Published in final edited form as:
Hum Genet. 2010 November ; 128(5): 515–527. doi:10.1007/s00439-010-0878-4.

A large-scale candidate-gene association study of age at menarche and age at natural menopause

Chunyan He,
Department of Public Health, Indiana University School of Medicine, Indianapolis, IN, USA, Phone: (317) 278-3033; Fax: (317) 278-2966

Peter Kraft,
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Daniel I. Chasman,
Donald W. Reynolds Center for Cardiovascular Research, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Julie E. Buring,
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. Donald W. Reynolds Center for Cardiovascular Research, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Constance Chen,
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Susan E. Hankinson,
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Guillaume Paré,
Donald W. Reynolds Center for Cardiovascular Research, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Stephen Chanock,
Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA

Paul M. Ridker, and
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. Donald W. Reynolds Center for Cardiovascular Research, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

David J. Hunter
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

Chunyan He: chunhe@iupui.edu

Correspondence to: Chunyan He, chunhe@iupui.edu.

Ethical Standards
The authors declare that the experiments comply with the current laws of the country in which they were performed.

Conflict of Interest
The authors declare that they have no conflict of interest.

Abstract

Recent genome-wide association (GWA) studies have identified several novel genetic loci associated with age at menarche and age at natural menopause. However, the stringent significance threshold used in GWA studies potentially lead to false negatives and true associations may have been overlooked. Incorporating biologically relevant information, we examined whether common genetic polymorphisms in candidate genes of 9 groups of biologically plausible pathways and related phenotypes are associated with age at menarche and age at natural menopause. A total of 18,862 genotyped and imputed single nucleotide polymorphisms (SNPs) in 278 genes were assessed for their associations with these two traits among a total of 24,341 women from the Nurses' Health Study (NHS, N=2,287) and the Women's Genome Health Study (WGHS, N=22,054). Linear regression was used to assess the marginal association of each SNP with each phenotype. We adjusted for multiple testing within each gene to identify statistically significant SNP associations at the gene level. To evaluate the overall evidence for an excess of statistically significant gene associations over the proportion expected by chance, we applied a one-sample test of proportion to each group of candidate genes. The steroid-hormone metabolism and biosynthesis pathway was found significantly associated with both age at menarche and age at natural menopause (p=0.040 and 0.011, respectively). Additionally, the group of genes associated with precocious or delayed puberty was found significantly associated with age at menarche (p=0.013), and the group of genes involved in premature ovarian failure with age at menopause (p=0.025).

Keywords

age at menarche; age at natural menopause; candidate genes; biologically plausible pathways

Introduction

Age at menarche and age at natural menopause have important implications for the health of women. An early onset of menarche and later menopause are well-established risk factors for the development of breast cancer (Kvale 1992; Peeters et al. 1995) and endometrial cancer (Kaaks et al. 2002; Xu et al. 2004). On the other hand, late menarche and early menopause increase the risk of osteoporosis (Ito et al. 1995; Kritz-Silverstein and Barrett-Connor 1993) and cardiovascular disease (Pechere-Bertschi and Burnier 2004; Rees 1995; van der Graaf et al. 1997; van der Schouw et al. 1996). These two traits also mark the beginning and the end of a woman's reproductive life (te Velde and Pearson 2002). Individual variability of age at menarche and age at natural menopause are partially under genetic control (de Bruin et al. 2001; Kaprio et al. 1995; Meyer et al. 1991; Snieder et al. 1998; Treloar et al. 1998; Treloar and Martin 1990). Twin and family studies have suggested relatively high heritability for age at menarche (53–74%) (Chie et al. 1997; Kaprio et al. 1995; Sharma 2002) and age at natural menopause (49–87%) (de Bruin et al. 2001; Murabito et al. 2005a; Snieder et al. 1998). Genome-wide linkage analyses using microsatellite markers have identified chromosomal regions that may harbor genes for menarche (Guo et al. 2006a; Pan et al. 2008; Rothenbuhler et al. 2006) and natural menopause (Murabito et al. 2005b; van Asselt et al. 2004). However, the specific causal genes remain mostly unidentified.

Despite the common belief that multiple genes are responsible for controlling the timing of menarche and natural menopause, very few genes have been identified that contain common genetic variants associated with age at menarche and age at natural menopause. Previous candidate gene association studies have focused on the genes involved in steroid-hormone metabolism and biosynthesis pathway, such as estrogen receptor genes (*ESR1* and *ESR2*) (Boot et al. 2004; Dvornyk et al. 2006; Gorai et al. 2003; He et al. 2007; Kok et al. 2005; Long et al. 2005; Stavrou et al. 2006; Stavrou et al. 2002; Weel et al. 1999), sex hormone binding globulin

gene (*SHBG*) (Xita et al. 2005), and genes involved in estrogen biosynthesis (*CYP17*, *CYP19*, and *HSD17*), hydroxylation (*CYP1A1*, *CYP1B1*, and *CYP3A4*), and inactivation of the reactive metabolites (*COMT*) (Gorai et al. 2003; Guo et al. 2006b; He et al. 2007; Hefler et al. 2005; Lai et al. 2001; Long et al. 2006; Mitchell et al. 2008). Association studies of candidate genes in other biologically plausible pathways are sparse, including *IGF1* (Zhao et al. 2007), *AMHR2* (Kevenaar et al. 2007), and genes associated with thrombophilia and vascular homeostasis such as *F5* (Tempfer et al. 2005; van Asselt et al. 2003), *APOE* (Koochmeshgi et al. 2004; Tempfer et al. 2005), *NOS3* (Hefler et al. 2002; Tempfer et al. 2005; Worda et al. 2004), *F2*, and *SERPINE1* (Tempfer et al. 2005). Recently, a few studies focused on genes associated with the extremes of these two phenotypes, for example, ten hypogonadotropic hypogonadism genes (*FGFR1*, *GNRH*, *GNRHR*, *GPR54/KISS1R*, *KAL1*, *KISS1*, *LEP*, *LEPR*, *PROK2*, and *PROKR2*) (Gajdos et al. 2008), and the *FMR1* gene (Ennis et al. 2006; Mallolas et al. 2001) have been examined in relation to age at menarche and age at menopause, respectively. While these studies have provided some promising results, a large-scale candidate-gene analysis has not been performed to evaluate common genetic variants in association with these two traits.

Most recently, genome-wide association (GWA) studies of age at menarche (He et al. 2009; Ong et al. 2009; Perry et al. 2009; Sulem et al. 2009) and natural menopause (He et al. 2009; Stolk et al. 2009) have identified several novel genetic loci. However, true associations may have been overlooked due to the stringent significance threshold used in these GWA studies (usually p-value at least $<1\times10^{-7}$). Incorporating biologically-relevant information, we used the genotyped and imputed SNP data from the Cancer Genetic Markers of Susceptibility (CGEMS) breast cancer project (Hunter et al. 2007) and the Women's Genome Health Study (WGHS) to examine whether common genetic variants of pre-specified candidate genes in 9 groups of genes in biologically plausible pathways and related phenotypes are associated with age at menarche and age at natural menopause. These pathways and phenotypes include: a) the steroid-hormone metabolism and biosynthesis pathway; b) the insulin-like growth factor (IGF) signaling pathway; c) the transforming growth factor-beta (TGF-β) superfamily and signaling pathway; d) the thrombophilia and vascular homeostasis pathway; e) obesity and obesity-related phenotypes; and for age at menarche only: f) precocious or delayed puberty; and for age at natural menopause only: g) premature ovarian failure; h) polycystic ovary syndrome (PCOS); and i) smoking and nicotine dependence. A total of 18,862 SNPs in 278 genes were assessed for their associations with age at menarche and age at natural menopause among 24,341 women from the Nurses' Health Study (NHS, N=2,287) and the WGHS (N=22,054).

