Human Molecular Genetics, 2013, Vol. 22, No. 7 **1465–1472**
doi:10.1093/hmg/dds551
Advance Access published on January 9, 2013

A genome-wide association study of early menopause and the combined impact of identified variants

John R. B. Perry[1,2,3,#], Tanguy Corre[4,5,6,#], Tõnu Esko[7,8,#], Daniel I. Chasman[10,11], Krista Fischer[7], Nora Franceschini[12], Chunyan He[14,15], Zoltan Kutalik[4,5], Massimo Mangino[3], Lynda M. Rose[10], Albert Vernon Smith[16,17], Lisette Stolk[18,21], Patrick Sulem[22], Michael N. Weedon[1], Wei V. Zhuang[23], Alice Arnold[24], Alan Ashworth[26], Sven Bergmann[4,5], Julie E. Buring[10,11,27], Andrea Burri[3], Constance Chen[28], Marilyn C. Cornelis[29], David J. Couper[13], Mark O. Goodarzi[30], Vilmundur Gudnason[16,17], Tamara Harris[31], Albert Hofman[19,21], Michael Jones[32], Peter Kraft[28,29,33], Lenore Launer[31], Joop S. E. Laven[20], Guo Li[25], Barbara McKnight[24], Corrado Masciullo[6], Lili Milani[7], Nicholas Orr[26], Bruce M. Psaty[34,35], ReproGen Consortium, Paul M. Ridker[10,11,27], Fernando Rivadeneira[18,21], Cinzia Sala[6], Andres Salumets[9,36], Minouk Schoemaker[32], Michela Traglia[6], Gérard Waeber[37], Stephen J. Chanock[38], Ellen W. Demerath[39], Melissa Garcia[31], Susan E. Hankinson[28,33], Frank B. Hu[28,29,33], David J. Hunter[28,29,33], Kathryn L. Lunetta[23], Andres Metspalu[7,8], Grant W. Montgomery[40], Joanne M. Murabito[41,42], Anne B. Newman[43], Ken K. Ong[44,45], Tim D. Spector[3], Kari Stefansson[22], Anthony J. Swerdlow[32], Unnur Thorsteinsdottir[22], Rob M. Van Dam[29,46], André G. Uitterlinden[18,21,19], Jenny A. Visser[18], Peter Vollenweider[37], Daniela Toniolo[6,47,#] and Anna Murray[1,#,*]

[1]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK [3]Department of Twin Research and Genetic Epidemiology, King's College London, Lambeth Palace Rd, London SE1 7EH, UK [4]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland [5]Swiss Institute of Bioinformatics, Lausanne, Switzerland [6]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy [7]Estonian Genome Center, [8]Institute of Molecular and Cell Biology, [9]Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia [10]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA [11]Harvard Medical School, Boston MA 02115, USA [12]Department of Epidemiology, [13]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA [14]Department of Public Health, Indiana University School of Medicine, Indianapolis, Indiana, USA [15]Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, USA [16]Icelandic Heart Association, Kopavogur, Iceland [17]University of Iceland, Reykjavik, Iceland [18]Department of Internal Medicine, [19]Department of Epidemiology, [20]Division of Reproductive Medicine, Erasmus MC, [21]Netherlands Consortium of Healthy Aging, Rotterdam, The Netherlands [22]Decode genetics, Sturlugata 8, Reykjavik 105, Iceland [23]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA [24]Department of Biostatistics, [25]Department of Medicine, University of Washington, Seattle, WA, USA [26]Breakthrough Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK [27]Harvard School of Public Health, Boston, MA 02115, USA [28]Department of Epidemiology, [29]Department of Nutrition, Harvard School of Public

[*]To whom correspondence should be addressed at: Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK EX1 2LU. Tel: +44 1392722976; Fax: +44 1392722926; Email: anna.murray@pms.ac.uk
[#]These authors contributed equally to this work.

Health, Boston, MA, USA [30]Division of Endocrinology, Diabetes & Metabolism, Cedars-Sinai Medical Center [31]National Institutes on Aging, NIH, Bethesda, MD, USA [32]Section of Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK [33]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA [34]Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA, USA [35]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA [36]Competence Centre on Reproductive Medicine and Biology, Tartu, Estonia [37]Division of Internal Medicine, Lausanne University Hospital, CHUV, Lausanne, Switzerland [38]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [39]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA [40]Queensland Institute of Medical Research, Brisbane, Queensland, Australia [41]National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA [42]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02118, USA [43]Departments of Epidemiology and Medicine, University of Pittsburgh, Pittsburgh, PA, USA [44]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK [45]Department of Paediatrics, University of Cambridge, Cambridge, UK [46]Saw Swee Hock School of Public Health and Yong Loo Lin School of Medicine, National University of Singapore, Singapore [47]Institute of Molecular Genetics, 27100 Pavia, Italy

Received July 24, 2012; Revised October 12, 2012; Accepted December 24, 2012

Early menopause (EM) affects up to 10% of the female population, reducing reproductive lifespan considerably. Currently, it constitutes the leading cause of infertility in the western world, affecting mainly those women who postpone their first pregnancy beyond the age of 30 years. The genetic aetiology of EM is largely unknown in the majority of cases. We have undertaken a meta-analysis of genome-wide association studies (GWASs) in 3493 EM cases and 13 598 controls from 10 independent studies. No novel genetic variants were discovered, but the 17 variants previously associated with normal age at natural menopause as a quantitative trait (QT) were also associated with EM and primary ovarian insufficiency (POI). Thus, EM has a genetic aetiology which overlaps variation in normal age at menopause and is at least partly explained by the additive effects of the same polygenic variants. The combined effect of the common variants captured by the single nucleotide polymorphism arrays was estimated to account for ~30% of the variance in EM. The association between the combined 17 variants and the risk of EM was greater than the best validated non-genetic risk factor, smoking.

