HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2012 September 01.

Published in final edited form as:

Nat Genet. ; 44(3): 260–268. doi:10.1038/ng.1051.

Meta-analyses identify 13 novel loci associated with age at menopause and highlights DNA repair and immune pathways

A full list of authors and affiliations appears at the end of the article.

Abstract

To identify novel loci for age at natural menopause, we performed a meta-analysis of 22 genome-wide association studies in 38,968 women of European descent, with replication in up to 14,435 women. In addition to four known loci, we identified 13 new age at natural menopause loci ($P < 5 \times 10^{-8}$). The new loci included genes implicated in DNA repair (*EXO1*, *HELQ*, *UIMC1*, *FAM175A*, *FANCI*, *TLK1*, *POLG*, *PRIM1*) and immune function (*IL11*, *NLRP11*, *BAT2*). Gene-set

Corresponding authors: Anna Murray, Peninsula Medical School, St Lukes, Magdalen Road, Exeter, EX1 2LU, UK, anna.murray@pms.ac.uk. Kathryn L Lunetta, Department of Biostatistics, Boston University School of Public Health, 801 Mass Avenue, Crosstown Center, CT313, Boston, MA 02118, klunetta@bu.edu.

*These authors contributed jointly to this work

#These senior authors jointly oversaw this work

URLS:

GWAS catalog: www.genome.gov/gwastudies/,

SCAN: scan.bsd.uchicago.edu.

IPA: www.ingenuity.com

Financial Disclosures:

None

Author Contributions

Individual Study design/management: DIC, CHe, EMB, PK, JSB, MBo, EB, HC, SJC, UdF, IJD, GVZD, SE, JGE, LuF, ARF, PG, CJMWvG, CG, DEG, PH, SEH, AH, EI, SLRK, DAL, PKEM, MMar, NGM, VM, NCOM, GPa, ANP, NLP, PHMP, OP, BMP, KR, IR, JMS, RS, VJMP, VG, GE, TBH, LJL, YVS, AT, YTvdS, PMV, GWa, HEW, JFW, BHRW, LL, DIB, JEB, LC, FBH, DJH, GWM, BO, PMR, DS, TDS, KSte, EAS, MU, CMvD, HV, SB, ASa, AMe. Data collection: BZA, SB, LBL, JSB, MBo, AB, HC, PdA, UdF, IJD, GVZD, SE, JGE, LiF, LuF, ARF, DEG, PH, SEH, ACH, AH, ACJWJ, IK, SL, DAL, PKEM, NGM, IMK, ABN, NLP, PHMP, OP, BMP, KR, IR, ASc, SNS, JMS, MGS, US, BT, LT, SU, VJMP, VG, GE, YTvdS, CHvG, JMV, GWa, HEW, AFW, TZ, LZ, MCZ, MZ, LL, AMA, JEB, EAS, AGU, JMM, AMe. Genotyping: JK, LS, TE, EB, FJB, SJC, PdA, GD, PD, CHa, ACH, JLi, BCJMF, SEM, VM, PN, DRN, NCOM, AP, ANP, BMP, JIR, ASi, KSti, AT, KT, YTvdS, MV, EW, TZ, LL, FBH, GWM, BO, UT. Genotype prep: JRBP, MBa, NF, EP, SYS, WVZ, LBL, PdA, GD, CJMWvG, CG, CHa, EI, SLRK, PKEM, IMK, CM, SEM, PN, DRN, NCOM, AP, GPi, ER, CS, JAS, HS, NS, YVS, AT, MV, EW, TZ, LL, FBH, DJH, SS, KLL, MH. Phenotype preparation: JRBP, LS, CHe, MMan, MBa, LB, EMB, FE, NF, DFG, JJH, PK, GZ, WVZ, LBL, MBo, EJCdG, IJD, GVZD, MGS SE, JGE, CHvG, LiF, KF, MH, CHa, EI, ACJWJ, SLRK, IK, JLa, SL, TL, DAL, LML, PKEM, IMK, CM, PHMP, GPi, OP, ER, CS, ASc, JAS, BT, SU, RMvD, VG, GE, TA, YTvdS, HW, GWi, BHRW, AFW, MCZ, LL, AMA, EWD, AMe, KLL, JMM. Analysis plan development: LS, DIC, CHe, EMB, PK, TC, PG, DK, DPK, NGM, DT, DJH, GWM, KLL, JMM, AMu. Analysis plan review: LS, DIC, CHe, AVS, ADC, NG, DK, DPK, BM, ABN, BMP, AMA, DJH, KLL, JMM, AMu. Study Data Analysis: LS, DIC, CHe, MMan, MBa, LB, EMB, FE, TE, NF, DFG, JJH, PK, PFM, EP, SYS, AVS, SvW, GZ, WVZ, EA, CC, MCC, TC, NG, TH, CHa, ZK, JLa, DAL, LML, DM, NCOM, GPa, GPi, ASP, VE, ER, JAS, MGS, SU, YTvdS, MV, LMY, LZ, SS, UT. Reviewed/Interpret analyses: JRBP, LS, DIC, CHe, PS, MBa, EMB, NF, PK, PFM, AVS, BZA, SJC, ADC, IJD, SEH, EI, DK, SLRK, DPK, LML, PKEM, NGM, GPa, ASP, JAS, HS, LMY, JEB, EWD, FBH, DJH, GWM, PMR, CMvD, HV, KLL, JMM, AMu. Meta analyses: LS, PS, KLL. Pathway/other analyses: JRBP, DIC, CHe, ADJ, MMan, GPa, JAV. Meno Study Design: JRBP, LS, DIC, CHe, DEG, PHMP, YTvdS, CHvG, AMe, KLL, JMM, AMu. Manuscript Preparation: JRBP, LS, DIC, CHe, TE, NF, ADC, DK, DPK, EWD, AMe, KLL, JMM, AMu, JAV. Manuscript review: JRBP, LS, DIC, CHe, NF, PK, PFM, AVS, BZA, TA, LBL, JSB, MBo, FJB, HC, SJC, CC, MCC, ADC, GD, UdF, IJD, GE, BCJMF, MEG, NG, DEG, PH, SEH, CHa, EI, DK, DPK, SLRK, JAS, IK, ZK, TL, TE, ASa, AMe, JSEL, JLi, LML, YVL, PKEM, IMK, BM, VM, PN, ABN, NCOM, GPa, ANP, NLP, PHMP, OP, BMP, JIR, IR, HS, JMS, RS, AT, RMvD, YTvdS, CHvG, PMV, MV, GWa, JFW, BHRW, AFW, LMY, TZ, LZ, MCZ, VJMP, LL, AMA, DIB, JEB, EWD, VG, TBH, FBH, DJH, LJL, PMR, TDS, EAS, HV, KLL, JMM, AMu, JAV. Oversaw Consortium: KLL, JMM, AMu, JAV.

