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, 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
7© The Author 2013. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Estonian Genome Center, University of Tartu, Tartu, Estonia

Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA 02215

Harvard Medical School, Boston MA 02115

Division of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC

Department of Public Health, Indiana University School of Medicine, Indianapolis, Indiana, USA

Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, US

Icelandic Heart Association, Kopavogur, Iceland

University of Iceland, Reykjavik, Iceland

Dept Internal Medicine, Erasmus MC Rotterdam, the Netherlands

Netherlands Consortium of Healthy Aging, Rotterdam, the Netherlands

Decode genetics, Sturlugata 8, Reykjavik 105, Iceland

Department of Biostatistics, Boston University School of Public Health, USA

Department of Biostatistics, University of Washington, Seattle, WA

Breakthrough Research Centre, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB

Harvard School of Public Health, Boston MA 02115

Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA

Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC

Division of Endocrinology, Diabetes & Metabolism Cedars-Sinai Medical Center

National Institutes on Aging, NIH, Bethesda, Maryland, USA

Dept of Epidemiology, Erasmus MC Rotterdam, the Netherlands

Section of Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG

Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital, and Harvard Medical School, Boston, Massachusetts, USA
Division of Reproductive Medicine, Erasmus MC Rotterdam, the Netherlands

Department of Medicine, University of Washington

Division of Genetics and Cell Biology: San Raffaele Scientific Institute, Milano

Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA

Group Health Research Institute, Group Health Cooperative, Seattle, WA

Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia

Competence Centre on Reproductive Medicine and Biology, Tartu, Estonia

Division of Internal Medicine, Lausanne University Hospital, CHUV, Lausanne, Switzerland

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN

Queensland Institute of Medical Research, Brisbane, Queensland, Australia

National Heart, Lung and Blood Institute’s Framingham Heart Study, Framingham, USA
Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02118 USA

Departments of Epidemiology and Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK

Department of Paediatrics, University of Cambridge, Cambridge, UK

Saw Swee Hock School of Public Health and Yong Loo Lin School of Medicine, National University of Singapore

Institute of Molecular Genetics, 27100 Pavia, Italy

Corresponding author: Anna Murray, Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, UK EX1 2LU Email: anna.murray@pms.ac.uk Tel: +44 (0) 1392 722976
Abstract

Early menopause 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 early menopause is largely unknown in the majority of cases. We have undertaken a meta-analysis of genome-wide association studies in 3493 early menopause cases and 13598 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 were also associated with early menopause and POI (Primary Ovarian Insufficiency). Thus early menopause 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 SNP arrays was estimated to account for approximately 30% of the variance in EM. The association between the combined 17 variants and the risk of early menopause 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 ovarian reserve is exhausted(1). It influences a woman’s well-being and early menopause 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 50-51 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 early menopause 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 (ART). Better understanding of the mechanisms that lead to early menopause, 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 to 65%, suggesting a substantial genetic component to the trait(8-12). Initial genome-wide association studies (GWAS) identified 4 loci associated with variation in age at natural menopause in the normal range (40-60 years)(13, 14) and more recent GWAS 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 one year (50.5 weeks) per allele and the 17 SNPs together explained 2.5-4.1% of the population variation in natural menopausal age.
Early menopause (EM), defined as menopause occurring before 45 years of age, occurs in about 5-10% of women and primary ovarian insufficiency (POI) when menstruation ceases before 40 years, affects about 1% of women(3, 16, 17). Premature ovarian aging may be the consequence of a precocious decline of the primordial follicle pool, which is established during fetal life, leading to 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). Early menopause may be caused by genetic defects (eg. Turner syndrome or FMR1 premutations), autoimmunity or iatrogenic (as the consequence of surgery, chemotherapy or radiation ) or might be the consequence environmental factors. Unexplained early menopause also has a substantial genetic component(19). A woman whose mother had an early menopause has approximately 6 fold increased risk of having early menopause(8, 20). However, in the majority of cases, the genes involved in early menopause 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 early menopause 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 early menopause.