INTRODUCTION

Menopause represents a major hormonal change, characterized by a decline in oestrogen and progesterone levels and cessation of female reproductive function as the ovarian reserve is exhausted (1). It influences a woman's well-being and early menopause (EM) is associated with increased risk of age-related diseases including cardiovascular disease, osteoarthritis and osteoporosis, but reduced risk of breast cancer (2).

The average age at natural menopause in women of Northern European descent is 50 to 51 years (3,4). Early entry into menopause has implications for women's fertility. Fertility starts to decrease on average at about age 30 years and is considerably diminished after age 35. It is estimated that natural fecundity ceases at a mean age of 41 years, i.e. 10 years before menopause (5). In recent decades, the average age at which a woman gives birth to her first child has increased from around 25 up to 30 years of age (6). As a consequence, women who are at risk of EM and who delay childbearing until their 30's are more likely to have problems conceiving (2). This tendency has led to an increase in age-related infertility, subsequently increasing the utilization of assisted reproductive technologies (ARTs). Better understanding of the mechanisms that lead to EM, and even the ability to predict it, could greatly improve family planning and reduce the need for invasive and costly ART treatments (5,7).

Heritability estimates for age at natural menopause, from twin and family studies, range from 44–65%, suggesting a substantial genetic component to the trait (8–12). Initial genome-wide association studies (GWASs) identified 4 loci associated with variation in age at natural menopause in the normal range (40–60 years) (13,14) and more recent GWASs have added a further 13 loci, bringing the total to 17, including genes implicated in DNA repair and immune function (15). The effect size ranged from 8.7 weeks to nearly 1 year (50.5 weeks) per allele and the 17 single nucleotide polymorphisms (SNPs) together explained 2.5–4.1% of the population variation in natural menopausal age.

EM, defined as menopause occurring before 45 years of age, occurs in ~5–10% of women and primary ovarian insufficiency (POI) when menstruation ceases before 40 years, affects ~1% of women (3,16,17). Premature ovarian ageing may be the consequence of a precocious decline of the primordial follicle pool, which is established during fetal life,

leading to a loss of negative feedback from ovarian sex steroids and inhibins on the hypothalamic–pituitary axis. Oocyte quality decreases with increasing age and EM may reflect the damage accumulated during reproductive life, and/or age-related changes in granulosa cell–oocyte communication (18). EM may be caused by genetic defects (eg. Turner syndrome or *FMR1* premutations), autoimmunity or iatrogenic (as a consequence of surgery, chemotherapy or radiation) or might be the consequence of environmental factors. Unexplained EM also has a substantial genetic component (19). A woman whose mother had an EM has ~6-fold increased risk of having EM (8,20). However, in the majority of cases, the genes involved in EM are largely unknown and may be different from the genes regulating age at menopause in the normal range.

We have addressed this issue by conducting a GWAS comparing EM cases with controls who had menopause at ages 50–60 years, in the *ReproGen* consortium. We find considerable overlap between the genetic variation that contributes to normal menopause age and EM.

RESULTS

To identify common genetic variants associated with EM, we followed a two-stage, case–control approach. From the *ReproGen* consortium cohorts with GWAS data, we selected cases as women with age at menopause before 45 years ($N = 3493$) and controls as women with age at menopause between 50 and 60 years ($N = 13598$). Only cohorts with ≥ 100 cases were included, giving 10 independent studies (Supplementary Material, Table S1). Meta-analysis of this EM discovery dataset identified four independent signals with *P*-values stronger than the genome-wide significant threshold of $P < 5 \times 10^{-8}$ (Supplementary Material, Table S2). All the four signals had been identified in the *ReproGen* quantitative trait (QT) GWAS of normal menopause age (15). A further four SNPs were borderline significant for EM ($P < 5 \times 10^{-7}$, Supplementary Material, Table S2) and two of these had not been previously identified in the QT GWAS: rs1867631 in *SGIP1* at chromosome 1p31.3 and rs1473307 near *NYAP2* at chromosome 2q36.3. Both SNPs were carried forward for replication by *de novo* genotyping or *in silico* analyses in an additional sample of 3412 cases and 4928 controls, from four cohorts (Supplementary Material, Table S1). For both SNPs the association *P*-value increased when the replication data were combined with the EM discovery data (Supplementary Material, Table S3); thus, we found no evidence for novel genetic loci associated with EM.

To estimate the proportion of variance explained by all common variants captured on the SNP arrays in a polygenic model, we used genome-wide complex trait analysis (analysis tools available at: http://www.complextraitgenomics.com/software/gcta/). We estimated the variance explained in the WGHS cohort, one of the largest cohorts used in the meta-analysis ($N = 10302$). For menopause as a QT, the SNPs explain 21% of the variance ($P = 1 \times 10^{-11}$, se = 0.03) in a model taking residuals of menopause age with body mass index, smoking and population eigenvectors. Using the same approach and assuming a population prevalence of 5 or 10%, heritability of EM due to the SNP array genotypes was estimated to be 27 and 33%, respectively ($P = 0.006$, se = 0.11; $P = 0.006$, se = 0.13, respectively).