Some of the authors declare competing financial interests: details accompany the full-text HTML version of the paper

enrichment pathway analyses using the full GWAS dataset identified exodeoxyribonuclease, NFκB signalling and mitochondrial dysfunction as biological processes related to timing of menopause.

Introduction

Menopause is the cessation of reproductive function of the human ovaries. This life stage is associated with one of the major hormonal changes of women characterized by a decline in secretion of estrogen, progesterone and, to a lesser degree, testosterone. It influences a woman's well-being and is associated with several major age-related diseases including cardiovascular disease, breast cancer, osteoarthritis, and osteoporosis[1]. Ovarian aging is reflected by the continuous decline of the primordial follicle pool, which is established during fetal life, subsequently leading to endocrine changes due to loss of the negative feedback from ovarian hormones on the hypothalamic-pituitary axis. In addition to follicle loss, oocyte quality diminishes with increasing age, which is believed to be due to increased meiotic nondisjunction[2]. Oocyte quality may be controlled at the time germ cells are formed during fetal life, but may also reflect accumulated damage during reproductive life, and/or age-related changes in granulosa cell-oocyte communication[3]. Although both oocyte quantity and quality decline with increasing age, it is not clear whether they are controlled by the same mechanisms and whether they decline in parallel.

The average age at natural menopause in women of Northern European descent is 50–51 years (range 40–60 years)[4]. Heritability estimates from twin and family studies for age at natural menopause range from 44 to 65%[5–8]. Thus far most genetic association studies regarding age at menopause have focussed on candidate genes[9] from the estrogen pathway[10,11], or vascular components[12,13]. Recently two genome-wide association studies (GWAS) identified five novel loci associated with age at natural menopause on chromosomes 5, 6, 13, 19 and 20[14,15]. These loci, however, explained <1.5% of the phenotypic variation of age at natural menopause, suggesting additional loci of small effect are likely to be discovered in larger samples. Therefore, we conducted a two-stage GWAS of women of European ancestry, combining the women from the two previous GWAS studies[14,15] with new participants for a total of 38,968 women from 22 studies in the discovery stage, and 14,435 women from 21 studies in the replication stage.

Results

In our discovery stage of 38,968 women with natural menopause between the ages of 40 and 60 (Supplementary Table 1, Supplementary Table 2), we identified 20 regions with SNPs meeting the genome-wide significance criterion $p<5 \times 10^{-8}$ (Figure 1a). Four of these loci confirmed prior reports of associations on chromosomes 5, 6, 19, and 20[14,15] (regions 5b, 6a, 19a, and 20, respectively, in Table 1) and 16 loci were novel. We failed to confirm one previously reported association on chromosome 13 (13q34, rs7333181, p=0.12). The overall genomic inflation factor was 1.03 (Figure 1b). Table 1 displays the SNP with the lowest p-value from each region. There was no between-study effect heterogeneity across discovery studies (p>0.05/20) for the 20 SNP associations presented. Within FHS, we tested for

differences in effect size for the 20 SNPs in retrospectively and prospectively collected menopause age, and also found no significant differences (data not shown). The effect sizes ranged from 0.17 years (8.7 weeks) to nearly one year (50.5 weeks) per each copy of the minor allele. We computed the effect sizes for dichotomized age at natural menopause in the WGHS women, dichotomizing at age 45 (N=745) versus later for early menopause and at 54 (N=1632) versus earlier for late menopause. The estimated odds ratios for early menopause for the menopause decreasing allele ranged from 1.01 to 2.03. The estimated odds ratios for late menopause for the menopause decreasing allele ranged from 0.52 to 0.96 (Supplementary Table 3). The top SNPs in regions 2c, 5a, and 19b were more than 400kb but less than 1Mb from the top SNP in another region on the same chromosome. The top SNP in each of these primary regions had low linkage disequilibrium ($r^2<0.5$) with the top SNP in the nearby region. To determine whether these associations were independent, we performed a conditional association analysis in the discovery study samples with the most significant SNP from each of the primary 17 regions included as covariates in the analysis. For regions 5a and 19b (rs890835, rs12461110), the effect estimates in the conditional analysis were unchanged compared to the discovery analysis (differences of 0.3% and 4%, respectively), and the p-values remained genome-wide significant. However, for region 2c the effect size was decreased by ~12.5% in the conditional analysis compared to the initial analysis, and the SNP p-value was no longer genome-wide significant ($p=9.8\times10^{-7}$) (Table 1), suggesting that the association with rs7606918 is not independent of the rs1018348 region 2b association. We attempted replication only for the 19 SNPs that represented independent regions that reached genome-wide significance ($p<5 \times 10^{-8}$) thus replication of rs7606918 was not pursued.

Replication

Twenty-one studies contributed 14,435 women for replication of the 19 SNPs defining the independent genome-wide significant regions from stage 1. Age at natural menopause was defined using the same criteria as in the discovery studies (Supplementary Table 1). Seventeen of these studies (N=6,639) were included in *in silico* replication (Supplementary Table 2); an additional 4 studies (N=7,796) contributed *de novo* genotypes for the 19 SNPs (Supplementary Table 2). Table 1 displays the effect sizes and p-values for replication and a combined meta-analysis of the discovery and replication samples. There was no evidence for effect heterogeneity among the replication studies (Table 1). Further, we tested for heterogeneity between the *in silico* and *de novo* genotyped studies, and found no evidence for heterogeneity of effect (data not shown), suggesting that for the significant SNPs, the genotype imputation methods did not result in significantly different effect size estimates than would have been obtained from direct genotyping. Seventeen of the 19 SNPs remained genome-wide significant and had lower p-values in combined meta-analysis of the discovery and replication samples. Regions 5a and 13a showed no evidence of association in the replication samples ($p>0.50$) and were not genome-wide significant in combined discovery and replication meta-analysis. Four of the 17 replicated regions were reported previously; thus our analysis identified 13 novel regions for age at natural menopause based on genome-wide significant discovery with replication. In the combined discovery and replication meta-analyses the effect estimates ranged from 8.2 to 49.3 weeks per minor allele. The estimated proportion of variance explained by the 17 replicated SNPs in the four replication studies

with *de novo* genotyped SNPs ranged from 2.5% (Osteos) to 3.7% (EPOS and BWHHS) to 4.1% (PROSPECT-EPIC).

