, with the exception of Malaysia and Singapore, in which the three distinct ethnic groups (Chinese, Indian, Malay) were treated as different strata

and the UK, which was treated as separate strata (SEARCH, from East Anglia and PROCAS from north-west England). Sensitivity analyses were carried out adjusting for BMI in women with recorded age at BMI, and among women without a diagnosis of breast cancer. Sensitivity analyses were also carried out defining non-carriers as women not harbouring PTVs in the five main genes or pathogenic MSVs in *BRCA1* and *BRCA2*.

The genome of the Icelandic population was characterised by whole-genome sequencing of 63,460 Icelanders using Illumina standard TruSeq methodology to a mean depth of 35x (SD 8x) with subsequent long-range phasing[99], and imputing the information into 173,025 individuals chip-genotyped employing multiple Illumina platforms[100].

The burden test associations are shown for three categories of rare variants with MAF < 2%; a) loss-of-function (LOF) variants, b) combination of LOF variants and predicted deleterious missense variants and c) combination of LOF variants and missense variants with CADD score ≥ 25. We furthermore show results for category c) using a more stringent frequency threshold, 0.1%. We included missense variants predicted to cause loss of function by two meta-predictors, MetaSVM and MetaLR[101], using variants available in dbNSFP v4.1c[102].

Phenome-wide association analysis

In order to test the association of ANM identified genes in other phenotypes, we processed additional reproductive ageing-related phenotypes, including age at menarche, cancer, telomere length (TL) and sex hormones (SH). All tested phenotypes were run as either continuous (age at menarche, TL and SH) or binary traits (cancer) corrected by age, age^2, sex, the first ten genetic principal components as calculated in Bycroft *et al*[91], and study participant ES batch as a categorical covariate (either 50k, 200k, or 450k). Phenotype definitions and processing used in this study are described in Supplementary Tables 12 and 13. Only the first instance (initial visit) was used for generating all phenotype definitions unless specifically noted in Supplementary Table 12. In case of cancer-specific analysis data from cancer registries, death records, hospital admissions and self-reported were harmonised to ICD10 coding. If a participant had a code for any of the cancers recorded in ICD10 (C00-C97) then they were

counted as a case for this phenotype. Minimal filtering was performed on the data, with only those cases where a diagnosis of sex-specific cancer was given in contrast to the sex data contained in UK Biobank record 31, was a diagnosis not used. For more information on cancer-specific analysis refer to Supplementary Tables 13 and 15.

Cancer PheWAS Associations

To test for an association between genes we identified as associated with menopause timing (Supplementary Table 4, Figure 1) and 90 individual cancers as included in cancer registries, death records, hospital admissions and self-reported data provided by UK Biobank (e.g. breast, prostate, etc.) we utilised a logistic model with identical covariates as used during gene burden testing (N = 2430 tests) (Supplementary Tables 13 and 15). As standard logistic regression can lead to inflated *P* value estimates in cases of severe case/control imbalance[103], we also performed a logistic regression with penalised likelihood estimation as described by Firth[41] (Supplementary Table 15). Models were run as discussed in Kosmidis *et al*.[104] using the 'brglm2' package implemented in R. brglm2 was run via the 'glm' function with default parameters other than "family" set to "binomial", "method" set to "brglmFit", and "type" set to "AS_mean".

WES sensitivity analysis using REGENIE.

To replicate the primary findings and account for potential bias that could be introduced by exclusively using one discovery approach, a second analyst independently derived the age at menopause phenotype using a previously published method[105] and conducted additional burden association analysis using the REGENIE regression algorithm (REGENIEv2.2.4; https://github.com/rgcgithub/regenie). REGENIE implements a generalised mixed-model region-based association test that can account for population stratification and sample relatedness in large-scale analyses. REGENIE runs in 2 steps[106], which we implemented on the UKBiobank RAP: In the first step, genetic variants are aggregated into gene specific units for each class of variant called masks. We selected variants in CCDS transcripts deemed to be high confidence by LOFTEE[95] with MAF<0.1% and annotated using VEP[93]. We created three masks, independently of primary analysis group: (1) loss-of-

function (LOF) variants (stop-gain, frameshift, or abolishing a canonical splice site (-2 or +2 bp from exon, excluding the ones in the last exon)) or missense variants with CADD score >30, (2) LOF or missense variants with CADD score >25, (3) all missense variants. In the second step, the three masks were tested for association with ANM. We applied an inverse normal rank transformation to ANM and included recruitment centre, sequence batch and 40 principal components as covariates. For each gene, we present results for the transcript with the smallest burden *P* value. The results for the sensitivity analysis performed via REGENIE are available in Supplementary Table 3.

Common variant GWAS lookups

Genes within 500kb upstream and downstream of the 290 lead SNPs from the latest GWAS of ANM[11] were extracted from the exome-wide analysis. There were a total of 2149 genes within the GWAS regions. Burden tests in these genes with a Bonferroni corrected *P* value of $<2.3*10^{-5}$ (0.05/2149) were highlighted. The results are available in Supplementary Table 8.

Analysis of GWAS and WES genes expression profiles in human female germ cells at various stages of development

We studied the mRNA abundance of WES genes during various stages of human female germ cell development using single-cell RNA sequencing data (Supplementary Tables 10 and 11). We used the processed single cell RNA resequencing datasets from two published studies. This included single-cell RNA sequencing data from foetal primordial germ cells of human female embryos (Accession code: GSE86146[107]), and from oocyte and granulosa cell fractions during various stages of follicle development (Accession code: GSE107746[108]). A pseudo score of 1 was added to all values before log transformation of the dataset. The samples from fetal germ cells (FGCs) were categorised into sub-clusters as defined in the original study. The study by Li *et al*[107] had identified 17 clusters by performing a t-distributed stochastic neighbour embedding (t-SNE) analysis and using expression profiles of known marker genes for various stages of fetal germ cell development. In our analysis we have included four clusters of female FGCs (Mitotic, Retinoic Acid (RA) responsive, Meiotic, Oogenesis) and four clusters containing somatic cells in the fetal gonads (Endothelial, Early_Granulosa, Mural_Granulosa,

Late_Granulosa). Software packages for R - tidyverse (https://www.tidyverse.org/), pheatmap, (https://CRAN.R-project.org/package=pheatmap), reshape2 (https://github.com/hadley/reshape), were used in processing and visualising the data.

Functional enrichment tests for *ZNF518A* transcription factor binding sites using fGWAS and SLDP

fGWAS (*v.0.3.6)*, a hierarchical model for joint analysis of GWAS and genomic annotations, was implemented to test the functional enrichment of ANM GWAS hits in *ZNF518A* transcription factor binding sites[33]. The fGWAS input file contained the ANM GWAS summary stats derived from the Reprogen study[11] annotated for *ZNF518A* binding sites. The *ZNF518A* annotation file was derived from the ENCODE ChIP-seq data[31,32] from human HEK293 cell line[109] the optimal independent discovery rate peak calling against hg19 [ENCFF415VBF] was used. The ANM GWAS hits were annotated for the presence/absence of the *ZNF518A* transcription factor binding sites in a binary way (0, 1), with '1' if the SNP falls within the transcription factor binding site and '0' otherwise. The fGWAS tool available from https://github.com/joepickrell/fgwas and was run in annotation mode "-w" for the described *ZNF518A* annotation. Detailed description of fGWAS methodology is available in Pickrell *et al*, 2014[33]. In short, the genome is split into independent blocks, which are allowed to contain either a single polymorphism that causally influences the trait or none. fGWAS then models the prior probability that any given block contains an association and the conditional prior probability that any given SNP in the block is the causal one, with probabilities allowed to vary according to functional annotations. The priors are then estimated using an empirical Bayes approach. The fGWAS output contained the maximum likelihood parameter estimates for each parameter in the model, in this case *ZNF518A*, with the lower and upper bound of the 95% confidence interval (CI) on the parameter. The P value was calculated from lower and upper CI in 3 following steps: (1) Standard error (SE) calculation: $SE = (Upper\ CI - Lower\ CI)/(2*1.96)$; (2) Test statistics calculation: $Z = Estimate / SE$; and (3) P value calculation: $P = exp(-0.717*Z - 0.416*Z^2)$.

Signed LD profile (SLDP) regression was applied to explore the directional effect of a signed functional annotation, *ZNF518A*, on a heritable trait like ANM

using GWAS summary statistics[11]. More specifically, we tested whether alleles that are predicted to increase the binding of the transcription factor *ZNF518A* have a genome-wide tendency to increase or decrease timing of menopause in women. The SLDP tool was installed from https://github.com/yakirr/sldp, with the comprehensive methodological steps described in Reshef *et al*, 2018[34]. For the analysis to be conducted, SLDP required GWAS summary statistics for ANM, signed LD profiles for *ZNF518A* binding, signed background model and reference panel in a SLDP compatible format. For the reference we used a 1000 Genomes Phase 3 European reference panel in *plink* format, which contained approximately 10M SNPs and 500 people and was available for download at the 'refpanel' page. The ANM GWAS summary statistics, available from our latest Reprogen study[11], was pre-processed using the '*preprocesspheno*' tool from the SLDP package. To conduct this step, we also obtained the list of regression SNPs along with the LD scores for the reference panel from the 'refpanel' page. The pre-processing step included filtering down to SNPs that are also present in the reference panel, harmonising alleles to the reference, and multiplying the summary statistics by the SLDP regression weights. In addition, we applied the '*preprocessrefpanel*' tool to compute a truncated singular value decomposition (SVD) for each LD block in the reference panel. These SVDs were later used to weight the SLDP regression. The *ZNF518A* annotation file was obtained from the ENCODE CHIP-seq analysis[31,32], as described above, and preprocessed using the '*preprocessannot*' tool that turns signed functional annotations into signed LD profiles. Prior to running SLDP, we also obtained the signed background LD profiles that enabled us to control for systematic signed effects of minor alleles, which could arise from either population stratification or negative selection. SLDP was then run on our data using '*sldp*' function. To explore the relevance of *ZNF518A* for menopause timing in comparison to other transcription regulators, we tested whether genome-wide sequence changes introduced by SNP alleles identified in ANM GWAS increase or decrease binding of additional 382 transcription factors (TFs). The preprocessed annotation files for 382 TFs derived from ENCODE CHIP-seq experiments[31,32], were available for download at the annotation data page. The results are available in Supplementary Table 9.

Functional analysis of *ZNF518A* binding sites

ZNF518A peaks were derived from unique genomic regions in ENCODE accession ENCFF415VBF described above. Quantification of ChIP-seq signal by aligning paired-end replicates (ENCFF174HBR, ENCFF574GQY, ENCFF808AJP, ENCFF453FDD) to the hg19 genome with Bowtie2 v2.3.5.1[110] with options "-I 0 -X 1000 –no-discordant –no-mixed", reads were filtered for those with MAPQ > 30 with samtools v1.10. Assessment of H3K27ac[36] and chromatin accessibility by ATAC-seq[37] in day 4 human primordial germ cell like cells (hPGCLCs) at *ZNF518A* peaks was performed. For H3K27ac single end reads from accessions GSM4257216, GSM4257217, GSM4257218 were obtained and aligned with Bowtie2 v2.3.5.1 with default settings and MAPQ > 30 reads retained as above. For ATAC-seq paired-end reads were obtained from accessions GSM3406938, GSM3406939 and mapped and filtered as *ZNF518A* reads above.

Quantification of ChIP-seq and ATAC-seq signals for peak heights, heatmaps was performed with https://github.com/owensnick/GenomeFragments.jl. Peak to TSS distances were calculated against Gencode v36 release liftover to hg19 using GenomicFeatures.jl and https://github.com/owensnick/ProximityEnrichment.jl. We consider four categories of peaks: TSS intersecting, TSS proximal (TSS < 2000kb, outside gene body), Gene body intersecting, Intergenic and Distal (TSS > 5kb).

To perform *de novo* motif discovery we used Homer v4.11.1[111] using findMotifsGenome.pl with options "hg19 -size 200". We ran this on all *ZNF518A* peaks, distal peaks and those intersecting TSS, we recovered a motif matching JASPAR[35] unvalidated motif UN0199.1 in all peak sets apart from those intersecting TSS. We then used https://github.com/exeter-tfs/MotifScanner.jl to quantify the occurrence of all instances of motif UN0199.1 in *ZNF518A* peaks.

We downloaded the 18-state ChromHMM[112] models for all 833 biosamples in Epimap[38] from http://compbio.mit.edu/epimap/. We calculated the intersection between each state in each biosample and either all *ZNF518A* peaks or distal *ZNF518A* peaks using GenomicFeatures.jl. We calculated odds ratios from contingency tables using the approximation of bedtools[113] and Giggle[114], by

estimating total genomic intervals as hg19 genome size divided by the sum of the mean *ZNF518A* peak size and the chromatin state interval size.

De novo mutation rate analyses

We calculated polygenic scores (PGSs) in participants from the rare disease programme of the 100,000 Genome Project (100kGP) v14. There are 77,901 individuals in the Aggregated Variant Calls (aggV2) after excluding participants whose genetically inferred sex is not consistent with their phenotypic sex. We restricted the PGS analysis to individuals of European ancestry, which was predicted by the Genomics England Bioinformatics team using a random forest model based on genetic principal components (PCs) generated by projecting aggV2 data onto the 1000 Genomes phase 3 PC loadings. We removed one sample in each pair of related probands with kinship coefficient > 1/(2^4.5), i.e. up to and including third degree relationships. Probands with the highest number of relatives were removed first. Similarly, we retained unrelated mothers and fathers of these unrelated probands. It left us with 8,089 mother-offspring duos and 8,029 father-offspring duos.

We used the lead variants (or proxies, as described below) for genome-wide significant loci previously reported for ANM[11] to calculate PGS in the parents. In 100kGP, we removed variants with minor allele frequency (MAF) <0.5% or missing rate >5% from the aggV2 variants prepared by the Genomics England bioinformatics team. For lead variants that did not exist in 100kGP, we used the most significant proxy variants with linkage disequilibrium (LD) r^2 >0.5 if available in 100kGP. This resulted in a PGS constructed from 287 of the 290 previously reported loci. We regressed out 20 genetic PCs that were calculated within the European subset from the PGS and scaled the residuals to have mean = 0 and standard deviation = 1. Higher PGS indicates later ANM.

De novo mutations (DNMs) were called in 10,478 parent offspring trios by the Genomics England Bioinformatics team. The detailed analysis pipeline is documented at: https://research-help.genomicsengland.co.uk/display/GERE/De+novo+variant+research+dataset . Extensive quality control (QC) and filtering were applied by Kaplanis *et al*. as described previously[3]. *De novo* single nucleotide variants (dnSNVs) were phased using a read-based approach based on heterozygous variants near the

DNM that were able to be phased to a parent. About one third of the dnSNVs were phased, of which three quarters were paternally phased (Supplementary Figure 1, Supplementary Table 1).

In association models, we accounted for parental age, the primary determinant of the number of DNMs, and various data quality metrics as described in[3]:

- Mean coverage for the child, mother and father (child_mean_RD, mother_mean_RD, father_mean_RD)
- Proportion of aligned reads for the child, mother and father (child_prop_aligned, mother_prop_aligned, father_prop_aligned)
- Number of SNVs called for child, mother and father (child_SNVs, mother_SNVs, father_SNVs)
- Median variant allele fraction of DNMs called in child (median_VAF)
- Median 'Bayes Factor' as outputted by Platypus for DNMs called in the child. This is a metric of DNM quality (median_BF).

We first tested the association between parental PGSs and total dnSNV count in the offspring in a Poisson regression:

$$dnSNVs\_total = \beta_0 + \beta_1 paternal\_PGS + \beta_2 maternal\_PGS + \beta_3 paternal\_age + \beta_4 maternal\_age + \beta_5 child\_mean\_RD + \beta_6 mother\_mean\_RD + \beta_7 father\_mean\_RD + \beta_8 child\_prop\_aligned + \beta_9 mother\_prop\_aligned + \beta_{10} father\_prop\_aligned + \beta_{11} child\_snvs + \beta_{12} mother\_snvs + \beta_{13} father\_snvs + \beta_{14} median\_VAF + \beta_{15} median\_BF$$

We also fitted Poisson regression models to test the association between the PGS of one of the parents and the dnSNVs in the offspring that were phased to the relevant parent.

The paternal model included paternal PGS, age, and data quality metrics that are related to the proband and the father:

$$dnSNVs\_paternal = \beta_0 + \beta_1 paternal\_PGS + \beta_2 paternal\_age +$$
$$\beta_3 child\_mean\_RD + \beta_4 father\_mean\_RD +$$
$$\beta_5 child\_prop\_aligned + \beta_6 father\_prop\_aligned +$$
$$\beta_7 child\_snvs + \beta_8 father\_snvs +$$
$$\beta_9 median\_VAF + \beta_{10} median\_BF$$

Similarly, the maternal model was as follows:

$$dnSNVs\_maternal = \beta_0 + \beta_1 maternal\_PGS + \beta_2 maternal\_age +$$
$$\beta_3 child\_mean\_RD + \beta_4 mother\_mean\_RD +$$
$$\beta_5 child\_prop\_aligned + \beta_6 mother\_prop\_aligned +$$
$$\beta_7 child\_snvs + \beta_8 mother\_snvs +$$
$$\beta_9 median\_VAF + \beta_{10} median\_BF$$

Finally, as a sanity check, we assessed the association between the maternal PGS and paternally phased dnSNVs, and vice versa:

$$dnSNVs\_paternal = \beta_0 + \beta_1 maternal\_PGS + \beta_2 paternal\_age +$$
$$\beta_3 child\_mean\_RD + \beta_4 father\_mean\_RD +$$
$$\beta_5 child\_prop\_aligned + \beta_6 father\_prop\_aligned +$$
$$\beta_7 child\_snvs + \beta_8 father\_snvs +$$
$$\beta_9 median\_VAF + \beta_{10} median\_BF$$

$$dnSNVs\_maternal = \beta_0 + \beta_1 paternal\_PGS + \beta_2 maternal\_age +$$
$$\beta_3 child\_mean\_RD + \beta_4 mother\_mean\_RD +$$
$$\beta_5 child\_prop\_aligned + \beta_6 mother\_prop\_aligned +$$
$$\beta_7 child\_snvs + \beta_8 mother\_snvs +$$
$$\beta_9 median\_VAF + \beta_{10} median\_BF$$

Mendelian Randomization

MR analysis was applied to examine the likelihood of a causal effect of polygenic score (PGS) of age at natural menopause on the risk of de novo mutation rates in the offspring (Supplementary Table 2). In this approach, genetic variants that are significantly associated with an exposure of interest are used as instrumental variables (IVs) to test the causality of that exposure on the outcome of interest[115–117]. For a genetic variant to be a reliable instrument, the following assumptions should be met: (1) the genetic instrument is associated with the exposure of interest, (2) the genetic instrument should not be associated with any other competing risk factor that is a confounder, and (3) the genetic instrument should not be associated with the outcome, except via the causal pathway that includes the exposure of interest[115,118]. Genotypes at all variants were aligned to designate the ANM PGS-increasing alleles as the effect alleles as described above and this was used as a genetic instrument of interest. The effect sizes of genetic instruments (genotypes in the mother) on maternally phased de novo SNVs in the offspring estimated in 8,089 duos were obtained from Genomics England.

MR Frameworks

The MR analysis was conducted using the inverse-variance weighted (IVW) model as the primary model due to the highest statistical power[119]. However, as it does not correct for heterogeneity in outcome risk estimates between individual variants[120], we applied a number of sensitivity MR methods that better account for heterogeneity[121]. These include MR Egger to identify and correct for unbalanced heterogeneity ('horizontal pleiotropy'), indicated by a significant Egger intercept (P<0.05)[122], and weighted median (WM) and penalised weighted median (PWM) models to correct for balanced heterogeneity[123]. In addition, we introduced the MR Radial method to exclude variants from each model in cases where they are recognized as outliers[124]. The results were considered as significant based on the *P* value significance consistency across different primary and sensitivity models applied. The results are available in Supplementary Table 2. Finally, in order to calculate the effect of ANM on offspring *de novo* mutation rate when comparing women with ANM at two extremes of the ANM distribution curve, we multiplied the effect obtained by MR

IVW, i.e. a de novo count beta per 1 year change in ANM, by 20, an arbitrary number that compares women with ANM 20 years apart.

Results

We sought to test the hypothesis that inter-individual variation in DDR processes would influence the mutation rate in germ cells and hence in the offspring. More specifically, we hypothesised that genetic susceptibility to earlier ovarian ageing would be associated with a higher DNM rate in the offspring. To test this, we analysed 8,089 whole-genome sequenced parent-offspring trios recruited in the rare disease programme of the 100,000 Genomes Project (100kGP, Supplementary Figure 1). We calculated a polygenic score (PGS) for ANM in the parents based on our previously identified 290 common variants[11] and tested this against the phased DNM rate in the offspring, adjusted for age. We found that maternal genetic susceptibility to earlier ANM was associated with an increased rate of maternally-derived DNMs in the offspring (rate ratio = 1.02 per SD of PGS, P=6.8*10^{-4}, N=8,089 duos with European ancestry; Supplementary Table 1). We confirmed this finding in sensitivity analyses using the same data, in a two-sample Mendelian Randomization (MR) framework that can better model the dose-response relationship of these variants (Supplementary Table 2). These results were highly concordant, with all models showing a significant result and no heterogeneity (P_{min}=6.3*10^{-5}). In contrast, the paternal PGS was not associated with paternally-derived DNMs (P=0.51, N=8,029) nor was the maternal PGS associated with paternally-derived DNMs (P=0.55).

To confirm these observations, we repeated these analyses using data from the Icelandic deCODE study[25]. We first restricted analyses to three generation families where we can trace the origin of the DNM to the transmitting parent via haplotype sharing[26], modelling the number of maternally transmitted mutations as function of maternal age at conception and GRS of menopause. Across these 1,096 three-generation probands, we found that a one standard deviation lower menopause GRS adds 0.41 mutations (P=1x10^{-4}) to the transmission from the mother to the child. This mutational load is comparable to the maternal age effect, where one additional maternal year at conception results in transmission of an extra 0.37 mutations to the offspring from the mother[26]. We expanded these analyses to a broader set of 5,945 trios, yielding an estimate of 0.2 extra mutations (P=7.7x10^{-3}) when using unphased mutations.