To identify associations at the gene level, where combinations of multiple SNPs may contribute in aggregate, we ran the Versatile Gene-Based Association Study' (VEGAS) test. Using our full discovery meta-analysis, VEGAS produced gene-level results for 17 580 genes. No genes passed our conservative Bonferroni correction at the 0.05 level ($P = 2.8 \times 10^{-6}$). There were 48 genes with $P < 0.001$ from the VEGAS analysis, we used GRAIL to identify any of these 48 genes which shared functional links with any gene within the 17 known menopause regions. Four genes reached a GRAIL $P < 0.05$; MCM6 (most similar to MCM8, top SNP rs2164210—$P = 7 \times 10^{-5}$), C6orf150 (similar to SYCP2L, top SNP rs311686—$P = 7 \times 10^{-4}$), CRHR1 (similar to UCN, top SNP rs4640231—$P = 2 \times 10^{-4}$), SLC25A13 (similar to POLG, top SNP rs2375044—$P = 2 \times 10^{-5}$). Pathway analysis with Magenta revealed no significant enrichment of biological pathways in EM.

The role of loci associated with variation in normal age at menopause in women with early menopause and POI

We next investigated the risk of EM for each of the 17 variants that were associated with normal variation in menopause age reported in the *ReproGen* QT GWAS. *In silico* data were available for 3840 individuals with EM (those with age at menopause 40–44 years were included in the previous QT GWAS[15], individuals with age at menopause <40 years have not been included previously). A further 1365 cases and 2475 controls from three studies not included in that QT GWAS were directly genotyped or had *in silico* data for the 17 SNPs. The odds ratios (ORs) for EM were in the same direction and of a similar magnitude in the discovery EM GWAS and in the meta-analysis of the three additional independent cohorts (Supplementary Material, Table S4). Combining both datasets, all 17 QT GWAS SNPs were nominally associated with EM (*P*-value <0.05) and were all directionally consistent with their effects on normal age at menopause (Table 1 and Supplementary Material, Table S4). The SNPs with the largest association with age at menopause in the normal range had the greatest OR for EM (Fig. 1).

In five of the studies (two from the discovery EM GWAS and three of the additional independent studies), there were more than 100 individuals with menopause before 40 years (Supplementary Material, Table S1). We tested the association of the 17 menopause SNPs in 1108 POI cases and 7727 controls who were not part of the sample for the QT GWAS. Despite limited power from the relatively small sample size, rs11668344 on chromosome 19 was significantly associated with POI in the meta-analysis [OR = 1.30 (CI 1.21–1.47), $P = 5.39 \times 10^{-8}$; after Bonferroni correction accounting for 17 tests]. Of the remaining 16 SNPs, all had an effect in the expected direction and eight were nominally associated with POI ($P < 0.05$) (Table 1). We also explored associations with EM and POI using dominant and recessive models for each the 17 menopause variants and found no evidence for any SNP acting in a non-additive fashion (Supplementary Material, Table S5).

Table 1. Effect of 17 SNPs, identified by the GWAS of normal menopause QT, in EM and POI cases versus controls