We used the largest study contributing data to our discovery GWAS (WGHS, N=11379) to explore whether substantial SNPxSNP interactions are present among the 17 replicated SNPs. We tested all 136 pairs of SNPs and found no evidence for interaction (all P>0.01).

Roles of genes at/near novel loci

All but two of the replicated SNPs are intronic or exonic to known genes (Table 2). The top SNPs in regions 6b, 12, 19b, and 20 are missense polymorphisms. Three of the four were predicted to have damaging protein function by SIFT[16], and one by PolyPhen2[17]. Using dbSNP and LocusZoom[18], we identified the genes underlying the novel top regions. We used SCAN (see URLs) to identify all genes with SNPs that are in linkage disequilibrium (LD) (r^2>0.5) with our SNPs (Table 2). We identified all SNPs with $r^2 \geq 0.8$ with our top SNPs and used several databases to determine if the SNPs were known to be associated with expression (Table 2).

The strongest novel signal was on chromosome 4 (region 4, rs4693089; $P=2.4\times10^{-19}$). The SNP is located in an intron of *HELQ*, the gene that encodes the protein HEL308, which is a DNA-dependent ATPase and DNA helicase [19]. The second strongest novel signal was on chromosome 12 (region 12, rs2277339; $P=2.5\times10^{-19}$). This SNP is a non-synonymous variant in exon 1 of *PRIM1*. The top SNP was significantly associated with expression of *PRIM1* in visual cortex, cerebellum, and pre-frontal cortex (Table 2).

Several other novel signals are located in introns of genes for which mouse models exist. These were region 8 in *ASH2L* (rs2517388; $P=9.3\times10^{-15}$), region 13b in *POLG* (rs2307449; $P=3.6\times10^{-13}$) and region 1b in *EXO1* (rs1635501; $P=8.5\times10^{-10}$). *ASH2L* codes for a trithorax group protein, and is involved in X chromosome inactivation in women[20]. *POLG* encodes the catalytic subunit of mitochondrial DNA polymerase, the enzyme responsible for replication and repair[21] of mitochondrial DNA. *EXO1* is a member of the RAD2 nuclease family of proteins involved in DNA replication, repair and recombination, and the top hit is in LD ($r^2=0.83$) with a functional polymorphism in *EXO1* that affects a transcription factor binding site in the promoter. Region 11 (rs12294104; $P=1.5\times10^{-11}$) is near and in LD ($r^2=0.92$) with SNPs in *FSHB*. Transcription of *FSHB* limits the rate of production of the heterodimeric follicle stimulating hormone (FSH), a key pituitary expressed hormone that stimulates maturation of follicles. Region 19a (rs11668344; $P=1.5\times10^{-59}$) is in tight LD with SNPs in *IL11*: this cytokine stimulates the T-cell-dependent development of immunoglobulin-producing B cells.

The top SNPs in two other novel regions are non-synonymous coding variants. Region 6b, rs1046089 ($P=1.6\times10^{-16}$) is in exon 22 of *BAT2* and was associated with expression of several transcripts in the HLA region in several tissues (Table 2). Region 19b, rs12461110 ($P=8.7\times10^{-10}$) is in exon 5 of *NLRP11*. *BAT2* encodes the gene for HLA-B associated transcript 2 and has several microsatellite repeats. *NLRP11* encodes for the NLR family, pyrin domain containing 11 protein, which is implicated in the activation of proinflammatory caspases[22].

Of the remaining five novel regions, the top SNPs for regions 1a, 2a, 2b, and 13b are located in introns. These were rs4246511 in *RHBDL2* (0.24 years/minor allele, $P=9.1\times10^{-17}$), which is thought to function as intramembrane serine proteases, rs2303369 in *FNDC4* coding for fibronectin type III domain containing 4 ($P=2.3\times10^{-12}$), rs10183486 in *TLK1* ($P=2.2\times10^{-14}$) a nuclear serine/threonine kinases that is potentially involved in the regulation of chromatin assembly, and rs4886238 in *TDRD3* ($P=9.5\times10^{-11}$). *TDRD3* is a known binding partner for *FMR1*, which has been associated with primary ovarian insufficiency. The top SNP in the final novel region, 16, is within 60kb of three genes: *TNFRSF17, GSPT1*, and *RUNDC2A*. It is in LD ($r^2>0.5$) with SNPs in these three genes as well as four others (rs10852344; $P=1.0\times10^{-11}$) (Table 2).

Pathway analyses

We used three independent pathway-based methods to identify connections among our single marker associations, and link them with broader biological processes. While all three approaches (Ingenuity Pathway Analysis(IPA, see URLs), MAGENTA[23] and GRAIL[24]) are based on published data, linking the gene products of our top hits to each other in functional pathways, each uses substantially different methodology and take different aspects of our results as input. Thus, we expect complementary results from the three approaches.

We used IPA(See URLs), to identify potential biological pathways common to the 17 replicated SNPs. Based on the genes physically nearest the 17 loci, four major functional networks were identified based on direct interactions only (Supplementary Table 4). Network 1, related to “lipid metabolism, molecular transport, and small molecule biochemistry”, contained 14 of the genes nearest to the menopause loci ($P=1\times10^{-30}$). Central to this network is the *HNF4A* gene, which is known to play a role in diabetes. Network 2, containing 12 of the input genes, relates to “cell cycle, cell death, and cancer” ($P=1\times10^{-24}$). The *ESR1* gene is central in this network, suggesting that genes in this network influence or are influenced by estrogen signaling. Network 3, also in part related to “cell death”, including TNF and NFκB ($P=1\times10^{-19}$). Network 4 relates to “infection mechanism, DNA replication, recombination, and repair, and gene expression” ($P=1\times10^{-12}$). Interestingly, several of the input genes included in network 1 (*EXO1*, *HELQ*) and network 2 (*UIMC1*, *FANCI*, *TLK1*) are also involved in DNA repair mechanisms.