Materials and Methods

Study Population

The NHS was initiated in 1976, when 121,700 United States registered nurses between the ages of 30 and 55, residing in 11 larger U.S. states, returned an initial questionnaire reporting medical histories and baseline health-related exposures, including information related to reproductive history (age at menarche, age at first birth, parity, age at menopause etc.), and exposure to exogenous hormones (oral contraception or post-menopausal hormone replacement therapy). Biennial questionnaires with collection of exposure information on risk factors have been collected prospectively, and outcome data with follow-up of reported disease events are collected. From May 1989 through September 1990, we collected blood samples from 32,826 participants in the NHS cohort. Subsequent follow-up has been greater than 99% for this subcohort.

The NHS nested breast cancer case-control study was derived from the 32,826 women in the blood subcohort who were free of diagnosed breast cancer at blood collection and followed

for incidence disease until June 1, 2004. Breast cancer follow-up in the NHS was conducted by personal mailings and searches of the National Death Index. Controls were women not diagnosed with breast cancer up to the time of case diagnosis, and were one-to-one matched to cases based on age at diagnosis, blood collection variables (time of day, season, and year of blood collection, as well as recent (<3 months) use of postmenopausal hormones), ethnicity (all cases and controls are self-reported Caucasians), and menopausal status (all cases were postmenopausal at diagnosis). The 2,287 NHS participants included in the present analysis were from this nested breast cancer case-control study and were self-described Caucasians with genotype data available from the National Cancer Institute's Cancer Genetic Marker of Susceptibility (CGEMS) Project (Hunter et al. 2007).

The WGHS cohort comprises 28,345 American women, who are participants in the ongoing Women's Health Study (WHS) who had no prior history of cardiovascular disease, cancer, or other major chronic illness, and who provided a blood sample at baseline with consent for analyses linking blood-derived observations with baseline risk factor profiles and incident disease events. These women were 45 years of age or older at baseline in 1993. Details of the rationale, design, and methodology of the WGHS and the WHS are described elsewhere (Ridker et al. 2008a; Ridker et al. 2005). There were 22,054 Caucasian WGHS participants with genotype data available for the present analysis.

Outcome Assessment

Age at menarche is defined as age at the first menstrual period (in years). This information was retrospectively ascertained by recall in the baseline questionnaires. In the NHS, the question was open-ended and asked "At what age did your menstrual periods begin?___ years of age". In the WHS, the question was not open-ended and asked, "At what age did your menstrual periods begin?" Response categories were: "9 or younger; 10; 11; 12; 13; 14; 15; 16; 17 or older". We excluded women whose age at menarche was reported as 18 years or greater in the NHS from the analysis (N=2), these women were more likely to have a pathological cause outside the spectrum of normal variation.

Age at natural menopause is defined as the age when menstrual periods ceased permanently and naturally (in years). Questions regarding to menopause status were asked in baseline and subsequent questionnaires. The questions were: "Have your menstrual periods ceased permanently?" If yes, "At what age did your natural periods cease?" and "For what reason did your periods cease?" Response categories were: "Surgical; Radiation or Chemotherapy; Natural". Age at natural menopause was assessed in the baseline questionnaire for postmenopausal women at baseline, and updated in subsequent questionnaires for premenopausal women at baseline. Women who had radiation/chemotherapy or surgically-induced menopause were excluded from the current study. We were unable to exclude women using hormone replacement therapy (HRT) before the onset of menopause, as the exact time when women started to use HRT relative to the precise onset of natural menopause cannot be determined through biennial questionnaires in either study. We also excluded women reported ages at menopause younger than 40 years or older than 60 years from the analysis as these extremes may reflect underlying pathologies.

Genotyping and Imputation

Genotyping in the NHS used the Illumina Infinium Sentrix HumanHap550 chip with high success rates and reproducibility. Detailed methods related to the genotyping were published previously (Hunter et al. 2007), including quality control, initial assessment of sample completion rates, assessment of SNP call rates, concordance rate, deviation from Hardy–Weinberg proportions in control DNA, and final sample selection and exclusion for association analysis.

Genotyping in the WGHS used either the Illumina HumanHap300 Duo-Plus chip or the combination of the HumanHap300 Duo and I-Select chips. A total of 363,808 SNPs was attempted. SNPs with call rates <90% were excluded from further analysis. All samples with percentage of missing genotypes higher than 2% were removed. Among retained samples, SNPs were further evaluated for deviation from Hardy-Weinberg equilibrium and other quality control measures (Ridker et al. 2008b). After quality control, a total of 336,108 SNPs were left.

For the NHS and the WGHS separately, we imputed total about 2.5 million SNP genotypes using the hidden Markov model implemented in MACH 1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html), using HapMap haplotypes as a reference (HapMap Phase II CEU population, Release 21, NCBI build 35). By adopting this strategy, we were able to increase the genome coverage and boost study power by combing data from different genotyping platforms in the NHS and the WGHS. To account for the uncertainty in the imputed genotypes, we used an "allele dosage", which is defined as the expected number of copies of the minor allele at each SNP and varies between 0 and 2. We excluded the imputed SNPs with an observed R square from MACH less than 0.30.

Candidate Gene and SNP Selection

Candidate genes were selected for analysis based on their membership in biologically plausible pathways or reported associations with related phenotypes. The rationale for selecting the pathways and related phenotype is provided in the Supplementary Methods. We assembled candidate genes in each of the 9 groups through literature search and by using Ingenuity Pathway Analysis (Ingenuity Systems: www.analysis.ingenuity.com). Detailed descriptions of the 278 genes evaluated are listed in Supplementary Table 1. For each gene, we included SNPs that are within 20kb upstream and 10kb downstream of the gene region. A total of 18,862 genotyped and imputed SNPs in 278 genes were assessed for their associations with age at menarche and age at natural menopause.

Statistical Analysis

Single-SNP association test—We performed linear regression to analyze the association between each of the SNPs (coded as counts of minor alleles) and age at menarche or age at natural menopause as a continuous variable, using ProABEL (Aulchenko et al.) and MACH2QTL(Li and Abecasis 2006) software in the NHS and the WGHS, respectively. SNPs with low MAF (< 1%) in either the NHS or the WGHS samples were excluded from analysis. To control for potential confounding by population stratification, we adjusted for the top principal components of genetic variation chosen for each study after excluding any admixed individuals clearly not of European ancestry (He et al. 2009; Patterson et al. 2006; Price et al. 2006; Pritchard and Rosenberg 1999). In the NHS analysis, controlling for breast cancer case-control status made no material difference to the results. To combine evidence for association with age at menarche or age at natural menopause and calculate a summary effect size across the NHS and the WGHS, we used fixed-effect joint analysis, and heterogeneity in effect size across two studies was assessed using the Q-test (Higgins and Thompson 2002).