SNPID	Chr	Location (bp)	Effect allele	Effect allele frequency	Normal menopause QT GWAS			EM cases versus controls			POI cases versus controls		
					Effect (years)	SE	*P*-value	OR [95% CI]	*P*-value	Dir	OR [95% CI]	*P*-value	Dir
rs16991615	20	5896227	g	0.93	−0.948	0.052	1.4E − 73	1.55 [1.41–1.71]	5.8E − 20	− − − − −	1.17 [0.99–1.39]	0.07	− − − − +
rs11668344	19	60525476	g	0.36	−0.416	0.026	1.5E − 59	1.33 [1.27–1.4]	2.2E − 32	? − − − −	1.34 [1.21–1.47]	3.7E − 09	− ? − − −
rs2517388	8	38096889	t	0.83	−0.262	0.034	9.3E − 16	1.23 [1.15–1.32]	1.5E − 09	? − + + +	1.22 [1.05–1.41]	0.008	+ ? − + −
rs2277339	12	55432336	g	0.1	−0.38	0.042	2.5E − 19	1.19 [1.1–1.28]	5.8E − 06	? − − − −	1.21 [1.04–1.4]	0.01	− ? − − −
rs12294104	11	30339475	c	0.83	−0.225	0.033	1.5E − 11	1.18 [1.11–1.26]	2.2E − 07	? − − − −	1.21 [1.07–1.38]	0.004	− ? − − −
rs1046089	6	31710946	a	0.35	−0.213	0.026	1.6E − 16	1.16 [1.11–1.22]	9.2E − 10	? − − + +	1.16 [1.05–1.28]	0.003	+ ? − + +
rs12461110	19	61012475	a	0.36	−0.158	0.026	8.7E − 10	1.16 [1.1–1.21]	4.0E − 09	? + + + +	1.14 [1.03–1.26]	0.01	+ ? + + +
rs4246511	1	39152972	c	0.73	−0.24	0.029	9.0E − 17	1.14 [1.08–1.21]	1.2E − 06	? + − − −	1.08 [0.97–1.2]	0.15	− ? + − +
rs4886238	13	60011740	g	0.67	−0.17	0.026	9.5E − 11	1.14 [1.08–1.2]	2.5E − 07	? − − − −	1.2 [1.08–1.33]	0.0006	− ? − − −
rs10852344	16	11924420	t	0.58	−0.168	0.025	1.0E − 11	1.13 [1.08–1.19]	2.0E − 07	? − + + +	1.03 [0.93–1.13]	0.60	− ? − + +
rs10183486	2	1.72E+08	t	0.37	−0.196	0.026	2.2E − 12	1.13 [1.08–1.19]	3.6E − 07	? + + + +	1.12 [1.02–1.23]	0.02	+ ? + + −
rs2153157	6	11005474	g	0.51	−0.165	0.024	7.8E − 12	1.12 [1.07–1.18]	6.2E − 07	? − − − −	1.10 [1.0–1.21]	0.06	− ? − + −
rs2303369	2	27568920	t	0.39	−0.175	0.025	2.3E − 12	1.12 [1.07–1.18]	1.1E − 06	? − + + +	1.09 [0.99–1.2]	0.09	+ ? − + +
rs2307449	15	87664932	g	0.4	−0.184	0.025	3.6E − 13	1.12 [1.07–1.17]	3.0E − 06	? + − − −	1.18 [1.07–1.3]	0.0009	− ? + − −
rs365132	5	1.76E+08	g	0.51	−0.287	0.025	9.1E − 32	1.11 [1.06–1.16]	1.1E − 05	? + − − −	1.05 [0.96–1.16]	0.27	+ ? − − −
rs1635501	1	2.4E+08	c	0.48	−0.164	0.027	8.5E − 10	1.1 [1.05–1.16]	0.0002	? − − − −	1.04 [0.94–1.14]	0.45	− ? + − +
rs4693089	4	84592646	a	0.51	−0.228	0.025	2.4E − 19	1.09 [1.04–1.14]	0.0005	? − + + +	1.04 [0.94–1.14]	0.45	+ ? − + +

SNPs are ordered by OR for EM. Direction of effects for individual studies given in the following order: BGS, Colaus, EGCUT, NIDO, discovery for EM and Aric, BGS, Colaus, NIDO, WGHS for POI. ? indicates that a study did not contribute data for that SNP, either because not genotyped or failed QC.

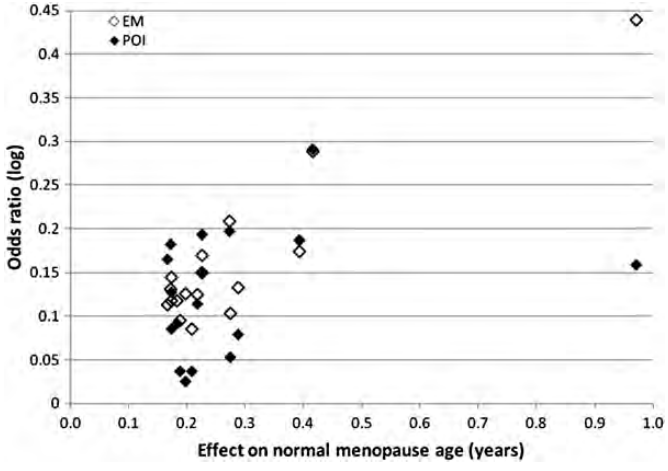

Figure 1. Effect on normal age of menopause as a QT plotted against the odds of EM (<45 years) or POI (<40 years) for each of 17 *ReproGen* age at menopause GWAS SNPs.

Observed versus expected estimates of the normal menopause range loci on early menopause risk

We estimated the association between the 17 QT normal menopause variants and the odds of having EM and POI, based on the associations with menopause age in the QT analysis (15) by comparing the expected with the observed odds (Supplementary Material, Table S6, Fig. S1). The expected odds were derived using published estimates for the incidence of POI and EM (1 and 5%, respectively) and using those to dichotomize a normal distribution of age at menopause. It was not possible to determine the actual incidence of EM and POI in participating cohorts because of the cross-sectional study design of most of the participating studies. We, therefore, conducted a sensitivity analysis taking cut-offs either side of 1 and 5% of the age at menopause distribution. The method assumes a normal distribution for menopause age, which may not be the case; thus, the results should be interpreted with caution. For the majority of SNPs, the effect on EM and POI was within the range expected from the QT study. However, there was evidence that one SNP varied significantly from expected. The allele associated with lower age at menopause at rs16991615 on chromosome 20 was significantly less strongly associated with POI than expected ($P = 1.04 \times 10^{-6}$). A significant difference between the observed and expected ORs for rs16991615 was seen at both 0.05 and 2.5% for the POI cases.