We used a gene set enrichment analysis (GSEA) implemented in MAGENTA[23] to explore pathway-based associations using the full GWAS results. Three pathways reached study-wide significance (FDR < 0.05): exodeoxyribonuclease (P = 0.0005), NFκB signalling (P = 0.0006) and mitochondrial dysfunction (P = 0.0001) (Supplementary Table 5).

Finally, we used the GRAIL method of literature based pathway analysis[24] to explore the connections between the genes near our top SNPs. Genes are considered related if they share informative words. GRAIL scores for genes associated with 3 of the replicating genome-wide significant SNPs were significant: *EXO1*, *FKBPL*, and *BRSK1*. When applied to a deeper set of 66 SNPs from the discovery meta-analysis with significance meeting FDR<0.05, 12 genes had significant GRAIL scores: *EXO1, MSH6, PARL, RHBDL2, FKBPL, TP53BP1, TLK1, RAD54L, CHEK2, H2AFX, APEX1, REV3L*. In addition, *BRSK1* was borderline significant with GRAIL FDR=0.06.

Candidate genes

Within the discovery GWAS 18,327 SNPs were within 60kb of the start and end of transcription of 125 candidate genes selected because of a reported relationship with ovarian function (Supplementary Table 6). After multiple testing correction, 101 SNPs in or near five of the candidate genes (*DMC1*, *EIF2B4*, *FSHB*, *POLG*, *RFPL4A*) were significantly associated with age at natural menopause after multiple testing correction. SNPs in or near four of these genes were already identified as genome-wide significant (*EIF2B4*, region 2a; *RFPL4A*, region 19b; *POLG*, region 15; and *FSHB*, region 11). For the other gene, *DMC1*, the most significant SNP was rs763121, with nominal $P=1.6\times10^{-7}$, (P=0.0009 corrected for candidate gene SNP analyses); age at natural menopause decreases by ~0.18 years per copy of the minor allele. *DMC1* encodes for a protein that is essential for meiotic homologous recombination and is regulated by *NOBOX*, mutations in which can cause Primary Ovarian Insufficiency (POI)[25–27]

Pleiotropy of primary hits

We examined overlap of our significant regions against published GWAS results for other traits (GWAS catalog; see URLs). Twelve menopause loci were within 1Mb of a previously published genome-wide significant SNP, but most of the co-localised SNPs were in low LD ($0 < r^2 < 0.21$) with our SNP in the region (Supplementary Table 7). The exception was at the *GCKR* locus on chromosome 2. Region 2a (rs2303369) was correlated ($r^2 \approx 0.5$) with four different SNPs reported to influence kidney function, type 2 diabetes, continuous glycaemic traits, as well as serum albumin, C reactive protein, serum urate, and triglycerides. These results increase the observed clustering of signals in complex trait genetics, whilst also adding to the increasing pleiotropy observed at the *GCKR* locus.

Discussion

In this large 2-stage GWAS we have confirmed four previously established menopause loci and identified and replicated 13 novel loci associated with age at natural menopause. Of these 17 hits, all but two are intronic or exonic to known genes. On average for associated SNPs in GWAS studies, 40% are intergenic, while only2% of our hits are intergenic. Further, our study finds two times more non-synonymous top hits as typically seen in GWAS (24% vs 12%)[28]. The 17 replicated loci function in diverse pathways including hormonal regulation, immune function and DNA repair. Together, they explained 2.5–4.1% of the population variation in menopausal age in independent replication samples. Biological pathway analysis of the genetic associations with age at natural menopause in this study using distinct algorithms and databases were in close agreement in emphasizing general biological pathways for mitochondrial function, DNA repair, cell cycle and cell death, and immune response.

Aging is thought to result from the accumulation of somatic damage[29]. Analysis of gene expression patterns in aging organs, such as heart and brain, identified changes in genes involved in inflammatory response, oxidative stress, and genome stability[30], processes also identified in analysis of age-related changes in mouse oocytes, including changes in mitochondrial function[31]. Comparisons of lifespans across species showed that there is a

general relationship between longevity and DNA repair function[32]. This notion is reinforced in the Werner and Bloom syndromes, which involve genome instability due to mutations in 3′–5′ DNA helicases of the RecQ family members, and are characterized by both premature aging and premature menopause[33]. Similarly, an increase in meiotic errors is associated with an age-related decline in oocyte quality, compounding the progress toward menopause due to follicle depletion[34].

In the biological pathways analysis, seven candidate genes identified by proximity to the 17 genome-wide significant associations with age at natural menopause are related to DNA damage repair and replication (*EXO1*, *HELQ*, *UIMC1*, *FAM175A*, *FANCI*, *TLK1*, *POLG*, *PRIM1*) (Supplementary Table 4). The protein encoded by *UIMC1* physically interacts with BRCA1 and estrogen receptor α (ERα) and is thought to recruit BRCA1 to DNA damage sites and to initiate G2/M checkpoint control. PRIM1 (Primase) is involved in DNA replication by synthesising RNA primers for Okazaki fragments during discontinuous replication[35]. A mutation in *POLG* can segregate with primary ovarian insufficiency (POI)[36]. *Polg* knock-in mice show reduced lifespan, premature ageing, and reduced fertility[37]. FANCI, as a second gene at the same locus adjacent to POLG, is a member of the Fanconi anemia (FA) complementation group. FA is a recessive disorder characterized by cytogenetic instability and defective DNA repair. FA patients experience irregular menstruations with menopause occurring around the age of 30[38]. The functional polymorphism correlated to our top hit in *EXO1*, has been associated with longevity in female centenarians[39]. Male and female *Exo1* knock-out mice are sterile, because the gene is essential for male and female meiosis[40]. In addition to the GWAS regions in/near genes previously associated with early menopause, we investigated a panel of candidate genes identified prior to the study, and found a SNP near the meiotic recombination gene *DMC1* to be significantly associated with age at menopause. How the DNA repair pathways contribute to menopause remains unclear. It is possible that with altered DNA repair mechanisms damage accumulates rendering poor quality oocytes for selection. On the other hand, the number of damaged follicles may increase with ageing, resulting in an increased rate of follicle loss through atresia. Next to this, our top hit in this study, one of the four known hits, is a non-synonymous SNP in *MCM8*, was not included in the IPA results, probably because the exact function of this protein is still unknown. The MCM family, however, is a key component of the pre-replication complex and its main function is to restrict DNA replication to one round per cell cycle[41].