Exome-wide gene burden associations with ANM

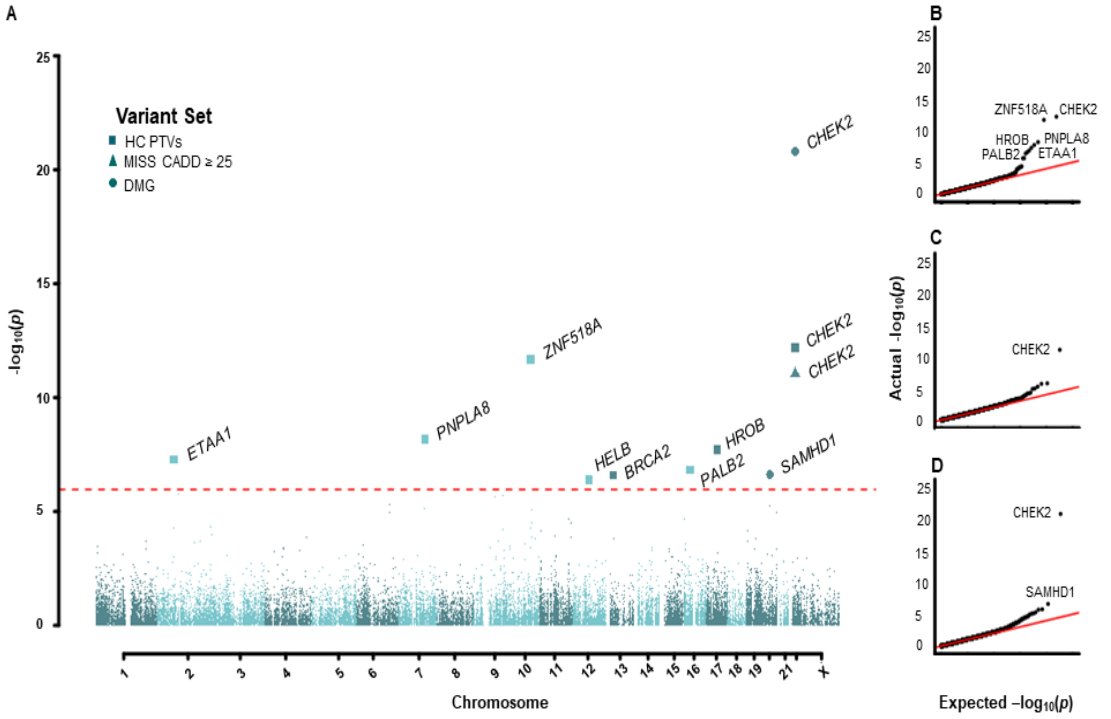
Previous studies have largely focussed on assessing the role of common genetic variation on ovarian ageing. Given the recently published observation that offspring germline hypermutations can be caused by their fathers carrying rare homozygous variants in key DNA repair genes[27], we sought to better understand the role of rare coding variation in ovarian ageing using WES data available in 106,973 post-menopausal UK Biobank female participants of European genetic-ancestry[28]. Individual gene burden association tests were conducted by collapsing genetic variants according to their predicted functional categories. We defined three categories of rare exome variants with minor allele frequency (MAF) < 0.1%: high-confidence Protein Truncating Variants (HC-PTVs), missense variants with Combined Annotation Dependent Depletion (CADD) score ≥ 25, and 'damaging' variants (defined as combination of HC-PTVs and missense variants with CADD ≥ 25). We analysed 17,475 protein-coding genes with the minimum of 10 rare allele carriers in at least one of the masks tested. The primary burden association analysis was conducted using BOLT-LMM[29] (Supplementary Table 3). The low exome-wide inflation scores (e.g. PTV λ=1.047) and the absence of significant association with synonymous variant burden for any gene indicate our statistical tests are well calibrated (Supplementary Figure 2).

We identified rare variation in nine genes associated with ANM at exome-wide significance ($P<1.08*10^{-6}$, Figures 1 and 2, Supplementary Figures 3 and 4). These were confirmed by an independent group of analysts using different quality control (QC) and analysis pipelines (Supplementary Tables 3, 4). Three of these genes have been previously reported in UKBB WES analysis[21] - we confirm the associations of *CHEK2* (beta=1.57 years, 95% CI: 1.23-1.92, $P=1.60*10^{-21}$, N=578 damaging allele carriers) and *HELB* (beta=1.84, 95% CI: 1.08-2.60, $P=4.20*10^{-7}$, N=120 HC-PTV carriers) with later ANM and a previously borderline association of *HROB* with earlier ANM (beta= -2.89 years, 95% CI: 1.86-3.92, $P=1.90*10^{-8}$, N=65 HC-PTV carriers). In addition, our previous ANM GWAS analyses[11] identified an individual low-frequency PTV variant in *BRCA2*, which we now extend to demonstrate that, in aggregate, *BRCA2* HC-PTV carriers exhibit 1.18 years earlier ANM (beta= -1.18, 95% CI:

0.72-1.65, P=2.60*10^{-7}, N=323). Rare variants in the remaining five genes – *ETAA1*, *ZNF518A*, *PNPLA8*, *PALB2* and *SAMHD1* have not been previously implicated in ovarian ageing. Effect sizes of these associations range from 5.61 years earlier ANM for HC-PTV carriers in *ZNF518A*, to 1.35 years later ANM for women carrying damaging alleles in *SAMHD1*. This contrasts with a maximum effect size of 1.06 years (median 0.12 years) for common variants (MAF>1%) identified by previous ANM GWAS[11].

We next attempted to replicate these findings using two independent datasets from the Icelandic deCODE study and the BRIDGES study (N=26,258). Firstly, despite the substantially smaller sample size and rarity of the alleles we were testing in the Icelandic population, we observed consistent effect estimates for all nine genes we identified (Supplementary Table 5). This included nominally significant associations at *BRCA2*, *CHEK2*, *ETAA1*, *HELB*, *SAMHD1* and *ZNF518A*. Secondly, we used data in up to 26,258 women with ANM from the BRIDGES study. As this study used a targeted sequencing approach of suspected breast cancer genes, it was only informative for *BRCA2*, *PALB2* and *CHEK2*. Despite the small sample size, for each of these genes we found effect estimates consistent with our discovery analyses, which were maintained when adjusting for cancer status and within women not diagnosed with breast cancer (Supplementary Table 6). Notably we replicated the novel association with *PALB2*, where the 78 women carrying protein-truncating variants experienced menopause 1.78 years earlier on average (P=4.8x10^{-4}).

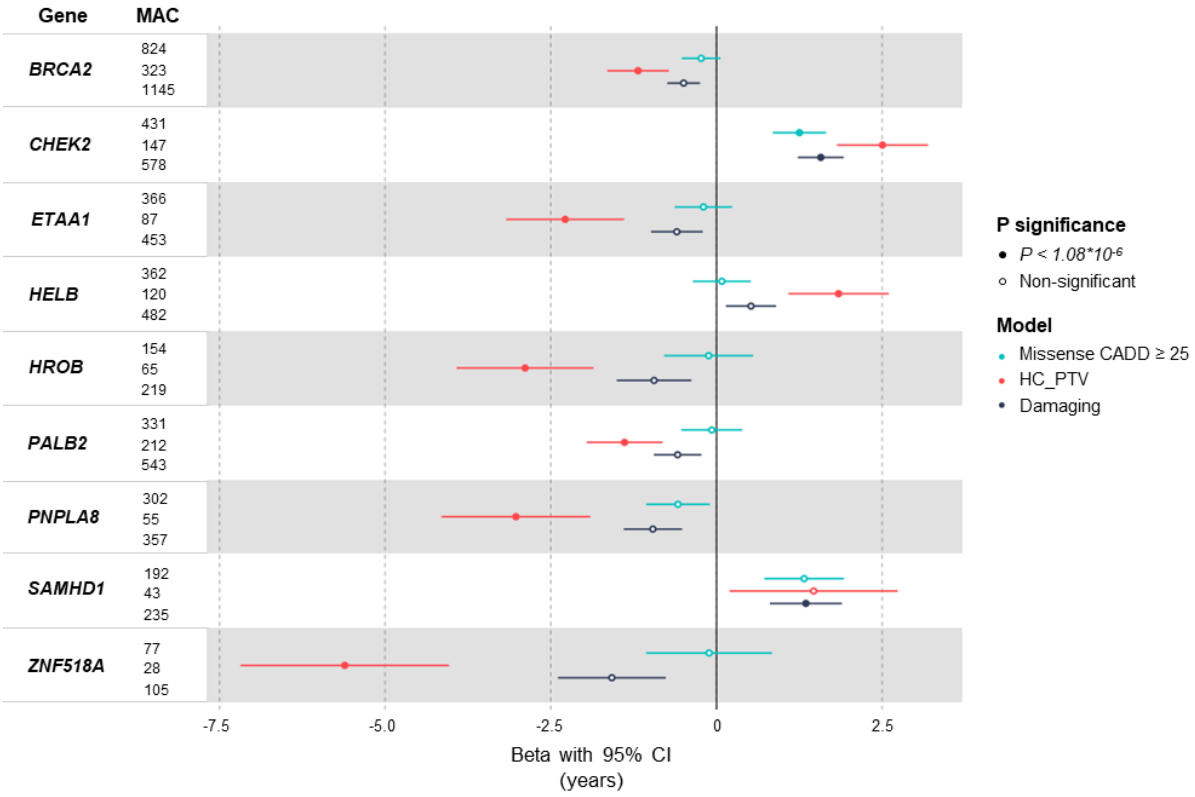
Figure 0–1: Exome-wide associations with age at natural menopause.



Note: ***(A)*** *Manhattan plot showing gene burden test results for age at natural menopause. Genes passing exome-wide significance ($P<1.08*10^{-6}$) are indicated, with point shape signifying the variant class-tested.* ***(B-E)*** *QQ plots for* ***(B)*** *high confidence PTVs,* ***(C)*** *CADD ≥ 25 missense variants* ***(D)*** *damaging variants.*

We next sought to understand why previous analyses of UKBB WES data missed the associations we report here, and conversely why we did not identify associations with other previously reported genes. Of the seven genes identified by Ward *et al.*[21], three were also identified by our study (*CHEK2*, *HELB* and *HROB*), three were recovered when we increased our burden test MAF threshold from 0.1% to 1% (*DCLRE1A*, *RAD54L*, *TOP3A*), and an additional gene fell just below our *P* value threshold when considering variants with <1% MAF (*CLPB; $P =1.2*10^{-5}$*). In contrast, our discovery of novel associations that were not reported by Ward *et al.* (*BRCA2*, *ETAA1*, *PALB2*, *PNPLA8*, *SAMHD1* and *ZNF518A*) were likely explained by differences in phenotype preparation, sample size, variant annotation and the statistical model used (see Supplementary Note and Supplementary Table 7).

Figure 0–2: Forest plot for gene burden associations with age at natural menopause.



Note: Exome-wide significant ($P < 1.08*10^{-6}$) genes are displayed. Points and error bars indicate beta and 95% CI for the variant category indicated. Betas, CIs, Minor Allele Counts (MAC) and P values are derived from BOLT-LMM.

Exploring common variant associations at identified ANM genes

To explore the overlap between common and rare variant association signals for ANM, we integrated our exome-wide results with data generated from the largest reported common variant GWAS of ANM[11]. Five of our nine identified WES genes (*CHEK2*, *BRCA2*, *ETAA1*, *HELB* and *ZNF518A)* mapped within 500kb of a common GWAS signal (Supplementary Table 8). Notably, we previously reported a common, predicted benign, missense variant (rs35777125-G439R, MAF=11%) in *ETAA1* associated with 0.26 years earlier ANM. In contrast, our WES analysis identified that carriers of rare HC-PTVs in *ETAA1* show a nearly 10-fold earlier ANM (beta= -2.28 years, 95% CI: 1.39-3.17, $P=5.30*10^{-8}$, N=87). Furthermore, three independent non-coding common GWAS signals ~150kb apart (MAF: 2.8-47.5%, beta: -0.28-0.28 years per minor allele) were reported proximal to *ZNF518A*, whereas gene burden testing finds that rare HC-PTV carriers show nearly 20-fold earlier ANM than common variant carriers (beta= -5.61 years, 95% CI: 4.04-7.18, $P=2.10*10^{-12}$, N=28).

ZNF518A is a poorly characterised C2H2 zinc finger transcription factor, which has been shown to associate with PRC2 and G9A-GLP repressive complexes along with its paralog *ZNF518B*, suggesting a potential role in transcriptional repression[30]. By integrating ChIP-Seq data, we demonstrate that common variants associated with ANM are enriched in the binding sites of *ZNF518A* (Supplementary Table 9; Supplementary Note), providing further support for the role of this gene in ovarian ageing.

In addition, there were two genes within 500kb of GWAS loci (*BRCA1* and *SLCO4A1*) that were associated with ANM by gene burden testing at $P<1.7*10^{-5}$. Effect sizes for common variant associations ranged from 0.07-0.24 years per allele at these loci, whereas gene burden tests for rarer variants at these same loci revealed much larger effect sizes: for *BRCA1,* 2.1 years earlier for PTVs ($P=2.4*10^{-6}$) and for *SLCO4A1,* 1.13 years earlier ANM for damaging variants ($P=1.1*10^{-5}$), with non-overlapping 95% confidence intervals between common and rare variant associations for *BRCA1.*

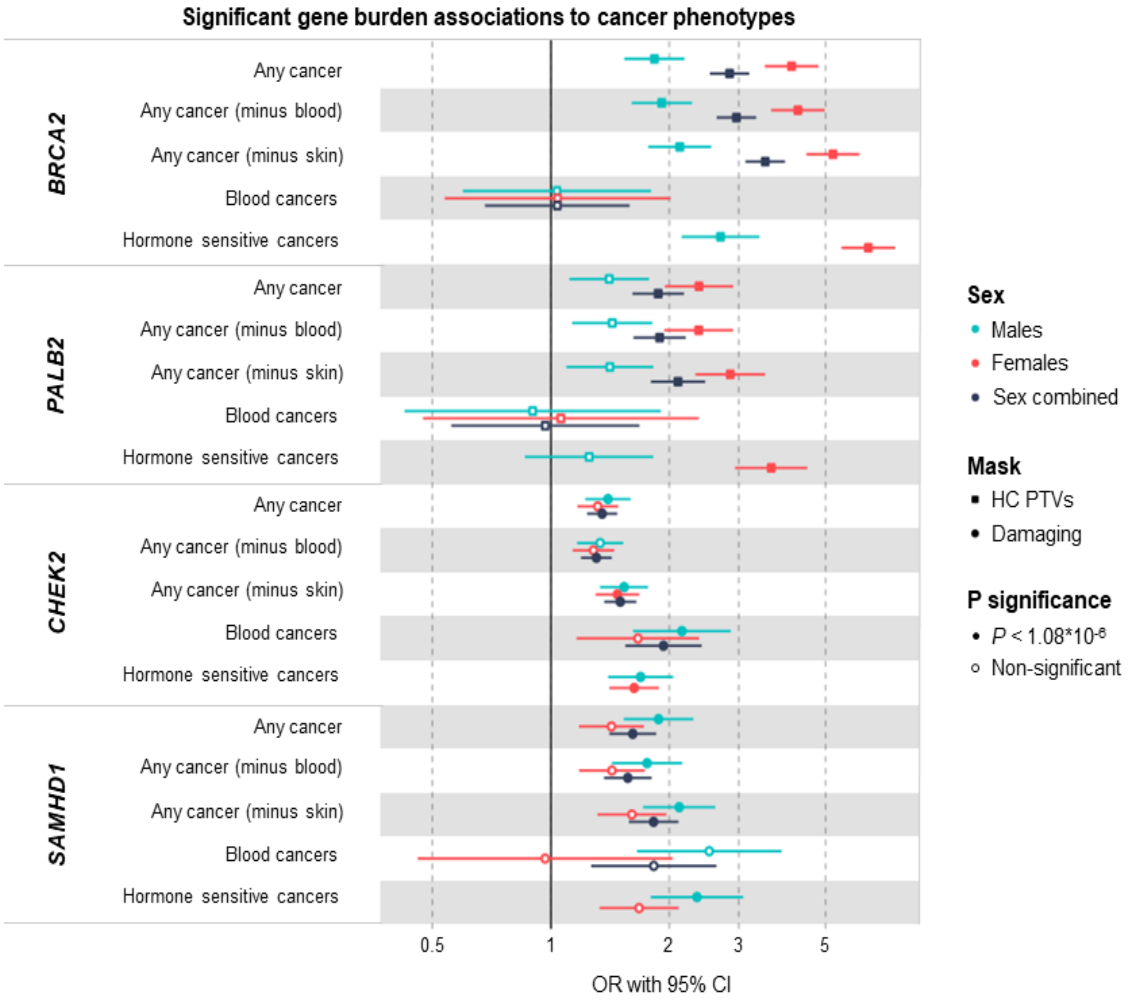
Identified genes influence other aspects of health and disease

Our genetic studies have previously shown that the genetic mechanisms regulating the end of reproductive life are largely distinct from those determining its beginning [39,40]. However, it is noteworthy that the largest reported GWAS for age at menarche identified a common variant signal at the *ZNF518A* locus for later puberty timing in girls (rs1172955, beta= 0.04 years, 95% CI: 0.03-0.05, $P=6.6*10^{-12}$), which appears nominally associated with earlier ANM (beta=-0.04, 95% CI: 0.01-0.06, $P=6.6*10^{-3}$)[39]. To extend this observation, we found that our identified *ZNF518A* PTVs were also associated with later age at menarche (0.56 years, 95% CI: 0.14-0.98, $P=9.2*10^{-3}$). Furthermore, using fGWAS and signed linkage disequilibrium profile (SLDP), we discovered that, similar to ANM (Supplementary Note), common variants that influence puberty in girls were enriched in transcriptional targets of *ZNF518A* (Supplementary Table 9). These data suggest that loss of *ZNF518A* shortens reproductive lifespan, by delaying puberty and reducing age at menopause.

We next explored what impact ANM-associated genes had on cancer outcomes, replicating previously reported associations with protein truncating variants in *BRCA2*, *CHEK2* and *PALB2* and cancer outcomes in males and

females (Supplementary Tables 12-14). We also identified a novel association of *SAMHD1* damaging variants and HC-PTVs with 'All cancer' in both males (Odds Ratio (OR)=2.12, 95% CI: 1.72-2.62, P=4.7*10^{-13}) and females (OR=1.61, 95% CI: 1.31-1.96, P=4*10^{-6}; Figure 3, Supplementary Table 12-14).

Figure 0–3: Forest plot for ANM WES genes with significant gene burden associations for cancer phenotypes.



Note: Exome-wide significant ($P < 1.08*10^{-6}$) genes are displayed, showing sex-stratified and combined results. Hormone sensitive cancers were only tested in males and females separately (Methods). The presented masks were selected based on the most significant association per gene and cancer type. Points and bars indicate OR and 95% CI for specific genes and their variant categories in cancer. Filled symbols indicate a result passing a Bonferroni-corrected significance threshold of $P < 1.08*10^{-6}$.

SAMHD1 associations with cancer appear to be driven by increased risk for multiple site-specific cancers, notably prostate cancer in males, mesothelioma in both males and females, and suggestive evidence for higher breast cancer susceptibility in females (Figure 4, Supplementary Table 15). Although the

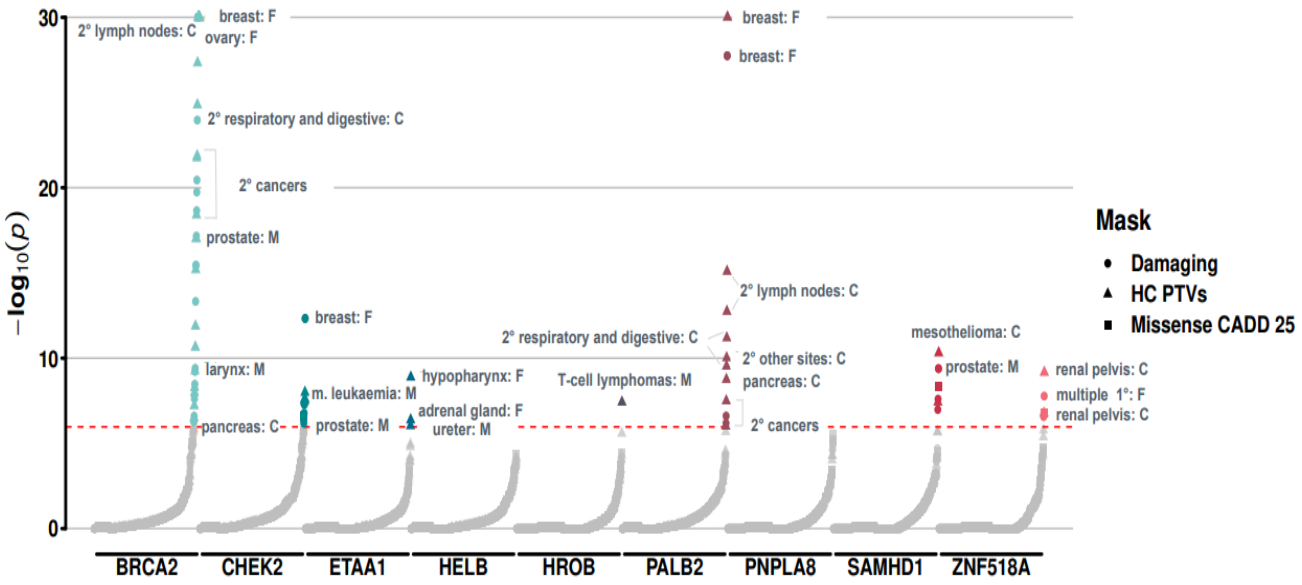
numbers of mutation carriers diagnosed with each site-specific cancer was small, the majority of these findings persisted using logistic regression with penalised likelihood estimation, which is more robust to extreme case/control imbalance[41] (Supplementary Table 15). Cancer risk-increasing alleles in *SAMHD1* were associated with later ANM, which is similar to the pattern demonstrated previously for *CHEK2*. This finding is consistent with a mechanism of disrupted DNA damage sensing and apoptosis, resulting in slowed depletion of the ovarian reserve[11]. We note however that other mechanisms of ovarian reserve depletion will exist, and future experimental work should seek to better understand this specific association. In addition, we provide robust evidence for a previously described rare variant association for *SAMHD1* with telomere length[42], highlighting that rare damaging variants cause longer telomere length (P=1.4*10^{-59}) (Supplementary Table 14, Supplementary Figure 5). Understanding the biological connection between *SAMHD1* and longer telomere length is paramount. As demonstrated in our genome-wide CRISPR–Cas9 functional telomere length screening, *SAMHD1*, a gene associated with deoxynucleoside triphosphohydrolase activity, emerged as a key determinant influencing telomere length[277]. Targeted genetic disruptions revealed that inactivation of *SAMHD1* led to telomere lengthening, suggesting a unique role for this gene in telomere maintenance.

Thymidine nucleotide metabolism, controlled by genes such as TYMS and *SAMHD1*, emerges as a critical pathway influencing human telomere length. *SAMHD1*, traditionally known for its role in regulating dNTP levels, surprisingly lengthened telomeres when disrupted. Furthermore, the abstract underscores the remarkable effect of dT supplementation in driving telomere elongation, implicating *SAMHD1* as a control point in this process.

These findings suggest that *SAMHD1*, through its role in thymidine nucleotide metabolism, plays a crucial and perhaps unexpected role in telomere length regulation. The substrate-independent stimulation of telomerase activity by thymidine triphosphate, as demonstrated in vitro, adds a layer of complexity to our understanding. This unique relationship between *SAMHD1* and telomere length may have therapeutic implications, particularly in the context of fatal degenerative diseases. Further exploration into the intricate interplay between *SAMHD1*'s biological function and telomere dynamics could potentially unveil

novel avenues for therapeutic interventions in diseases associated with telomere dysfunction.

Figure 0–4: Genetic susceptibility to premature ovarian ageing and increased risk for diverse cancer types.



Note: Plot showing the association between loss of ANM genes identified in this study and risk of 90 site specific cancers among UK Biobank participants. Summary statistics for cancer associations were obtained using a logistic regression with penalised likelihood estimation that controls for case/control imbalance (Methods)[34]. Associations highlighted in text passed exome-wide significance ($P < 1.08*10^{-6}$). The y-axis is capped at $-\log_{10}(P) = 30$ for visualisation purposes; un-capped summary statistics can be found in ***Supplementary Table 15****. **F:** females, **M:** males, **C:** sex-combined. **1°:** primary cancer, **2°:** secondary cancer. Different colour shades denote associations with each of the 9 genes.*

Discussion

Previous studies have demonstrated that parental age is strongly associated with the number of *de novo* mutations in offspring[43], with the majority of these mutations arising from the high rate of spermatogonial stem cell divisions that underlie spermatogenesis throughout adult life of males[44]. Our current study provides the first direct evidence that maternal mutation rate is heritable, with women at higher genetic risk of earlier menopause transmitting an increased rate of *de novo* mutations to offspring. This could have direct implications for the health of future generations given the widely reported link between *de novo* mutations and increased risk of psychiatric disease and developmental disorders[45–48]. We speculate that if genetic susceptibility to earlier menopause influences *de novo* mutation rate, it is possible that non-genetic risk factors for earlier ANM, such as smoking and alcohol intake, would likely have the same effect[49]. Our observations makes conceptual sense given that menopause timing appears to be primarily driven by the genetic integrity of oocytes and their ability to sustain, detect, repair and respond to acquired DNA damage[11]. These observations also build on earlier work in mice and humans that *BRCA1*/*2* deficiency increases the rate of double strand breaks in oocytes and reduces ovarian reserve[50–52].

