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

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1093/hmg/dds551</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:10642218">http://nrs.harvard.edu/urn-3:HUL.InstRepos:10642218</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
A genome-wide association study of early menopause and the combined impact of identified variants


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

#These authors contributed equally to this work.

*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.

© The Author 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/ 3.0/), which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permission@oup.com
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 \( \sim 30\% \) 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 30s 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 \( \sim 5–10\% \) of women and primary ovarian insufficiency (POI) when menstruation ceases before 40 years, affects \( \sim 1\% \) 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), autoimmune 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 \( \text{de novo} \) genotyping or \textit{in silico} analyses in an additional sample of 3412 cases and 4928 controls, from four cohorts (Supplementary Material, Table S1). Both SNPs 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 (\( N = 10 \; 302 \)). 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 (\( P = 0.006, \text{se} = 0.11; P = 0.006, \text{se} = 0.13, \text{respectively} \)).

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—\( P = 2 \times 10^{-3} \)). 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. \textit{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), 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 \textit{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

<table>
<thead>
<tr>
<th>SNPID</th>
<th>Chr</th>
<th>Location (bp)</th>
<th>Effect frequency</th>
<th>Effect (years)</th>
<th>SE</th>
<th>P \text{-value}</th>
<th>Dir OR \text{ [95% CI]}</th>
<th>P \text{-value}</th>
<th>Dir OR \text{ [95% CI]}</th>
<th>P \text{-value}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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. ?

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.

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.

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.

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.

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.

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.

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.

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.
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/replcation, 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/S000292710003125), but briefly, test statistics across a UCSC gene region (+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 GWAS15. 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(expected OR)] of ln(1 + 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 type 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.
REFERENCES