The pathway analyses highlighted additional candidate genes with functions in DNA repair (exodeoxyribonuclease), but with sub-genome-wide levels of significance for association with age of natural menopause. These 12 candidates (Supplementary Table 5) included the Werner (WRN) helicase gene, mutations in which cause Werner syndrome, a classic progeria with advanced ageing phenotype, and ovarian aging[42]. Estrogen can enhance WRN expression preventing cell senescence, suggesting its involvement in menopause[43]. The identification of DNA repair as one of the biological pathways involved in menopause may also provide an explanation for the association between smoking and an earlier age at menopause. Damage caused by smoking activates several different DNA repair mechanisms. Indeed, a polymorphism in Exo1, one of our top loci, has been associated with

colorectal adenomas in smokers only[44]. Future studies will reveal whether smoking status modifies the association between age at natural menopause and polymorphisms in DNA repair genes, as has been observed for various cancers.

The pathways-based analysis also emphasized that genes related to auto-immune disease also influence age at natural menopause. This link has not been reported before, however, in a proportion (2–10%) of women with POI, ovarian auto-immunity can play a role[45]. In addition, POI is frequently associated with additional autoimmune diseases, such as type 1 diabetes mellitus [46]. The top SNP in region 19a is near *IL11*, which binds the IL11 receptor alpha chain (IL-11Rα). Female mice with null mutations in *IL-11Rα* are infertile due to defective uterine decidualization, the process necessary for successful embryo implantation[47]. *NLRP11* (region 19b) is a member of the *NLRP* family of genes that play important roles in the innate immune system and reproductive system. Several *NLRP* genes show an oocyte specific expression pattern [46], while *NLRP5* has been implicated in POI, and serves as an autoantigen in a mouse model of autoimmune POI[48,49]. Many autoimmune conditions are associated with a particular HLA type, but no such association has been reported for POI[50,51]. One of our top menopause associations (rs1046089) is a missense substitution in *BAT2* (HLA-B associated transcript), which is in the HLA class III complex on chromosome 6 and has been associated with type I diabetes mellitus and rheumatoid arthritis. Multiple phenotypes have been associated with *BAT2* SNPs in GWAS studies including BMI, neonatal lupus, HIV control and height (Supplementary Table 7), but the SNPs have low correlation with our top hit. Expression data for rs1046089 demonstrated that the polymorphism was associated with altered expression of HLA-DRB4 in monocytes and HLA-DQA1 in lymphoblastoid cell lines (Table 2). Thus, this gene is an excellent candidate for a pro-inflammatory component to oocyte depletion affecting menopause age. Indeed, the enrichment of the genes involved in NFκB signalling (*TNF*, *TNFRSF17*, *CSNK2B*) in the biological pathways analysis suggests that susceptibility to inflammation, which often accompanies immunosenescence in aging, may also affect ovarian aging. The finding that the innate immune response can be upregulated in response to DNA damage[52] suggests that interplay between the two main pathways we have identified (DNA repair and inflammation), may contribute to variation in the age at natural menopause.

Three of the 17 regions can be linked to hormonal regulation, an additional route to follicle pool exhaustion. The top region 11 SNP (rs12294104) is in high LD with SNPs in *FSHB* (r^2=0.92, Table 2), a gene which limits the rate of production of follicle stimulating hormone (FSH), a key pituitary expressed hormone that stimulates maturation of follicles. FSH-deficient female mice are infertile[53]. Transgenic mice with FSH over-expression show premature infertility due to postimplantation reduction of embryo-fetal survival[54]. FSH concentrations rise in women approaching menopause, which might be related to a decrease in growing follicles[55]. Mutations in *FSHB* cause hypogonadism and primary amenorrhea in women[56] and raised FSH levels and infertility in males[57]. The latter observation is due to a promoter polymorphism that may be causal[58] and is in LD (r^2 = 0.7) with our most significant SNP. Although *STAR*, encoding steroidogenic acute regulatory protein (StAR) was not the nearest gene to the top SNP in region 8 (rs2517388), its functional role in cleavage of cholesterol to pregnenolone in response to tropic hormones makes it the likely

functional candidate, and our top SNP is in high LD with SNPs in that gene (r^2=0.81, Table 2). Pregnenolone is a precursor for several steroid hormones, such as estrogen and progesterone, and mutations in the *StAR* gene are associated with congenital lipoid adrenal hyperplasia and POI[59]. Furthermore *STAR* is a target of FOXL2, for which truncating mutations are preferentially associated with POI[60]. Similarly, known biological function suggests that *BCAR4*, encoding the breast cancer antiestrogen resistance 4 protein, is the best candidate gene near region 16. *BCAR4* is expressed only in placenta and oocytes and may play a role in hormonal stimulation in the ovary. In breast cancer treatment, tumours expressing higher levels of *BCAR4* are more resistant to tamoxifen treatment[61], reinforcing its role in transduction of hormonal signals.

In summary, our findings demonstrate the role of genes which regulate DNA repair and immune function, as well as genes affecting neuroendocrine pathways of ovarian function, indicating the process of ageing as a shared player in both somatic and germ line ageing.

We expect a substantial number of additional common variants with small effects on ANM are yet to be identified, and that many of them will be in genes that are in the pathways identified in this study. Sequencing and exome chip studies to determine whether low frequency and rare variants of large effect also contribute to ANM are underway or being planned in many of the cohorts involved in this GWAS. A collaboration of several consortia is currently examining the contribution of common genetic variants to ANM in African American (AA) women, and will allow us to determine whether the genetic variation that affects ANM in AA women are the same or substantially different than for women of primarily European descent. We are currently conducting a study of women with primary ovarian insufficiency to determine whether the variants that are associated with ANM within the normal range also contribute to disease conditions related tothe early menopause phenotype.

Methods

Discovery

Age at natural menopause was defined as the age at the last menstrual period which occured naturally with at least 12 consecutive months of amenorrhea. This analysis included women with natural menopause between the ages of 40 and 60 years. 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. Most cohorts collected the age at natural menopause retrospectively; in the Framingham Offspring, ARIC, NHS and WGHS studies, some women became menopausal under study observation. Study specific questions, mean age at menopause and age at interview are shown in Supplementary Table 1. Genotyping and imputation information for the discovery cohorts are shown in Supplementary Table 2. Descriptions of each study are in the Supplementary Note.