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 50-60 years (N=13598). Only cohorts with 100 or more cases were included, giving ten independent studies (suppl Table 1). Meta-analysis of this EM discovery data set identified four independent signals with p-values stronger than the genome-wide significant threshold of p<5x10^-8 (suppl Table 2). All 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<5x10^{-7}, suppl Table 2) and two of these had not been previously identified in the quantitative trait GWAS: rs1867631 in SGIPI 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 4 cohorts (suppl Table 1). For both SNPs the association p value increased when the replication data were combined with the EM discovery data (suppl Table 3); thus we found no evidence for novel genetic loci associated with early menopause.

To estimate the proportion of variance explained by all common variants captured on the SNP arrays in a polygenic model, we used GCTA (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 quantitative trait the SNPs explain 21% of the variance (p=1x10^{-11}, se=0.03) in a model taking residuals of menopause age with BMI, smoking and population eigenvectors. Using the same approach and assuming a population prevalence of 5% or 10%, heritability of early menopause due to the SNP array genotypes was estimated as 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 = 2.8x10^{-6}). Using a VEGAS P< 0.001, we used GRAIL to identify any genes which shared text-based gene homology to 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=7x10^{-5}), C6orf150 (similar to SYCP2L, top SNP rs311686 - P = 7x10^{-4}), CRHR1 (similar to UCN, top SNP rs4640231 - P = 2x10^{-4}), SLC25A13 (similar to POLG, top SNP rs2375044 - P = 2x10^{-5}). Pathway analysis with Magenta revealed no significant enrichment of biological pathways in early menopause.
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 early menopause 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, individuals with age at menopause <40 years have not been included previously). A further 1365 cases and 2475 controls from 3 studies not included in that QT GWAS, were directly genotyped or had *in silico* data for the 17 SNPs. The odds ratios for EM were in the same direction and of similar magnitude in the discovery EM GWAS and in the meta-analysis of the 3 additional independent cohorts (*suppl table 4*). 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 suppl table 4*). The SNPs with the largest association with age at menopause in the normal range had the greatest odds ratio for EM (*Figure 1*).

In five of the studies (2 from the discovery EM GWAS and 3 of the additional independent studies) there were more than 100 individuals with menopause before 40 years (*suppl Table 1*). 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, [1.21-1.47], p=5.39 x10^-8; after Bonferroni correction accounting for 17 tests). Of the remaining 16 SNPs, all had an effect in the expected direction and 8 were nominally associated with POI (p-value <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 (*suppl table 5*).
Observed vs 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 to the observed odds (suppl table 6, Figure 1). The expected odds was derived using published estimates for the incidence of POI and EM (1% and 5% respectively) and using those to dichotomise a normal distribution of age at menopause. It was not possible to determine 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 early menopause 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 x10^{-6}). A significant difference between observed and expected odds ratios 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 figure 2. A per risk allele OR for EM of 1.13 ([1.08-1.17], P=7.75x10^{-10}) was observed in the NIDO cohort, which was similar to the estimate in EGCUT (OR = 1.14 ([1.08-1.19], P=6x10^{-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([1.94-3.14], P=2.7x10^{-13}) for EM risk was observed when comparing the top 20%, with the most EM risk alleles, to the bottom 20%. This difference was higher when combined with smoking status. Smoking status alone (current vs former/never smokers) was associated with a doubling in risk for EM (OR 1.96 [1.51-2.56], P=6x10^{-7}). Those women with the combination of the top 20% EM risk allele group plus current smoking had an OR 3.38 ([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 was a significant improvement on 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 study13,(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**

Recent GWAS has identified 17 loci associated with age at natural menopause in the normal range (40-60 years), explaining about 4% of the variation in menopause age(15). However this GWAS excluded women who had menopause before 40 years (POI), a condition affecting about 1% of the female population. Early menopause 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 (~2.5SDs 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 early menopause has distinct genetic aetiology compared to 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. This ensured 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 can not be assessed by our current GWAS approach.