Gene-based analysis—To take into account the number of SNPs genotyped in each gene and their underlying linkage disequilibrium (LD) structure, we applied the approach of Gao *et al.* (Gao et al. 2008) to obtain the effective number of independent tests (M_{eff}) represented in the composite LD correlation matrix, by use of the EIGEN function in R software (v 2.7.0) (R Development Core Team 2008). P-values from single-SNP association tests were further adjusted for multiple comparisons at the gene level with Bonferroni correction based on M_{eff}. We tested the significance of the gene association by assessing the significance of the

adjusted minimal p-value in each gene. A gene was considered significantly associated with the phenotype if it had an adjusted p-value less than 0.05.

Pathway/group-based analysis—For each pathway/group, we evaluated the overall evidence for an excess of statistically significant gene associations over the proportion expected by chance by calculating the ratio of observed versus the expected number of significant genes. To test the significance of each pathway/group, the proportion of statistically significant genes for each pathway/group was compared with a nominal 5% significant level using the one-sample test for proportions (Gauvreau 2006).

All statistical methods were performed using SAS software, version 9.1 (SAS Institute, Cary, NC), unless stated otherwise.

Results

Among the total 24,341 study participants, 24,298 women with data on age at menarche remained for the analysis of age at menarche. The mean age at menarche was 12.5 years with a standard deviation of 1.4 years; only 12,723 women had natural menopause between the ages of 40–60 and remained for analysis of age at natural menopause. The mean age at natural menopause was 50.6 years with a standard deviation of 3.6 years. The means and standard deviations were similar across the NHS and the WGHS.

We proposed 9 groups of candidate genes that might play a role in age at menarche and age at natural menopause, including genes involved in biologically plausible pathways and for related phenotypes (Table 1). As we observed no evidence of systematic heterogeneity in genetic association with either trait between the NHS and the WGHS, we chose to combine evidence for association to optimize power and calculate a summary estimate across the two studies. Single-SNP analyses identified 1,155 of the 12,901 SNPs (9.0%) and 1,135 of the 17,519 SNPs (6.5%) tested with nominal p-value <0.05 for the association with age at menarche and age at natural menopause, respectively. For each gene, we estimated the effective number of independent tests (M_{eff}, Supplementary Table 2) and adjusted the significance threshold using Bonferroni correction based on M_{eff}. We observed that 42 SNPs in 9 genes (*FSHB, LHCGR, POMC, UGT2B4, GHRH, CD40LG, FGFR1, KISS1, NKX2-1*) were significantly associated with age at menarche (Table 2a), but we did not find significant association between age at menarche and any SNP among the TGF-β signaling pathway genes or the obesity-related genes. There were significant associations for 86 SNPs in 16 genes (*CYP19A1, FSHB, LHCGR, PGR, SRD5A1, IGF1, IGF2R, SMAD7, TGFBR1, PCSK1, PPARG, TNF, EIF2B4, POLG, NBN, ANKK1*) with age at natural menopause (Table 2b). We found no evidence that SNPs in vascular homeostasis genes are associated with age at natural menopause.

The results of candidate gene pathways/groups are summarized in Table 3. We found an excess of statistically significant gene associations over the proportion expected by chance in steroid-hormone metabolism and biosynthesis pathway for both age at menarche and age at natural menopause (p=0.040 and 0.011, respectively). Additionally, the group of genes associated with precocious or delayed puberty were found significantly associated with age at menarche (p= 0.013), and the group of genes involved in premature ovarian failure was significantly associated with age at menopause (p=0.025).

Discussion

We conducted a large-scale candidate gene association study to identify common genetic variants underlying age at menarche and age at natural menopause. To our knowledge, this is the largest and most comprehensive candidate gene association study for age at menarche and

age at natural menopause to date, including genes that have rarely been considered in previous studies. We identified 9 groups of candidate genes in different biological pathways and for related phenotypes that are potentially important to onset of menarche and natural menopause. Through a pathway/group-based analysis, we found that the steroid-hormone metabolism and biosynthesis pathway was significantly associated with both traits, and the group of genes related to precocious or delayed puberty was significantly associated with age at menarche, and the group of genes related to premature ovary failure with age at natural menopause.

In previous studies conducted to investigate genetic polymorphisms in candidate genes in relation to age at menarche and age at natural menopause, polymorphisms in 11 genes have been reported to be associated with these two phenotypes (Table 5). The genes involved in estrogen metabolism and biosynthesis pathway, especially estrogen receptors (*ESR1* and *ESR2*) and the CYP family (*CYP17*, *CYP19*, and *CYP1B1*), have been the most popular candidates in previous association studies, but with inconsistent results (Boot et al. 2004; Dvornyk et al. 2006; Gorai et al. 2003; Guo et al. 2006b; He et al. 2007; Hefler et al. 2005; Kok et al. 2005; Lai et al. 2001; Long et al. 2006; Long et al. 2005; Mitchell et al. 2008; Stavrou et al. 2006; Stavrou et al. 2002; Weel et al. 1999). Several possible explanations for this inconsistency include small sample sizes that fail to detect moderate effect of multiple susceptible genes, failure to take into account gene-gene and gene-environmental interactions and multiple comparisons, different study designs, and population differences between studies. In this study, we replicated one finding for *ESR2* (rs4986938) that has previously been reported associated with age at menarche, with G allele associated with early onset (Stavrou et al. 2006) (Table 5). Although rs4986938 lies in an intron of the *ESR2* gene and does not lead to an amino acid change in the ERS2 protein, it is possible that it is in linkage disequilibrium (LD) with other regulatory sequence variations that may affect gene expression or function (Yaich et al. 1992). Alternatively, this polymorphism may affect gene expression through a structural change in mRNA and thus lead to an alteration in estrogenic biological activity. This SNP has also been associated with ovulatory dysfunctions in patients with menstrual disorders (Sundarrajan et al. 2001).

In the past few years, genome-wide association (GWA) studies have been successful in identifying common genetic polymorphisms associated with complex diseases and human traits. As an agnostic approach, a GWA study surveys most of the genome for genetic variants associated with the phenotype under study, not relying on a prior hypothesis of underlying mechanisms. Recent GWA studies have identified several novel genetic loci associated with age at menarche (He et al. 2009; Ong et al. 2009; Perry et al. 2009; Sulem et al. 2009) and age at natural menopause (He et al. 2009; Stolk et al. 2009). However, the massive number of comparisons involved requires very stringent P values (usually at least $< 1\times10^{-7}$) and potentially leads to few true associations being declared as positive with many false negatives. None of the previous candidate genes has exceeded this significance threshold in the GWA studies. It is likely that true modest associations have been overlooked due to the very stringent P value threshold used in the initial GWA scan and failed to be included in further replication study. On the other hand, the candidate-gene association study incorporates biological knowledge and may allow a less stringent significance threshold (Hirschhorn and Altshuler 2002), serving as a complementary approach to GWA study. Our study provided an evaluation of common variants of biologically plausible genes pre-specified by their membership in physiological pathways and their associations with menarche/menopause-related phenotypes.