Replication

A total of 14,435 women from 21 studies meeting the same inclusion and exclusion criteria as the discovery analysis women were included in the replication analysis. The women had mean and standard deviation of age at natural menopause similar to the discovery set (Supplementary Table 1). Genotyping and imputation methods for the *in silico* replication cohorts are shown in Supplementary Table 2. Genotyping information for the studies that genotyped the SNPs *de novo* is shown in Supplementary Table 2. Descriptions of each study are in the Supplementary Note.

The 19 independent genome-wide significant SNPs were tested for association with age at natural menopause using linear regression models.

Meta-analysis

Inverse variance weighted meta-analysis of the studies was performed using METAL using genomic control[62]. A SNP within a study was omitted if the minor allele frequency (MAF) was < 1% or imputation quality score was < 0.2. The discovery meta-analysis included 2,551,160 autosomal SNPs and 38,968 samples.

eQTL analysis

For each of the genome-wide significant menopause SNPs (Table 1), all proxy SNPs with r^2>0.8 were determined in HapMap CEU release 22. Each SNP and its proxies were searched against a collected database of expression quantitative trait locus (eQTL) results, including the following tissues: fresh lymphocytes[63], fresh leukocytes[64], leukocyte samples in individuals with Celiac disease[65], lymphoblastoid cell lines (LCL) derived from asthmatic children[66], HapMap LCL from three populations[67], a separate study on HapMap CEU LCL[68], fibroblasts, T cells and LCL derived from cord blood[69], two studies on peripheral blood monocytes[70,71], CD4+ lymphocytes[72], adipose and blood samples[73], two studies on brain cortex[70,74], three large studies of brain regions including prefrontal cortex, visual cortex and cerebellum (Emilsson, personal communication), a study of cerebellum, frontal cortex, temporal cortex and caudal pons[75], a separate study on prefrontal cortex[76], liver[77], and osteoblasts[78]. The collected eQTL results met criteria for statistical significance for association with gene transcript levels as described in the original papers. Table 2 provides a summary of eQTL findings for replicated GWAS SNPs.

Conditional analysis

On each chromosome, the lowest p-value SNPs that met genome-wide significance were identified. Genome-wide significant SNPs more than 250,000 base pairs and less than 1 million base pairs apart that also had pairwise HapMap CEU linkage disequilibrium values of r^2<0.5 were considered potentially independent regions. Potential independent regions that were within 1 million base pairs of a second region with more significant p-value were tested for independence using conditional analysis. In this analysis, the most significant SNP in the most significant region on each chromosome was used as a covariate in a genome-wide analysis. The second region on the chromosome was then re-tested for independent association.

Pathway analyses

Ingenuity pathway analysis (IPA) Knowledge Base 8.8 (see URLs) was used to explore the functional relationship between proteins encoded by the 17 replicated menopause loci. The IPA Knowledge Base contains millions of findings curated from the literature. All reference genes (n=61) within 60kb potentially encoded by the 17 loci (Table 2) were entered into the Ingenuity database. Fifty-one genes were eligible for pathway analysis. These eligible 'focus genes' were analyzed for direct interactions only. Networks were generated with a maximum size of 35 genes and shown as graphical representations of the molecular relationships between genes or gene products. Proteins are depicted as nodes in various shapes representing the functional class of the protein. Lines depict the biological relationships between nodes. To determine the probability of the analyzed gene to be found together in a network from Ingenuity Pathways Knowledge Base due to random chance alone, IPA applies a Fisher's exact test. The network score or p-value represents the significance of the focus gene enrichment. Enrichment of focus genes to diseases and functional categories was also evaluated in the IPA Knowledge Base. The p-value, based on a right-tailed Fisher's exact test, considers the number of identified focus genes and the total number of molecules known to be associated with these categories in the IPA knowledge database.

MAGENTA was used to explore pathway-based associations in the full GWAS dataset. MAGENTA implements a GSEA-based approach, the methodology of which has been previously described[23]. Briefly, 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. Genes within the HLA region were excluded from analysis due to difficulties in accounting for gene density and LD patterns. 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 (Supplementary Table 5). In total, 2,580 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with age at natural menopause.

GRAIL is designed to provide evidence for related biological function among a set of candidate genes. The method is based on connections between gene names and informative words extracted from PubMed abstracts by automated language processing techniques. Genes are considered related, and achieve a high similarity score, if they share informative words. For this analysis, the input for GRAIL was a list of candidate SNPs associated with age at natural menopause. From among candidate genes mapping near the candidate SNPs, GRAIL identifies genes with associated informative words that are significantly similar to informative words from other candidate genes. Genes with significant similarity scores are thus consistent with the set of candidate genes as a whole in having greater sharing of

informative words than expected by chance, which may suggest shared biological functions or even biological pathways. GRAIL was first applied to the lead SNPs from each of the replicating genome-wide significant loci using the database of genes and informative words from 2006. Separately, GRAIL was applied to a list of 66 SNPs, one from each locus that had at least one SNP meeting a false discovery rate (FDR) threshold of 0.05 from the QVALUE software in R[79]. For the age at natural menopause meta-analysis, the FDR < 0.05 threshold implied a p-value < $2.8\times10^{-}5$.

Candidate gene analysis

We explored the association of natural age of menopause with 125 candidate genes selected because of a reported relationship with ovarian function, including animal models where gene mutations affect ovarian function (N=37), human studies of menopause or isolated primary ovarian insufficiency (N=48), syndromes which include ovarian failure (N=4), or genes expressed in the ovary or female germ cells (N=38) (see Supplementary Table 6). For each gene, the start and end of transcription was defined by the transcripts that span the largest portion of the genome. NCBI36/hg18 positions taken from the UCSC genome browser were used to define gene and SNP locations. Using the correlation measured from a set of ~850 independent Framingham Heart Study participants, we computed the effective number of independent SNPs for each chromosome [80], and used the total (5,774) in a Bonferroni correction for multiple testing.