Testing the combined effect of the 17 age at menopause SNPs on EM risk

We sought to assess the combined association of the 17 QT age at menopause SNPs with the risk of EM, in two datasets independent of the QT and EM discovery samples (NIDO—691 cases, 1394 controls, and EGCUT—647 cases, 848 controls). The number of age at menopause lowering alleles carried per individual was calculated, and the distribution of these alleles in cases and controls is shown in Fig. 2. A per risk allele OR for EM of 1.13 ([CI 1.08–1.17],

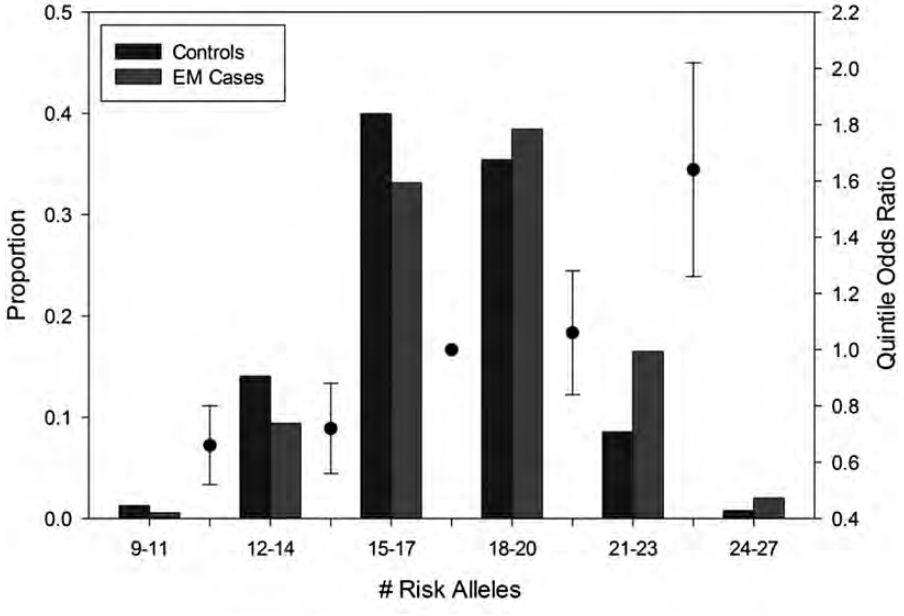

Figure 2. Distribution of the age at menopause-lowering allele score (quintiles) in women with EM and controls and ORs (95% CIs) for EM. Data shown are from the two replication cohorts combined. OR's are calculated relative to the median quintile.

$P = 7.75 \times 10^{-10}$) was observed in the NIDO cohort, which was similar to the estimate in EGCUT (OR = 1.14 ([CI 1.08–1.19], $P = 6 \times 10^{-8}$).

We divided the case–control samples into risk quintiles, based on the number of risk alleles they carried, weighted by the relative effect sizes of those alleles from the EM discovery + replication GWAS meta-analysis. The risk of EM associated with being in each quintile relative to the median quintile is shown in Figure 2. An OR of 2.47([CI 1.94–3.14], $P = 2.7 \times 10^{-13}$) for EM risk was observed when comparing the top 20%, with the most EM risk alleles, with the bottom 20%. This difference was higher when combined with smoking status. The smoking status alone (current versus former/never smokers) was associated with a doubling in risk for EM (OR 1.96 [CI 1.51–2.56], $P = 6 \times 10^{-7}$). Those women with the combination of the top 20% EM risk allele group plus current smoking had an OR of 3.38 ([CI 1.74–6.59], $P = 0.003$) higher risk of EM than those in the lowest 20% EM risk allele group who were former/never smokers.

We tested the ability of the 17 SNPs to discriminate EM cases from controls by calculating a receiver operating characteristic (ROC) area under the curve (AUC), using individual's weighted EM risk allele score and smoking status. Data from the NIDO and EGCUT cohorts gave highly concordant results, with an AUC of 0.60 for the 17 SNPs. This showed a significant improvement over smoking status alone (AUC = 0.55). Combining genetic and smoking risk factors gave an AUC of 0.63 (sensitivity = 35.4%, specificity = 81.3%).

Prior to the recent identification of 13 new variants associated with normal age at menopause, there were four loci reported, which were replicated in the more recent study, (13,14, 15). The AUC for the first four published loci associated with age at menopause was 0.55.

DISCUSSION

Shared aetiology of EM/POI and normal menopause

A recent GWAS has identified 17 loci associated with age at natural menopause in the normal range (40–60 years), explaining ~4% of the variation in menopause age (15). However, this GWAS excluded women who had menopause before 40 years (POI), a condition affecting ~1% of the female population. EM leads to short reproductive lifespan and is also associated with several harmful health outcomes including increased risk of cardiovascular diseases (21). Up to 30% of POI cases have an affected relative suggesting a substantial genetic burden in these women, but candidate gene studies have been unable to determine a genetic cause in the majority of cases (22). The definition of POI is arbitrarily based on the population distribution of menopause age, affected women representing the extreme 1% tail (~2.5 SDs from the mean), rather than distinct clinical characteristics. A small proportion of women with POI spontaneously conceive and thus, it is a heterogeneous condition. We hypothesized that very EM has distinct genetic aetiology compared with menopause age within the normal range, caused by either independent deleterious variants in the known age at menopause genes, or by variants at different loci, which have a larger effect on menopause age. In order to understand the genetic aetiology of menopause at the extreme of the age distribution, we performed a GWAS in women with menopause before 45 years of age. This ensured that we captured the full spectrum of ovarian insufficiency and gave us a large enough sample size to make it feasible to perform a GWAS; however, a clinical diagnosis of POI was not recorded in any of our studies. It is also possible that rare variants, poorly captured by the SNP chips, are

more prevalent in individuals at the extreme of the menopause distribution, but these cannot be assessed by our current GWAS approach.