Our study also extends the number of genes implicated in ovarian ageing through the identification of rare, protein-coding variants. Effect sizes ranged from 5.61 years earlier ANM for HC-PTV carriers in *ZNF518A*, to 1.35 years later ANM for women carrying damaging variants in *SAMHD1* compared to a maximum effect size of 1.06 years (median 0.12 years) reported for common variants (MAF>1%)[11]. Several of these effect estimates were comparable to those conferred by *FMR1* premutations, which are currently used as part of the only routinely applied clinical genetics test for premature ovarian insufficiency (POI)[53]. Deleterious variants in three genes (*CHEK2*, *HELB* and *SAMHD1*) were associated with an increase in ANM and therefore represent potential therapeutic targets for enhancing ovarian stimulation in women undergoing *in vitro* fertilisation (IVF) treatment through short-term apoptotic inhibition. Seven of the nine ANM genes identified have known roles in DNA damage repair, and three of these are linked to ANM for the first time (*PALB2*, *ETAA1* and *HROB*): *PALB2* is involved in *BRCA2* localization and stability and compound

heterozygous mutations result in Fanconi anaemia and predispose to childhood malignancies[54]. *ETAA1* accumulates at DNA damage sites in response to replication stress[55–58] and *HROB* is involved in homologous recombination by recruiting the *MCM8-MCM9* helicase to sites of DNA damage to promote DNA synthesis[59,60]. Homozygous loss-of-function of *HROB* is associated with POI[61] and infertility in both sexes in mouse models[59].

Novel biological mechanisms of ovarian ageing were revealed by finding associations with two non-DDR genes (*PNPLA8* and *ZNF518A*): *PNPLA8* is a calcium-independent phospholipase[62–64] and a recessive cause of neurodegenerative mitochondrial disease and mitochondrial myopathy[65–69]; an association with reproductive phenotypes has not been described previously. *ZNF518A* belongs to the zinc finger protein family and is likely a transcriptional regulator for a large number of genes[30]. We found that female carriers of rare protein truncating variants in *ZNF518A* have shorter reproductive lifespan due to delayed puberty timing and earlier menopause. Enrichment of GWAS signals at *ZNF518A* binding sites suggests that *ZNF518A* regulates the genes involved in reproductive longevity by repression of elements distal to transcription start sites.

While mutation in *SAMHD1* is a common somatic event in a variety of cancers[70–80], it has not been described as a germline risk factor previously. Recessive inheritance of *SAMHD1* missense and PTV variants have been associated with Aicardi–Goutieres syndrome, a congenital autoimmune disease[81]. Our identified damaging variants in *SAMHD1* increased risk of ‘All cancer’ in males and females, as well as in sex-specific cancers, highlighting *SAMHD1* as a novel risk factor for prostate cancer in males and hormone-sensitive cancers in females. *SAMHD1* has a role in preventing the accumulation of excess deoxynucleotide triphosphates (dNTPs), particularly in non-dividing cells[82]. A regulated dNTP pool is important for the fidelity of DNA repair, thus highlighting additional roles of this gene in facilitation of DNA end resection during DNA replication and repair[82–87]. *SAMHD1* deficiency leads to resistance to apoptosis[88,89], suggesting that delayed ANM might originate from slowed depletion of ovarian reserve due to disrupted apoptosis, analogous to the mechanism for *CHEK2* that has been reported previously.

One major limitation of our work is that, based on data availability, these analyses have been restricted to women of European ancestry, making it difficult to evaluate how generalizable these findings may be to other populations. We anticipate this will be addressed in future studies, in addition to experimental work that will build on our identified genetic associations to help further our understanding of the underlying biological mechanisms governing ovarian ageing.

Acknowledgements

This work was funded by the Medical Research Council (Unit programs: MC_UU_12015/2, MC_UU_00006/2, MC_UU_12015/1, and MC_UU_00006/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. This research was conducted using the UK Biobank Resource under application 9905 (University of Cambridge) and 9072 and 871 (University of Exeter). A full list of funding sources and acknowledgements can be found in the supplementary information.

Disclosures

John Perry is an employee and shareholder of Adrestia Therapeutics.

Data availability

All data used in discovery analyses are available upon application from the UK Biobank study and Genomics England.

References

1. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433–438 (2017).

2. Kong, A. *et al*. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471–475 (2012).

3. Kaplanis, J. *et al*. Genetic and chemotherapeutic influences on germline hypermutation. *Nature* 605, 503–508 (2022).

4. Rahbari, R. *et al*. Timing, rates and spectra of human germline mutation. *Nat Genet*. 48, 126–133 (2016).

5. Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet*. 1, 40–47 (2000).

6. Gao, Z. *et al*. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc Natl Acad Sci. U. S. A*. 116, 9491–9500 (2019).

7. Jónsson, H. *et al*. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 549, 519–522 (2017).

8. Yauk, C. L. *et al*. Mainstream tobacco smoke causes paternal germ-line DNA mutation. *Cancer Res*. 67, 5103–5106 (2007).

9. Goldmann, J. M. *et al*. Differences in the number of de novo mutations between individuals are due to small family-specific effects and stochasticity. *Genome Res*. 31, 1513–1518 (2021).

10. Hwang, S., Neville, M. D. C., Genomics England Research Consortium, Day, F. R. & Scally, A. Heritability of de novo germline mutation reveals a contribution from paternal but not maternal genetic factors. bioRxiv 2022.12.17.520885 (2022) doi:10.1101/2022.12.17.520885.

11. Ruth, K. S. *et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393–397 (2021).

12. He, C. *et al.* Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat Genet* **41**, 724–728 (2009).

13. Horikoshi, M. *et al.* Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nat Commun* **9**, (2018).

14. Stolk, L. *et al.* Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat Genet* **41**, 645–647 (2009).

15. Hartge, P. Genetics of reproductive lifespan. *Nat Genet* **41**, 637–638 (2009).

16. Perry, J. R. B. *et al.* A genome-wide association study of early menopause and the combined impact of identified variants. *Hum Mol Genet* **22**, 1465–1472 (2013).

17. Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* **47**, 1294–1303 (2015).

18. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* **44**, 260–268 (2012).

19. Perry, J. R. B. *et al.* DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. *Hum Mol Genet* **23**, 2490–2497 (2014).

20. Murray, A. *et al.* Common genetic variants are significant risk factors for early menopause: results from the Breakthrough Generations Study. *Hum Mol Genet* **20**, 186–192 (2011).

21. Ward, L. D. *et al.* Rare coding variants in DNA damage repair genes associated with timing of natural menopause. *HGG Adv* **3**, (2021).

22. Perry, J. R. B., Murray, A., Day, F. R. & Ong, K. K. Molecular insights into the aetiology of female reproductive ageing. *Nat Rev Endocrinol* **11**, 725–734 (2015).

23. te Velde, E. R. & Pearson, P. L. The variability of female reproductive ageing. *Hum Reprod Update* **8**, 141–154 (2002).

24. te Velde, E. R., Dorland, M. & Broekmans, F. J. Age at menopause as a marker of reproductive ageing. *Maturitas* **30**, 119–125 (1998).

25. Halldorsson, B. *et al*. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, 6425 (2019).

26. Jonsson, H. *et al*. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549** (7673), 519-522 (2017).

27. Kaplanis, J. *et al*. Genetic and chemotherapeutic influences on germline hypermutation. *Nature* **605** (7910), 503-508 (2022).

28. Szustakowski, J. D. *et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet* **53**, 942–948 (2021).

29. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284–290 (2015).

30. Maier, V. K. *et al.* Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol Cell Proteomics* **14**, 1435–1446 (2015).

31. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

32. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794–D801 (2018).

33. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559–573 (2014).

34. Reshef, Y. A. *et al.* Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. *Nat Genet* **50**, 1483–1493 (2018).

35. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165–D173 (2022).

36. Chen, D. *et al.* Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. *Cell Rep* **29**, 4568-4582.e5 (2019).

37. Chen, D. *et al.* The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. *Cell Rep* **25**, 3591-3602.e5 (2018).

38. Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300–307 (2021).

39. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834–841 (2017).

40. Day, F. R. *et al.* Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun* **6**, (2015).

41. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38 (1993).

42. Codd, V. *et al.* Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet* **53**, 1425–1433 (2021).

43. Rahbari, R. *et al.* Timing, rates and spectra of human germline mutation. *Nat Genet* **48**, 126–133 (2016).

44. Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* **1**, 40–47 (2000).

45. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

46. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).

47. Wang, W., Corominas, R. & Lin, G. N. De novo Mutations From Whole Exome Sequencing in Neurodevelopmental and Psychiatric Disorders: From Discovery to Application. *Front Genet* **10**, (2019).

48. Coe, B. P. *et al.* Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet* **51**, 106–116 (2019).

49. Linschooten, J. O. *et al.* Paternal lifestyle as a potential source of germline mutations transmitted to offspring. *FASEB J* **27**, 2873–2879 (2013).

50. Titus, S. *et al.* Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci Transl Med* **5**, (2013).

51. Miao, Y. *et al.* BRCA2 deficiency is a potential driver for human primary ovarian insufficiency. *Cell Death Dis* **10**, (2019).

52. Lin, W., Titus, S., Moy, F., Ginsburg, E. S. & Oktay, K. Ovarian Aging in Women With BRCA Germline Mutations. *J Clin Endocrinol Metab* **102**, 3839–3847 (2017).

53. Sherman, S. L. Premature ovarian failure in the fragile X syndrome. *Am J Med Genet* **97**, 189–194 (2000).

54. Reid, S. *et al.* Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* **39**, 162–164 (2007).

55. Bass, T. E. *et al.* ETAA1 acts at stalled replication forks to maintain genome integrity. *Nat Cell Biol* **18**, 1185–1195 (2016).

56. Feng, S. *et al.* Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks. *J Biol Chem* **291**, 21956–21962 (2016).

57. Haahr, P. *et al.* Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat Cell Biol* **18**, 1196–1207 (2016).

58. Saldivar, J. C. *et al.* An intrinsic S/G 2 checkpoint enforced by ATR. *Science* **361**, 806–810 (2018).

59. Hustedt, N. *et al.* Control of homologous recombination by the HROB-MCM8-MCM9 pathway. *Genes and Development* **33**, 1397–1415 (2019).

60. Huang, J. W. *et al.* MCM8IP activates the MCM8-9 helicase to promote DNA synthesis and homologous recombination upon DNA damage. *Nat Commun* **11**, (2020).

61. Tucker, E. J. *et al.* Meiotic genes in premature ovarian insufficiency: variants in HROB and REC8 as likely genetic causes. *Eur J Hum Genet* **30**, 219–228 (2022).

62. Hara, S., Yoda, E., Sasaki, Y., Nakatani, Y. & Kuwata, H. Calcium-independent phospholipase A 2 γ (iPLA 2 γ) and its roles in cellular functions and diseases. *Biochim Biophys Acta Mol Cell Biol Lipids* **1864**, 861–868 (2019).

63. Liu, G. Y. *et al.* The phospholipase iPLA 2 γ is a major mediator releasing oxidized aliphatic chains from cardiolipin, integrating mitochondrial bioenergetics and signaling. *J Biol Chem* **292**, 10672–10684 (2017).

64. Mancuso, D. J. *et al.* Genetic ablation of calcium-independent phospholipase A2gamma leads to alterations in mitochondrial lipid metabolism and function resulting in a deficient mitochondrial bioenergetic phenotype. *J Biol Chem* **282**, 34611–34622 (2007).

65. Shukla, A., Saneto, R. P., Hebbar, M., Mirzaa, G. & Girisha, K. M. A neurodegenerative mitochondrial disease phenotype due to biallelic loss-of-function variants in PNPLA8 encoding calcium-independent phospholipase A2γ. *Am J Med Genet A* **176**, 1232–1237 (2018).

66. Mancuso, D. J. *et al.* Genetic ablation of calcium-independent phospholipase A2gamma leads to alterations in hippocampal cardiolipin content and molecular species distribution, mitochondrial degeneration, autophagy, and cognitive dysfunction. *J Biol Chem* **284**, 35632–35644 (2009).

67. Mancuso, D. J. *et al.* Genetic ablation of calcium-independent phospholipase A2gamma prevents obesity and insulin resistance during high fat feeding by mitochondrial uncoupling and increased adipocyte fatty acid oxidation. *J Biol Chem* **285**, 36495–36510 (2010).

68. Saunders, C. J. *et al.* Loss of function variants in human PNPLA8 encoding calcium-independent phospholipase A2 γ recapitulate the mitochondriopathy of the homologous null mouse. *Hum Mutat* **36**, 301–306 (2015).

69. Masih, S., Moirangthem, A. & Phadke, S. R. Homozygous Missense Variation in PNPLA8 Causes Prenatal-Onset Severe Neurodegeneration. *Mol Syndromol* **12**, 174–178 (2021).

70. Clifford, R. *et al.* SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* **123**, 1021–1031 (2014).

71. Schott, K. *et al.* SAMHD1 in cancer: curse or cure? *J Mol Med (Berl)* **100**, 351–372 (2022).

72. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).

73. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435–439 (2011).

74. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).

75. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).

76. Herold, N. *et al.* With me or against me: Tumor suppressor and drug resistance activities of SAMHD1. *Exp Hematol* **52**, 32–39 (2017).

77. Coggins, S. A., Mahboubi, B., Schinazi, R. F. & Kim, B. SAMHD1 Functions and Human Diseases. *Viruses* **12**, (2020).

78. Chen, Z., Hu, J., Ying, S. & Xu, A. Dual roles of SAMHD1 in tumor development and chemoresistance to anticancer drugs. *Oncol Lett* **21**, (2021).

79. Xagoraris, I. *et al.* Expression of the novel tumour suppressor sterile alpha motif and HD domain-containing protein 1 is an independent adverse prognostic factor in classical Hodgkin lymphoma. *Br J Haematol* **193**, 488–496 (2021).

80. Merati, M. *et al.* Aggressive CD8(+) epidermotropic cutaneous T-cell lymphoma associated with homozygous mutation in SAMHD1. *JAAD Case Rep* **1**, 227–229 (2015).

81. Rice, G. I. *et al.* Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat Genet* **41**, 829–832 (2009).

82. Franzolin, E. *et al.* The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci U S A* **110**, 14272–14277 (2013).

83. Kumar, D. *et al.* Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Res* **39**, 1360–1371 (2011).

84. Coquel, F. *et al.* SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature* **557**, 57–61 (2018).

85. Daddacha, W. *et al.* SAMHD1 Promotes DNA End Resection to Facilitate DNA Repair by Homologous Recombination. *Cell Rep* **20**, 1921–1935 (2017).

86. Mathews, C. K. Deoxyribonucleotide metabolism, mutagenesis and cancer. *Nat Rev Cancer* **15**, 528–539 (2015).

87. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* **7**, 2902–2906 (2008).

88. Bonifati, S. *et al.* SAMHD1 controls cell cycle status, apoptosis and HIV-1 infection in monocytic THP-1 cells. *Virology* **495**, 92–100 (2016).

89. Kodigepalli, K. M., Li, M., Liu, S. L. & Wu, L. Exogenous expression of SAMHD1 inhibits proliferation and induces apoptosis in cutaneous T-cell lymphoma-derived HuT78 cells. *Cell Cycle* **16**, 179–188 (2017).

90. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).

91. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

92. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

93. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, (2016).

94. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* **13**, (2021).

95. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

96. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284–290 (2015).

97. Seabold, S. P. J. Statsmodels: Econometric and Statistical Modeling with Python. *SCIPY* (2010).

98. Breast Cancer Association Consortium *et al*. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* **384**(5), 428–439 (2021).

99. Kong, A. *et al*. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* **40**, 1068-1075 (2008).

100. Gudbjartsson, D. F. *et al*. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* **47**, 435-444 (2015).

101. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* **37**, 235-241 (2016).

102. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med* **12**, 103 (2020).

103. Ma, C., Blackwell, T., Boehnke, M. & Scott, L. J. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* **37**, 539–550 (2013).

104. Kosmidis, I. K. P. E. C. S. N. Mean and median bias reduction in generalized linear models. *Statistics and Computing* **30**, 43–59 (2019).

105. Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci Rep* **6**, (2016).

106. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097–1103 (2021).

107. Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873.e4 (2017).

108. Zhang, Y. *et al.* Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. *Mol Cell* **72**, 1021-1034.e4 (2018).

109. Lou, S. *et al.* TopicNet: a framework for measuring transcriptional regulatory network change. *Bioinformatics* **36**, I474–I481 (2020).

110. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

111. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589 (2010).

112. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215–216 (2012).

113. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

114. Layer, R. M. *et al.* GIGGLE: a search engine for large-scale integrated genome analysis. *Nat Methods* **15**, 123–126 (2018).

115. Smith, G. D. & Ebrahim, S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1–22 (2003).

116. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat Commun* **11**, (2020).

117. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133–1163 (2008).

118. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* **4**, (2020).

119. Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet Epidemiol* **44**, 313–329 (2020).

120. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* **36**, 1783–1802 (2017).

121. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* **28**, 30–42 (2017).

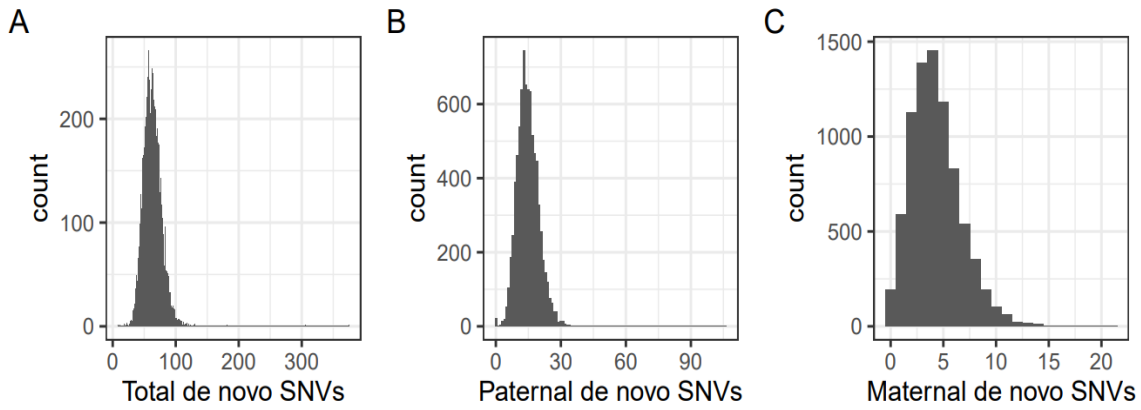
122. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512–525 (2015).

123. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304–314 (2016).

124. Bowden, J. *et al.* Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int J Epidemiol* **47**, 1264–1278 (2018).

Supplementary Figures

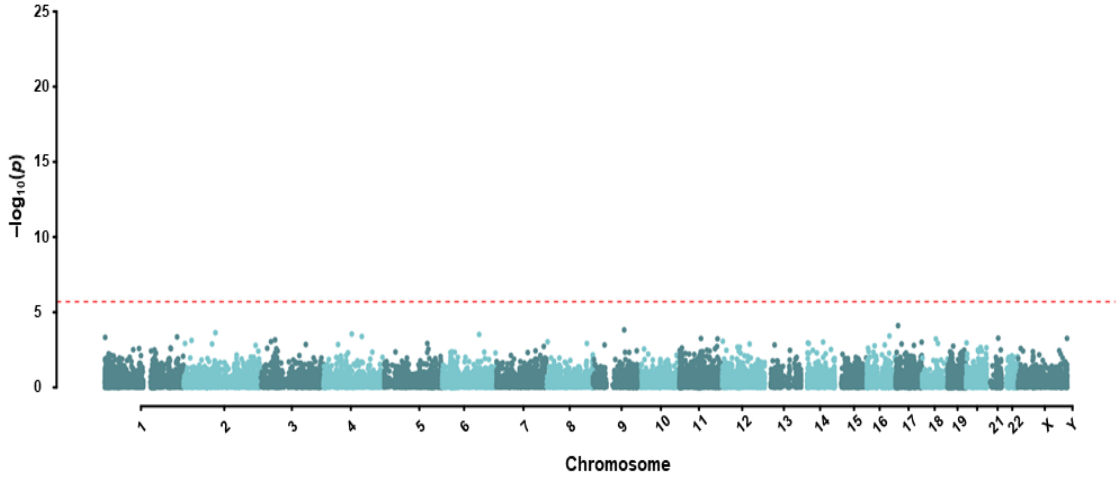
Figure 0–5: Distribution of de novo single nucleotide variants (dnSNVs).



Supplementary Figure 1

*Note: The histogram shows the number of **(A)** total dnSNVs, **(B)** paternally derived dnSNVs and **(C)** maternally derived dnSNVs in unrelated probands with European ancestry from the 100,000 Genomes Project.*

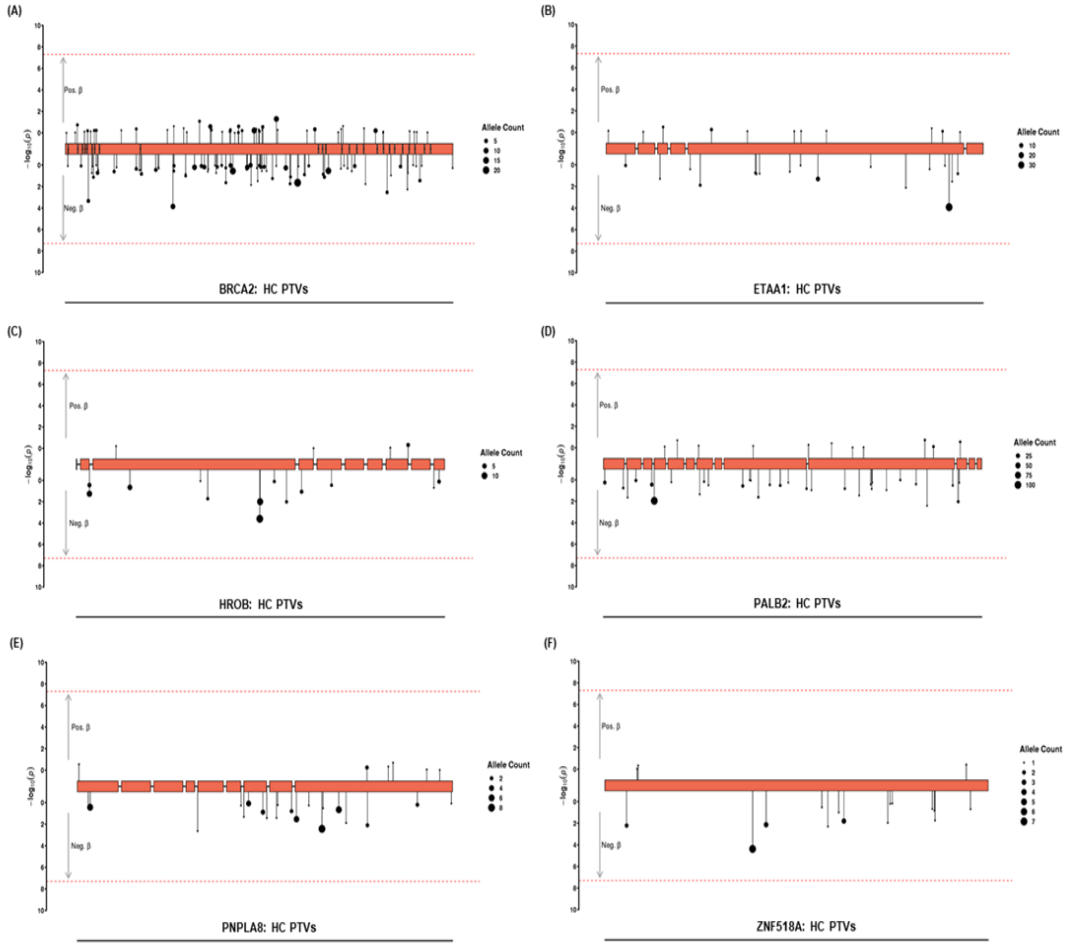
Figure 0–6: Exome-wide association results for synonymous variants.