Pleiotropy of primary hits

Allelic pleiotropy was explored by comparing the genome-wide significant menopause signals to the online catalogue of published genome-wide association studies (GWAS catalog; see URLs). All reported associations that reached $P<5\times10^{-8}$ and were within 1Mb of the menopause signal were considered. LD estimates between the SNP pairs were assessed using HapMap (CEU, release#27). The results are presented in Supplementary Table 7.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Lisette Stolk[1,2,*], John RB Perry[3,4,*], Daniel I Chasman[5,6], Chunyan He[7,8], Massimo Mangino[9], Patrick Sulem[10], Maja Barbalic[11], Linda Broer[12], Enda M Byrne[13], Florian Ernst[14], Tõnu Esko[15,16,17], Nora Franceschini[18], Daniel F Gudbjartsson[10], Jouke-Jan Hottenga[19], Peter Kraft[20,21], Patick F McArdle[22], Eleonora Porcu[23], So-Youn Shin[24], Albert V Smith[25,26], Sophie van Wingerden[12], Guangju Zhai[9,27], Wei V Zhuang[28], Eva Albrecht[29], Behrooz Z Alizadeh[30], Thor Aspelund[25,26], Stefania Bandinelli[31], Lovorka Barac Lauc[32], Jacques S Beckmann[33,34], Mladen Boban[35], Eric Boerwinkle[11], Frank J Broekmans[36], Andrea Burri[9], Harry Campbell[37], Stephen J Chanock[38], Constance Chen[20,39], Marilyn C Cornelis[39], Tanguy Corre[40], Andrea D Coviello[41,42], Pio d'Adamo[43,44], Gail Davies[45], Ulf de Faire[46], Eco JC de

Geus[19,47], Ian J Deary[45,48], George VZ Dedoussis[49], Panagiotis Deloukas[24], Shah Ebrahim[50], Gudny Eiriksdottir[25], Valur Emilsson[25], Johan G Eriksson[51,52,53,54,55], Bart CJM Fauser[36], Liana Ferreli[23], Luigi Ferrucci[56], Krista Fischer[15], Aaron R Folsom[57], Melissa E Garcia[58], Paolo Gasparini[43,44], Christian Gieger[29], Nicole Glazer[41], Diederick E Grobbee[59], Per Hall[60], Toomas Haller[15], Susan E Hankinson[20,61], Merli Hass[15], Caroline Hayward[62], Andrew C Heath[63], Albert Hofman[2,12], Erik Ingelsson[60], A Cecile JW Janssens[12], Andrew D Johnson[42], David Karasik[42,64], Sharon LR Kardia[65], Jules Keyzer[66], Douglas P Kiel[42,64], Ivana Kolcic[35], Zoltán Kutalik[33,67], Jari Lahti[68], Sandra Lai[23], Triin Laisk[69], Joop SE Laven[70], Debbie A Lawlor[71], Jianjun Liu[72], Lorna M Lopez[45,48], Yvonne V Louwers[70], Patrik KE Magnusson[60], Mara Marongiu[23], Nicholas G Martin[13], Irena Martinovic Klaric[73], Corrado Masciullo[40], Barbara McKnight[74], Sarah E Medland[13], David Melzer[3], Vincent Mooser[75], Pau Navarro[62], Anne B Newman[76], Dale R Nyholt[13], N. Charlotte Onland-Moret[59], Aarno Palotie[24,77,78], Guillaume Paré[5,6,79], Alex N Parker[80,81], Nancy L Pedersen[60], Petra HM Peeters[59,82], Giorgio Pistis[40], Andrew S Plump[83], Ozren Polasek[35], Victor JM Pop[84], Bruce M Psaty[85,86], Katri Räikkönen[68], Emil Rehnberg[60], Jerome I Rotter[87], Igor Rudan[35,37], Cinzia Sala[40], Andres Salumets[15,69,88], Angelo Scuteri[89], Andrew Singleton[90], Jennifer A Smith[65], Harold Snieder[30,91], Nicole Soranzo[9,24], Simon N Stacey[10], John M Starr[48,92], Maria G Stathopoulou[49,93], Kathleen Stirrups[24], Ronald P Stolk[30,91], Unnur Styrkarsdottir[10], Yan V Sun[94], Albert Tenesa[62,95], Barbara Thorand[96], Daniela Toniolo[40,97], Laufey Tryggvadottir[26,98], Kim Tsui[80], Sheila Ulivi[43], Rob M van Dam[39,99], Yvonne T van der Schouw[59], Carla H van Gils[59], Peter van Nierop[100], Jacqueline M Vink[19], Peter M Visscher[48,101], Marlies Voorhuis[36,59], Gérard Waeber[102], Henri Wallaschofski[103], H Erich Wichmann[104,105,106], Elisabeth Widen[77], Colette JM Wijnands-van Gent[107], Gonneke Willemsen[19], James F Wilson[37], Bruce HR Wolffenbuttel[91,108], Alan F Wright[62], Laura M Yerges-Armstrong[22], Tatijana Zemunik[35], Lina Zgaga[37,109], M. Carola Zillikens[1], Marek Zygmunt[110], The LifeLines Cohort Study[91], Alice M Arnold[74], Dorret I Boomsma[19,47], Julie E. Buring[5,6,111], Laura Crisponi[23], Ellen W Demerath[57], Vilmundur Gudnason[25,26], Tamara B Harris[58], Frank B Hu[20,39,61], David J Hunter[20,39,61,21], Lenore J Launer[58], Andres Metspalu[15,16,17,88], Grant W Montgomery[13], Ben A Oostra[112], Paul M Ridker[5,6,111,113], Serena Sanna[23], David Schlessinger[114], Tim D Spector[9], Kari Stefansson[10,26], Elizabeth A Streeten[22], Unnur Thorsteinsdottir[10,26], Manuela Uda[23], André G Uitterlinden[1,2,12], Cornelia M van Duijn[2], Henry Völzke[115], Anna Murray[3,#], Joanne M Murabito[41,42,#], Jenny A Visser[1,#], and Kathryn L Lunetta[28,42,#]

Affiliations

[1]Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands [2]Netherlands Consortium of Healthy Aging, Rotterdam, the Netherlands [3]Peninsula Medical School, University of Exeter, UK [4]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK [5]Division of Preventive Medicine, Brigham and Women's Hospital, Boston USA [6]Harvard Medical School, Boston, USA [7]Department of Public Health, Indiana University School of Medicine,