In this study, we further define 9 sets of genes based upon their common biological attributes (e.g. biological pathways or related phenotypes). We then measured the degree of overrepresentation or “enrichment” of each gene set among nominally phenotype-associated SNP markers. Such “pathway/group-based” analysis can extract more biologically-meaningful information than might be apparent from examination of only statistically significant single

SNP findings, as genes harboring one phenotype-associated polymorphism are likely to harbor more than one (McClellan et al. 2007) and locus and allelic heterogeneity appear substantial in complex traits such as age at menarche and age at natural menopause. Additionally, pathway/group-based analyses conducted in independent samples may provide gene set replication whether or not the single marker findings replicate. Thus, this analysis approach can complement and extend findings derived from single marker replication studies, yielding information relevant to our understanding of genetic determinants of complex traits.

We observed an excess of gene associations in the steroid-hormone metabolism and biosynthesis pathway in relation to both age at menarche and age at natural menopause, supporting the important role of steroid- hormones in the differentiation, maturation, and function of the reproductive system. Although we tested the same genes and SNPs in this gene set, we found the two traits were generally associated with different SNPs in the same gene or different genes, with the exception of a SNP, rs11031010, located at 5′-untranslated region (UTR) of *FSHB* gene. We also observed that the gene sets related to the extremes of the phenotypes, i.e. precocious or delayed puberty and premature ovarian failure, are overrepresented among phenotype-associated single SNP markers. These results are in line with the hypothesis that the genes underlying the extremes of the phenotypes may also influence phenotypic variation in the general population. Specifically, severe mutations in a gene can cause extreme phenotypes, while less pathogenic and common variants of the same genes may influence the normal variation of the phenotypes.

Our results should be interpreted considering several limitations. First, candidate genes in the study were selected based on educated guesses on which genes or pathways are most likely to be associated with phenotypes. Because our knowledge of the expression and function of the human genome is incomplete, this approach may fail to discover important, novel genes and fully characterize pathways. Secondly, larger genes with more SNPs are more likely to be associated with phenotypes compared to small genes with few SNPs. We guarded against this bias through calculating an effective number of independent tests for each gene and adjusting significance threshold with a modified Bonferroni correction. Third, the genomic coverage may not be optimal for some genes and important genetic variants may not be included and tested. This could potentially lead to bias in our results of gene-phenotype associations. Fourth, we didn't adjust for the number of pathways/groups we have investigated for the P value of each pathway/group. However, this is appropriate for tests of promising hypotheses where the pathway/group is believed to play a direct role in disease etiology. Fifth, population stratification can lead to spurious associations in genetic association studies. We utilized a large subset of unlinked genetic markers to adjust for population structure. However, there is little evidence of population stratification in our study, as controlling for the top principal components of genetic variation does not make material changes to the results. Finally, age at menarche and age at natural menopause were self-reported, and age at menarche was determined retrospectively. Thus, both traits might be measured with error, although this error is likely to be random with respect to genotype. Age at menarche and age at natural menopause are associated with breast cancer and other endpoints in both the NHS and the WHS with magnitudes and directions of association that are consistent with other studies, attesting to the relative validity of the measurements. In addition, Colditz *et al.* found a correlation of 0.82 for reported ages at natural menopause between two questionnaires in the NHS (Colditz et al. 1987). Must *et al.* observed a correlation of 0.79 between contemporarily documented and recalled age at menarche after 30 years (Must et al. 2002). Moreover, random measurement error, if present, would be expected to attenuate associations in our study, rather than give rise to false positive associations.

In summary, we conducted a large-scale candidate gene association study to investigate whether common genetic variants of genes in biologically plausible pathways and for related

phenotypes are associated with age at menarche and age at natural menopause. We found that common variants in the steroid-hormone metabolism and biosynthesis pathways were significantly associated with both age at menarche and age at natural menopause; and the groups of genes involved in extremes of the phenotypes, i.e. precocious or delayed puberty and premature ovarian failure, were significantly associated with age at menarche and age at natural menopause, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Miletich and A. Parker as well as the technical staff at Amgen, Inc (Cambridge MA) for their collaboration and scientific support in performing the genotyping for the WGHS. The NHS GWAS was performed as part of the Cancer Genetic Markers of Susceptibility initiative of the NCI. We particularly acknowledge the contributions of R. Hoover, A. Hutchinson, K. Jacobs and G. Thomas. We thank H. Ranu and P. Soule of the DF/HCC High Throughput Polymorphism Detection Laboratory for assistance. The WGHS is supported by HL 043851 and HL69757 from the National Heart Lung and Blood Institute and CA 047988 from the National Cancer Institute (Bethesda, MD), the Donald W Reynolds Foundation (Las Vegas, NV), the Fondation Leducq (Paris FR), with collaborative scientific support and funding for genotyping provided by Amgen, Inc. The NHS is supported by CA 40356, CA 87969, and U01-CA98233 from the National Cancer Institute. We acknowledge the study participants in the NHS and the WGHS for their contribution in making this study possible.

References

Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11:134. [PubMed: 20233392]

Boot AM, van der Sluis IM, de Muinck Keizer-Schrama SM, van Meurs JB, Krenning EP, Pols HA, Uitterlinden AG. Estrogen receptor alpha gene polymorphisms and bone mineral density in healthy children and young adults. Calcif Tissue Int 2004;74:495–500. [PubMed: 15354856]

Chie WC, Liu YH, Chi J, Wu V, Chen A. Predictive factors for early menarche in Taiwan. J Formos Med Assoc 1997;96:446–450. [PubMed: 9216169]

Colditz GA, Stampfer MJ, Willett WC, Stason WB, Rosner B, Hennekens CH, Speizer FE. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. Am J Epidemiol 1987;126:319–325. [PubMed: 3605058]

de Bruin JP, Bovenhuis H, van Noord PA, Pearson PL, van Arendonk JA, te Velde ER, Kuurman WW, Dorland M. The role of genetic factors in age at natural menopause. Hum Reprod 2001;16:2014–2018. [PubMed: 11527915]

Dvornyk V, Long JR, Liu PY, Zhao LJ, Shen H, Recker RR, Deng HW. Predictive factors for age at menopause in Caucasian females. Maturitas 2006;54:19–26. [PubMed: 16213680]

Ennis S, Ward D, Murray A. Nonlinear association between CGG repeat number and age of menopause in FMR1 premutation carriers. Eur J Hum Genet 2006;14:253–255. [PubMed: 16251893]

Gajdos ZK, Butler JL, Henderson KD, He C, Supelak P, Euyud M. Association studies of common variants in ten hypogonadotropic hypogonadism genes with age at menarche. J Clin Endocrinol Metab. 2008 in press.

Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 2008;32:361–369. [PubMed: 18271029]

Gauvreau K. Hypothesis testing: proportions. Circulation 2006;114:1545–1548. [PubMed: 17015806]

Gorai I, Tanaka K, Inada M, Morinaga H, Uchiyama Y, Kikuchi R, Chaki O, Hirahara F. Estrogen-metabolizing gene polymorphisms, but not estrogen receptor-alpha gene polymorphisms, are associated with the onset of menarche in healthy postmenopausal Japanese women. J Clin Endocrinol Metab 2003;88:799–803. [PubMed: 12574216]

Guo Y, Shen H, Xiao P, Xiong DH, Yang TL, Guo YF, Long JR, Recker RR, Deng HW. Genomewide linkage scan for quantitative trait loci underlying variation in age at menarche. J Clin Endocrinol Metab 2006a;91:1009–1014. [PubMed: 16394082]

Guo Y, Xiong DH, Yang TL, Guo YF, Recker RR, Deng HW. Polymorphisms of estrogen-biosynthesis genes CYP17 and CYP19 may influence age at menarche: a genetic association study in Caucasian females. Hum Mol Genet 2006b;15:2401–2408. [PubMed: 16782804]

He C, Kraft P, Chen C, Buring JE, Pare G, Hankinson SE, Chanock SJ, Ridker PM, Hunter DJ, Chasman DI. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat Genet. 2009

He LN, Xiong DH, Liu YJ, Zhang F, Recker RR, Deng HW. Association study of the oestrogen signalling pathway genes in relation to age at natural menopause. J Genet 2007;86:269–276. [PubMed: 18305346]

Hefler LA, Grimm C, Heinze G, Schneeberger C, Mueller MW, Muendlein A, Huber JC, Leodolter S, Tempfer CB. Estrogen-metabolizing gene polymorphisms and age at natural menopause in Caucasian women. Hum Reprod 2005;20:1422–1427. [PubMed: 15774541]

Hefler LA, Worda C, Huber JC, Tempfer CB. A polymorphism of the Nos3 gene and age at natural menopause. Fertil Steril 2002;78:1184–1186. [PubMed: 12477509]

Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–1558. [PubMed: 12111919]

Hirschhorn JN, Altshuler D. Once and again-issues surrounding replication in genetic association studies. J Clin Endocrinol Metab 2002;87:4438–4441. [PubMed: 12364414]

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–874. [PubMed: 17529973]

Ito M, Yamada M, Hayashi K, Ohki M, Uetani M, Nakamura T. Relation of early menarche to high bone mineral density. Calcif Tissue Int 1995;57:11–14. [PubMed: 7671158]

Kaaks R, Lukanova A, Kurzer MS. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. Cancer Epidemiol Biomarkers Prev 2002;11:1531–1543. [PubMed: 12496040]

Kaprio J, Rimpela A, Winter T, Viken RJ, Rimpela M, Rose RJ. Common genetic influences on BMI and age at menarche. Hum Biol 1995;67:739–753. [PubMed: 8543288]

Kevenaar ME, Themmen AP, Rivadeneira F, Uitterlinden AG, Laven JS, van Schoor NM, Lips P, Pols HA, Visser JA. A polymorphism in the AMH type II receptor gene is associated with age at menopause in interaction with parity. Hum Reprod 2007;22:2382–2388. [PubMed: 17636279]

Kok HS, Onland-Moret NC, van Asselt KM, van Gils CH, van der Schouw YT, Grobbee DE, Peeters PH. No association of estrogen receptor alpha and cytochrome P450c17alpha polymorphisms with age at menopause in a Dutch cohort. Hum Reprod 2005;20:536–542. [PubMed: 15539439]

Koochmeshgi J, Hosseini-Mazinani SM, Morteza Seifati S, Hosein-Pur-Nobari N, Teimoori-Toolabi L. Apolipoprotein E genotype and age at menopause. Ann N Y Acad Sci 2004;1019:564–567. [PubMed: 15247087]

Kritz-Silverstein D, Barrett-Connor E. Early menopause, number of reproductive years, and bone mineral density in postmenopausal women. Am J Public Health 1993;83:983–988. [PubMed: 8328621]

Kvale G. Reproductive factors in breast cancer epidemiology. Acta Oncol 1992;31:187–194. [PubMed: 1622633]

Lai J, Vesprini D, Chu W, Jernstrom H, Narod SA. CYP gene polymorphisms and early menarche. Mol Genet Metab 2001;74:449–457. [PubMed: 11749050]

Li Y, Abecasis GR. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. Am J Hum Genet 2006;S79:2290.

Long JR, Shu XO, Cai Q, Cai H, Gao YT, Jin F, Zheng W. Polymorphisms of the CYP1B1 gene may be associated with the onset of natural menopause in Chinese women. Maturitas 2006;55:238–246. [PubMed: 16766147]

Long JR, Xu H, Zhao LJ, Liu PY, Shen H, Liu YJ, Xiong DH, Xiao P, Liu YZ, Dvornyk V, Li JL, Recker RR, Deng HW. The oestrogen receptor alpha gene is linked and/or associated with age of menarche in different ethnic groups. J Med Genet 2005;42:796–800. [PubMed: 16199553]

Mallolas J, Duran M, Sanchez A, Jimenez D, Castellvi-Bel S, Rife M, Mila M. Implications of the FMR1 gene in menopause: study of 147 Spanish women. Menopause 2001;8:106–110. [PubMed: 11256870]

McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. Br J Psychiatry 2007;190:194–199. [PubMed: 17329737]

Meyer JM, Eaves LJ, Heath AC, Martin NG. Estimating genetic influences on the age-at-menarche: a survival analysis approach. Am J Med Genet 1991;39:148–154. [PubMed: 2063916]

Mitchell ES, Farin FM, Stapleton PL, Tsai JM, Tao EY, Smith-DiJulio K, Woods NF. Association of estrogen-related polymorphisms with age at menarche, age at final menstrual period, and stages of the menopausal transition. Menopause 2008;15:105–111. [PubMed: 17589376]

Murabito JM, Yang Q, Fox C, Wilson PW, Cupples LA. Heritability of age at natural menopause in the Framingham Heart Study. J Clin Endocrinol Metab 2005a;90:3427–3430. [PubMed: 15769979]

Murabito JM, Yang Q, Fox CS, Cupples LA. Genome-wide linkage analysis to age at natural menopause in a community-based sample: the Framingham Heart Study. Fertil Steril 2005b;84:1674–1679. [PubMed: 16359963]

Must A, Phillips SM, Naumova EN, Blum M, Harris S, Dawson-Hughes B, Rand WM. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? Am J Epidemiol 2002;155:672–679. [PubMed: 11914195]

Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U, Gillson CJ, Glaser B, Golding J, Hardy R, Khaw KT, Kuh D, Luben R, Marcus M, McGeehin MA, Ness AR, Northstone K, Ring SM, Rubin C, Sims MA, Song K, Strachan DP, Vollenweider P, Waeber G, Waterworth DM, Wong A, Deloukas P, Barroso I, Mooser V, Loos RJ, Wareham NJ. Genetic variation in LIN28B is associated with the timing of puberty. Nat Genet. 2009

Pan F, Xiao P, Guo Y, Liu YJ, Deng HY, Recker RR, Deng HW. Chromosomal regions 22q13 and 3p25 may harbor quantitative trait loci influencing both age at menarche and bone mineral density. Hum Genet 2008;123:419–427. [PubMed: 18379822]

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190. [PubMed: 17194218]

Pechere-Bertschi A, Burnier M. Female sex hormones, salt, and blood pressure regulation. Am J Hypertens 2004;17:994–1001. [PubMed: 15485766]

Peeters PH, Verbeek AL, Krol A, Matthyssen MM, de Waard F. Age at menarche and breast cancer risk in nulliparous women. Breast Cancer Res Treat 1995;33:55–61. [PubMed: 7749133]