In our sample of ~3500 cases, we found no evidence for novel genetic associations with EM that reached genome-wide significance thresholds. We did, however, find genome-wide significant associations with four loci previously identified in the normal menopause age QT GWAS (15). Our study was, therefore, powered to detect associations with ORs of 1.17–1.59, depending on the minor allele frequency. There was, however, considerable overlap between the samples used in the normal menopause QT GWAS and the current EM GWAS, which may have increased our chances of detecting such signals due to the winner's curse phenomenon. Despite following up two borderline signals in replication cohorts, we were unable to detect any new variants for EM. With our sample size of ~3500 cases and ~13500 controls, we had ~80% power to detect ORs of 1.2 with 30% minor allele frequency SNPs. We estimated that all common variants captured by our SNP arrays account for ~20% of the variance in natural age at menopause, thus a significant proportion of the genetic component to the trait is likely to be due to rarer or complex variants not captured by the SNP arrays. We did not include non-genetic variables in our association analyses and it is possible that that there are genetic interactions with known environmental risk factors for EM, e.g. smoking. EM is a heterogeneous trait and it is possible that clinical classification of sub-types would increase our power to detect genetic factors associated with the condition. We found no evidence for a distinct genetic aetiology in EM cases. If there were a genetically distinct group of individuals at the extreme end of the distribution, by choosing a relatively broad extreme category, representing ~10% of the menopause age distribution, there may be too much overlap with the normal range of menopause age, thus masking any differences. We did not have sufficient number of cases with menopause at ages <40 years to perform a GWAS on this category, but we were able to investigate the role of known QT menopause signals in this group of women representing the extreme ~1% tail of the distribution.

We tested the 17 variants identified in the *ReproGen* QT GWAS of normal menopause, in cases of EM and POI. For all 17 variants, the allele that was associated with younger menopause age was also associated with increased risk of EM and POI. Only four SNPs reached genome-wide levels of significance for EM, but all 17 for EM and 3 for POI were below the Bonferroni-corrected *P*-value of <0.0015, assuming 34 independent tests. There was some evidence that the association with POI was weaker than expected for the SNP with largest effect on normal menopause, but this requires a formal confirmation. Stolk *et al*. determined common pathways for the variants associated with age at menopause in the normal range and highlighted DNA repair/replication, hormonal regulation and immune function as key pathways (15). However, there was no evidence that genes from a particular biological pathway were more important in EM or POI. Our data support the hypothesis that EM and POI represent the tail of the menopause distribution and thus have overlapping polygenic aetiology, with individuals carrying more age at menopause-lowering variants having increased risk of EM and POI.

New SNPs increase discriminative power over previous four SNPs

By combining the effect of the 17 variants in a weighted allele score, we demonstrated a larger effect on EM risk than the best-known non-genetic risk factor, smoking (23,24). However, the increased OR for EM for carriers of the most risk alleles compared with the fewest was 2.47, which is still significantly lower than the OR associated with having a mother with EM, which is about six in most reported studies (8, 20). However, the current 17 variants only explain <5% of the variance in menopause age and thus as more genetic variants are discovered the discriminative power is likely to increase. We observed a significant improvement in discriminative power for EM when the 13 most recently described variants were added to the first four previously published signals (25).

In conclusion, while much of the genetic aetiology of EM is yet to be discovered, we have demonstrated that the combined effect of multiple genes involved in determining the age at normal menopause plays a role. This of course does not exclude the possibility that rarer variants with larger effects are also involved, as these may not have been well captured by the SNP arrays used in the GWAS. Genetic markers of ovarian ageing are present throughout life and thus may be superior to current best predictors, e.g. AMH, inhibin B and FSH levels, which are only reliable indicators up to about 5–10 years prior to menopause. As more genetic components of this trait are discovered, we will be able to include additional genetic data in predictive models for menopause age, giving women information about potential reproductive lifespan and enabling them to make informed reproductive choices.

MATERIALS AND METHODS

GWAS for EM

EM cases were selected from studies which contributed to the ReproGen GWAS of normal menopause (15). EM cases were defined as women who had menopause before 45 years of age, and controls were women with age at menopause from 50 to 60 years. Age at menopause was assessed through questionnaires, as detailed in Supplementary Material, Table S1. Women of self-reported non-European ancestry were excluded, as were women with menopause due to hysterectomy and/or bilateral ovariectomy, or chemotherapy/irradiation, if validated by medical records, and women using hormone replacement therapy before menopause. Other variables associated with age at menopause, e.g. smoking, were not excluded. We only included studies which had >100 EM cases. There were 10 studies included in the meta-analysis, from the *ReproGen* consortium, with a total of 3493 cases and 13598 controls (Supplementary Material, Table S7). All samples were of European ancestry. All cohorts performed SNP array genotyping followed by imputation to HapMapII, to generate a common set of ~2.5 million autosomal SNPs with a minor allele frequency of >1% (Supplementary Material, Table S7). Each individual study performed their own quality control for imputation quality, deviation from Hardy–Weinberg equilibrium, SNP call rate and lambda GC correction (Supplementary Material,

Table S7). Meta-analysis was performed using inverse variance weighting in METAL with genomic control correction. Heterogeneity between studies was assessed using Cochrane's Q statistic test in METAL. Replication was carried out in four independent cohorts, including 3412 cases and 4928 controls (Supplementary Material, Table S1). *In silico* genome-wide SNP data were available from COLAUS, while the other three studies performed *de-novo* genotyping by Taqman SNP assay.