Indianapolis, Indiana, USA [8]Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, USA [9]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK [10]deCODE Genetics, Reykjavik, Iceland [11]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA [12]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands [13]Queensland Institute of Medical Research, Brisbane, Australia [14]Interfakultäres Institut für Genomforschung, Universität Greifswald, Germany [15]Estonian Genome Center, University of Tartu, Tartu, Estonia [16]Estonian Biocenter, Tartu, Estonia [17]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia [18]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA [19]Dept Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands [20]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA [21]Broad Institute of Harvard and MIT, USA [22]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA [23]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cagliari, Italy [24]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK [25]Icelandic Heart Association, Kopavogur, Iceland [26]Faculty of Medicine, University of Iceland, Reykjavik, Iceland [27]Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada [28]Department of Biostatistics, Boston University School of Public Health, Boston Massachusetts, USA [29]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany [30]Department of Epidemiology, University Medical Center Groningen, University of Groningen, the Netherlands [31]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy [32]Croation Science Foundation, Zagreb, Croatia [33]Department of Medical Genetics, University of Lausanne, Switzerland [34]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois (CHUV), University Hospital, Lausanne, Switzerland [35]Faculty of Medicine, University of Split, Split, Croatia [36]Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht, Utrecht, the Netherlands [37]Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK [38]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA [39]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA [40]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy [41]Sections of General Internal Medicine, Preventive Medicine and Epidemiology, Department of Medicine, Boston University School of Medicine, Boston MA, USA [42]NHLBI Framingham Heart Study, Framingham, MA, USA [43]Institute for Maternal and Child Health, IRCCS "Burlo Garofolo" Trieste, Italy [44]University of Trieste, Trieste, Italy [45]Department of Psychology, The University of Edinburgh, Edinburgh, UK [46]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden [47]EMGO+ Institute, VU Medical Centre, Amsterdam, The Netherlands [48]Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK

[49]Department of Nutrition and Dietetics, Harokopio University, Athens, Greece [50]Department of Epidemiology & Population Healths, London School of Hygiene & Tropical Medicine, UK [51]National Institute for Health and Welfare, Finland [52]Department of General Practice and Primary Health Care, University of Helsinki, Finland [53]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland [54]Folkhalsan Research Centre, Helsinki, Finland [55]Vasa Central Hospital, Vasa, Finland [56]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA [57]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA [58]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, NIH, Bethesda, MD, USA [59]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands [60]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [61]Channing Laboratory, Department of Medicine, Brigham and Women.s Hospital Harvard Medical School, Boston, Massachusetts, USA [62]MRC Human Genetics Unit at the Medical Research Council Institute of Genetics and Molecular Medicine at the University of Edinburgh, Western General Hospital, Edinburgh, UK [63]Washington University St.Louis, St. Louis, MO, USA [64]Hebrew SeniorLife Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, USA [65]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA [66]Diagnostic GP laboratory Eindhoven, Eindhoven, the Netherlands [67]Swiss Institute of Bioinformatics, Switzerland [68]Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland [69]Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia [70]Division of Reproductive Medicine, Department of Obstetrics & Gynaecology, Erasmus MC, Rotterdam, the Netherlands [71]MRC Centre for Causal Analysis in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, UK [72]Human genetic, Genome Institute of Singapore, Singapore [73]Institute for Migration and Ethnic Studies, Zagreb, Croatia [74]Department of Biostatistics, University of Washington, Seattle, WA, USA [75]Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania, USA [76]Departments of Epidemiology and Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania [77]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland [78]Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland [79]Genetic and Molecular Epidemiology Laboratory, McMaster University, Hamilton, ON Canada [80]Amgen, Cambridge, MA USA [81]Foundation Medicine, Inc., Cambridge MA USA [82]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK [83]Cardiovascular Disease, Merck Research Laboratory, Rahway, NJ, USA [84]Department of Clinical Health Psychology, University of Tilburg, Tilburg, the Netherlands [85]Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA USA [86]Group Health Research Institute, Group Health Cooperative, Seattle, WA USA [87]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA [88]Competence Centre on Reproductive Medicine and Biology, Tartu,

Estonia [89]Lab Cardiovascular Sciences - NIA - NIH, Baltimore, USA [90]Laboratory of Neurogenetics, National Institute of Ageing, Bethesda, MD, USA [91]LifeLines Cohort Study & Biobank, University Medical Center Groningen, University of Groningen, the Netherlands [92]Geriatric Medicine Unit, University of Edinburgh, Edinburgh, UK [93]Cardiovascular Genetics Research Unit, EA4373, Université Henri Poincaré - Nancy 1, Nancy, France [94]Department of Epidemiology, Emory University, Atlanta, GA, USA [95]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, UK [96]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany [97]Institute of Molecular Genetics-CNR, Pavia, Italy [98]Icelandic Cancer Registry, Reykjavik, Iceland [99]Saw Swee Hock School of Public Health and Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore [100]Municipal Health Service Brabant-Zuidoost, Helmond, the Netherlands [101]Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia [102]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), University Hospital, Lausanne, Switzerland [103]Institute for Clinical Chemistry and Laboratory Medicine, University of Greifswald [104]Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany [105]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany [106]Klinikum Grosshadern, Munich, Germany [107]POZOB Veldhoven, Veldhoven, the Netherlands [108]Department of Endocrinology, University Medical Center Groningen, University of Groningen, the Netherlands [109]Andrija Stampar School of Public Health, Medical School, University of Zagreb, Zagreb, Croatia [110]Klinik für Gynäkologie und Geburtshilfe, Universität Greifswald, Germany [111]Harvard School of Public Health, Boston, MA USA [112]Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands [113]Division of Cardiology, Brigham and Women's Hospital, Boston, MA USA [114]National Institute on Aging, Intramural Research Program, Baltimore, MD, USA [115]Institut für Community Medicine, Universität Greifswald, Germany

Acknowledgments

The authors are very grateful to the study participants and staff from all cohorts involved in this study. Extended acknowledgements per cohort can be found in the Supplement.

References

1. Burger HG. The menopausal transition. Baillieres Clin Obstet Gynaecol. 1996; 10:347–59. [PubMed: 8931899]
2. Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. Endocr Rev. 2009; 30:465–93. [PubMed: 19589949]
3. te Velde ER, Pearson PL. The variability of female reproductive ageing. Hum Reprod Update. 2002; 8:141–54. [PubMed: 12099629]
4. te Velde ER, Dorland M, Broekmans FJ. Age at menopause as a marker of reproductive ageing. Maturitas. 1998; 30:119–25. [PubMed: 9871906]

5. 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. 2005; 90:3427–30. [PubMed: 15769979]
6. 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; 83:1875–80. [PubMed: 9626112]
7. van Asselt KM, et al. Heritability of menopausal age in mothers and daughters. Fertil Steril. 2004; 82:1348–51. [PubMed: 15533358]
8. Vink JM, Boomsma DI. Modeling age at menopause. Fertil Steril. 2005; 83:1068. [PubMed: 15820835]
9. He C, et al. A large-scale candidate gene association study of age at menarche and age at natural menopause. