Common genetic variants are significant risk factors for early menopause: results from the Breakthrough Generations Study

Anna Murray1,∗, Claire E. Bennett1, John R.B. Perry1, Michael N. Weedon1, ReproGen Consortium, Patricia A. Jacobs2, Danielle H. Morris3, Nicholas Orr4, Minouk J. Schoemaker3, Michael Jones3, Alan Ashworth4 and Anthony J. Swerdlow3

1Peninsula Medical School, University of Exeter, St Lukes, Exeter EX1 2LU, UK, 2Wessex Regional Genetics, Salisbury District Hospital, Salisbury SP2 8BJ, UK, 3Section of Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK and 4Breakthrough Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Received May 24, 2010; Revised and Accepted September 20, 2010

INTRODUCTION

Earlier menopause is associated with a decreased risk of breast cancer, but an increased risk of osteoporosis and cardiovascular disease (1). There is also a significant impact on fertility associated with early menopause (EM), which is particularly relevant to current populations where delaying childbearing has become more prevalent. The number of births per 1000 British women is now greater for those in their early 30s than it is for women in their early 20s (2,3). Fertility decreases long before the onset of menopause, beginning on average at about the age of 30 years. It is estimated that natural fecundity ceases at a mean age of 41 years, i.e. 10 years before menopause, and therefore women who are destined to have an EM and who delay childbearing until their 30s are more likely to have problems conceiving (4).

Natural, non-surgical, menopause occurs at a mean age of 51 years in Caucasian populations, with a roughly normal distribution between 40 and 60 years, but a tail below 40 years (5,6). Menopause before the age of 40, or premature ovarian failure (POF), occurs in 1% of the population (6). Menopause before 45 years occurs in ~5% of women and is often termed ‘EM’. Menopause is initiated by a fall in the number of oocytes in the ovary below a threshold level of about 1000

*To whom correspondence should be addressed. Tel: +44 1392262976; Fax: +44 1392262926; Email: anna.murray@pms.ac.uk

© The Author 2010. 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/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
However, loss of oocytes occurs throughout female life with maximal numbers present before birth: approximately 6 million oocytes are present at 6 months gestation. By puberty, the number has decreased to ~400,000. Only a small proportion of oocytes are lost through ovulation, and the majority of the reduction is by atresia. The rate of atresia increases with age, particularly in the 10 years prior to menopause (8). The current methods for predicting age at menopause are reliant on detecting the perimenopausal changes in oocyte number and are therefore poor long-range predictors (9). Hormonal serum levels alter prior to menopause, including follicle-stimulating hormone (FSH), antimullerian hormone (AMH) and inhibin B. Of these, AMH is the best long-term predictor, with levels decreasing approximately 10 years before menopause (10,11). In addition to endocrine markers, other markers of ovarian reserve are antral follicle count and ovarian volume (9). Genetic predictors of menopausal age have the obvious advantage of being present from birth and thus have the potential to offer women advice about their reproductive lifespan from an early age, enabling them to make informed reproductive choices.

The heritability of menopausal age has been estimated to be between 30 and 85%, indicating a substantial genetic component to this complex trait, and a significant proportion (15–30%) of POF cases are familial, suggesting a genetic aetiology (12–15). Recent genome-wide association studies (GWAS) have been very successful in identifying genetic loci for many complex traits. Two independent studies, involving GWAS have been very successful in identifying genetic loci associated with normal variation in menopausal age in controls.

### RESULTS

#### Four common genetic variants influence menopausal age by between 0.7 and 11 months per allele

Using menopausal age as a quantitative trait, the directions of association and effect sizes of all four single nucleotide polymorphisms (SNPs) were consistent with those published in the GWAS (Table 1). The effect sizes for all SNPs were slightly lower than that in the discovery GWAS, as expected due to ‘winners curse’, but the chromosome 19 and 20 hits in particular have substantial effects on age at menopause with a reduction in the menopausal age of 3 months (0.257 years) and an increase of 11 months (0.924 years) per allele, respectively. Using the adjusted $R^2$ from the regression model, we calculated that together the four variants explained 1.4% of the variance in menopausal age in controls.

#### Menopause variants are associated with EM

All four SNPs were associated with an increased risk of menopause before the age of 46, with the risk allele consistent with the GWAS, i.e. being the allele associated with decreasing menopausal age. The odds ratios (ORs) per allele for each SNP ranged from 1.13 to 1.85, the non-synonymous SNP on chromosome 20 having the largest effect (Table 2). When comparing homozygote groups, the ORs ranged from 1.35 to 2.8. Three percent of women were homozygous for all four risk variants; of these 97 women, 66 (68%) were in the EM group and 31 (32%) were controls. We calculated the expected ORs for EM based on the quantitative trait estimates from the ReproGen GWAS and the BGS controls, and there was evidence that rs4806660 had a larger OR for EM than predicted [observed OR = 1.45 (CI 1.32–1.59) versus expected OR = 1.20 (CI 1.17–1.23), $P = 0.0001$] (Supplementary Material, Table S2).

#### Menopause variants are associated with POF

There were 260 women in the BGS cohort with POF and we determined the association with the menopause SNPs in these women. We were not well powered to detect these effects, but for all four SNPs there was evidence that the odds of being a POF case, per risk allele, were not significantly different from the odds of being an EM case excluding POF ($P > 0.05$) (Table 2). The smallest $P$-value in this analysis was for rs16991615, with $P = 0.051$. There was also nominal evidence that the rs16991615 SNP had a lower OR for POF than would be expected from the quantitative trait estimates [observed

### Table 1. Association of GWAS menopause SNPs with age of natural menopause in BGS controls, i.e. menopause >45 years, excluding those with surgical menopause ($n = 1261$)

<table>
<thead>
<tr>
<th>SNP</th>
<th>chr</th>
<th>Minor allele</th>
<th>Allele 2</th>
<th>MAF in controls (%)</th>
<th>ReproGen GWAS, per-allele effect</th>
<th>BGS controls Per-allele effect (se)</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4806660</td>
<td>19</td>
<td>C</td>
<td>T</td>
<td>36.5</td>
<td>$-0.406 (0.03)$</td>
<td>$-0.257 (0.12)$</td>
<td>0.027</td>
</tr>
<tr>
<td>rs16991615</td>
<td>20</td>
<td>A</td>
<td>G</td>
<td>7.0</td>
<td>$0.971 (0.0624)$</td>
<td>$0.924 (0.23)$</td>
<td>0.000056</td>
</tr>
<tr>
<td>rs9379896</td>
<td>6</td>
<td>C</td>
<td>T</td>
<td>18.5</td>
<td>$0.242 (0.0377)$</td>
<td>$0.121 (0.14)$</td>
<td>0.39</td>
</tr>
<tr>
<td>rs244715</td>
<td>5</td>
<td>A</td>
<td>G</td>
<td>45.9</td>
<td>$0.291 (0.0334)$</td>
<td>$0.059 (0.12)$</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Effect sizes are in years and are per copy of the minor allele.
OR = 1.3 (CI 0.88–1.92) versus expected OR = 2.08 (CI 1.90–2.29), P = 0.02 (Supplementary Material, Table S2).

Increased odds of having EM when menopause risk alleles are combined

Combining all risk alleles into an allele score gave a per-allele increased odds of being in the case group of 1.34 (95% CI 1.26–1.43, P = 2.2 × 10^{-20}) (Fig. 1). Weighting each risk allele by effect size did not appreciably alter the OR (OR = 1.38, 95% CI 1.29–1.48, P = 1.1 × 10^{-19}); therefore, results are presented for the unweighted alleles.

When risk alleles were combined, there were significantly increased odds of having EM in individuals with eight risk alleles compared with the median number in the control population of five risk alleles (OR = 2.02, 95% CI 1.30–3.15, P = 6.48 × 10^{-7}). Comparing the 4.5% of individuals with the lowest number of risk alleles (two or three) with the 3.0% with the highest number (eight risk alleles), the OR was 4.1 (95% CI 2.4–7.1, P = 4.0 × 10^{-7}). Comparing the 18.8% of women with less than five risk alleles with the 19.1% with more than six, the difference was also significant (OR = 2.9, 95% CI, 2.3–3.6, P = 4.6 × 10^{-19}).

Combined risk alleles have 60% discriminatory power

The discriminative power of the four menopause SNPs for EM was calculated by determining the area under the curve (AUC) in a receiver operator characteristic (ROC) analysis. Combining all four SNPs gave an AUC of 0.6 (Fig. 3).

DISCUSSION

It is well established that EM can have a genetic aetiology, but few loci have been identified. Four loci associated with variation in the normal age of menopause were recently identified by two independent GWAS with very robust statistical evidence (16,17) and we investigated the role of these common variants in EM.