In our sample of about 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 odds ratios of 1.2 with 30% MAF SNPs. We estimated that all common variants captured by our SNP arrays account for about 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 there are genetic interactions with known environmental risk factors for early menopause, eg. smoking. Early menopause is a heterogeneous trait and it 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 about 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 numbers of cases with menopause at ages less than 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 three for POI were below the Bonferroni corrected p value of $p<0.0015$, 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 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 4 SNPs.** By combining the effect of the 17 variants in a weighted allele score we demonstrated a larger effect on early menopause risk than the best known non-genetic risk factor, smoking(23, 24). However the increased odds ratio for early menopause for carriers of the most risk alleles compared to the fewest was 2.47, which is still significantly lower than the odds ratio associated with having a mother with early menopause, which is about 6 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 early menopause is yet to be discovered, we have demonstrated that the combined effect of multiple genes involved in determining age at normal menopause play 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 GWAS. Genetic markers of ovarian ageing are present throughout life and thus may be superior to current best predictors, eg. AMH, inhibin B and FSH levels, which are only reliable indicators up to about 10-5 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

Early menopause 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 old, and controls were women with age at menopause from 50-60 years. Age at menopause was assessed through questionnaires, as detailed in supplementary table 1. 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 HRT before menopause. Other variables associated with age at menopause, eg. 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 (suppl table 7). 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 minor allele frequency >1% (suppl table 7). Each individual study performed their own quality control for imputation quality, deviation from Hardy Weinberg equilibrium, SNP call rate and lambda GC correction (suppl table 7). 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 (suppl table 1). In silico genome-wide SNP data were available from COLAUS, while the three other 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: 1. 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 to 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 < 0.05 in either analysis. In total, 2,580 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with early menopause status.

2. We searched for evidence of multiple-SNP signal enrichment at the gene level using the ‘Versatile Gene-Based Association Study’ (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 (+/- 50kb) 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 vs observed OR**

We estimated the expected odds ratio for both Early Menopause (<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\textsuperscript{15}. We calculated the expected odds ratios for both the point estimate QT coefficient and the upper and lower 95% CI intervals, 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 a 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 Early Menopause (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 \( \ln(\text{observed OR}) - \ln(\text{expected OR})/\)
\[\sqrt{\text{se}^2_{\text{observed OR}} + \text{se}^2_{\text{expected OR}}}].\] 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. Firstly, the number of EM risk alleles carried per individual was calculated using the ‘score’ command in PLINK. Any individuals with less than half 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 Receiver operating characteristic (ROC) curve statistics using the ‘lroc’ command in Stata. Results were repeated using a raw risk allele sum score from only individuals with all genotypes present. Total sample sizes available with 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. Odds Ratios were calculated for the risk of EM based on quintile membership, relative to the median (3rd) quintile. The two cohorts were combined in this analysis, with adjustment for cohort as an additional dichotomous trait in the logistic regression model.

**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 organising conference calls and face-to-face meetings of *ReproGen*. A full list of members of *ReproGen* members is available in supplementary information. J.R.B.P is funded by the Wellcome Trust as a Sir Henry Wellcome Postdoctoral Research Fellow (092447/Z/10/Z).
References


Table and figure legends

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

**Figure 2.** Distribution of the age at menopause-lowering allele score (quintiles) in women with early menopause (EM) and controls and odds ratios (95% confidence intervals) for EM. Data are from the two replication cohorts combined. OR’s are calculated relative to the median quintile.