Supplementary Figure 2

Note: Plotted are per-gene burden results for synonymous variants. The red line indicates the exome-wide significant P value after Bonferroni correction of $1.08*10^{-6}$.

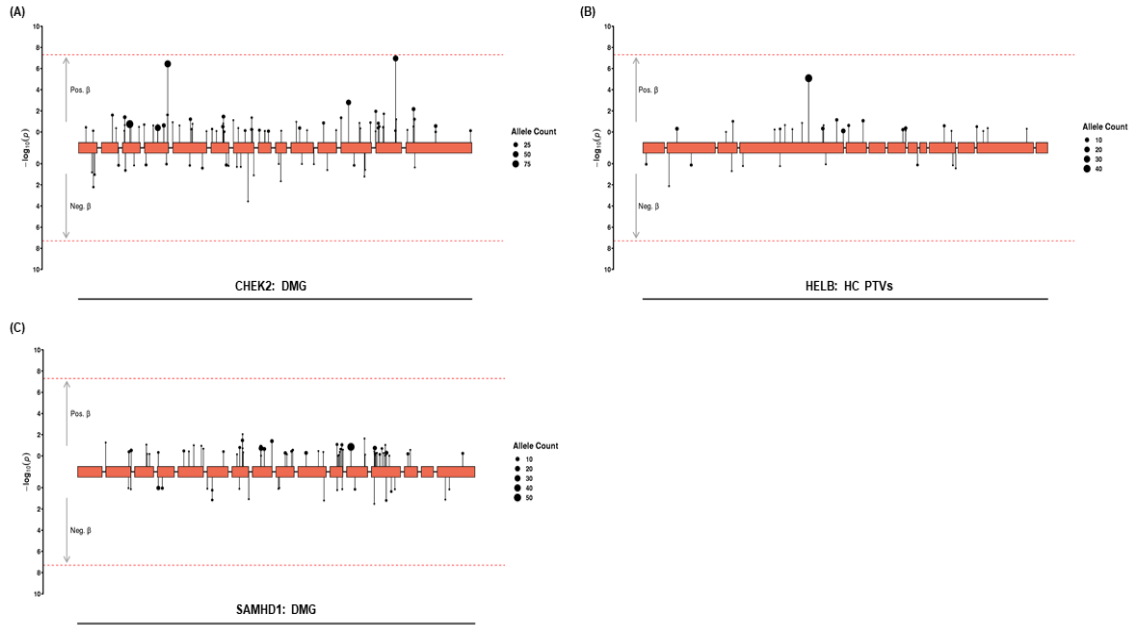
Figure 0–7: Variant level associations for ANM decreasing WES genes.



Supplementary Figure 3

Note: Lolliplot plots show the variants clustered for the best performing functional mask in a gene that went into the gene burden test for ANM using BOLT-LMM. These include: ***(A)*** *BRCA2 HC PTV mask;* ***(B)*** *ETAA1, HC PTV mask;* ***(C)*** *HROB, HC PTV mask;* ***(D)*** *PALB2, HC PTV mask;* ***(E)*** *PNPLA8, HC PTV mask; and* ***(F)*** *ZNF518A, HC PTV mask. The arrows pointing upwards represent the variants positively associated with ANM, while the downwards ones show the negatively associated variants. The size of the point indicates the allele count in carriers.*

Figure 0–8: Variant level associations for ANM increasing WES genes.



Supplementary Figure 4

Note: Lolliplot plots show the variants clustered for the best performing functional mask in a gene that went into the gene burden test for ANM using BOLT-LMM. These include: ***(A)*** *CHEK2, damaging mask;* ***(B)*** *HELB, HC PTV mask; and* ***(C)*** *SAMHD1, damaging mask. The arrows pointing upwards represent the variants positively associated with ANM, while the downwards ones show the negatively associated variants. The size of the point indicates the allele count in carriers.*

Supplementary notes

Comparison with previously published results on ANM WES

A previous ANM analysis of the 450K UKBB exome data, published by Ward and colleagues, identified seven genes: *CHEK2*, *DCLRE1A*, *HELB*, *CLPB*, *TOP3A*, *RAD54L* and *HROB*[21]. Our results replicate the association with *CHEK2* and *HELB* and provide more robust evidence ($P = 1.9*10^{-8}$) for the previously described suggestive *HROB* association ($P = 2.9*10^{-6}$). In addition, we identified six genes, which were not captured by Ward *et al*.: *ZNF518A*, *BRCA2*, *ETAA1*, *PALB2*, *PNPLA8* and *SAMHD1* (Supplementary Table 7). We investigated potential study design differences that could account for variation in the findings as the same data was used in both studies.

Associations not captured in the current analysis.

First, we investigated potential analytical parameters that could account for differences in the findings. Associations with *DCLRE1A*, *RAD54L*, *TOP3A* and *CLPB* were not identified in our study because we restricted our analysis to variants with a MAF <0.1%, rather than <1%. We re-analysed our data with a burden test MAF threshold of <1% and three of the four associations were replicated: *DCLRE1A* ($P_{\text{MAF 1\%}} = 3.8*10^{-8}$, N: 1056), *RAD54L* ($P_{\text{MAF 1\%}} = 6.4*10^{-7}$, N: 1892) and *TOP3A* ($P_{\text{MAF 1\%}} = 1.5*10^{-7}$, N: 2001). *RAD54L and TOP3A* were genes highlighted by GWAS and the exome association in *TOP3A* was driven by a single, relatively common variant (rs34001746, MAF=0.7%, $P=1.63*10^{-10}$). This variant was in linkage disequilibrium with the previously reported lead GWAS SNP (rs569145577, $r^2=0.92$), with little evidence for association after its exclusion ($P=0.50$ for all other missense and PTVs). *CLPB* just missed our *P*-value threshold, but again a single variant (rs150343959, $P=8.22*10^{-6}$) was largely driving the association signal - when excluded in leave-one-out analysis, the *CLPB* burden association dropped ($P=1.19*10^{-2}$). By including relatively common variants in gene burden masks, single variants can dominate the general functional effect being tested, which could be contributed to by LD with non-exomic functional variants. Therefore in order to be able to make a stronger link between genetic variants and individual genes, we chose to restrict our analysis to rarer variants with MAF <0.1%.

Associations not captured by Ward *et al.*

Differences in MAF thresholds did not explain why our study identified an additional six genes (*BRCA2*, *ETAA1*, *PALB2*, *PNPLA8*, *SAMHD1* and *ZNF518A*) compared with Ward *et al*[21]. We therefore tested differences in the phenotype preparation, tools and variant masks used to test the associations. Four of our six gene burden associations (*BRCA2*, *PALB2*, *PNPLA8*, and *SAMHD1*) were relatively near the borderline of the significance threshold in our analyses in the primary BOLT-LMM pipeline, although *ETAA1* and *PALB2* were just below the threshold in the REGENIE pipeline (Supplementary Table 7).

We had a ~20% larger sample size (106,973 post-menopausal women) in comparison to Ward *et al*.[21] (78,311 unrelated post-menopausal women), which would have resulted in more statistical power in our analyses (Supplementary Table 7). This was particularly important for *BRCA2*, *ETAA1* and *PALB2* - Ward *et al.* included 63 (19.5%) fewer *BRCA2* and 46 (21.7%) fewer *PALB2* carriers of rare damaging variants (Ward *et al.* REGENIE analysis vs. our main analysis) and identified the ANM association at these genes only at the borderline of exome-wide significance, $P=1.55*10^{-6}$ and $P=7.47*10^{-5}$, respectively. Similarly for *SAMHD1*, Ward *et al.* captured 57 (24.3%) fewer carriers in their linear regression model compared with our main analysis, which resulted in association *P* values of $6.38*10^{-4}$ in the linear regression model and $P=8.02*10^{-6}$ in the time to event analysis. Ward *et al.* used only unrelated individuals in their primary analyses, where we used linear mixed models and were therefore able to include an additional ~19,000 related individuals. Secondly, Ward *et al.* excluded ~2,300 women with ANM <40 and >60 years, while we used the full natural menopause distribution. Finally, the difference in sample size was partly due to differences in phenotype preparation (resulting in an additional ~7,400 women); specifically, we took into account four instances where questions regarding ANM were asked, whereas Ward *et al.* used data from the baseline visit in their main analysis.

As sensitivity analyses and to better replicate the methods of Ward *et al.*, for four of the genes (*CHEK2* and *DCLRE1A* found in both analyses, and the novel associations in *ZNF518A* and *PNPLA8*) we compared results from linear regression in unrelated individuals using MAF<1% with those from a truncated

menopause distribution (ANM 40-60 years) and a time-to-event Cox proportional hazards model (Supplementary Table 7). For the truncated distribution, all four genes passed the threshold of exome-wide association and, for all with the exception of *ZNF518A*, the association *P* value was larger (*CHEK2*: $3.3*10^{-35}$, *DCLRE1A*: $1.3*10^{-7}$, *ZNF518A*: $8.4*10^{-11}$, *PNPLA8*: $3*10^{-8}$) than for analyses based on the full range of ANM. Association statistics from the Cox model (*CHEK2*: $2.4*10^{-39}$, *DCLRE1A*: $6*10^{-8}$, *ZNF518A*: $1.4*10^{-9}$, *PNPLA8*: $5.7*10^{-10}$) were comparable to those from linear regression models based on the full range of ANM (*CHEK2*: $3.1*10^{-46}$, *DCLRE1A*: $2.5*10^{-8}$, *ZNF518A*: $1.2*10^{-9}$, *PNPLA8*: $1.9*10^{-9}$).

Finally, differences in variant annotation may also explain some inconsistencies between studies. *ZNF518A* was not reported by Ward *et al.*, which may be because all variants are in the last and only coding exon of the gene, and in some annotations, such variants would inappropriately be excluded from being considered as loss of function. We note that another single coding-exon gene (*NFIL3*) was not included in the Ward *et al.* publication but was in our analysis.

Detailed comparisons between our study and Ward *et al.* are available in Supplementary Table 7.

Common ANM-associated variants are enriched in *ZNF518A* binding sites.

Heterozygous loss of function of *ZNF518A* had the largest effect on ANM of the genes we identified. *ZNF518A* is a poorly characterised C2H2 zinc finger transcription factor, which has been shown to associate with PRC2 and G9A-GLP repressive complexes along with its paralog *ZNF518B*, suggesting a potential role in transcriptional repression[30]. *ZNF518A* localises robustly to 18,706 sites in the genome, based on ChIP-seq data available from ENCODE[31,32] and binds primarily to gene promoters, with 33.5% (6,263) of *ZNF518A* binding sites within 2kb of a transcription start site (TSS) (Supplementary Note Figure 1a-c). Common variants associated with ANM[11] were enriched in the transcriptional targets of *ZNF518A* (P=$1.32*10^{-4}$) using fGWAS[33]. We further tested functional enrichment using signed linkage disequilibrium profile (SLDP) regression[34]. This confirmed the enrichment of *ZNF518A* binding sites near to loci associated with ANM and showed that its transcriptional repression is associated with earlier ANM (P=0.02), consistent

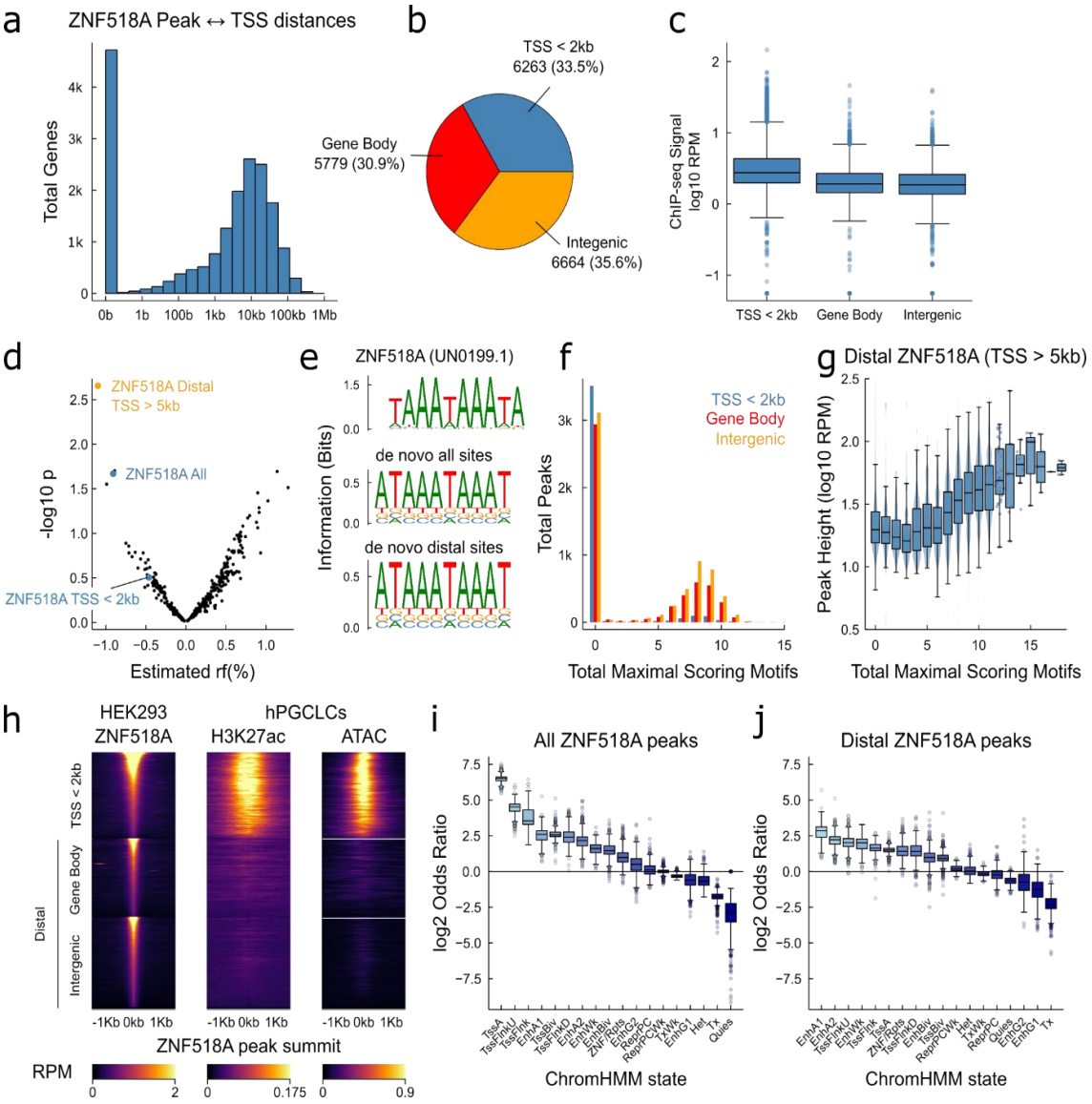
with evidence from rare variant burden tests. Separating *ZNF518A* sites by those proximal (< 2Kb) and distal (>5kb) from a TSS, demonstrated this association was due to *ZNF518A* binding at regulatory regions distal to the TSS (proximal TSS *P*=0.3, distal *ZNF518A P*=0.002). Notably, these regulatory *ZNF518A* bound loci produce the largest association amongst an SLDP catalogue of 382 transcription factors and regulators (Supplementary Table 9, Supplementary Note Figure 1d). These results suggest a different functional role for *ZNF518A* at TSS and more distal regulatory regions. In order to explore this further we assessed the sequence determinants of *ZNF518A* binding. *De novo* motif discovery identified an AT-rich motif enriched at distal regulatory *ZNF518A* binding sites, but not at TSS bound by *ZNF518A*. This AT-rich motif was centrally enriched within *ZNF518A* ChIP-seq peaks, and matched an unvalidated motif present in the JASPAR transcription factor motif database[35] (Supplementary Note Figure 1e). We found the number of perfect instances of this AT-rich motif to be strongly associated with *ZNF518A* occupancy as assessed by *ZNF518A* ChIP-seq signal at distal regions but not at TSS (Supplementary Note Figure 1f,g). At distal regions, the maximal association between peaks greater than the median height was found at least seven motif instances (Hypergeometric right tail $P < 10^{-389}$, Odds Ratio 7.41). These data suggest that *ZNF518A* is recruited by DNA sequence at distal sites, but at TSS may be recruited to gene promoters by interaction with another DNA binding factor.

We next employed public *in vitro* differentiated human primordial germ like-cell data[36,37] to assess the chromatin state at *ZNF518A* bound loci, directly comparing distal regions with TSS. *ZNF518A* bound TSS showed chromatin accessibility[37] and were marked with H3K27ac[36]. In contrast, distal regions lacked H3K27ac and showed minimal chromatin accessibility (Supplementary Note Figure 1h). Extending this comparison to the Epimap chromatin states[38], we find that overall *ZNF518A* bound loci are enriched in active TSS and that distal *ZNF518A* regions are variously enriched in active and repressed chromatin (Supplementary Note Figure 1i,j). Consistent with previous data which has found *ZNF518A* in repressive complexes, these data suggest that *ZNF518A* is recruited by DNA sequence to distal regulatory regions where it acts to repress local chromatin.

While *ZNF518A* is known to have diverse tissue expression including the ovary, we found that it was highly expressed in fetal germ cells at both the mitotic and meiotic stages (Supplementary Tables 10 and 11; Supplementary Note Figures 2 and 3). The eight other WES genes identified in this study were expressed at varying levels in fetal gonadal cells, oocytes and granulosa cells across different developmental stages (Supplementary Note Figures 2 and 3).

Supplementary Figures for Supplementary Note

Figure 0–9: Functional analysis of ZNF518A bound loci.

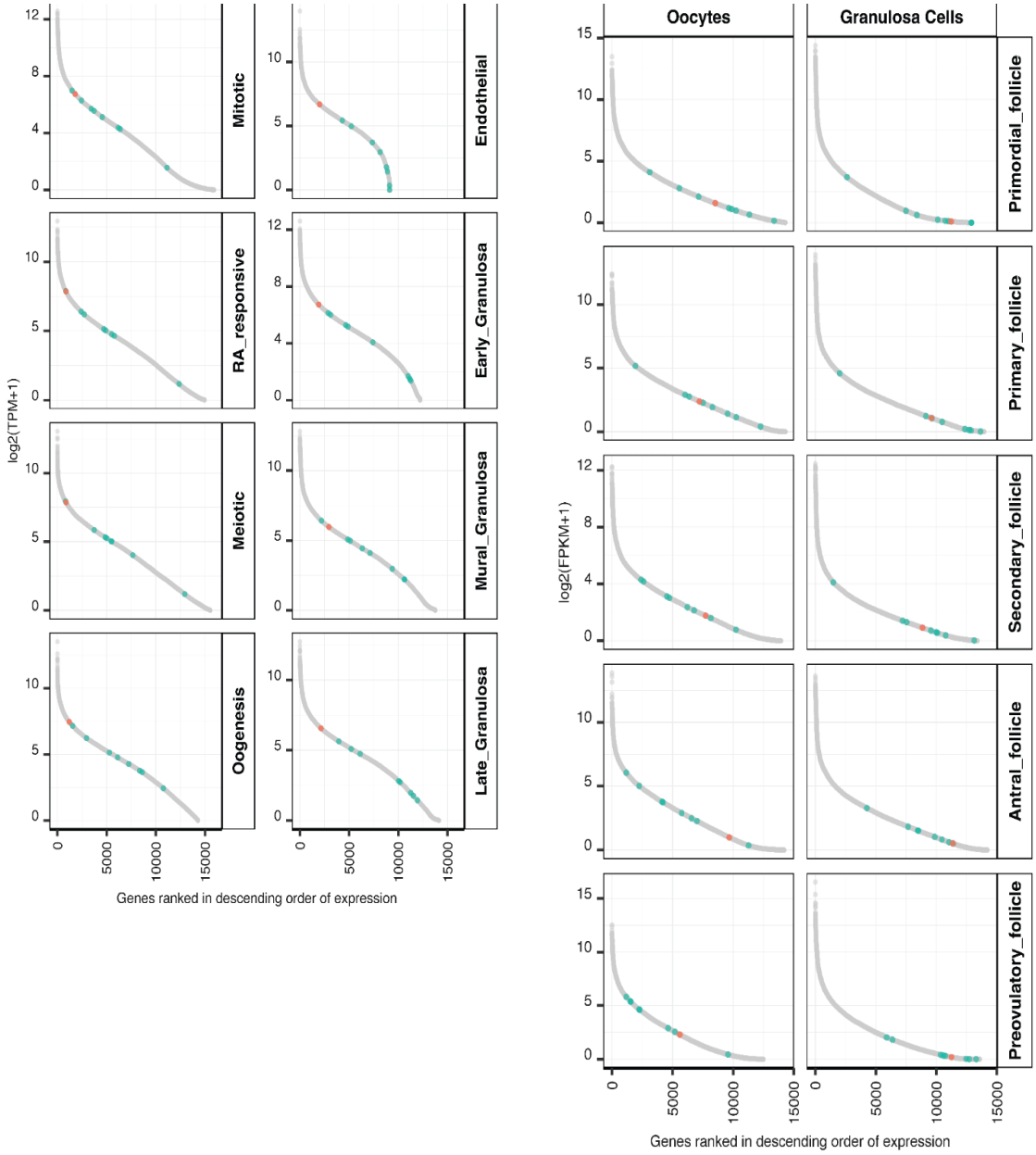


Supplementary Note Figure 1

*Note: **(a)** Histogram of log10-scale distances between ZNF518A and nearest gene transcription start site (TSS). **(b)** Proportion of ZNF18A peaks falling proximal to TSS (TSS < 2kb), within gene bodies and in intergenic regions. **(c)** Boxplots showing total normalised reads per million (RPM) for every peak for categories TSS < 2kb, gene body and intergenic - ZNF518A peaks have greater signal at proximal to TSS. **(d)** SLDP association between ANM GWAS variants and ZNF518A peaks, stratified by all peaks, proximal (< 2kb) from a TSS, and distal (> 5kb) from a TSS. The association between ANM variants and ZNF518A peaks appears due to distal ZNF518A peaks (either gene body or intergenic, > 5kb TSS) and not proximal TSS binding. Numerical results are reported in **Supplementary Table 9**. **(e)** De novo motif discovery recovers unvalidated JASPAR motif for ZNF518A UN0199.1. Homer enrichment statistics: all sites P= 10^{-6451} motif in 31.2% of targets (1.15% background); distal sites P= 10^{-4590} motif in 47.3% of targets (1.81 % background). **(f)** Proportion of maximal scoring instances of UN0199.1 (sequences that exactly match motif consensus) by ZNF518A peak category. Many distal peaks*

contain multiple perfect instances of the motif. ***(g)*** *Boxplots, violin plots and dot plots depicting the relationship between ZNF518A ChIP-seq peak height and number of maximal scoring motifs present in peak. A strong relationship between peak height and number of motif instances can be observed.* ***(h)*** *Heatmaps depicting ZNF518A ChIP-seq, H3K27ac ChIP-seq in hPGCLCs, and chromatin accessibility by ATAC-seq in hPGCLCs. Signal shown over all ZNF518A peaks in RPM +/- 1kb of ZNF518A peak summit. ZNF518A bound promoters (TSS < 2kb) are accessible and are marked with H3K27ac, distal regions either in gene bodies or intergenic regions show no H3K27ac or chromatin accessibility, suggestive that ZNF518A represses these regulatory regions.* ***(i,j)*** *Association shown in odds ratios of ChromHMM states over 833 tissues/cell types from Epimap, boxplots with outliers shown, each boxplot summarises the distribution of associations over all tissues/cell types for a given chromatin state.* ***(i)*** *All ZNF518A peaks;* ***(j)*** *ZNF518A peaks distal from TSS.*