Perry JR, Stolk L, Franceschini N, Lunetta KL, Zhai G, McArdle PF, Smith AV, Aspelund T, Bandinelli S, Boerwinkle E, Cherkas L, Eiriksdottir G, Estrada K, Ferrucci L, Folsom AR, Garcia M, Gudnason V, Hofman A, Karasik D, Kiel DP, Launer LJ, van Meurs J, Nalls MA, Rivadeneira F, Shuldiner AR, Singleton A, Soranzo N, Tanaka T, Visser JA, Weedon MN, Wilson SG, Zhuang V, Streeten EA, Harris TB, Murray A, Spector TD, Demerath EW, Uitterlinden AG, Murabito JM. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat Genet. 2009

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909. [PubMed: 16862161]

Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65:220–228. [PubMed: 10364535]

Rees M. The age of menarche. Orgyn 1995:2–4. [PubMed: 12319855]

Ridker PM, Chasman DI, Zee RY, Parker A, Rose L, Cook NR, Buring JE. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. Clin Chem 2008a;54:249–255. [PubMed: 18070814]

Ridker PM, Cook NR, Lee IM, Gordon D, Gaziano JM, Manson JE, Hennekens CH, Buring JE. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med 2005;352:1293–1304. [PubMed: 15753114]

Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. Am J Hum Genet 2008b;82:1185–1192. [PubMed: 18439548]

Rothenbuhler A, Fradin D, Heath S, Lefevre H, Bouvattier C, Lathrop M, Bougneres P. Weight-adjusted genome scan analysis for mapping quantitative trait Loci for menarchal age. J Clin Endocrinol Metab 2006;91:3534–3537. [PubMed: 16804047]

Sharma K. Genetic basis of human female pelvic morphology: a twin study. Am J Phys Anthropol 2002;117:327–333. [PubMed: 11920368]

Snieder H, MacGregor AJ, Spector TD. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. J Clin Endocrinol Metab 1998;83:1875–1880. [PubMed: 9626112]

Stavrou I, Zois C, Chatzikyriakidou A, Georgiou I, Tsatsoulis A. Combined estrogen receptor alpha and estrogen receptor beta genotypes influence the age of menarche. Hum Reprod 2006;21:554–557. [PubMed: 16210384]

Stavrou I, Zois C, Ioannidis JP, Tsatsoulis A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. Hum Reprod 2002;17:1101–1105. [PubMed: 11925413]

Stolk L, Zhai G, van Meurs JB, Verbiest MM, Visser JA, Estrada K, Rivadeneira F, Williams FM, Cherkas L, Deloukas P, Soranzo N, de Keyzer JJ, Pop VJ, Lips P, Lebrun CE, van der Schouw YT, Grobbee DE, Witteman J, Hofman A, Pols HA, Laven JS, Spector TD, Uitterlinden AG. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. Nat Genet. 2009

Sulem P, Gudbjartsson DF, Rafnar T, Holm H, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Alexandersen P, Feenstra B, Boyd HA, Aben KK, Verbeek AL, Roeleveld N, Jonasdottir A, Styrkarsdottir U, Steinthorsdottir V, Karason A, Stacey SN, Gudmundsson J, Jakobsdottir M, Thorleifsson G, Hardarson G, Gulcher J, Kong A, Kiemeney LA, Melbye M, Christiansen C, Tryggvadottir L, Thorsteinsdottir U, Stefansson K. Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. Nat Genet. 2009

Sundarrajan C, Liao WX, Roy AC, Ng SC. Association between estrogen receptor-beta gene polymorphisms and ovulatory dysfunctions in patients with menstrual disorders. J Clin Endocrinol Metab 2001;86:135–139. [PubMed: 11231990]

te Velde ER, Pearson PL. The variability of female reproductive ageing. Hum Reprod Update 2002;8:141–154. [PubMed: 12099629]

Tempfer CB, Riener EK, Keck C, Grimm C, Heinze G, Huber JC, Gitsch G, Hefler LA. Polymorphisms associated with thrombophilia and vascular homeostasis and the timing of menarche and menopause in 728 white women. Menopause 2005;12:325–330. [PubMed: 15879922]

Treloar SA, Do KA, Martin NG. Genetic influences on the age at menopause. Lancet 1998;352:1084–1085. [PubMed: 9798581]

Treloar SA, Martin NG. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. Am J Hum Genet 1990;47:137–148. [PubMed: 2349942]

van Asselt KM, Kok HS, Peeters PH, Roest M, Pearson PL, te Velde ER, Grobbee DE, van der Schouw YT. Factor V Leiden mutation accelerates the onset of natural menopause. Menopause 2003;10:477–481. [PubMed: 14501610]

van Asselt KM, Kok HS, Putter H, Wijmenga C, Peeters PH, van der Schouw YT, Grobbee DE, te Velde ER, Mosselman S, Pearson PL. Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative trait loci influencing variation in human menopausal age. Am J Hum Genet 2004;74:444–453. [PubMed: 14872408]

van der Graaf Y, de Kleijn MJ, van der Schouw YT. Menopause and cardiovascular disease. J Psychosom Obstet Gynaecol 1997;18:113–120. [PubMed: 9219107]

van der Schouw YT, van der Graaf Y, Steyerberg EW, Eijkemans JC, Banga JD. Age at menopause as a risk factor for cardiovascular mortality. Lancet 1996;347:714–718. [PubMed: 8602000]

Weel AE, Uitterlinden AG, Westendorp IC, Burger H, Schuit SC, Hofman A, Helmerhorst TJ, van Leeuwen JP, Pols HA. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. J Clin Endocrinol Metab 1999;84:3146–3150. [PubMed: 10487678]

Worda C, Walch K, Sator M, Eppel W, Tempfer CB, Schneeberger C, Huber JC, Hefler LA. The influence of Nos3 polymorphisms on age at menarche and natural menopause. Maturitas 2004;49:157–162. [PubMed: 15474760]

Xita N, Tsatsoulis A, Stavrou I, Georgiou I. Association of SHBG gene polymorphism with menarche. Mol Hum Reprod 2005;11:459–462. [PubMed: 15879463]

Xu WH, Xiang YB, Ruan ZX, Zheng W, Cheng JR, Dai Q, Gao YT, Shu XO. Menstrual and reproductive factors and endometrial cancer risk: Results from a population-based case-control study in urban Shanghai. Int J Cancer 2004;108:613–619. [PubMed: 14696129]

Yaich L, Dupont WD, Cavener DR, Parl FF. Analysis of the PvuII restriction fragment-length polymorphism and exon structure of the estrogen receptor gene in breast cancer and peripheral blood. Cancer Res 1992;52:77–83. [PubMed: 1345763]

Yang F, Xiong DH, Guo Y, Shen H, Xiao P, Zhang F, Jiang H, Recker RR, Deng HW. The chemokine (C-C-motif) receptor 3 (CCR3) gene is linked and associated with age at menarche in Caucasian females. Hum Genet 2007;121:35–42. [PubMed: 17146638]

Zhao J, Xiong DH, Guo Y, Yang TL, Recker RR, Deng HW. Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in caucasian females. Hum Reprod 2007;22:1789–1794. [PubMed: 17376820]

Table 1

The proposed 9 groups of candidate genes in biological pathways or for related-phenotypes.

Group	Description
	Age at Menarche and Age at Natural Menopause
1	Steroid-hormone metabolism and biosynthesis pathway
2	Insulin-like growth factor (IGF) pathway
3	Transforming growth factor-beta (TGF-β) superfamily and signaling pathway
4	Thrombophilia and vascular homeostasis pathway
5	Obesity and obesity-related phenotypes
	Age at Menarche Only
6	Precocious or delayed puberty
	Age at Natural Menopause Only
7	Premature ovarian failure
8	Polycystic ovary syndrome
9	Smoking and nicotine dependence

Table 2a

Statistically significant associations at gene level for age at menarche.

Group	Gene	Chr.	SNP	Function	Position (bp)[a]	Major, Minor alleles	MAFb	Betac	s.e.	Raw p-valued	Adjusted p-valuee	Imputation quality (NHS/WGHS)	Other statistically significant SNPs in high linkage disequilibrium with the listed SNP (pairwise r2>0.8)
1	FSHB	11	rs555621	5′-UTR	30193120	T,C	0.41	0.059	0.016	0.00018	0.0011	1.00/0.97	rs601681, rs622356, rs639403, rs609896, rs1716022, rs560078, rs6169, rs506197, rs615577, rs618611, rs590833, rs489096, rs572883, rs543401, rs595496, rs518357, rs601640
			rs11031010	5′-UTR	30196754	C,A	0.15	0.079	0.023	0.00061	0.0037	0.99/0.87	
			rs1782507	5′-UTR	30200444	G,T	0.35	0.046	0.016	0.0035	0.021	1.00/1.00	
	LHCGR	2	rs4953616	Intronic	48853079	C,T	0.28	−0.065	0.017	0.00010	0.0060	1.00/1.00	rs4318432, rs6737881
			rs6729809	Intronic	48850313	T,C	0.33	0.063	0.017	0.00017	0.0099	0.98/0.95	
			rs7579411	Intronic	48850007	C,T	0.43	0.058	0.016	0.00022	0.013	1.00/1.00	rs4574159
			rs4374421	Intronic	48847043	T,C	0.32	0.060	0.017	0.00030	0.017	1.00/1.00	
	POMC	2	rs2164808	3′-UTR	25288827	C,T	0.45	0.046	0.016	0.0034	0.038	0.98/0.98	
			rs7589318	3′-UTR	25290023	G,A	0.32	0.048	0.017	0.0042	0.047	0.98/0.97	
	UGT2B4	4	rs2013573	Intronic	70535238	T,C	0.21	0.060	0.019	0.0012	0.019	1.00/1.00	
			rs13111134	Intronic	70535571	A,G	0.28	0.055	0.018	0.0026	0.042	0.95/0.87	
2	GHRL	20	rs1073768	3′-UTR	35310424	T,C	0.43	−0.046	0.016	0.0032	0.045	1.00/1.00	
4	CD40LG	X	rs5930973	Intronic	135457668	G,A	0.06	0.101	0.031	0.0013	0.0067	1.00/0.99	
			rs3092921	3′-UTR	135468520	T,C	0.09	0.074	0.028	0.0074	0.037	1.00/0.98	
6	FGFR1	8	rs2288696	Intronic	38405382	A,G	0.22	−0.058	0.019	0.0016	0.033	1.00/0.99	
	KISS1	1	rs7538038	Intronic	200892683	A,G	0.27	−0.061	0.018	0.00090	0.012	0.92/0.92	rs12097666, rs4951316, rs11240695, rs2510, rs4951319
	NKX2-1	14	rs999460	3′-UTR	36054262	C,T	0.36	0.046	0.016	0.0039	0.039	1.00/1.00	

[a] Positions based on NCBI build 35;

[b] Minor allele frequency among combined samples of the NHS and the WGHS;

[c] The regression parameter beta refers to the mean change in age at menarche or age at natural menopause per copy of the SNP minor allele;

[d] The unadjusted p-values are from linear regression with additive genetic coding;

[e] The p values are adjusted for multiple comparisons in each gene using the estimated effective number of independent markers. UTR, untranslated region.

Table 2b

tically significant associations at gene level for age at natural menopause.

p	Gene	Chr.	SNP	Function	Position (bp)[a]	Major, Minor alleles	MAF[b]	Beta[c]	s.e.	Raw p-value[d]	Adjusted p-value[e]	Imputation quality (NHS/WGHS)	Other statistically significant SNPs in high linkage disequilibrium with the listed SNP (pairwise r2>0.8)
	CYP19A1	15	rs11856927	Intronic	49335997	T,G	0.40	0.19	0.055	0.00051	0.023	0.93/0.93	
	FSHB	11	rs621686	5′-UTR	30203352	G,A	0.076	0.32	0.10	0.0012	0.0071	1.00/1.00	
			rs7951733	3′-UTR	30218544	G,A	0.082	−0.32	0.11	0.0026	0.016	0.87/0.73	
			rs11031010	5′-UTR	30196754	C,A	0.15	0.21	0.076	0.0058	0.035	0.99/0.87	
	LHCGR	2	rs1464729	3′-UTR	48817607	C,T	0.32	0.23	0.056	6.2E-05	0.0036	1.00/0.99	
	PGR	11	rs619487	Intronic	100477546	G,T	0.34	0.19	0.056	0.00090	0.026	1.00/0.97	rs560291, rs596223, rs1145463, rs480851, rs485283, rs493957, rs493220, rs550778, rs506487
	SRD5A1	5	rs494958	Intronic	6692179	T,A	0.16	0.23	0.072	0.0014	0.037	1.00/1.00	rs1651074, rs535981
	IGF1	12	rs1019731	Intronic	101366892	A,C	0.14	−0.28	0.076	0.00018	0.0053	1.00/1.00	
	IGF2R	6	rs9457827	Intronic	160474067	C,T	0.064	0.37	0.11	0.00063	0.038	1.00/1.00	
			rs2297362	Intronic	160469976	T,C	0.062	0.38	0.11	0.00068	0.041	0.99/1.00	rs1803989
	SMAD7	18	rs4939833	Intronic	44721958	C,T	0.057	−0.40	0.12	0.0010	0.029	0.92/0.88	rs2337104, rs2337109
	TGFBR1	9	rs1590	3′-UTR	98995720	G,T	0.26	0.18	0.061	0.0026	0.042	0.99/0.97	rs334353, rs10819639, rs334348, rs7042852, rs334349
	PCSK1	5	rs271924	Intronic	95763180	A,T	0.38	−0.18	0.055	0.0013	0.034	0.97/0.94	
	PPARG	3	rs4135280	Intronic	12423994	C,T	0.037	0.54	0.14	0.00014	0.0048	0.84/1.00	rs2292101, rs12489347, rs4135283
	TNF	6	rs909253	5′-UTR	31648292	G,A	0.34	0.20	0.055	0.00022	0.0034	1.00/1.00	rs1041981
	EIF2B4	2	rs7586601	3′-UTR	27496317	A,G	0.44	−0.19	0.055	0.00041	0.0020	0.95/0.92	
			rs12476704	5′-UTR	27524682	C,A	0.38	−0.18	0.054	0.00093	0.0047	1.00/0.99	rs13472, rs7602534, rs2280737, rs7594812, rs1647284, rs1528533, rs6760828
	POLG	15	rs2351002	Intronic	87670333	G,C	0.40	−0.18	0.055	0.0013	0.019	0.98/0.96	rs12899409, rs2072266, rs758130, rs2302084, rs2307449, rs2246900, rs976072
	NBN	8	rs2697679	Intronic	91019056	A,C	0.35	−0.24	0.055	1E-05	0.00021	1.00/0.97	rs2734823, rs1881469, rs2021881, rs2021882, rs9995, rs2142097, rs1063054, rs2735386, rs1063053, rs2735384, rs2735385, rs2735383
			rs2735387	Intronic	91020448	C,G	0.31	0.23	0.059	0.00013	0.0027	0.95/0.91	rs2735388, rs12680687
			rs7011299	5′-UTR	91082415	C,G	0.31	0.21	0.056	0.00020	0.0042	1.00/0.99	rs926092, rs2697677, rs1805794, rs2293775, rs1805833, rs16786,