**Table 2.** Association of GWAS menopause variants in EM and POF cases versus controls

<table>
<thead>
<tr>
<th>SNP</th>
<th>chr</th>
<th>Risk allele frequency in control</th>
<th>N</th>
<th>Risk allele frequency in EM including POF (menopause &lt;46) versus controls</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4806660</td>
<td>19</td>
<td>C</td>
<td>1898</td>
<td>0.365</td>
<td>2067</td>
<td>1.45</td>
<td>1.32–1.59</td>
</tr>
<tr>
<td>rs16991615</td>
<td>20</td>
<td>G</td>
<td>1879</td>
<td>0.930</td>
<td>2041</td>
<td>1.15</td>
<td>1.01–1.29</td>
</tr>
<tr>
<td>rs9379896</td>
<td>6</td>
<td>T</td>
<td>1902</td>
<td>0.815</td>
<td>2072</td>
<td>1.13</td>
<td>1.01–1.27</td>
</tr>
<tr>
<td>rs244715</td>
<td>5</td>
<td>G</td>
<td>1886</td>
<td>0.541</td>
<td>1733</td>
<td>1.20</td>
<td>1.09–1.32</td>
</tr>
</tbody>
</table>

ORs are per risk allele.
The four published variants had similar effects on age at menopause as observed in the two published discovery studies and our unpublished data from ReproGen, although in each case the effect was smaller, which is consistent with other GWAS, where the discovery effect is larger than in replication cohorts due to over-inflation or 'winners curse'. Two of these variants, on chromosomes 19 and 20, were significant at \( P < 0.05 \), but we had \(< 20\%\) power to detect the observed effect sizes at \( P < 0.05 \) for the SNPs on chromosomes 5 and 6. Together, the four SNPs explain about 1.4% of the variation in age at menopause. This is comparable to other complex traits such as height and BMI: 20 variants explain about 3% of the variation in height (18) and the 17 variants published to date for BMI only explain about 1% of the variation (19).

Our data suggest that variants associated with normal menopausal age are also significant risk factors for EM. The non-synonymous SNP in \( MCM8 \) (rs16991615) increased the risk of EM by 85% per allele, the rare allele being protective for EM. The chromosome 19 variant near \( TMEM224 \) (rs4806660) also has a substantial effect on risk of EM, increasing the OR by 45%. These ORs are higher than many reported for complex traits, which generally have per-allele ORs less than 1.5, with the exception of autoimmune disorders, macular degeneration and pigmentation loci (20) (http://www.genome.gov/gwastudies); however, further replication in additional EM cohorts would be beneficial in order to confirm the effects.

It has been hypothesized that EM cases may have a different aetiology from normal menopause. Our data suggest that there is at least some common aetiology between EM and normal menopause as the same genetic loci are associated with both traits. However, the OR for the chromosome 19 SNP was significantly greater than predicted, based on the effect on menopause age in the normal range in the quantitative trait analysis, suggesting that this locus may have a non-linear effect on menopause age. We investigated the aetiology of EM further in women at the extreme of the menopause distribution, i.e. menopause before 40 years. Women who have menopause before 40 are clinically classified as having POF and are often investigated for a genetic aetiology, most commonly by cytogenetic screening and \( FMR1 \) mutation testing. The published menopause variants were originally identified in women excluding those with POF. The chromosome 20 SNP was the only variant where there was a suggestion that the effect might be different in POF cases compared with other EM cases, as the OR for POF cases was lower than predicted from the quantitative trait estimates and did not overlap the 95% confidence interval for EM, but with the caveat that this SNP is rare (MAF = 7%) and the number of POF cases was relatively small. However, these data suggest that the effect of the chromosome 20 SNP is not as strong in POF cases and suggest that POF may have a different aetiology.

Our data therefore provide evidence that while some EM cases represent the tail of the normal distribution, some may have a different aetiology. The role of menopause variants in POF requires replication in additional independent studies, but these preliminary data support the necessity to look for EM-specific genes as these may not overlap with genes for menopause in the remainder of the age distribution.