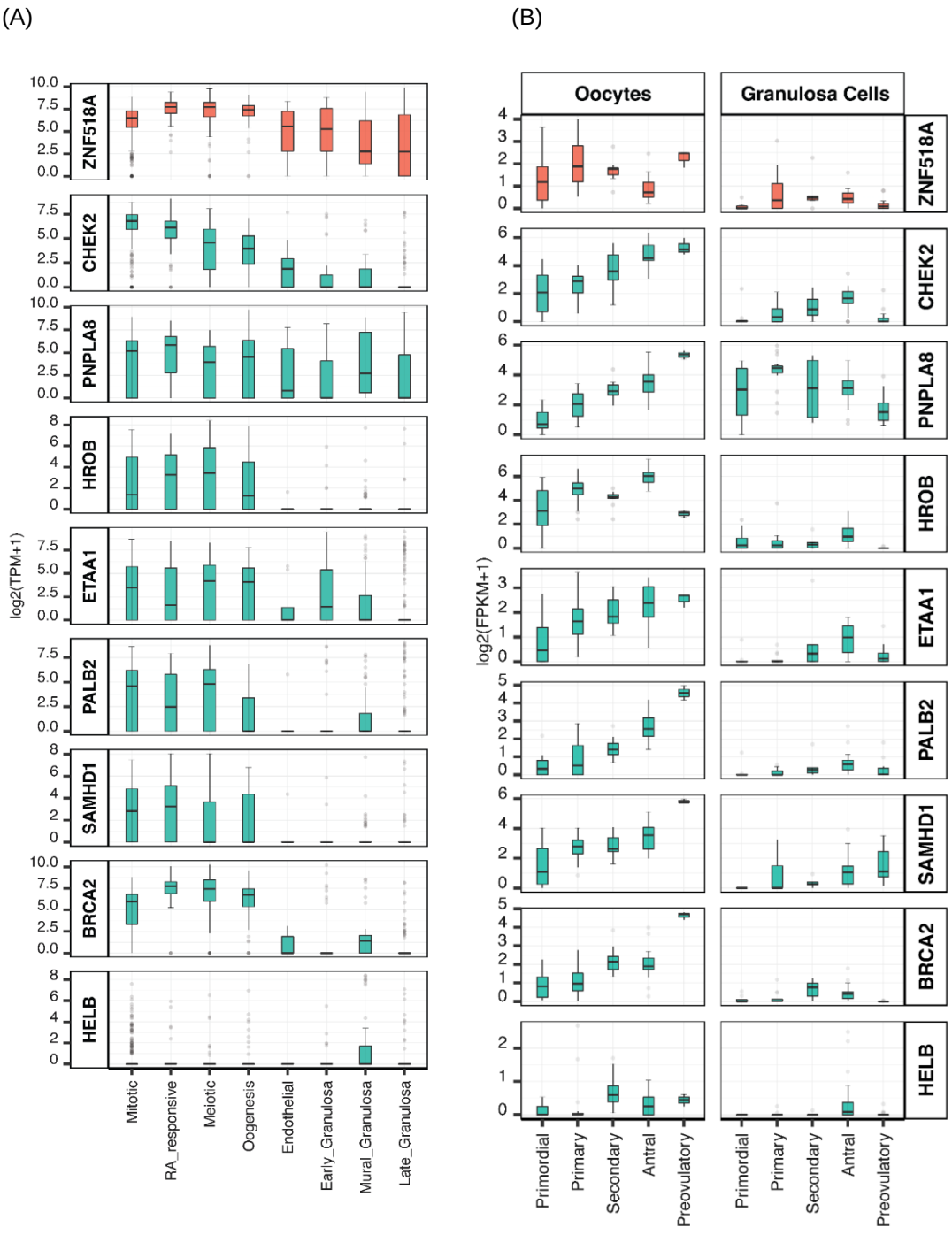
Figure 0–10: Expression levels of genes across various stages of female germ cell development.



Supplementary Note Figure 2

Note: In the X-axis, genes are ranked according to their average expression at each stage (Y-axis) ***(A)*** *in human foetal primordial germ cells and* ***(B)*** *in granulosa cells in adult follicles. Genes identified as novel ANM genes in WES analysis are coloured in green and all other genes in the genome are in grey. ZNF518A is depicted in orange for the ease of comparison with other genes.*

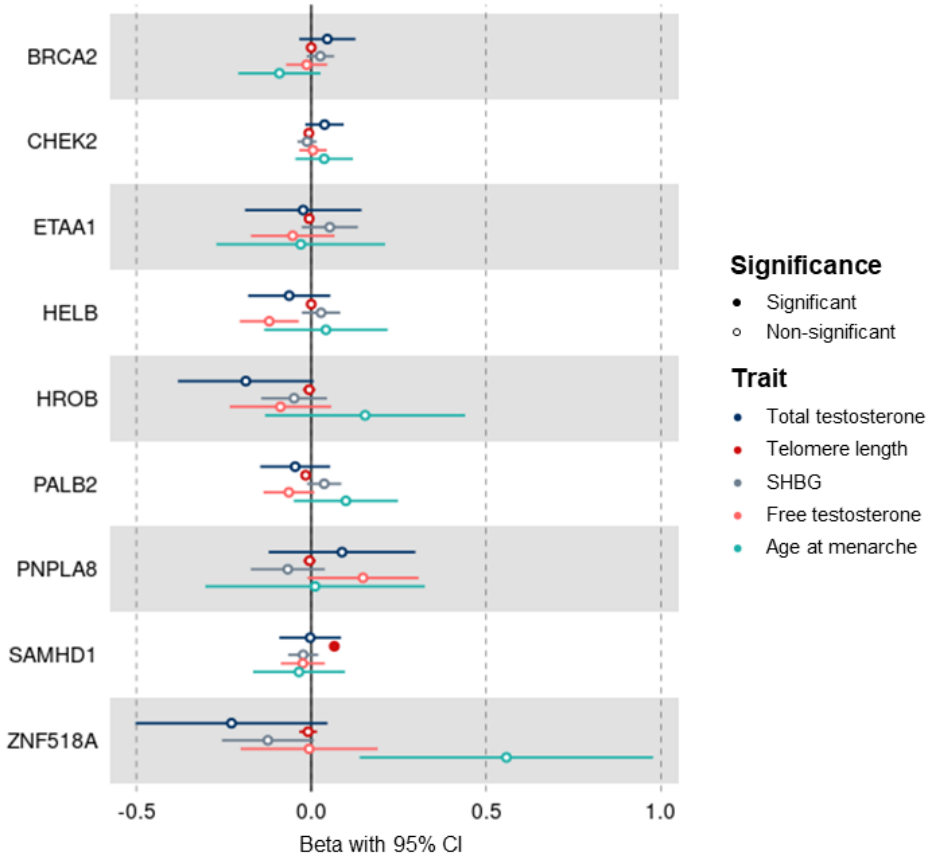
Figure 0–11: mRNA expression of WES genes during foetal stages and folliculogenesis.



Supplementary Note Figure 3

Note: Box and whisker plots of mRNA expression of the WES genes at different stages of germ cell development. The plots represent the interquartile range of TPM values, the line at the centre of the box representing the median, error bars indicate the 95% confidence interval and outliers shown as dots. ***(A)*** *The sub-clusters from single foetal cells from week 5 to 26 post-fertilisation are on the X-axis with the average TPM expression values log2(TPM+1) on the Y-axis.* ***(B)*** *Different stages of folliculogenesis in oocytes and granulosa cells are represented on the X-axis with their average expression values log2(FPKM+1) on the Y-axis.*

Figure 0–12: ANM gene burden associations with reproductive ageing-related traits of interest in females only.



Supplementary Figure 5

Note: The coefficients and 95% CIs were female-specific and plotted for the quantitative traits only. The association was tested using BOLT-LMM. Male-specific and sex combined associations could be found in ***Supplementary Table 14****.*

Chapter 5:

Discussion

This thesis encompasses several studies that significantly enhance our comprehension of the rare and very rare genetic elements that impact the female reproductive lifespan, specifically focusing on POI and premature menopause. In this section, I delve into the effectiveness of these genetic investigations, the broader implications of their findings on our collective knowledge, and recurring patterns evident in the results, and propose potential avenues for future research exploration.

Leveraging Population Cohort Data: Reclassifying Genetic Variants and Uncovering True Penetrance

The issue of pathogenicity and penetrance in genetic variants poses a significant challenge to our understanding of disease development and clinical outcomes. Penetrance signifies the likelihood of a specific genetic variant causing the anticipated trait, while expressivity denotes the varying intensity of that trait among individuals with the same genetic alteration[264]. The exploration of penetrance and expressivity has revealed that genetic variants do not always lead to the expected clinical phenotype, and individuals with the same genotype can manifest a wide range of symptoms[278]. Various factors, including common variants, regulatory region variants, epigenetic influences, environmental factors, and lifestyle choices, contribute to this variability[278,279].

It is essential to recognise that the previous identification of genetic variants causing monogenic disorders relied on small clinical and family studies, potentially leading to an overestimation of their penetrance and expressivity in the general population[280-282]. Understanding the concepts of incomplete penetrance and variable expressivity plays a pivotal role in comprehending the diverse phenotypes observed in individuals affected by genetic disorders. By delving into these concepts, we gain valuable insights into the complexities surrounding the penetrance and expressivity of genetic variants in disease pathogenesis.

Population cohort data, such as UKB, has proved invaluable in analysing the penetrance and expressivity of genetic variants on a larger scale[274]. Leveraging

such data, researchers have reclassified variants previously considered completely penetrant. For example, in Chapters 2 and 3, we analysed over 500 unique single variants associated with POI and found that these variants are unlikely to be the sole cause of the condition due to their high frequency in the population cohort[283].

Notably, this observation has important implications for clinical practice. The discrepancy between the prevalence of these variants in population cohorts and their presumed association with POI challenges the traditional classification of certain variants as solely pathogenic in clinical panels. This discrepancy prompts a revaluation of the clinical utility of these variants and underscores the necessity of incorporating population-based data into the interpretation of genetic test results. The evolving understanding of penetrance and expressivity from large-scale population studies informs clinicians about the true impact of specific genetic variants, guiding more accurate patient counselling and management strategies.

Additionally, investigations into the penetrance of diabetes in individuals with MODY and developmental disorders have consistently indicated that the penetrance of rare Mendelian disease variants can be lower in population-based cohorts compared to clinically ascertained cohorts[279,284-286]. These findings emphasize the substantial impact of population-based studies in accurately understanding the true penetrance of genetic variants[287].

Combining data from population cohorts and clinical studies can enhance our understanding of the penetrance of rare and complex diseases, providing a thorough insight into diseases' aetiology and their impact on individuals. This integration enables a deeper understanding of the incomplete penetrance and variable expressivity observed in genetic variants.

Furthermore, the variation in penetrance and expressivity is influenced by several factors. For instance, genetic modifiers can interact with the primary gene to enhance or suppress its effects, leading to different outcomes in individuals with the same variant[288,289]. Additionally, mosaicism, where different cells within an individual's body carry different genetic variants, can result in varied expressivity of the trait[278,279]. Moreover, polygenic factors involving multiple genes can collectively contribute to the overall expression of the trait,

further adding to the complexity of the observed variations in penetrance and expressivity[290]. To illustrate, in the context of a specific genetic disorder, some individuals may exhibit more severe symptoms due to the presence of other modifying genes or cellular mosaicism, while others with the same variant may display milder symptoms owing to the interaction of various genetic factors[291]. All these factors contribute to the complexity of the pathogenicity and penetrance issue[279,292].

While studying the penetrance and expressivity of genetic variants is crucial, researchers encounter various challenges. Access to comprehensive and diverse population cohort data is essential for obtaining accurate estimations of penetrance and expressivity. The complex nature of genetic interactions and the influence of environmental factors pose difficulties in establishing definitive causal relationships. Additionally, to differentiate between true causal effects and chance associations when interpreting data from clinical studies and population cohorts, various approaches and expertise are employed. These include assessing statistical significance through P values, conducting replication studies to validate findings across multiple independent datasets, and incorporating experimental evidence to establish causality and biological mechanisms. Such rigorous methodologies are essential in ensuring the reliability and accuracy of the insights gained from these studies.

The Role of Cohorts in Unravelling Genetic Variant Effects on Penetrance and Pathogenicity

When studying the penetrance and pathogenicity of genetic variants associated with female reproductive aging, incorporating diverse and well-defined cohorts is essential for gaining a deeper understanding of these factors and their complexities. By analysing various cohorts, we can assess how genetic variants impact reproductive traits across different populations, considering the diversity of phenotypic expressions and the influence of various ethnic backgrounds[293]. Leveraging electronic health records (EHRs) and carefully considering specific traits related to female reproductive aging allows us to obtain valuable insights into the penetrance and pathogenicity of these genetic variants, providing a comprehensive view of their effects on reproductive lifespan.

Phenotypes play a crucial role in studying penetrance and pathogenicity[294,295]. These phenotypes encompass a diverse set of observable characteristics

resulting from the interplay between an individual's genetic composition and their environment. In the context of female reproductive aging, genetic variants can manifest in a variety of phenotypic outcomes. For instance, they may lead to variations in age at natural menopause, irregular menstrual cycles, hormonal imbalances, fertility issues, and the onset of menopausal symptoms such as hot flashes[296,297]. Additionally, some genetic variants might be associated with increased risk for specific reproductive disorders, such as POI or polycystic ovary syndrome, while others may have no discernible impact on reproductive traits. Hence, it is imperative to investigate cohorts with a wide range of reproductive phenotypes to encompass the entire spectrum of effects linked to particular genetic variants[298]. By including diverse phenotypes in the analysis, we can obtain a more accurate assessment of penetrance and pathogenicity.

Electronic health records (EHRs) provide a valuable resource for studying penetrance and pathogenicity[299]. EHRs contain comprehensive patient data, including medical history, diagnoses, treatments, and outcomes[300]. Analysing large cohorts such as global Biobanks through EHRs enables the identification of associations between specific genetic variants and the development of particular diseases or conditions. However, EHRs pose several challenges for reproductive studies. These challenges include potential missing or inaccurate data impacting reliability, difficulties in integrating data from different EHR systems due to varying formats and coding, strict privacy regulations due to sensitive health information, sample selection bias affecting generalisability, limited reproductive-specific data reducing study scope, historical or follow-up data limitations impacting assessment of reproductive outcomes over time, diverse data collection methods hindering data combination and analysis, balancing data accessibility with patient confidentiality, the need for sophisticated methods to extract insights from complex EHR data, and consideration of existing health disparities and biases in healthcare that may influence research outcomes and interpretation[301-304].

Ethnicity is a critical factor to consider in the study of penetrance and pathogenicity. Different ethnic groups exhibit variations in genetic background, environmental exposures, and cultural practices, which can influence the penetrance and pathogenicity of genetic variants[305-307]. By analysing cohorts from diverse ethnicities, we can gain valuable insights into the population-

specific effects of genetic variants and understand the intricate interplay between different genes and variants. Fine mapping and exploring the genetic architecture provide a more detailed understanding of how specific genes or variants contribute to complex traits across various ethnic groups, shedding light on the complex mechanisms underlying these associations. This knowledge improves risk assessment and genetic counselling for individuals within these populations, promoting personalised healthcare[308].

Therefore, the investigation of penetrance and pathogenicity of genetic variants requires the utilization of diverse cohorts. Incorporating cohorts with different phenotypes, leveraging EHRs, and considering the influence of different ethnicities contribute to a comprehensive understanding of genetic variant effects. Population-based cohorts, disease-specific cohorts, and biobanks provide valuable resources for studying penetrance and pathogenicity across a range of conditions and populations. By employing these approaches, researchers can enhance their knowledge of genetic variant effects and improve clinical decision-making and genetic counselling for individuals affected by these variants.

Navigating the Challenges of Low Penetrance in Premature Ovarian Insufficiency: Considerations for Genetic Testing and New-born Screening

The implications of low penetrance in POI and the advent of genetic testing, including direct-to-consumer (DTC) testing through whole-genome sequencing (WGS), are crucial areas of study. Understanding the impact of low penetrance on genetic testing can have significant implications for disease diagnosis, risk assessment, and genetic counselling. Additionally, pilot efforts for new-born screening in the UK offer a potential avenue for early detection and intervention. This discussion explores these implications and their significance in the context of POI and genetic testing[309].

DTC testing, which provides individuals with access to their genetic information without the involvement of healthcare professionals, has gained popularity in recent years[310]. WGS, a comprehensive genetic testing method, offers individuals the opportunity to explore their genetic predispositions and potential disease risks. However, when it comes to conditions like POI, the implications of low penetrance become evident. In the case of POI, low penetrance can lead

to challenges in accurately predicting an individual's risk based solely on genetic testing. While DTC testing can provide valuable insights into an individual's genetic makeup, the interpretation of results must consider the impact of low penetrance and the limitations it poses for predictive purposes.

New-born screening programs are instrumental in identifying specific conditions in infants early on, leading to timely interventions and improved outcomes. Currently, there have been no specific pilot efforts in the UK targeting low penetrance genetic variants associated with certain conditions in new-born screening, particularly for conditions like POI, which may have a more complex genetic basis. However, it is crucial to approach new-born screening for POI with caution. While comprehensive screening for POI may not be applicable due to its multifactorial nature, targeted screening for particular allele risks could be considered. Factors such as feasibility, cost-effectiveness, and the availability of accurate screening tests must be thoroughly evaluated before implementing such screening programs. Furthermore, establishing robust protocols for counselling and support is essential to provide families with accurate information about screening results and their potential implications. Addressing ethical aspects, including informed consent and privacy, is also of utmost importance in the implementation of any screening program.

From GWAS to exWAS: Advancing our Understanding of Genetic Influences on Menopause Timing

In a previous study by Ruth et al. (2021), a GWAS involving a large cohort of over 200,000 women revealed the presence of 290 common genetic variants associated with ANM[38]. These variants collectively accounted for 10% to 12% of the variation in ANM and exhibited effects ranging from approximately 3.5 weeks to 1.3 years. Building upon the findings of the GWAS, this study delved deeper into the genetic architecture of ANM by investigating the role of both common and rare variants. Through the analysis UKB WES and exome-wide gene burden association tests, we identified additional genetic factors associated with ANM and gained insights into novel biological pathways.

In our study in chapter 3, we conducted an exome-wide association study (exWAS) and identified several genes significantly associated with ANM. These genes include *CHEK2*, *HELB*, *HROB*, *BRCA2*, *ETAA1*, *ZNF518A*, *PNPLA8*, *PALB2*, and *SAMHD1*. Each of these genes contributes to the regulation of

ovarian aging and exerts distinct influences on menopause timing. The effect sizes of these gene associations varied across the identified genes. For instance, carriers of high-confidence protein truncating variants (HC-PTVs) in *ZNF518A* exhibited an effect size of 5.61 years earlier ANM, while carriers of damaging alleles in *SAMHD1* experienced an effect size of 1.35 years later ANM. Notably, these effect sizes surpassed the effect sizes of common variants previously identified in GWAS studies of ANM.

Our study explored rare genomic variants and their association with ovarian aging, revealing novel aspects not investigated in previous GWAS studies. We confirmed the roles of known DNA repair genes like *CHEK2* and *BRCA2* in ANM and identified new associations with genes such as *ETAA1*, *PNPLA8*, and *HROB*, which are involved in DNA damage response and repair processes. These findings deepen our understanding of the mechanisms influencing ovarian reserve and menopause timing. Moreover, our study revealed the enrichment of *ZNF518A*, a poorly characterized C2H2 zinc finger transcription factor, in binding sites of common variants associated with ANM. This novel finding suggests that transcription factors like *ZNF518A* may be essential in regulating menopause timing and may open avenues for investigating other transcription factors' roles in menopause and related traits.

By combining the findings from GWAS and exWAS, we underscore the complexity of genetic influences on menopause timing. The GWAS identified common genetic variants associated with ANM, providing a broad understanding of the genetic basis. In contrast, the exWAS expanded on this knowledge by uncovering rare genetic variants, identifying additional genes associated with ANM, and revealing novel biological pathways involved in ovarian aging. Indeed, the exclusivity of our analyses to UKB and European cohorts presents a limitation to the generalisability of our findings. To improve the validity and reliability of our results, it is crucial to expand our investigations to include cohorts with diverse ethnic backgrounds, access data from other Biobanks, and incorporate population data from various sources. This broader approach will allow for a more comprehensive understanding of the genetic determinants of reproductive aging and enhance the applicability of our findings to different populations.

Exploring the Role of *SAMHD1* in Cancer Susceptibility: Potential Mechanisms and Implications

In chapter 3, we investigated the impact of ANM-associated genes on cancer outcomes, with a specific focus on *SAMHD1*. While confirming known associations between genes like *BRCA2*, *CHEK2*, and *PALB2* [38,311-314], our novel finding of an association between *SAMHD1* damaging variants and "All cancer" in both males and females highlights the role of *SAMHD1* in cancer susceptibility, particularly for specific site-specific cancers such as prostate cancer in males, mesothelioma in both sexes, and a potential link to increased breast cancer susceptibility in females.

The replicated associations between cancer risk-increasing alleles in *SAMHD1* and later ANM suggest a potential mechanism involving disrupted DNA damage sensing and apoptosis, leading to a slowed depletion of the ovarian reserve, thereby influencing menopause timing and contributing to cancer risk. Additional experimental work is required to fully comprehend this association, considering other mechanisms of ovarian reserve depletion. Furthermore, our study provides robust evidence of rare damaging variants in *SAMHD1* affecting telomere length, a factor associated with increased cancer risk[315-317]. This finding underscores the potential impact of rare variants in *SAMHD1* on telomere biology and cancer susceptibility.

The implications of these findings are significant, offering insights into the genetic basis linking reproductive timing and cancer risk[38,41,296]. Understanding the specific genes associated with both processes' sheds light on the underlying mechanisms driving these associations. This knowledge can inform future research and clinical strategies for cancer prevention, risk assessment, and treatment.

Looking ahead, deeper investigations are warranted to explore the biology of these genes and the broader themes of DNA damage response, reproductive aging, and cancer. It would be valuable to explore the biology of genetic variations that increase cancer risk and consider whether similar mechanisms could impact ovarian reserve. Additionally, examining estradiol exposure versus risk from disrupted DNA damage response could provide further insights into the interplay of reproductive aging and cancer susceptibility.

To generalise these findings to diverse populations, larger cohort studies and multi-ethnic analyses are essential. Such efforts will contribute to a comprehensive understanding of the complex relationship between reproductive timing and cancer, leading to improved strategies for cancer prevention, early detection, and personalised management of cancer risk. Notably, focusing on the link between cancer and late menopause will guide future research in this field.