Analysis of 17 QT menopause SNPs in EM and POI

Association statistics for the 17 SNPs previously identified to influence normal age at menopause were extracted from the EM data. This included the 10 EM discovery GWAS cohorts and 3 of the 4 replication cohorts, giving a combined sample size of 5205 EM cases and 16926 controls. Four studies had genotype data for the 17 SNPs on more than 100 cases with POI (ARIC, WGHS, COLAUS, NIDO), giving a total of 1108 POI cases and 7727 controls. BGS had data on a subset of SNPs in 2121 EM and 260 POI cases and were added to the meta-analysis for those SNPs. Meta-analyses were carried in METAL.

Pathway analysis

We implemented two methods to assess whether particular gene pathways were enriched in our EM GWAS data: (i) We used a GSEA-based approach with MAGENTA (26), where each gene in the genome is mapped to a single index SNP with the lowest *P*-value within a 110 kb upstream, 40 kb downstream window. This *P*-value, representing a gene score, is then corrected for confounding factors such as gene size, SNP density and LD-related properties in a regression model. Each mapped gene in the genome is then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, is calculated. This observed statistic is then compared with 1 000 000 randomly permuted pathways of identical size. This generates an empirical GSEA *P*-value for each pathway. Significance was determined when an individual pathway reached a false discovery rate of <0.05 in either analysis. In total, 2580 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with EM status.

(ii) We searched for evidence of multiple-SNP signal enrichment at the gene level using the (VEGAS algorithm (http://gump.qimr.edu.au/VEGAS/). This method is described in detail by Liu *et al*. (http://www.cell.com/AJHG/retrieve/pii/S0002929710003125), but briefly, test statistics across a UCSC gene region ($\pm$ 50 kb) are collapsed into a single statistic representing the gene. The statistic is adjusted for confounding factors such as gene size, LD and SNP density. The analysis was run on the full discovery meta-analysis summary statistics, using the default settings of the online tool. Genes reaching a *P*-value of <0.001 were analysed by GRAIL (http://www.broadinstitute.org/mpg/grail/) for literature-based homology to the genes within the 17 known menopause regions. A nominal GRAIL similarity $P < 0.05$ was chosen to highlight genes of interest.

Expected versus observed OR

We estimated the expected OR for both EM ($<$46 years) and POI ($<$40 years) for each of the 17 variants, based on the coefficient estimate from the QT effect size in the normal menopause age GWAS[15]. We calculated the expected ORs for both the point estimate QT coefficient and the upper and lower 95% CIs, by using the 'Case–Control for threshold-selected QTs' analysis on the Genetic Power Calculator website (http://pngu.mgh.harvard.edu/~purcell/gpc/). Using the proportion of variation explained by an SNP and the allele frequency, the program generates expected allele frequencies in cases and controls, where cases and controls are defined by standard deviation thresholds. We tested three standard deviation thresholds for EM (equivalent to the 2.5, 5 and 10% tail of the menopause distribution) and three for POI (0.05, 1 and 2.5% tails). We then tested for heterogeneity by a *Z*-test of ln(observed_OR)-ln(expected_OR)/sqrt[se_squared(observed_OR) + se_squared(expected_OR)]. The method assumes that menopause age is normally distributed, but this was not tested in individual studies.

Risk prediction

Two datasets independent of the *ReproGen* discovery studies were used to assess the predictive impact of the 17 menopause SNPs—NIDO and EGCUT. First, the number of EM risk alleles carried per individual was calculated using the 'score' command in PLINK. Any individuals with less than half of the genotyped SNPs missing were excluded from analysis. The same command then creates a genotypic score for each individual, imputing any missing genotypes based on the sample allele frequency and gives a weighting based on SNP effect sizes from the combined EM + replication meta-analysis (Table 1). This score was then used to calculate the ROC curve statistics using the 'lroc' command in Stata. The results were repeated using a raw risk allele sum score from only individuals with all genotypes present. Total sample sizes available with a genotypic risk score and phenotype were 691 cases and 1394 controls (NIDO) and 647 cases and 848 controls (EGCUT). Smoking status was available in the EGCUT samples, indicating 'current', 'former' or 'never' smoking based on questionnaire data. The individuals in these datasets were additionally partitioned into quintiles based on their genotypic risk score. ORs were calculated for the risk of EM based on the quintile membership, relative to the median (third) quintile. The two cohorts were combined in this analysis, with adjustment for cohort as an additional dichotomous trait in the logistic regression model.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are grateful to the study participants and staff from all cohorts involved in this study. We would like to acknowledge *ReproGen* consortium members for initial analyses which were not included in the final draft of the paper and the CHARGE consortium for organizing conference calls and face-to-face meetings of *ReproGen*. A full list of members of *ReproGen* members is available as supplementary information.

Conflict of Interest statement. None declared.