EM has a significant impact on female health and results in early infertility. Current techniques have good predictive power for the end of female reproductive life, but only in the immediate pre-menopausal period when ovarian reserve is already diminished and natural conception is likely to be difficult or impossible. Commercial over-the-counter tests are available that measure hormone levels, usually FSH and/or AMH, but they need to be repeated every 2 years and are not good long-range predictors. It would be beneficial for women to be able to predict the timing of the end of their reproductive life in their early 20s, so that they can decide whether they want to risk the chance of infertility by delaying childbearing. The high heritability of menopausal age makes the potential for a genetic test extremely attractive. Very few genes have been demonstrated to be common causes of EM, with the possible exception of the \( FMR1 \) premutation gene which accounts for \(~ 5\%\) of idiopathic POF cases (21), but the utility of \( FMR1 \) as a genetic predictor has yet to be proved (22). We therefore studied the discriminative power of the four menopause loci identified by GWAS. The AUC for the ROC analysis was 0.6 for the four variants in combination, which is still some way from the power of tests such as the Framingham risk score for predicting coronary heart disease, where the AUC is typically near 0.8. The genetic power to discriminate EM cases is, however, comparable to other genetic risk predictors, e.g. recent estimates for breast cancer, where 10 variants have an AUC of 59.7% (23) and diabetes, where the AUC is 55–60% (24,25). Although the predictive power of the current variants is limited, as more variants are discovered this should increase, and in the absence of other good predictors for menopause, a genetic score would have value, because it could be carried out early and would be relatively inexpensive.

**MATERIALS AND METHODS**

**Study population**

We selected 2118 women with natural menopause before the age of 46 years and 1941 controls with menopause after 45 years, from the BGS. The BGS is a prospective...
epidemiological cohort study launched in September 2004, the primary objective of which is to investigate the environmental, behavioural, hormonal and genetic causes of breast cancer and which is also investigating the causes of other cancers and diseases (http://www.breakthroughgenerations.org.uk/). The cohort consists of over 110,000 women from the general population of the UK aged 16 and older at the date of entry. Recruitment is through volunteers connected with the charity Breakthrough Breast Cancer, volunteers responding to publicity and via them their friends, family members and other contacts. Each participant completes a questionnaire and most provide a blood sample for analysis of genomic, hormonal and other blood factors. Participants are asked questions that include detailed menstrual histories, thus enabling identification for the present analyses of a group. Natural menopause was defined as absent menstruation for at least 6 months without known cause. Women were excluded if periods stopped because of pregnancy, breastfeeding, surgery, hormonal contraceptive use and other types of medical treatment or if there was a medical condition or illness that could have caused amenorrhoea (e.g. polycystic ovary syndrome). We selected one control for each EM case, matched for date of birth (within 12 months), ethnicity, year of questionnaire completion and source of recruitment. Women were eligible as controls if they were post-menopausal at entry to the study with a menopausal age of 46 and over (25.7%). Menopause could be natural or surgically induced provided there was evidence they were still menstruating after age 45. There were 182 women who had surgical menopause in the control group.

### Genotyping

Following the publication of the four variants associated with age at menopause, the research groups involved formed a consortium and have pooled data and meta-analysed results. Thus, the signals have been refined, and for three of the four signals, a different SNP became the strongest association signal in the region and was selected for this analysis (ReproGen Consortium, unpublished data). The linkage disequilibrium between the tested and published SNPs was as follows: $R^2 = 0.677$, $R^2 = 0.965$, $R^2 = 0.194$. All samples were therefore typed for the following four SNPs: rs16991615 (chr 20, position 5896227), rs9379896 (chr 6, position 10994935), rs4806660 (chr 19, position 60516446), rs244715 (chr 5, position 176436169). Genotyping was performed in-house using TaqMan PCR assays designed by Applied Biosystems. Genotype frequencies were in Hardy–Weinberg equilibrium ($P > 0.1$), call rates were >93%, with >99% concordance of 288 duplicates.

### Analysis

In order to determine the effect size of the published genetic variants (16,17) on menopausal age without the bias of the ‘winner’s curse’ (26), we analysed the association of the SNPs with normal menopausal age as a quantitative trait, in the BGS controls (menopausal age >45 years), who had a natural menopause. Thus, we excluded the 182 women who had surgical menopause >45 years and the 498 women who were not yet menopausal, leaving a cohort of 1261 women...

**Figure 3.** ROC plot modelling the discriminatory power of all four menopause SNPs, for EM ($≤ 45$). On the $y$-axis is the true-positive rate or sensitivity of the test for predicting EM and on the $x$-axis is the false-positive rate or specificity of the SNPs for predicting EM.
with natural menopause from the control group. Linear regression was used to determine the effect of the minor allele for each SNP on menopausal age. A combination of all four variants in one model was used to estimate the variance in menopause age explained by the SNPs, using the adjusted $R^2$ value.