Unravelling the Genetic Complexity of Reproductive Lifespan through Whole-Genome Sequencing

To further advance gene discovery in reproductive aging, we must explore several key areas, including addressing the missing heritability, investigating gene interactions, exploring non-coding rare variants, and delving into potential epigenetic influences[318,319]. Currently known genetic variants explain only a fraction of the observed heritability, making it crucial to uncover additional genetic factors and comprehend their contributions to reproductive aging. These endeavours are essential for advancing our understanding of the underlying mechanisms involved in reproductive aging processes.

Future research should prioritize investigating rare non-coding variants, as they have received relatively less attention compared to coding and noncoding common variants. Despite the significant influence of coding causal variants on menopause (chapter 4), recent studies have demonstrated the important role of non-coding variants in various traits, such as developmental disorders and congenital hyperinsulinism[320,321]. Thus, non-coding variants may also play a crucial role in reproductive aging. Exploring these non-coding variants can lead to a more comprehensive understanding of the genetic architecture affecting menopause timing and related reproductive traits. By identifying and studying these variants, researchers can gain valuable insights into the underlying mechanisms driving reproductive aging processes.

The utilization of WGS represents a pivotal advancement in gene discovery for reproductive aging. WGS offers a comprehensive and detailed view of the entire genome, allowing for the identification of rare variants, non-coding variants, and other genomic elements that may be instrumental in determining menopause timing and related reproductive traits. Several WGS efforts in other areas have significantly contributed to our understanding of various conditions. For

instance, WGS studies have elucidated the genetic basis of rare diseases, cancer subtypes, and complex traits, providing valuable insights into disease mechanisms and potential therapeutic targets[322-325]. The availability of advanced WGS technologies and data analysis tools further facilitates the incorporation of WGS into future reproductive aging studies, offering unparalleled opportunities to gain deeper insights into the genetic determinants of reproductive aging. By harnessing the power of WGS, we can uncover novel genetic factors and potential pathways involved in reproductive aging, ultimately leading to improved personalised healthcare and targeted interventions.

By focusing on these next steps, we can uncover additional genetic factors, elucidate the missing heritability, explore rare non-coding variants, and leverage advanced sequencing technologies to further unravel the genetic basis of reproductive aging. This knowledge will contribute to a deeper understanding of menopause timing and its implications for women's health and fertility.

Conclusion

Overall, these findings significantly contribute to our understanding of the genetic determinants of menopause timing and have important implications for reproductive health. The identified genes and pathways present potential targets for therapeutic interventions aimed at enhancing ovarian stimulation in fertility treatments. Additionally, they may offer insights into the relationship between menopause timing and other health outcomes, such as cancer susceptibility. Further research investigating these genetic factors and pathways will deepen our understanding of ovarian aging and its broader implications.

While the positive aspects of this study have been underscored, it is essential to recognise the limitations associated with the use of biobanks, such as the UKBB, in shaping these findings. The inherent biases within biobank data, including selection biases in participant demographics, must be acknowledged to ensure the generalizability of our conclusions. Additionally, ethical considerations related to consent, privacy, and data use in biobanks warrant careful consideration. By addressing these aspects, future research endeavors can refine methodologies, foster transparency, and strengthen the robustness

of conclusions drawn from large-scale population studies like those conducted with biobank data.

In moving forward, further investigations into the identified genetic factors and pathways are imperative. Not only will this deepen our understanding of ovarian aging, but it will also unveil the intricate connections between genetic determinants of menopause timing and broader health implications. Ongoing research endeavors should leverage these insights to drive advancements in reproductive medicine, contributing to more effective fertility treatments and holistic approaches to women's health as they age.

References

1 Variations in reproductive events across life: a pooled analysis of data from 505 147 women across 10 countries. *Hum Reprod* **34**, 881-893 (2019). https://doi.org:10.1093/humrep/dez015

2 Murray, A., Ennis, S., MacSwiney, F., Webb, J. & Morton, N. E. Reproductive and menstrual history of females with fragile X expansions. *Eur J Hum Genet* **8**, 247-252 (2000). https://doi.org:10.1038/sj.ejhg.5200451

3 Broekmans, F. J., Soules, M. R. & Fauser, B. C. Ovarian aging: mechanisms and clinical consequences. *Endocr Rev* **30**, 465-493 (2009). https://doi.org:10.1210/er.2009-0006

4 Wallace, W. H. & Kelsey, T. W. Human ovarian reserve from conception to the menopause. *PLoS One* **5**, e8772 (2010). https://doi.org:10.1371/journal.pone.0008772

5 Mandl, A. M. & Zuckerman, S. The relation of age to numbers of oocytes. *J Endocrinol* **7**, 190-193 (1951). https://doi.org:10.1677/joe.0.0070190

6 Allen, J. (Oxford, NY: Oxford University Press, 2010).

7 Gruhn, J. R. *et al.* Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science* **365**, 1466-1469 (2019). https://doi.org:10.1126/science.aav7321

8 Kerr, J. B. *et al.* Cisplatin-induced primordial follicle oocyte killing and loss of fertility are not prevented by imatinib. *Nat Med* **18**, 1170-1172; author reply 1172-1174 (2012). https://doi.org:10.1038/nm.2889

9 Reed, B. G. & Carr, B. R. in *Endotext* (eds K. R. Feingold *et al.*) (MDText.com, Inc.

Copyright © 2000-2023, MDText.com, Inc., 2000).

10 Hirshfield, A. N. Development of follicles in the mammalian ovary. *Int Rev Cytol* **124**, 43-101 (1991). https://doi.org:10.1016/s0074-7696(08)61524-7

11 Faddy, M. J., Gosden, R. G., Gougeon, A., Richardson, S. J. & Nelson, J. F. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod* **7**, 1342-1346 (1992). https://doi.org:10.1093/oxfordjournals.humrep.a137570

12 Hansen, K. R. *et al.* A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod* **23**, 699-708 (2008). https://doi.org:10.1093/humrep/dem408

13 Perheentupa, A. & Huhtaniemi, I. Aging of the human ovary and testis. *Mol Cell Endocrinol* **299**, 2-13 (2009). https://doi.org:10.1016/j.mce.2008.11.004

14 Organization, W. H. (2023).

15 Sobotka, T., Skirbekk, V. & Philipov, D. Economic recession and fertility in the developed world. *Population and development review* **37**, 267-306 (2011).

16 Bunting, L. & Boivin, J. Decision-making about seeking medical advice in an internet sample of women trying to get pregnant. *Hum Reprod* **22**, 1662-1668 (2007). https://doi.org:10.1093/humrep/dem057

17 Schmidt, L., Sobotka, T., Bentzen, J. G. & Nyboe Andersen, A. Demographic and medical consequences of the postponement of parenthood. *Hum Reprod Update* **18**, 29-43 (2012). https://doi.org:10.1093/humupd/dmr040

18 Lundberg, S. & Rose, E. Child gender and the transition to marriage. *Demography* **40**, 333-349 (2003). https://doi.org:10.1353/dem.2003.0015

19 Wu, L. & Wolfe, B. *Out of wedlock: Causes and consequences of nonmarital fertility*. (Russell Sage Foundation, 2001).

20 Booth, A. & Crouter, A. C. *The new population problem: Why families in developed countries are shrinking and what it means*. (Psychology Press, 2005).

21 Kushnir, V. A., Barad, D. H., Albertini, D. F., Darmon, S. K. & Gleicher, N. Systematic review of worldwide trends in assisted reproductive technology 2004-2013. *Reprod Biol Endocrinol* **15**, 6 (2017). https://doi.org:10.1186/s12958-016-0225-2

22 Nargund, G. Declining birth rate in Developed Countries: A radical policy re-think is required. *Facts Views Vis Obgyn* **1**, 191-193 (2009).
23 Ombelet, W. Global access to infertility care in developing countries: a case of human rights, equity and social justice. *Facts Views Vis Obgyn* **3**, 257-266 (2011).
24 Miller, W. B., Severy, L. J. & Pasta, D. J. A framework for modelling fertility motivation in couples. *Popul Stud (Camb)* **58**, 193-205 (2004). https://doi.org:10.1080/0032472042000213712
25 Song, M. *et al.* Differences in Family Planning and Fertility Among Female and Male Gynecologic Oncologists. *Womens Health Rep (New Rochelle)* **2**, 78-84 (2021). https://doi.org:10.1089/whr.2020.0046
26 Singh, V., Rajpal, G., Majumdar, S., Mandal, S. & Sahu, R. Awareness, knowledge, and psycho-emotional aspect of infertility among women seeking fertility assistance at a tertiary care hospital in Chhattisgarh, a cross-sectional study. *J Educ Health Promot* **12**, 3 (2023). https://doi.org:10.4103/jehp.jehp_581_22
27 Mishra, G. D. *et al.* Early menarche, nulliparity and the risk for premature and early natural menopause. *Hum Reprod* **32**, 679-686 (2017). https://doi.org:10.1093/humrep/dew350
28 Mishra, G. D. *et al.* EMAS position statement: Predictors of premature and early natural menopause. *Maturitas* **123**, 82-88 (2019). https://doi.org:10.1016/j.maturitas.2019.03.008
29 Kuh, D. L., Wadsworth, M. & Hardy, R. Women's health in midlife: the influence of the menopause, social factors and health in earlier life. *Br J Obstet Gynaecol* **104**, 923-933 (1997). https://doi.org:10.1111/j.1471-0528.1997.tb14352.x
30 Zhu, D. *et al.* Body mass index and age at natural menopause: an international pooled analysis of 11 prospective studies. *Eur J Epidemiol* **33**, 699-710 (2018). https://doi.org:10.1007/s10654-018-0367-y
31 Zhu, D. *et al.* Relationships between intensity, duration, cumulative dose, and timing of smoking with age at menopause: A pooled analysis of individual data from 17 observational studies. *PLoS Med* **15**, e1002704 (2018). https://doi.org:10.1371/journal.pmed.1002704
32 Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci Rep* **6**, 24710 (2016). https://doi.org:10.1038/srep24710
33 Christensen, K., Carlson, L. M. & Lehmann, G. M. The role of epidemiology studies in human health risk assessment of polychlorinated biphenyls. *Environ Res* **194**, 110662 (2021). https://doi.org:10.1016/j.envres.2020.110662
34 Perry, J. R. *et al.* DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. *Hum Mol Genet* **23**, 2490-2497 (2014). https://doi.org:10.1093/hmg/ddt620
35 Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* **44**, 260-268 (2012). https://doi.org:10.1038/ng.1051
36 He, C. *et al.* A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum Genet* **128**, 515-527 (2010). https://doi.org:10.1007/s00439-010-0878-4
37 Murabito, J. M., Yang, Q., Fox, C., Wilson, P. W. & Cupples, L. A. Heritability of age at natural menopause in the Framingham Heart Study. *J Clin Endocrinol Metab* **90**, 3427-3430 (2005). https://doi.org:10.1210/jc.2005-0181
38 Ruth, K. S. *et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393-397 (2021). https://doi.org:10.1038/s41586-021-03779-7
39 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. *J Clin Endocrinol Metab* **83**, 1875-1880 (1998). https://doi.org:10.1210/jcem.83.6.4890

40 Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97 (2014). https://doi.org:10.1038/nature13545
41 Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* **47**, 1294-1303 (2015). https://doi.org:10.1038/ng.3412
42 Stochholm, K., Juul, S., Juel, K., Naeraa, R. W. & Gravholt, C. H. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. *J Clin Endocrinol Metab* **91**, 3897-3902 (2006). https://doi.org:10.1210/jc.2006-0558
43 Jacobs, P. A. *et al.* Evidence for the existence of the human "super female". *Lancet* **2**, 423-425 (1959). https://doi.org:10.1016/s0140-6736(59)90415-5
44 Otter, M., Schrander-Stumpel, C. T. & Curfs, L. M. Triple X syndrome: a review of the literature. *Eur J Hum Genet* **18**, 265-271 (2010). https://doi.org:10.1038/ejhg.2009.109
45 Tassone, F. *et al.* Clinical involvement and protein expression in individuals with the FMR1 premutation. *Am J Med Genet* **91**, 144-152 (2000). https://doi.org:10.1002/(sici)1096-8628(20000313)91:2<144::aid-ajmg14>3.0.co;2-v
46 Pembrey, M., Winter, R., Davies, K., Opitz, J. M. & Reynolds, J. F. A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. *American journal of medical genetics* **21**, 709-717 (1985).
47 Ennis, S., Ward, D. & Murray, A. Nonlinear association between CGG repeat number and age of menopause in FMR1 premutation carriers. *Eur J Hum Genet* **14**, 253-255 (2006). https://doi.org:10.1038/sj.ejhg.5201510
48 Verkerk, A. J. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905-914 (1991).
49 Hagerman, P. J. & Hagerman, R. J. The Fragile-X Premutation: A Maturing Perspective. *The American Journal of Human Genetics* **74**, 805-816 (2004). https://doi.org:https://doi.org/10.1086/386296
50 Jacquemont, S. *et al.* Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet* **72**, 869-878 (2003). https://doi.org:10.1086/374321
51 Sherman, S., Pletcher, B. A. & Driscoll, D. A. Fragile X syndrome: diagnostic and carrier testing. *Genet Med* **7**, 584-587 (2005). https://doi.org:10.1097/01.gim.0000182468.22666.dd
52 Wesevich, V., Kellen, A. N. & Pal, L. Recent advances in understanding primary ovarian insufficiency. *F1000Res* **9** (2020). https://doi.org:10.12688/f1000research.26423.1
53 Golezar, S., Ramezani Tehrani, F., Khazaei, S., Ebadi, A. & Keshavarz, Z. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric* **22**, 403-411 (2019). https://doi.org:10.1080/13697137.2019.1574738
54 Rudnicka, E. *et al.* Premature ovarian insufficiency - aetiopathology, epidemiology, and diagnostic evaluation. *Prz Menopauzalny* **17**, 105-108 (2018). https://doi.org:10.5114/pm.2018.78550
55 Coulam, C. B., Adamson, S. C. & Annegers, J. F. Incidence of premature ovarian failure. *Obstet Gynecol* **67**, 604-606 (1986).
56 Harlow, B. L. & Signorello, L. B. Factors associated with early menopause. *Maturitas* **35**, 3-9 (2000). https://doi.org:10.1016/s0378-5122(00)00092-x
57 Goswami, D. & Conway, G. S. Premature ovarian failure. *Hum Reprod Update* **11**, 391-410 (2005). https://doi.org:10.1093/humupd/dmi012
58 Szeliga, A. *et al.* Autoimmune Diseases in Patients with Premature Ovarian Insufficiency-Our Current State of Knowledge. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms22052594
59 Fortuño, C. & Labarta, E. Genetics of primary ovarian insufficiency: a review. *J Assist Reprod Genet* **31**, 1573-1585 (2014). https://doi.org:10.1007/s10815-014-0342-9

60 Venturella, R. *et al.* The Genetics of Non-Syndromic Primary Ovarian Insufficiency: A Systematic Review. *Int J Fertil Steril* **13**, 161-168 (2019). https://doi.org:10.22074/ijfs.2019.5599
61 van Kasteren, Y. M. *et al.* Familial idiopathic premature ovarian failure: an overrated and underestimated genetic disease? *Hum Reprod* **14**, 2455-2459 (1999). https://doi.org:10.1093/humrep/14.10.2455
62 Qin, Y. *et al.* ESR1, HK3 and BRSK1 gene variants are associated with both age at natural menopause and premature ovarian failure. *Orphanet J Rare Dis* **7**, 5 (2012). https://doi.org:10.1186/1750-1172-7-5
63 Zhou, Y. *et al.* Novel alanyl-tRNA synthetase 2 (AARS2) homozygous mutation in a consanguineous Chinese family with premature ovarian insufficiency. *Fertil Steril* **112**, 569-576.e562 (2019). https://doi.org:10.1016/j.fertnstert.2019.05.005
64 Dallabona, C. *et al.* Novel (ovario) leukodystrophy related to AARS2 mutations. *Neurology* **82**, 2063-2071 (2014). https://doi.org:10.1212/wnl.0000000000000497
65 Zheng, W. B. *et al.* A novel variant in AIRE causing a rare, non-classical autoimmune polyendocrine syndrome type 1. *Mol Med Rep* **22**, 1285-1294 (2020). https://doi.org:10.3892/mmr.2020.11227
66 Alavi, A. *et al.* Mutation in ALOX12B likely cause of POI and also ichthyosis in a large Iranian pedigree. *Mol Genet Genomics* **295**, 1039-1053 (2020). https://doi.org:10.1007/s00438-020-01663-z
67 Alvaro Mercadal, B. *et al.* AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Hum Reprod* **30**, 1196-1202 (2015). https://doi.org:10.1093/humrep/dev042
68 Wang, Y. *et al.* Pathogenic variants of meiotic double strand break (DSB) formation genes PRDM9 and ANKRD31 in premature ovarian insufficiency. *Genet Med* **23**, 2309-2315 (2021). https://doi.org:10.1038/s41436-021-01266-y
69 Chatterjee, S. *et al.* Longer CAG repeat length in the androgen receptor gene is associated with premature ovarian failure. *Hum Reprod* **24**, 3230-3235 (2009). https://doi.org:10.1093/humrep/dep296
70 Patiño, L. C. *et al.* New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Hum Reprod* **32**, 1512-1520 (2017). https://doi.org:10.1093/humrep/dex089
71 Delcour, C. *et al.* ATG7 and ATG9A loss-of-function variants trigger autophagy impairment and ovarian failure. *Genet Med* **21**, 930-938 (2019). https://doi.org:10.1038/s41436-018-0287-y
72 Liu, H. *et al.* Whole-exome sequencing in patients with premature ovarian insufficiency: early detection and early intervention. *J Ovarian Res* **13**, 114 (2020). https://doi.org:10.1186/s13048-020-00716-6
73 Ding, X. *et al.* DNA double-strand break genetic variants in patients with premature ovarian insufficiency. *Journal of Ovarian Research* **16**, 135 (2023). https://doi.org:10.1186/s13048-023-01221-2
74 Arora, H. *et al.* Bloom syndrome. *Int J Dermatol* **53**, 798-802 (2014). https://doi.org:10.1111/ijd.12408
75 Di Pasquale, E., Beck-Peccoz, P. & Persani, L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. *Am J Hum Genet* **75**, 106-111 (2004). https://doi.org:10.1086/422103
76 Dixit, H. *et al.* Missense mutations in the BMP15 gene are associated with ovarian failure. *Hum Genet* **119**, 408-415 (2006). https://doi.org:10.1007/s00439-006-0150-0
77 Gónzalez, A. *et al.* Association of genetic markers within the BMP15 gene with anovulation and infertility in women with polycystic ovary syndrome. *Fertil Steril* **90**, 447-449 (2008). https://doi.org:10.1016/j.fertnstert.2007.06.083

78 Fonseca, D. J. *et al.* BMP15 c.-9C>G promoter sequence variant may contribute to the cause of non-syndromic premature ovarian failure. *Reprod Biomed Online* **29**, 627-633 (2014). https://doi.org:10.1016/j.rbmo.2014.07.018
79 Rossetti, R. *et al.* BMP15 mutations associated with primary ovarian insufficiency cause a defective production of bioactive protein. *Hum Mutat* **30**, 804-810 (2009). https://doi.org:10.1002/humu.20961
80 Bouilly, J. *et al.* Identification of Multiple Gene Mutations Accounts for a new Genetic Architecture of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **101**, 4541-4550 (2016). https://doi.org:10.1210/jc.2016-2152
81 Di Pasquale, E. *et al.* Identification of new variants of human BMP15 gene in a large cohort of women with premature ovarian failure. *J Clin Endocrinol Metab* **91**, 1976-1979 (2006). https://doi.org:10.1210/jc.2005-2650
82 Tiotiu, D. *et al.* Variants of the BMP15 gene in a cohort of patients with premature ovarian failure. *Hum Reprod* **25**, 1581-1587 (2010). https://doi.org:10.1093/humrep/deq073
83 Laissue, P. *et al.* Mutations and sequence variants in GDF9 and BMP15 in patients with premature ovarian failure. *Eur J Endocrinol* **154**, 739-744 (2006). https://doi.org:10.1530/eje.1.02135
84 Patiño, L. C. *et al.* BMP15 Mutations Associated With Primary Ovarian Insufficiency Reduce Expression, Activity, or Synergy With GDF9. *J Clin Endocrinol Metab* **102**, 1009-1019 (2017). https://doi.org:10.1210/jc.2016-3503
85 Wang, B. *et al.* Analyses of growth differentiation factor 9 (GDF9) and bone morphogenetic protein 15 (BMP15) mutation in Chinese women with premature ovarian failure. *Clin Endocrinol (Oxf)* **72**, 135-136 (2010). https://doi.org:10.1111/j.1365-2265.2009.03613.x
86 Santos, M. *et al.* Association of BMP15 and GDF9 variants to premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 2163-2169 (2019). https://doi.org:10.1007/s10815-019-01548-0
87 Zhang, W. *et al.* A novel homozygous mutation of bone morphogenetic protein 15 identified in a consanguineous marriage family with primary ovarian insufficiency. *Reprod Biomed Online* **36**, 206-209 (2018). https://doi.org:10.1016/j.rbmo.2017.10.104
88 Mayer, A., Fouquet, B., Pugeat, M. & Misrahi, M. BMP15 "knockout-like" effect in familial premature ovarian insufficiency with persistent ovarian reserve. *Clin Genet* **92**, 208-212 (2017). https://doi.org:10.1111/cge.12970
89 Rossetti, R. *et al.* Fundamental role of BMP15 in human ovarian folliculogenesis revealed by null and missense mutations associated with primary ovarian insufficiency. *Hum Mutat* **41**, 983-997 (2020). https://doi.org:10.1002/humu.23988
90 Renault, L. *et al.* BMPR1A and BMPR1B Missense Mutations Cause Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgz226
91 Zhang, D. *et al.* Basonuclin 1 deficiency is a cause of primary ovarian insufficiency. *Hum Mol Genet* **27**, 3787-3800 (2018). https://doi.org:10.1093/hmg/ddy261
92 Turchetti, D., Zuntini, R., Tricarico, R. & Bellacosa, A. BRCA2 in Ovarian Development and Function. *N Engl J Med* **380**, 1086-1087 (2019). https://doi.org:10.1056/NEJMc1813800
93 Chen, Q. *et al.* Rare deleterious BUB1B variants induce premature ovarian insufficiency and early menopause. *Hum Mol Genet* **29**, 2698-2707 (2020). https://doi.org:10.1093/hmg/ddaa153
94 Fan, S. *et al.* Homozygous mutations in C14orf39/SIX6OS1 cause non-obstructive azoospermia and premature ovarian insufficiency in humans. *Am J Hum Genet* **108**, 324-336 (2021). https://doi.org:10.1016/j.ajhg.2021.01.010