p	Gene	Chr.	SNP	Function	Position (bp)[a]	Major, Minor alleles	MAF[b]	Beta[c]	s.e.	Raw p-value[d]	Adjusted p-value[e]	Imputation quality (NHS/WGHS)	Other statistically significant SNPs in high linkage disequilibrium with the listed SNP (pairwise r2>0.8)
													rs2097825, rs6985793, rs6470523, rs6470524, rs7840099, rs3736639
	ANKK1	11	rs6279	3′-UTR	112786283	C,G	0.30	0.17	0.057	0.0032	0.044	0.99/0.98	rs2734842

ns based on NCBI build 35;

allele frequency among combined samples of the NHS and the WGHS;

gression parameter beta refers to the mean change in age at menarche or age at natural menopause per copy of the SNP minor allele;

adjusted p-values are from linear regression with additive genetic coding;

values are adjusted for multiple comparisons in each gene using the estimated effective number of independent markers. UTR, untranslated region.

Table 3

Summary of pathway/group results for age at menarche and age at natural menopause.

Group	Biological Pathway or Related-phenotype	Number of Genes	Number of SNPs	Unadjusted p-value of Most Significant SNP (Associated Gene)	Observed Number of Significant Genes[a]	Expected Number of Significant Genes[b]	Observed: Expected	p-value for Group[c]
				Age at Menarche				
1	Steroid-hormone metabolism and biosynthesis	38	2,627	0.00010 (LHCGR)	4	1.9	2.1	**0.040**
2	IGF pathway	24	1,547	0.0032 (GHRH)	1	1.2	0.8	0.34
3	TGF-β superfamily and signaling pathway	49	2,878	0.0022 (INHA)	0	2.5	0.0	0.92
4	Thrombophilia and vascular homeostasis	32	2,069	0.00098(AGT)	1	1.6	0.6	0.48
5	Obesity and obesity-related phenotypes	36	2,437	0.00043 (FTO)	0	1.8	0.0	0.84
6	Precocious or delayed puberty	19	1,343	0.00074 (BCAT1)	3	1.0	3.2	**0.013**
Total		198	12,901	0.00010 (LHCGR)	9	9.9	0.9	0.53
				Age at Natural Menopause				
1	Steroid-hormone metabolism and biosynthesis	38	2,627	6.2E-05(LHCGR)	5	1.9	2.6	**0.011**
2	IGF pathway	24	1,547	0.00018 (IGF1)	2	1.2	1.7	0.12
3	TGF-β superfamily and signaling pathway	49	2,878	0.0010(SMAD7)	2	2.5	0.8	0.45
4	Thrombophilia and vascular homeostasis	32	2,069	0.0014 (ADAMTS9)	0	1.6	0.0	0.81
5	Obesity and obesity-related phenotypes	36	2,437	0.00041 (EIF2B4)	3	1.8	1.7	0.10
7	Premature ovary failure	13	648	0.0087 (EIF2B5)	2	0.7	3.1	**0.025**
8	Polycystic ovary syndrome	18	2,013	1.0E-05 (NBN)	1	0.9	1.1	0.23
9	Smoking and nicotine dependence	49	3,300	0.00064 (GABBR2)	1	2.5	0.4	0.71
Total		259	17,519	1.0E-05 (NBN)	16	13.0	1.2	0.16

[a] The number of genes with significant associations at gene level;

[b] The expected number of significant genes under the null hypothesis at conventional significance level of 0.05;

[c] p-values from the proportion test comparing the observed number of significant genes with the expected number of significant genes.

Table 4

Comparison of results from the current study on genes previously associated with age at menarche and age at natural menopause.

Gene	SNP	Ref. Allele	Effect Allele	Previous Study			Current Study		
				Direction[a]	p-value	Reference	Imputation Quality (NHS/WGHS)	Direction[a]	p-value[b]
Age at Menarche									
CCR3	rs3091309	G	A	ND	0.006	(Yang et al. 2007)	0.93/0.76	+	0.42
CCR3	rs6441948	G	A	ND	0.009	(Yang et al. 2007)	1.00/0.93	+	0.91
CYP17	rs743572	C	T	−	0.027	(Gorai et al. 2003)	NA	NA	N/A
CYP19	rs2445761	A	G	+	1.2×10^{-6}	(Guo et al. 2006b)	1.00/1.00	+	0.79
CYP19	rs2470144	A	G	ND	5.9×10^{-6}	(Guo et al. 2006b)	0.91/0.90	+	0.85
CYP19	rs6493496	A	T	ND	0.002	(Guo et al. 2006b)	NA	NA	N/A
CYP1B1	rs1056827	T	G	+	0.03	(Mitchell et al. 2008)	NA	NA	N/A
ESR1	rs2228480	G	A	+	0.03	(Long et al. 2005)	1.00/0.98	+	0.27
ESR1	rs3778082	G	A	−	0.03	(Long et al. 2005)	1.00/1.00	−	0.82
ESR1	rs9340799 (Xba I)	T	C	−	0.017	(Stavrou et al. 2002)	0.99/0.99	−	0.76
ESR2	rs4986938	A	G	−	0.005	(Stavrou et al. 2006)	0.97/0.92	−	**0.037**
IGF1	rs6214	G	A	ND	0.015	(Zhao et al. 2007)	1.00/1.00	−	0.86
Age at Natural Menopause									
AMHR2	rs2002555	A	G	−	0.054	(Kevenaar et al. 2007)	0.99/0.98	−	0.68
APOE	rs7412	C	T	−	0.03	(Tempfer et al. 2005)	NA	NA	N/A
CYP1B1	rs10012	C	G	+	0.02	(Long et al. 2006)	NA	NA	N/A
CYP1B1	rs1056827	G	T	+	0.04	(Long et al. 2006)	NA	NA	N/A
CYP1B1	rs1056836	C	G	−	0.004	(Long et al. 2006)	0.99/0.49	−	0.82
CYP1B1	rs1800440	A	G	−	0.018	(Hefler et al. 2005)	1.00/1.00	−	0.25
ESR1	rs2234693 (Pvu II)	T	C	+	0.03	(Weel et al. 1999)	0.99/1.00	+	0.12
F5	rs6020	G	A	−	0.03	(Tempfer et al. 2005)	NA	NA	N/A
HSDB1	rs2830	A	G	+	0.03	(Mitchell et al. 2008)	0.99/0.95	+	0.20

ND = not determined; NA = not available.

[a] Direction of the association. "+" means later onset of menarche or natural menopause when carrying the effect allele; "−" means earlier onset of menarche or natural menopause when carrying the effect allele.

[b] unadjusted p-value from linear regression with additive genetic coding.