We performed additional analysis by subdividing the EM cohort group into: (i) POF, i.e. women with menopause <40 years; (ii) women with menopause between 40 and 45 years inclusive (Table 3). We compared each case group with the controls, i.e. women who were either non-menopausal but over 45 years at entry into study or had gone through menopause aged over 45. Logistic regression was used to determine the effect of menopause-lowering alleles of each SNP on the odds of being in the case group, assuming an additive genetic model. We repeated the logistic regression excluding non-white individuals ($n = 39$). In addition, we performed conditional logistic regression to account for matched pairs of case–control samples. The ORs were very similar in the conditional regression and excluding non-whites, when compared with the unconditional regression including non-whites (Supplementary Material, Table S1); we therefore present results for unconditional logistic regression and included all individuals in subsequent analyses.

We estimated the expected OR for each of the four variants for both the EM (<46 years) and POF (<40 years) groups based on the beta estimate from the ReproGen Consortium GWAS (unpublished data) and compared with the ORs we observed in the BGS. We also calculated the expected ORs based on the quantitative trait analysis in BGS controls, but as the menopause distribution is truncated at 45 years in this cohort and the sample size is relatively small, we consider the GWAS estimate to be more accurate. We calculated the expected ORs for both the point estimate quantitative trait beta and the upper and lower 95% CI intervals, by using the ‘Case–Control for threshold-selected quantitative traits’ 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 then tested for heterogeneity (using Cochran’s $Q$-test in StatsDirect) between the expected and observed ORs.

The total number of risk alleles for EM, across all four SNPs, was calculated and those with menopause below 46 years were compared with controls. Individuals were only included if they were successfully genotyped for all four SNPs ($n = 3242$, of which 48% were controls). We determined the increased likelihood of being a case depending on the number of risk alleles using logistic regression. In addition, we compared individuals at the extreme 5 and 20% of the risk allele distribution. We also compared ORs for the number of risk alleles compared with the median number of alleles in controls, i.e. 5. In a further analysis, we calculated a weighted risk score based on the effect size of each variant: the number of risk alleles at each locus was multiplied by its per-allele effect size calculated from the quantitative trait analysis in controls. The weighted score was then re-scaled to reflect the number of SNPs tested, by dividing the weighted score by the sum of the effects (1.348) and multiplying by the number of SNPs (i.e. 4).

The power of the SNPs to discriminate EM cases was calculated by determining the AUC in an ROC analysis in Stata, including all SNPs as separate linear terms in the model, against menopause status, i.e. case ($\leq 45$) versus control. The AUC measures how well the model discriminates between cases and controls, such that a perfect test gives an AUC of 1.0 and a test with no predictive power gives an AUC of 0.5.

Adjusting by smoking status, a significant predictor of menopause status ($27,28$), did not influence our results; therefore, all data are presented without correction for smoking status. Data were analysed in Stata v10.1 (http://www.stata.com/).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at *HMG* online.

**ACKNOWLEDGEMENTS**

We thank the women who participated in the study, our colleagues in the BGS Team, particularly the research nurses Alison Butlin, Alison Hart, Margo Pelerin and Jill Wood who collected blood. We would like to acknowledge the contribution of the participants and investigators from the ReproGen Consortium for providing data from unpublished meta-analyses. ReproGen includes the following studies: AGES, ARIC, CHS, DeCode, EGPUT, ERF, FHS, HAPI, 2011, Vol. 20, No. 1 191

**Table 3.** Individuals selected from the BGS cohort for inclusion into the study

<table>
<thead>
<tr>
<th></th>
<th>Controls (menopause $&gt;45$ years)</th>
<th>EM Menopause 40–45 years inclusive</th>
<th>Menopause $&lt;40$ years</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>1941</td>
<td>1858</td>
<td>260</td>
</tr>
<tr>
<td>Mean menopausal age</td>
<td>51.5 years (sd = 2.9)</td>
<td>43.3 years (sd = 1.6)</td>
<td>35.7 years (sd = 4.3)</td>
</tr>
<tr>
<td>Mean age at recruitment</td>
<td>58.7 (sd = 8.5)</td>
<td>59.0 (sd = 8.4)</td>
<td>53.3 (sd = 11.0)</td>
</tr>
<tr>
<td>Smoking at menopause</td>
<td>159 (8.2%)</td>
<td>336 (18.1%)</td>
<td>64 (24.6%)</td>
</tr>
<tr>
<td>White ethnicity (%)</td>
<td>1921 (99%)</td>
<td>1841 (99%)</td>
<td>256 (98%)</td>
</tr>
<tr>
<td>Risk allele frequency for rs244715</td>
<td>0.541</td>
<td>0.589</td>
<td>0.568</td>
</tr>
</tbody>
</table>

Human Molecular Genetics, 2011, Vol. 20, No. 1
REFERENCES