95 Hou, D. *et al.* Variations of C14ORF39 and SYCE1 Identified in Idiopathic Premature Ovarian Insufficiency and Nonobstructive Azoospermia. *J Clin Endocrinol Metab* **107**, 724-734 (2022). https://doi.org:10.1210/clinem/dgab777
96 Chasseloup, F. *et al.* Germline CDKN1B Loss-of-Function Variants Cause Pediatric Cushing's Disease With or Without an MEN4 Phenotype. *J Clin Endocrinol Metab* **105**, 1983-2005 (2020). https://doi.org:10.1210/clinem/dgaa160
97 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 (2015). https://doi.org:10.1016/j.jns.2015.04.038
98 Dursun, F. *et al.* A Novel Missense Mutation in the CLPP Gene Causing Perrault Syndrome Type 3 in a Turkish Family. *J Clin Res Pediatr Endocrinol* **8**, 472-477 (2016). https://doi.org:10.4274/jcrpe.2717
99 Al-Agha, A. E. *et al.* Primary Ovarian Insufficiency and Azoospermia in Carriers of a Homozygous PSMC3IP Stop Gain Mutation. *J Clin Endocrinol Metab* **103**, 555-563 (2018). https://doi.org:10.1210/jc.2017-01966
100 Tšuiko, O. *et al.* Copy number variation analysis detects novel candidate genes involved in follicular growth and oocyte maturation in a cohort of premature ovarian failure cases. *Hum Reprod* **31**, 1913-1925 (2016). https://doi.org:10.1093/humrep/dew142
101 Jiao, W., Zhao, S., Liu, R., Guo, T. & Qin, Y. CPEB1 deletion is not a common explanation for premature ovarian insufficiency in a Chinese cohort. *J Ovarian Res* **13**, 49 (2020). https://doi.org:10.1186/s13048-020-00630-x
102 Rouen, A. *et al.* Whole exome sequencing in a cohort of familial premature ovarian insufficiency cases reveals a broad array of pathogenic or likely pathogenic variants in 50% of families. *Fertil Steril* **117**, 843-853 (2022). https://doi.org:10.1016/j.fertnstert.2021.12.023
103 Carvalho, L. C. *et al.* Clinical, hormonal, ovarian, and genetic aspects of 46,XX patients with congenital adrenal hyperplasia due to CYP17A1 defects. *Fertil Steril* **105**, 1612-1619 (2016). https://doi.org:10.1016/j.fertnstert.2016.02.008
104 Bione, S. *et al.* Mutation analysis of two candidate genes for premature ovarian failure, DACH2 and POF1B. *Hum Reprod* **19**, 2759-2766 (2004). https://doi.org:10.1093/humrep/deh502
105 Genesio, R. *et al.* Short stature and primary ovarian insufficiency possibly due to chromosomal position effect in a balanced X;1 translocation. *Mol Cytogenet* **8**, 50 (2015). https://doi.org:10.1186/s13039-015-0154-3
106 Rizzolio, F. *et al.* Epigenetic control of the critical region for premature ovarian failure on autosomal genes translocated to the X chromosome: a hypothesis. *Hum Genet* **121**, 441-450 (2007). https://doi.org:10.1007/s00439-007-0329-z
107 Mandon-Pépin, B. *et al.* Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur J Endocrinol* **158**, 107-115 (2008). https://doi.org:10.1530/eje-07-0400
108 He, W. B. *et al.* DMC1 mutation that causes human non-obstructive azoospermia and premature ovarian insufficiency identified by whole-exome sequencing. *J Med Genet* **55**, 198-204 (2018). https://doi.org:10.1136/jmedgenet-2017-104992
109 Fogli, A. *et al.* Ovarian failure related to eukaryotic initiation factor 2B mutations. *Am J Hum Genet* **72**, 1544-1550 (2003). https://doi.org:10.1086/375404
110 Zhao, M., Feng, F., Chu, C., Yue, W. & Li, L. A novel EIF4ENIF1 mutation associated with a diminished ovarian reserve and premature ovarian insufficiency identified by whole-exome sequencing. *J Ovarian Res* **12**, 119 (2019). https://doi.org:10.1186/s13048-019-0595-0

111 Chatzispyrou, I. A. *et al.* A homozygous missense mutation in ERAL1, encoding a mitochondrial rRNA chaperone, causes Perrault syndrome. *Hum Mol Genet* **26**, 2541-2550 (2017). https://doi.org:10.1093/hmg/ddx152
112 Qin, Y. *et al.* CSB-PGBD3 Mutations Cause Premature Ovarian Failure. *PLoS Genet* **11**, e1005419 (2015). https://doi.org:10.1371/journal.pgen.1005419
113 Yang, X. *et al.* Rare variants in FANCA induce premature ovarian insufficiency. *Hum Genet* **138**, 1227-1236 (2019). https://doi.org:10.1007/s00439-019-02059-9
114 França, M. M. & Mendonca, B. B. Genetics of ovarian insufficiency and defects of folliculogenesis. *Best Pract Res Clin Endocrinol Metab* **36**, 101594 (2022). https://doi.org:10.1016/j.beem.2021.101594
115 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* **37**, 609-635 (2016). https://doi.org:10.1210/er.2016-1047
116 Yang, Y. *et al.* FANCL gene mutations in premature ovarian insufficiency. *Hum Mutat* **41**, 1033-1041 (2020). https://doi.org:10.1002/humu.23997
117 Fouquet, B. *et al.* A homozygous FANCM mutation underlies a familial case of non-syndromic primary ovarian insufficiency. *Elife* **6** (2017). https://doi.org:10.7554/eLife.30490
118 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 (2020). https://doi.org:10.1016/j.maturitas.2020.06.004
119 Zhao, H. *et al.* Transcription factor FIGLA is mutated in patients with premature ovarian failure. *Am J Hum Genet* **82**, 1342-1348 (2008). https://doi.org:10.1016/j.ajhg.2008.04.018
120 Chen, B. *et al.* Consanguineous familial study revealed biallelic FIGLA mutation associated with premature ovarian insufficiency. *J Ovarian Res* **11**, 48 (2018). https://doi.org:10.1186/s13048-018-0413-0
121 Mei, L. *et al.* Mutations in FIGLA Associated With Premature Ovarian Insufficiency in a Chinese Population. *Front Med (Lausanne)* **8**, 714306 (2021). https://doi.org:10.3389/fmed.2021.714306
122 Li, J. *et al.* Heterozygous FMN2 missense variant found in a family case of premature ovarian insufficiency. *J Ovarian Res* **15**, 31 (2022). https://doi.org:10.1186/s13048-022-00960-y
123 Yoo, S., Yoon, J. Y., Keum, C. & Cheon, C. K. The first case of novel variants of the FSHR mutation causing primary amenorrhea in 2 siblings in Korea. *Ann Pediatr Endocrinol Metab* **28**, 54-60 (2023). https://doi.org:10.6065/apem.2142116.058
124 Crisponi, L. *et al.* The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* **27**, 159-166 (2001). https://doi.org:10.1038/84781
125 Laissue, P. *et al.* Functional evidence implicating FOXL2 in non-syndromic premature ovarian failure and in the regulation of the transcription factor OSR2. *J Med Genet* **46**, 455-457 (2009). https://doi.org:10.1136/jmg.2008.065086
126 Yang, X. W. *et al.* Novel FOXL2 mutations cause blepharophimosis-ptosis-epicanthus inversus syndrome with premature ovarian insufficiency. *Mol Genet Genomic Med* **6**, 261-267 (2018). https://doi.org:10.1002/mgg3.366
127 Nomura, R. *et al.* Nr5a1 suppression during the murine fetal period optimizes ovarian development by fine-tuning Notch signaling. *J Cell Sci* **132** (2019). https://doi.org:10.1242/jcs.223768
128 Grzechocińska, B., Warzecha, D., Wypchło, M., Ploski, R. & Wielgoś, M. Premature ovarian insufficiency as a variable feature of blepharophimosis, ptosis, and epicanthus inversus syndrome associated with c.223C > T p.(Leu75Phe) FOXL2 mutation: a case report. *BMC Med Genet* **20**, 132 (2019). https://doi.org:10.1186/s12881-019-0865-0

129 Fonseca, D. J. *et al.* Screening for mutations of the FOXO4 gene in premature ovarian failure patients. *Reprod Biomed Online* **24**, 339-341 (2012). https://doi.org:10.1016/j.rbmo.2011.11.017

130 Eskenazi, S. *et al.* Next Generation Sequencing Should Be Proposed to Every Woman With "Idiopathic" Primary Ovarian Insufficiency. *J Endocr Soc* **5**, bvab032 (2021). https://doi.org:10.1210/jendso/bvab032

131 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. *J Clin Endocrinol Metab* **88**, 3491-3498 (2003). https://doi.org:10.1210/jc.2003-030217

132 Liu, H. *et al.* A novel homozygous mutation in the FSHR gene is causative for primary ovarian insufficiency. *Fertil Steril* **108**, 1050-1055.e1052 (2017). https://doi.org:10.1016/j.fertnstert.2017.09.010

133 Fridovich-Keil, J. L. *et al.* Ovarian function in girls and women with GALT-deficiency galactosemia. *J Inherit Metab Dis* **34**, 357-366 (2011). https://doi.org:10.1007/s10545-010-9221-4

134 Dixit, H. *et al.* Mutational screening of the coding region of growth differentiation factor 9 gene in Indian women with ovarian failure. *Menopause* **12**, 749-754 (2005). https://doi.org:10.1097/01.gme.0000184424.96437.7a

135 Wang, T. T. *et al.* Identification of a mutation in GDF9 as a novel cause of diminished ovarian reserve in young women. *Hum Reprod* **28**, 2473-2481 (2013). https://doi.org:10.1093/humrep/det291

136 Verma, K. P. *et al.* A homozygous truncating variant in GDF9 in siblings with primary ovarian insufficiency. *J Assist Reprod Genet* **38**, 1539-1543 (2021). https://doi.org:10.1007/s10815-021-02144-x

137 Marinakis, N. M. *et al.* Ovarian insufficiency and secondary amenorrhea in a patient with a novel variant within GDF9 gene. *Menopause* **29**, 491-495 (2022). https://doi.org:10.1097/gme.0000000000001928

138 Foley, A. R. *et al.* GGPS1 Mutations Cause Muscular Dystrophy/Hearing Loss/Ovarian Insufficiency Syndrome. *Ann Neurol* **88**, 332-347 (2020). https://doi.org:10.1002/ana.25772

139 Pierce, S. B. *et al.* Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. *Proc Natl Acad Sci U S A* **108**, 6543-6548 (2011). https://doi.org:10.1073/pnas.1103471108

140 Wang, J., Zhang, W., Jiang, H. & Wu, B. L. Mutations in HFM1 in recessive primary ovarian insufficiency. *N Engl J Med* **370**, 972-974 (2014). https://doi.org:10.1056/NEJMc1310150

141 Pu, D. *et al.* Association analysis between HFM1 variation and primary ovarian insufficiency in Chinese women. *Clin Genet* **89**, 597-602 (2016). https://doi.org:10.1111/cge.12718

142 Tucker, E. J. *et al.* Meiotic genes in premature ovarian insufficiency: variants in HROB and REC8 as likely genetic causes. *Eur J Hum Genet* **30**, 219-228 (2022). https://doi.org:10.1038/s41431-021-00977-9

143 Pierce, S. B. *et al.* Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet* **87**, 282-288 (2010). https://doi.org:10.1016/j.ajhg.2010.07.007

144 Chen, K. *et al.* A homozygous missense variant in HSD17B4 identified in a consanguineous Chinese Han family with type II Perrault syndrome. *BMC Med Genet* **18**, 91 (2017). https://doi.org:10.1186/s12881-017-0453-0

145 Felipe-Medina, N. *et al.* A missense in HSF2BP causing primary ovarian insufficiency affects meiotic recombination by its novel interactor C19ORF57/BRME1. *Elife* **9** (2020). https://doi.org:10.7554/eLife.56996

146 Li, S. *et al.* Pathogenic Variations of Homologous Recombination Gene HSF2BP Identified in Sporadic Patients With Premature Ovarian Insufficiency. *Front Cell Dev Biol* **9**, 768123 (2021). https://doi.org:10.3389/fcell.2021.768123
147 Jolly, A. *et al.* Exome Sequencing of a Primary Ovarian Insufficiency Cohort Reveals Common Molecular Etiologies for a Spectrum of Disease. *J Clin Endocrinol Metab* **104**, 3049-3067 (2019). https://doi.org:10.1210/jc.2019-00248
148 Wu, H. *et al.* Homozygous missense mutation in CCDC155 disrupts the transmembrane distribution of CCDC155 and SUN1, resulting in non-obstructive azoospermia and premature ovarian insufficiency in humans. *Hum Genet* **141**, 1795-1809 (2022). https://doi.org:10.1007/s00439-022-02459-4
149 Wang, B. *et al.* Sequence variants of KHDRBS1 as high penetrance susceptibility risks for primary ovarian insufficiency by mis-regulating mRNA alternative splicing. *Hum Reprod* **32**, 2138-2146 (2017). https://doi.org:10.1093/humrep/dex263
150 Pierce, S. B. *et al.* Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. *Am J Hum Genet* **92**, 614-620 (2013). https://doi.org:10.1016/j.ajhg.2013.03.007
151 Latronico, A. C. *et al.* Brief report: testicular and ovarian resistance to luteinizing hormone caused by inactivating mutations of the luteinizing hormone-receptor gene. *N Engl J Med* **334**, 507-512 (1996). https://doi.org:10.1056/nejm199602223340805
152 Zielen, A. C. *et al.* A novel homozygous frame-shift variant in the LHCGR gene is associated with primary ovarian insufficiency in a Pakistani family. *Clin Genet* **94**, 396-397 (2018). https://doi.org:10.1111/cge.13406
153 McPherson, E. *et al.* Ovarian failure and dilated cardiomyopathy due to a novel lamin mutation. *Am J Med Genet A* **149a**, 567-572 (2009). https://doi.org:10.1002/ajmg.a.32627
154 AlAsiri, S. *et al.* Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J Clin Invest* **125**, 258-262 (2015). https://doi.org:10.1172/jci78473
155 Tenenbaum-Rakover, Y. *et al.* Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. *J Med Genet* **52**, 391-399 (2015). https://doi.org:10.1136/jmedgenet-2014-102921
156 Dou, X. *et al.* Minichromosome maintenance complex component 8 mutations cause primary ovarian insufficiency. *Fertil Steril* **106**, 1485-1489.e1482 (2016). https://doi.org:10.1016/j.fertnstert.2016.08.018
157 Desai, S. *et al.* MCM8 and MCM9 Nucleotide Variants in Women With Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **102**, 576-582 (2017). https://doi.org:10.1210/jc.2016-2565
158 Bouali, N. *et al.* New MCM8 mutation associated with premature ovarian insufficiency and chromosomal instability in a highly consanguineous Tunisian family. *Fertil Steril* **108**, 694-702 (2017). https://doi.org:10.1016/j.fertnstert.2017.07.015
159 Zhang, Y. X. *et al.* Novel loss-of-function mutation in MCM8 causes premature ovarian insufficiency. *Mol Genet Genomic Med* **8**, e1165 (2020). https://doi.org:10.1002/mgg3.1165
160 Heddar, A., Beckers, D., Fouquet, B., Roland, D. & Misrahi, M. A Novel Phenotype Combining Primary Ovarian Insufficiency Growth Retardation and Pilomatricomas With MCM8 Mutation. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgaa155
161 Wang, F., Guo, S. & Li, P. Two novel mutations in the MCM8 gene shared by two Chinese siblings with primary ovarian insufficiency and short stature. *Mol Genet Genomic Med* **8**, e1396 (2020). https://doi.org:10.1002/mgg3.1396
162 Jin, H. *et al.* Identification of potential causal variants for premature ovarian failure by whole exome sequencing. *BMC Med Genomics* **13**, 159 (2020). https://doi.org:10.1186/s12920-020-00813-x

172

163 Wood-Trageser, M. A. *et al.* MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am J Hum Genet* **95**, 754-762 (2014). https://doi.org:10.1016/j.ajhg.2014.11.002
164 Fauchereau, F. *et al.* A non-sense MCM9 mutation in a familial case of primary ovarian insufficiency. *Clin Genet* **89**, 603-607 (2016). https://doi.org:10.1111/cge.12736
165 Yang, X. *et al.* Gene variants identified by whole-exome sequencing in 33 French women with premature ovarian insufficiency. *J Assist Reprod Genet* **36**, 39-45 (2019). https://doi.org:10.1007/s10815-018-1349-4
166 Guo, T. *et al.* Novel pathogenic mutations in minichromosome maintenance complex component 9 (MCM9) responsible for premature ovarian insufficiency. *Fertil Steril* **113**, 845-852 (2020). https://doi.org:10.1016/j.fertnstert.2019.11.015
167 Alvarez-Mora, M. I. *et al.* An exome-wide exploration of cases of primary ovarian insufficiency uncovers novel sequence variants and candidate genes. *Clin Genet* **98**, 293-298 (2020). https://doi.org:10.1111/cge.13803
168 França, M. M. *et al.* Screening of targeted panel genes in Brazilian patients with primary ovarian insufficiency. *PLoS One* **15**, e0240795 (2020). https://doi.org:10.1371/journal.pone.0240795
169 Shen, J. *et al.* Genetic etiologic analysis in 74 Chinese Han women with idiopathic premature ovarian insufficiency by combined molecular genetic testing. *J Assist Reprod Genet* **38**, 965-978 (2021). https://doi.org:10.1007/s10815-021-02083-7
170 Caburet, S. *et al.* A truncating MEIOB mutation responsible for familial primary ovarian insufficiency abolishes its interaction with its partner SPATA22 and their recruitment to DNA double-strand breaks. *EBioMedicine* **42**, 524-531 (2019). https://doi.org:10.1016/j.ebiom.2019.03.075
171 Chen, A. *et al.* Mutations in the mitochondrial ribosomal protein MRPS22 lead to primary ovarian insufficiency. *Hum Mol Genet* **27**, 1913-1926 (2018). https://doi.org:10.1093/hmg/ddy098
172 Carlosama, C. *et al.* A homozygous donor splice-site mutation in the meiotic gene MSH4 causes primary ovarian insufficiency. *Hum Mol Genet* **26**, 3161-3166 (2017). https://doi.org:10.1093/hmg/ddx199
173 Akbari, A. *et al.* Rare missense variant in MSH4 associated with primary gonadal failure in both 46, XX and 46, XY individuals. *Hum Reprod* **36**, 1134-1145 (2021). https://doi.org:10.1093/humrep/deaa362
174 Guo, T. *et al.* Mutations in MSH5 in primary ovarian insufficiency. *Hum Mol Genet* **26**, 1452-1457 (2017). https://doi.org:10.1093/hmg/ddx044
175 Wu, X. *et al.* A NANOS3 mutation linked to protein degradation causes premature ovarian insufficiency. *Cell Death Dis* **4**, e825 (2013). https://doi.org:10.1038/cddis.2013.368
176 Santos, M. G. *et al.* Homozygous inactivating mutation in NANOS3 in two sisters with primary ovarian insufficiency. *Biomed Res Int* **2014**, 787465 (2014). https://doi.org:10.1155/2014/787465
177 Tucker, E. J. *et al.* Identification of variants in pleiotropic genes causing "isolated" premature ovarian insufficiency: implications for medical practice. *Eur J Hum Genet* **26**, 1319-1328 (2018). https://doi.org:10.1038/s41431-018-0140-4
178 Li, G. *et al.* Haploinsufficiency in non-homologous end joining factor 1 induces ovarian dysfunction in humans and mice. *J Med Genet* **59**, 579-588 (2022). https://doi.org:10.1136/jmedgenet-2020-107398
179 Qin, Y. *et al.* NOBOX homeobox mutation causes premature ovarian failure. *Am J Hum Genet* **81**, 576-581 (2007). https://doi.org:10.1086/519496
180 Bouilly, J., Bachelot, A., Broutin, I., Touraine, P. & Binart, N. Novel NOBOX loss-of-function mutations account for 6.2% of cases in a large primary ovarian insufficiency cohort. *Hum Mutat* **32**, 1108-1113 (2011). https://doi.org:10.1002/humu.21543

181 Bouali, N. *et al.* NOBOX is a strong autosomal candidate gene in Tunisian patients with primary ovarian insufficiency. *Clin Genet* **89**, 608-613 (2016). https://doi.org:10.1111/cge.12750
182 Ferrari, I. *et al.* Impaired protein stability and nuclear localization of NOBOX variants associated with premature ovarian insufficiency. *Hum Mol Genet* **25**, 5223-5233 (2016). https://doi.org:10.1093/hmg/ddw342
183 Li, L. *et al.* A homozygous NOBOX truncating variant causes defective transcriptional activation and leads to primary ovarian insufficiency. *Hum Reprod* **32**, 248-255 (2017). https://doi.org:10.1093/humrep/dew271
184 Sassi, A. *et al.* Compound heterozygous null mutations of NOBOX in sisters with delayed puberty and primary amenorrhea. *Mol Genet Genomic Med* **9**, e1776 (2021). https://doi.org:10.1002/mgg3.1776
185 Kosaki, K. *et al.* Premature ovarian failure in a female with proximal symphalangism and Noggin mutation. *Fertil Steril* **81**, 1137-1139 (2004). https://doi.org:10.1016/j.fertnstert.2003.08.054
186 Patiño, L. C. *et al.* Functional evidence implicating NOTCH2 missense mutations in primary ovarian insufficiency etiology. *Hum Mutat* **40**, 25-30 (2019). https://doi.org:10.1002/humu.23667
187 Li, L. *et al.* NOTCH2 variant D1853H is mutated in two non-syndromic premature ovarian insufficiency patients from a Chinese pedigree. *J Ovarian Res* **13**, 41 (2020). https://doi.org:10.1186/s13048-020-00645-4
188 Lourenço, D. *et al.* Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med* **360**, 1200-1210 (2009). https://doi.org:10.1056/NEJMoa0806228
189 Warman, D. M. *et al.* Three new SF-1 (NR5A1) gene mutations in two unrelated families with multiple affected members: within-family variability in 46,XY subjects and low ovarian reserve in fertile 46,XX subjects. *Horm Res Paediatr* **75**, 70-77 (2011). https://doi.org:10.1159/000320029
190 Janse, F. *et al.* Limited contribution of NR5A1 (SF-1) mutations in women with primary ovarian insufficiency (POI). *Fertil Steril* **97**, 141-146.e142 (2012). https://doi.org:10.1016/j.fertnstert.2011.10.032
191 Lakhal, B. *et al.* Mutational screening of SF1 and WNT4 in Tunisian women with premature ovarian failure. *Gene* **509**, 298-301 (2012). https://doi.org:10.1016/j.gene.2012.08.007
192 Philibert, P. *et al.* NR5A1 (SF-1) gene variants in a group of 26 young women with XX primary ovarian insufficiency. *Fertil Steril* **99**, 484-489 (2013). https://doi.org:10.1016/j.fertnstert.2012.10.026
193 Safari, S. *et al.* An Iranian family with azoospermia and premature ovarian insufficiency segregating NR5A1 mutation. *Climacteric* **17**, 301-303 (2014). https://doi.org:10.3109/13697137.2013.847079
194 Jiao, X. *et al.* Novel NR5A1 missense mutation in premature ovarian failure: detection in han chinese indicates causation in different ethnic groups. *PLoS One* **8**, e74759 (2013). https://doi.org:10.1371/journal.pone.0074759
195 Eggers, S. *et al.* Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1. *Eur J Hum Genet* **23**, 486-493 (2015). https://doi.org:10.1038/ejhg.2014.130
196 Oral, E. *et al.* Clinical and Genetic Investigation of Premature Ovarian Insufficiency Cases from Turkey. *J Gynecol Obstet Hum Reprod* **48**, 817-823 (2019). https://doi.org:10.1016/j.jogoh.2019.04.007
197 Jaillard, S. *et al.* Analysis of NR5A1 in 142 patients with premature ovarian insufficiency, diminished ovarian reserve, or unexplained infertility. *Maturitas* **131**, 78-86 (2020). https://doi.org:10.1016/j.maturitas.2019.10.011

198 Bertrand-Delepine, J. *et al.* In cases of familial primary ovarian insufficiency and disorders of gonadal development, consider NR5A1/SF-1 sequence variants. *Reprod Biomed Online* **40**, 151-159 (2020). https://doi.org:10.1016/j.rbmo.2019.10.002

199 Cattoni, A. *et al.* The Potential Synergic Effect of a Complex Pattern of Multiple Inherited Genetic Variants as a Pathogenic Factor for Ovarian Dysgenesis: A Case Report. *Front Endocrinol (Lausanne)* **11**, 540683 (2020). https://doi.org:10.3389/fendo.2020.540683