**Table 1.** Effect of 17 SNPs, identified by GWAS of normal menopause QT, in EM and POI cases versus controls. SNPs are ordered by OR for EM. Direction of effects for individual studies given in 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.
<table>
<thead>
<tr>
<th>SNPID</th>
<th>Chr</th>
<th>Location (bp)</th>
<th>allele</th>
<th>effect allele frequency</th>
<th>Normal Menopause QT GWAS</th>
<th>Early Menopause Cases vs Controls</th>
<th>POI Cases vs Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect (years)</td>
<td>SE</td>
<td>P-value</td>
</tr>
<tr>
<td>rs16991615</td>
<td>20</td>
<td>5896227</td>
<td>g</td>
<td>0.93</td>
<td>-0.948</td>
<td>0.052</td>
<td>1.4E-73</td>
</tr>
<tr>
<td>rs11668344</td>
<td>19</td>
<td>60525476</td>
<td>g</td>
<td>0.36</td>
<td>-0.416</td>
<td>0.026</td>
<td>1.5E-59</td>
</tr>
<tr>
<td>rs2517388</td>
<td>8</td>
<td>38096889</td>
<td>t</td>
<td>0.83</td>
<td>-0.262</td>
<td>0.034</td>
<td>9.3E-16</td>
</tr>
<tr>
<td>rs2277339</td>
<td>12</td>
<td>55432336</td>
<td>g</td>
<td>0.1</td>
<td>-0.38</td>
<td>0.042</td>
<td>2.5E-19</td>
</tr>
<tr>
<td>rs12294104</td>
<td>11</td>
<td>30339475</td>
<td>c</td>
<td>0.83</td>
<td>-0.225</td>
<td>0.033</td>
<td>1.5E-11</td>
</tr>
<tr>
<td>rs1046089</td>
<td>6</td>
<td>31710946</td>
<td>a</td>
<td>0.35</td>
<td>-0.213</td>
<td>0.026</td>
<td>1.6E-16</td>
</tr>
<tr>
<td>rs12461110</td>
<td>19</td>
<td>61012475</td>
<td>a</td>
<td>0.36</td>
<td>-0.158</td>
<td>0.026</td>
<td>8.7E-10</td>
</tr>
<tr>
<td>rs4246511</td>
<td>1</td>
<td>39152972</td>
<td>c</td>
<td>0.73</td>
<td>-0.24</td>
<td>0.029</td>
<td>9.0E-17</td>
</tr>
<tr>
<td>rs4886238</td>
<td>13</td>
<td>60011740</td>
<td>g</td>
<td>0.67</td>
<td>-0.17</td>
<td>0.026</td>
<td>9.5E-11</td>
</tr>
<tr>
<td>rs10852344</td>
<td>16</td>
<td>11924420</td>
<td>t</td>
<td>0.58</td>
<td>-0.168</td>
<td>0.025</td>
<td>1.0E-11</td>
</tr>
<tr>
<td>rs10183486</td>
<td>2</td>
<td>1.7E+08</td>
<td>t</td>
<td>0.37</td>
<td>-0.196</td>
<td>0.026</td>
<td>2.2E-12</td>
</tr>
<tr>
<td>rs2153157</td>
<td>6</td>
<td>11005474</td>
<td>g</td>
<td>0.51</td>
<td>-0.165</td>
<td>0.024</td>
<td>7.8E-12</td>
</tr>
<tr>
<td>rs2303369</td>
<td>2</td>
<td>27568920</td>
<td>t</td>
<td>0.39</td>
<td>-0.175</td>
<td>0.025</td>
<td>2.3E-12</td>
</tr>
<tr>
<td>rs2307449</td>
<td>15</td>
<td>87664932</td>
<td>g</td>
<td>0.4</td>
<td>-0.184</td>
<td>0.025</td>
<td>3.6E-13</td>
</tr>
<tr>
<td>rs365132</td>
<td>5</td>
<td>1.7E+08</td>
<td>g</td>
<td>0.51</td>
<td>-0.287</td>
<td>0.025</td>
<td>9.1E-32</td>
</tr>
<tr>
<td>rs1635501</td>
<td>1</td>
<td>2.4E+08</td>
<td>c</td>
<td>0.48</td>
<td>-0.164</td>
<td>0.027</td>
<td>8.5E-10</td>
</tr>
<tr>
<td>rs4693089</td>
<td>4</td>
<td>84592646</td>
<td>a</td>
<td>0.51</td>
<td>-0.228</td>
<td>0.025</td>
<td>2.4E-19</td>
</tr>
</tbody>
</table>
Figure 1.
Figure 2.