FUNDING

J.R.B.P. is funded by the Wellcome Trust as a Sir Henry Wellcome Postdoctoral Research Fellow (092447/Z/10/Z). Funding details for individual studies is provided in supplementary information. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

REFERENCES

1. Santoro, N. (2005) The menopausal transition. *Am. J. Med.*, **118**(Suppl 12B), 8–13.
2. Broekmans, F.J., Soules, M.R. and Fauser, B.C. (2009) Ovarian aging: mechanisms and clinical consequences. *Endocr. Rev.*, **30**, 465–493.
3. Coulam, C.B., Adamson, S.C. and Annegers, J.F. (1986) Incidence of premature ovarian failure. *Obstet. Gynecol.*, **67**, 604–606.
4. Luoto, R., Kaprio, J. and Uutela, A. (1994) Age at natural menopause and sociodemographic status in Finland. *Am. J. Epidemiol.*, **139**, 64–76.
5. Lambalk, C.B., van Disseldorp, J., de Koning, C.H. and Broekmans, F.J. (2009) Testing ovarian reserve to predict age at menopause. *Maturitas*, **63**, 280–291.
6. Breart, G. (1997) Delayed childbearing. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **75**, 71–73.
7. Henderson, K.D., Bernstein, L., Henderson, B., Kolonel, L. and Pike, M.C. (2008) Predictors of the timing of natural menopause in the Multiethnic Cohort Study. *Am. J. Epidemiol.*, **167**, 1287–1294.
8. Morris, D.H., Jones, M.E., Schoemaker, M.J., Ashworth, A. and Swerdlow, A.J. (2012) Familial concordance for age at natural menopause: results from the Breakthrough Generations Study. *Menopause*, **18**, 956–961.
9. Murabito, J.M., Yang, Q., Fox, C., Wilson, P.W. and Cupples, L.A. (2005) Heritability of age at natural menopause in the Framingham Heart Study. *J. Clin. Endocrinol. Metab.*, **90**, 3427–3430.
10. Snieder, H., MacGregor, A.J. and Spector, T.D. (1998) 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.
11. van Asselt, K.M., Kok, H.S., Pearson, P.L., Dubas, J.S., Peeters, P.H., Te Velde, E.R. and van Noord, P.A. (2004) Heritability of menopausal age in mothers and daughters. *Fertil. Steril.*, **82**, 1348–1351.
12. Vink, J. and Boomsma, D. (2005) Modeling age at menopause. *Fertil. Steril.*, **83**, 1068.
13. He, C., Kraft, P., Chen, C., Buring, J.E., Pare, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D.J. and Chasman, D.I. (2009) Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.*, **41**, 724–728.
14. Stolk, L., Zhai, G., van Meurs, J.B.J., Verbiest, M.M.P.J., Visser, J.A., Estrada, K., Rivadeneira, F., Williams, F.M., Cherkas, L., Deloukas, P. *et al.* (2009) Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat. Genet.*, **41**, 645–647.
15. Stolk, L., Perry, J.R., Chasman, D.I., He, C., Mangino, M., Sulem, P., Barbalic, M., Broer, L., Byrne, E.M., Ernst, F. *et al.* (2012) Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat. Genet*, **44**, 260–268.
16. Davis, C.J., Davison, R.M. and Conway, G.S. (1998) Genetic basis of premature ovarian failure. *Hum. Fertil. (Camb)*, **1**, 20–22.
17. Shelling, A.N. Premature ovarian failure. *Reproduction*, **140**, 633–641.
18. te Velde, E.R. and Pearson, P.L. (2002) The variability of female reproductive ageing. *Hum Reprod Update*, **8**, 141–154.
19. Cordts, E., Christofolini, D., dos Santos, A., Bianco, B. and Barbosa, C. Genetic aspects of premature ovarian failure: a literature review. *Arch. Gynecol. Obstet.*, **283**, 635–643.
20. Torgerson, D.J., Thomas, R.E. and Reid, D.M. (1997) Mothers and daughters menopausal ages: is there a link? *Eur. J. Obstet. Gynecol. Reprod. Biol.*, **74**, 63–66.
21. Kok, H.S., van Asselt, K.M., van der Schouw, Y.T., van der Tweel, I., Peeters, P.H., Wilson, P.W., Pearson, P.L. and Grobbee, D.E. (2006) Heart disease risk determines menopausal age rather than the reverse. *J. Am. Coll. Cardiol.*, **47**, 1976–1983.
22. Goswami, D. and Conway, G.S. (2005) Premature ovarian failure. *Hum. Reprod. Update*, **11**, 391–410.
23. Augood, C., Duckitt, K. and Templeton, A.A. (1998) Smoking and female infertility: a systematic review and meta-analysis. *Hum. Reprod.*, **13**, 1532–1539.
24. Cramer, D.W., Harlow, B.L., Xu, H., Fraer, C. and Barbieri, R. (1995) Cross-sectional and case-controlled analyses of the association between smoking and early menopause. *Maturitas*, **22**, 79–87.
25. Murray, A., Bennett, C.E., Perry, J.R., Weedon, M.N., Consortium, R., Jacobs, P.A., Morris, D.H., Orr, N., Schoemaker, M.J., Jones, M. *et al.* (2011) Common genetic variants are significant risk factors for early menopause: results from the Breakthrough Generations Study. *Hum. Mol. Genet.*, **20**, 186–192.
26. Segre, A.V., Consortium, D., investigators, M., Groop, L., Mootha, V.K., Daly, M.J. and Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*, **6**, e1001058.