200 Brauner, R., Pierrepont, S., Bignon-Topalovic, J., McElreavey, K. & Bashamboo, A. Etiology of primary ovarian insufficiency in a series young girls presenting at a pediatric endocrinology center. *Eur J Pediatr* **174**, 767-773 (2015). https://doi.org:10.1007/s00431-014-2457-5

201 Weinberg-Shukron, A. *et al.* A mutation in the nucleoporin-107 gene causes XX gonadal dysgenesis. *J Clin Invest* **125**, 4295-4304 (2015). https://doi.org:10.1172/jci83553

202 Ren, Y. *et al.* Functional study of a novel missense single-nucleotide variant of NUP107 in two daughters of Mexican origin with premature ovarian insufficiency. *Mol Genet Genomic Med* **6**, 276-281 (2018). https://doi.org:10.1002/mgg3.345

203 Qin, Y., Jiao, X., Simpson, J. L. & Chen, Z. J. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update* **21**, 787-808 (2015). https://doi.org:10.1093/humupd/dmv036

204 Peng, T. *et al.* Novel PMM2 missense mutation in a Chinese family with non-syndromic premature ovarian insufficiency. *J Assist Reprod Genet* **37**, 443-450 (2020). https://doi.org:10.1007/s10815-019-01675-8

205 Masunaga, Y. *et al.* Primary ovarian insufficiency in a female with phosphomannomutase-2 gene (PMM2) mutations for congenital disorder of glycosylation. *Endocr J* **68**, 605-611 (2021). https://doi.org:10.1507/endocrj.EJ20-0706

206 Lacombe, A. *et al.* Disruption of POF1B binding to nonmuscle actin filaments is associated with premature ovarian failure. *Am J Hum Genet* **79**, 113-119 (2006). https://doi.org:10.1086/505406

207 Luoma, P. *et al.* Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. *Lancet* **364**, 875-882 (2004). https://doi.org:10.1016/s0140-6736(04)16983-3

208 Mancuso, M., Filosto, M., Oh, S. J. & DiMauro, S. A novel polymerase gamma mutation in a family with ophthalmoplegia, neuropathy, and Parkinsonism. *Arch Neurol* **61**, 1777-1779 (2004). https://doi.org:10.1001/archneur.61.11.1777

209 Hakonen, A. H. *et al.* Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *Am J Hum Genet* **77**, 430-441 (2005). https://doi.org:10.1086/444548

210 Pagnamenta, A. T. *et al.* Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. *Hum Reprod* **21**, 2467-2473 (2006). https://doi.org:10.1093/humrep/del076

211 Blok, M. J. *et al.* The unfolding clinical spectrum of POLG mutations. *J Med Genet* **46**, 776-785 (2009). https://doi.org:10.1136/jmg.2009.067686

212 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 (2012). https://doi.org:10.1016/j.gene.2012.02.034

213 Chen, B. *et al.* Identification of the first homozygous POLG mutation causing non-syndromic ovarian dysfunction. *Climacteric* **21**, 467-471 (2018). https://doi.org:10.1080/13697137.2018.1467891

214 Yang, H. J., Park, G., Nam-Goong, I. S., Ahn, J. W. & Weon, Y. C. The First Korean Siblings With Adult-Onset 4H Leukodystrophy Related to Nonsynonymous POLR3B Mutations. *Neurol Genet* **8**, e667 (2022). https://doi.org:10.1212/nxg.0000000000000667

215 Franca, M. M. *et al.* Exome Sequencing Reveals the POLR3H Gene as a Novel Cause of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **104**, 2827-2841 (2019). https://doi.org:10.1210/jc.2018-02485
216 Wang, J. *et al.* New candidate gene POU5F1 associated with premature ovarian failure in Chinese patients. *Reprod Biomed Online* **22**, 312-316 (2011). https://doi.org:10.1016/j.rbmo.2010.11.008
217 Zangen, D. *et al.* XX ovarian dysgenesis is caused by a PSMC3IP/HOP2 mutation that abolishes coactivation of estrogen-driven transcription. *Am J Hum Genet* **89**, 572-579 (2011). https://doi.org:10.1016/j.ajhg.2011.09.006
218 Mei, L. *et al.* Two novel biallelic mutations in PSMC3IP in a patient affected by premature ovarian insufficiency. *Mol Med Rep* **25** (2022). https://doi.org:10.3892/mmr.2021.12561
219 Sirchia, F., Giorgio, E., Cucinella, L., Valente, E. M. & Nappi, R. E. Biallelic mutations in PSMC3IP are associated with secondary amenorrhea: expanding the spectrum of premature ovarian insufficiency. *J Assist Reprod Genet* **39**, 1177-1181 (2022). https://doi.org:10.1007/s10815-022-02471-7
220 Luo, W. *et al.* Variants in Homologous Recombination Genes EXO1 and RAD51 Related with Premature Ovarian Insufficiency. *J Clin Endocrinol Metab* **105** (2020). https://doi.org:10.1210/clinem/dgaa505
221 Coppieters, F. *et al.* Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. *Am J Hum Genet* **99**, 470-480 (2016). https://doi.org:10.1016/j.ajhg.2016.06.017
222 Huang, C., Guo, T. & Qin, Y. Meiotic Recombination Defects and Premature Ovarian Insufficiency. *Front Cell Dev Biol* **9**, 652407 (2021). https://doi.org:10.3389/fcell.2021.652407
223 Wang, B. *et al.* Mutational analysis of SAL-Like 4 (SALL4) in Han Chinese women with premature ovarian failure. *Mol Hum Reprod* **15**, 557-562 (2009). https://doi.org:10.1093/molehr/gap046
224 Wang, Q. *et al.* Whole-exome sequencing reveals SALL4 variants in premature ovarian insufficiency: an update on genotype-phenotype correlations. *Hum Genet* **138**, 83-92 (2019). https://doi.org:10.1007/s00439-018-1962-4
225 Faridi, R. *et al.* Mutations of SGO2 and CLDN14 collectively cause coincidental Perrault syndrome. *Clin Genet* **91**, 328-332 (2017). https://doi.org:10.1111/cge.12867
226 Zhao, S. *et al.* Transcription factor SOHLH1 potentially associated with primary ovarian insufficiency. *Fertil Steril* **103**, 548-553.e545 (2015). https://doi.org:10.1016/j.fertnstert.2014.11.011
227 Qin, Y. *et al.* Novel variants in the SOHLH2 gene are implicated in human premature ovarian failure. *Fertil Steril* **101**, 1104-1109.e1106 (2014). https://doi.org:10.1016/j.fertnstert.2014.01.001
228 Yao, C. *et al.* Bi-allelic SPATA22 variants cause premature ovarian insufficiency and nonobstructive azoospermia due to meiotic arrest. *Clin Genet* **101**, 507-516 (2022). https://doi.org:10.1111/cge.14129
229 Smirin-Yosef, P. *et al.* A Biallelic Mutation in the Homologous Recombination Repair Gene SPIDR Is Associated With Human Gonadal Dysgenesis. *J Clin Endocrinol Metab* **102**, 681-688 (2017). https://doi.org:10.1210/jc.2016-2714
230 Heddar, A., Guichoux, N., Auger, N. & Misrahi, M. A SPIDR homozygous nonsense pathogenic variant in isolated primary ovarian insufficiency with chromosomal instability. *Clin Genet* **101**, 242-246 (2022). https://doi.org:10.1111/cge.14080
231 Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *N Engl J Med* **370**, 943-949 (2014). https://doi.org:10.1056/NEJMoa1309635
232 Le Quesne Stabej, P. *et al.* STAG3 truncating variant as the cause of primary ovarian insufficiency. *Eur J Hum Genet* **24**, 135-138 (2016). https://doi.org:10.1038/ejhg.2015.107

233 He, W. B. *et al.* Whole-exome sequencing identifies a homozygous donor splice-site mutation in STAG3 that causes primary ovarian insufficiency. *Clin Genet* **93**, 340-344 (2018). https://doi.org:10.1111/cge.13034
234 Colombo, R., Pontoglio, A. & Bini, M. A STAG3 missense mutation in two sisters with primary ovarian insufficiency. *Eur J Obstet Gynecol Reprod Biol* **216**, 269-271 (2017). https://doi.org:10.1016/j.ejogrb.2017.08.005
235 França, M. M. *et al.* Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency. *Eur J Med Genet* **62**, 186-189 (2019). https://doi.org:10.1016/j.ejmg.2018.07.008
236 Heddar, A., Dessen, P., Flatters, D. & Misrahi, M. Novel STAG3 mutations in a Caucasian family with primary ovarian insufficiency. *Mol Genet Genomics* **294**, 1527-1534 (2019). https://doi.org:10.1007/s00438-019-01594-4
237 Xiao, W. J. *et al.* In-Frame Variants in STAG3 Gene Cause Premature Ovarian Insufficiency. *Front Genet* **10**, 1016 (2019). https://doi.org:10.3389/fgene.2019.01016
238 Mellone, S. *et al.* A Long Contiguous Stretch of Homozygosity Disclosed a Novel STAG3 Biallelic Pathogenic Variant Causing Primary Ovarian Insufficiency: A Case Report and Review of the Literature. *Genes (Basel)* **12** (2021). https://doi.org:10.3390/genes12111709
239 Akbari, A. *et al.* Novel STAG3 variant associated with primary ovarian insufficiency and non-obstructive azoospermia in an Iranian consanguineous family. *Gene* **821**, 146281 (2022). https://doi.org:10.1016/j.gene.2022.146281
240 Gómez-Rojas, S., Aristizábal-Duque, J. E., Muñoz-Fernández, L. F., Sarmiento-Ramón, M. P. & Pereira-Gómez, M. D. P. New STAG3 gene variant as a cause of premature ovarian insufficiency. *Rev Colomb Obstet Ginecol* **73**, 142-148 (2022). https://doi.org:10.18597/rcog.3806
241 de Vries, L. *et al.* Exome sequencing reveals SYCE1 mutation associated with autosomal recessive primary ovarian insufficiency. *J Clin Endocrinol Metab* **99**, E2129-2132 (2014). https://doi.org:10.1210/jc.2014-1268
242 He, W. B. *et al.* Homozygous variants in SYCP2L cause premature ovarian insufficiency. *J Med Genet* **58**, 168-172 (2021). https://doi.org:10.1136/jmedgenet-2019-106789
243 Tucker, E. J. *et al.* TP63-truncating variants cause isolated premature ovarian insufficiency. *Hum Mutat* **40**, 886-892 (2019). https://doi.org:10.1002/humu.23744
244 Mathorne, S. W. *et al.* Novel phenotype of syndromic premature ovarian insufficiency associated with TP63 molecular defect. *Clin Genet* **97**, 779-784 (2020). https://doi.org:10.1111/cge.13725
245 Karlberg, S., Tiitinen, A., Alfthan, H. & Lipsanen-Nyman, M. Premature ovarian insufficiency and early depletion of the ovarian reserve in the monogenic Mulibrey nanism disorder. *Hum Reprod* **33**, 1254-1261 (2018). https://doi.org:10.1093/humrep/dey103
246 Morino, H. *et al.* Mutations in Twinkle primase-helicase cause Perrault syndrome with neurologic features. *Neurology* **83**, 2054-2061 (2014). https://doi.org:10.1212/wnl.0000000000001036
247 Demain, L. A. *et al.* Expanding the genotypic spectrum of Perrault syndrome. *Clin Genet* **91**, 302-312 (2017). https://doi.org:10.1111/cge.12776
248 Lerat, J. *et al.* An Application of NGS for Molecular Investigations in Perrault Syndrome: Study of 14 Families and Review of the Literature. *Hum Mutat* **37**, 1354-1362 (2016). https://doi.org:10.1002/humu.23120
249 Ołdak, M. *et al.* Novel neuro-audiological findings and further evidence for TWNK involvement in Perrault syndrome. *J Transl Med* **15**, 25 (2017). https://doi.org:10.1186/s12967-017-1129-4
250 Domínguez-Ruiz, M. *et al.* Perrault syndrome with neurological features in a compound heterozygote for two TWNK mutations: overlap of TWNK-related recessive disorders. *J Transl Med* **17**, 290 (2019). https://doi.org:10.1186/s12967-019-2041-x

251 Gotta, F. *et al.* A novel mutation of Twinkle in Perrault syndrome: A not rare diagnosis? *Ann Hum Genet* **84**, 417-422 (2020). https://doi.org:10.1111/ahg.12384
252 Zhou, Y. *et al.* Wdr62 is involved in female meiotic initiation via activating JNK signaling and associated with POI in humans. *PLoS Genet* **14**, e1007463 (2018). https://doi.org:10.1371/journal.pgen.1007463
253 Rossi, M. L., Ghosh, A. K. & Bohr, V. A. Roles of Werner syndrome protein in protection of genome integrity. *DNA Repair (Amst)* **9**, 331-344 (2010). https://doi.org:10.1016/j.dnarep.2009.12.011
254 Wang, H. *et al.* Novel WT1 Missense Mutations in Han Chinese Women with Premature Ovarian Failure. *Sci Rep* **5**, 13983 (2015). https://doi.org:10.1038/srep13983
255 Wang, Y. *et al.* Whole exome sequencing identified a rare WT1 loss-of-function variant in a non-syndromic POI patient. *Mol Genet Genomic Med* **10**, e1820 (2022). https://doi.org:10.1002/mgg3.1820
256 McGlacken-Byrne, S. M. *et al.* Pathogenic variants in the human m6A reader YTHDC2 are associated with primary ovarian insufficiency. *JCI Insight* **7** (2022). https://doi.org:10.1172/jci.insight.154671
257 McGlacken-Byrne, S. M. *et al.* ZSWIM7 Is Associated With Human Female Meiosis and Familial Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **107**, e254-e263 (2022). https://doi.org:10.1210/clinem/dgab597
258 Yatsenko, S. A. *et al.* Pathogenic Variants in ZSWIM7 Cause Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* **107**, e2359-e2364 (2022). https://doi.org:10.1210/clinem/dgac090
259 Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-69 (2012). https://doi.org:10.1126/science.1219240
260 Zuk, O. *et al.* Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455-E464 (2014). https://doi.org:doi:10.1073/pnas.1322563111
261 McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356-369 (2008). https://doi.org:10.1038/nrg2344
262 Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31-46 (2010). https://doi.org:10.1038/nrg2626
263 Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**, 363-376 (2011). https://doi.org:10.1038/nrg2958
264 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009). https://doi.org:10.1038/nature08494
265 Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**, 415-425 (2010). https://doi.org:10.1038/nrg2779
266 Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet* **48**, 314-317 (2016). https://doi.org:10.1038/ng.3507
267 Stasa, S. *et al.* Genetic susceptibility to earlier ovarian ageing increases de novo mutation rate in offspring. *medRxiv*, 2022.2006.2023.22276698 (2022). https://doi.org:10.1101/2022.06.23.22276698
268 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001). https://doi.org:10.1038/35057062
269 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015). https://doi.org:10.1038/nature15393

270 Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* **44**, 243-246 (2012). https://doi.org:10.1038/ng.1074
271 Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A* **108**, 11983-11988 (2011). https://doi.org:10.1073/pnas.1019276108
272 Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* **49**, 1421-1427 (2017). https://doi.org:10.1038/ng.3954
273 Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet* **100**, 635-649 (2017). https://doi.org:10.1016/j.ajhg.2017.03.004
274 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018). https://doi.org:10.1038/s41586-018-0579-z
275 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021). https://doi.org:10.1038/s41586-021-03205-y
276 Gravel, S. Population Genetics Models of Local Ancestry. *Genetics* **191**, 607-619 (2012). https://doi.org:10.1534/genetics.112.139808
277 Mannherz, W. & Agarwal, S. Thymidine nucleotide metabolism controls human telomere length. *Nature Genetics* **55**, 568-580 (2023). https://doi.org:10.1038/s41588-023-01339-5
278 Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* **104**, 275-286 (2019). https://doi.org:10.1016/j.ajhg.2018.12.015
279 Kingdom, R. & Wright, C. F. Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Front Genet* **13**, 920390 (2022). https://doi.org:10.3389/fgene.2022.920390
280 Kraft, P., Zeggini, E. & Ioannidis, J. P. Replication in genome-wide association studies. *Stat Sci* **24**, 561-573 (2009). https://doi.org:10.1214/09-sts290
281 Park, S. *et al.* Adjusting heterogeneous ascertainment bias for genetic association analysis with extended families. *BMC Med Genet* **16**, 62 (2015). https://doi.org:10.1186/s12881-015-0198-6
282 Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H. & Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res* **15**, 1496-1502 (2005). https://doi.org:10.1101/gr.4107905
283 Shekari, S. *et al.* Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency. *Nature Medicine* (2023). https://doi.org:10.1038/s41591-023-02405-5
284 Weedon, M. N. *et al.* The impact of Mendelian sleep and circadian genetic variants in a population setting. *PLoS Genet* **18**, e1010356 (2022). https://doi.org:10.1371/journal.pgen.1010356
285 Mirshahi, U. L. *et al.* Reduced penetrance of MODY-associated HNF1A/HNF4A variants but not GCK variants in clinically unselected cohorts. *Am J Hum Genet* **109**, 2018-2028 (2022). https://doi.org:10.1016/j.ajhg.2022.09.014
286 Locke, J. M. *et al.* Association of birthweight and penetrance of diabetes in individuals with HNF4A-MODY: a cohort study. *Diabetologia* **65**, 246-249 (2022). https://doi.org:10.1007/s00125-021-05581-6
287 Rees, E. & Owen, M. J. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Medicine* **12**, 43 (2020). https://doi.org:10.1186/s13073-020-00734-5
288 Gratten, J. & Visscher, P. M. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med* **8**, 78 (2016). https://doi.org:10.1186/s13073-016-0332-x

289 Visscher, P. M. & Yang, J. A plethora of pleiotropy across complex traits. *Nat Genet* **48**, 707-708 (2016). https://doi.org:10.1038/ng.3604
290 Ritchie, M. D. & Van Steen, K. The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. *Ann Transl Med* **6**, 157 (2018). https://doi.org:10.21037/atm.2018.04.05
291 Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* **132**, 1077-1130 (2013). https://doi.org:10.1007/s00439-013-1331-2
292 Boycott, K. M. & Innes, A. M. When One Diagnosis Is Not Enough. *N Engl J Med* **376**, 83-85 (2017). https://doi.org:10.1056/NEJMe1614384
293 van Rooij, J. *et al.* Reduced penetrance of pathogenic ACMG variants in a deeply phenotyped cohort study and evaluation of ClinVar classification over time. *Genet Med* **22**, 1812-1820 (2020). https://doi.org:10.1038/s41436-020-0900-8
294 Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* **19**, 249-255 (2017). https://doi.org:10.1038/gim.2016.190
295 Elfatih, A., Mohammed, I., Abdelrahman, D. & Mifsud, B. Frequency and management of medically actionable incidental findings from genome and exome sequencing data: a systematic review. *Physiol Genomics* **53**, 373-384 (2021). https://doi.org:10.1152/physiolgenomics.00025.2021
296 Ruth, K. S. *et al.* Genome-wide association study of anti-Müllerian hormone levels in pre-menopausal women of late reproductive age and relationship with genetic determinants of reproductive lifespan. *Hum Mol Genet* **28**, 1392-1401 (2019). https://doi.org:10.1093/hmg/ddz015
297 Ruth, K. S. *et al.* Genetic evidence that lower circulating FSH levels lengthen menstrual cycle, increase age at menopause and impact female reproductive health. *Hum Reprod* **31**, 473-481 (2016). https://doi.org:10.1093/humrep/dev318
298 Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31 (2019). https://doi.org:10.1016/j.cell.2019.02.048
299 Zong, N. *et al.* Leveraging Genetic Reports and Electronic Health Records for the Prediction of Primary Cancers: Algorithm Development and Validation Study. *JMIR Med Inform* **9**, e23586 (2021). https://doi.org:10.2196/23586
300 Hripcsak, G. & Albers, D. J. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* **20**, 117-121 (2013). https://doi.org:10.1136/amiajnl-2012-001145
301 Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics* **13**, 395-405 (2012). https://doi.org:10.1038/nrg3208
302 Honeyford, K. *et al.* Challenges and recommendations for high quality research using electronic health records. *Front Digit Health* **4**, 940330 (2022). https://doi.org:10.3389/fdgth.2022.940330
303 Richesson, R. L., Sun, J., Pathak, J., Kho, A. N. & Denny, J. C. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. *Artif Intell Med* **71**, 57-61 (2016). https://doi.org:10.1016/j.artmed.2016.05.005
304 Vassy, J. L. *et al.* The Impact of Whole-Genome Sequencing on the Primary Care and Outcomes of Healthy Adult Patients: A Pilot Randomized Trial. *Ann Intern Med* **167**, 159-169 (2017). https://doi.org:10.7326/m17-0188
305 Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161-164 (2016). https://doi.org:10.1038/538161a

306 Bustamante, C. D., Burchard, E. G. & De la Vega, F. M. Genomics for the world. *Nature* **475**, 163-165 (2011). https://doi.org:10.1038/475163a
307 Shriner, D. & Rotimi, C. N. Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *Am J Hum Genet* **102**, 547-556 (2018). https://doi.org:10.1016/j.ajhg.2018.02.003
308 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019). https://doi.org:10.1038/s41588-018-0269-7
309 Fraison, E., Crawford, G., Casper, G., Harris, V. & Ledger, W. Pregnancy following diagnosis of premature ovarian insufficiency: a systematic review. *Reprod Biomed Online* **39**, 467-476 (2019). https://doi.org:10.1016/j.rbmo.2019.04.019
310 Horton, R. *et al.* Direct-to-consumer genetic testing. *Bmj* **367**, l5688 (2019). https://doi.org:10.1136/bmj.l5688
311 Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* **52**, 572-581 (2020). https://doi.org:10.1038/s41588-020-0609-2
312 Cybulski, C. *et al.* Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-646 (2015). https://doi.org:10.1038/ng.3284
313 Kuchenbaecker, K. B. *et al.* Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama* **317**, 2402-2416 (2017). https://doi.org:10.1001/jama.2017.7112
314 Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-59 (2002). https://doi.org:10.1038/ng879
315 Okamoto, K. & Seimiya, H. Revisiting Telomere Shortening in Cancer. *Cells* **8** (2019). https://doi.org:10.3390/cells8020107
316 Kroupa, M. *et al.* The dynamics of telomere length in primary and metastatic colorectal cancer lesions. *Scientific Reports* **13**, 9097 (2023). https://doi.org:10.1038/s41598-023-35835-9
317 Brown, J. C. *et al.* The effects of exercise and diet on oxidative stress and telomere length in breast cancer survivors. *Breast Cancer Research and Treatment* **199**, 109-117 (2023). https://doi.org:10.1007/s10549-023-06868-5
318 Bocher, O., Willer, C. J. & Zeggini, E. Unravelling the genetic architecture of human complex traits through whole genome sequencing. *Nature Communications* **14**, 3520 (2023). https://doi.org:10.1038/s41467-023-39259-x
319 Wainschtein, P. *et al.* Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nature Genetics* **54**, 263-273 (2022). https://doi.org:10.1038/s41588-021-00997-7
320 Wakeling, M. N. *et al.* Non-coding variants disrupting a tissue-specific regulatory element in HK1 cause congenital hyperinsulinism. *Nat Genet* **54**, 1615-1620 (2022). https://doi.org:10.1038/s41588-022-01204-x
321 Wright, C. F. *et al.* Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. *Am J Hum Genet* **108**, 1083-1094 (2021). https://doi.org:10.1016/j.ajhg.2021.04.025
322 Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102 (2020). https://doi.org:10.1038/s41586-020-2434-2
323 Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175-180 (2017). https://doi.org:10.1038/nature22071
324 Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016). https://doi.org:10.1038/nature17676
325 Herberts, C. *et al.* Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature* **608**, 199-208 (2022). https://doi.org:10.1038/s41586-022-04975-9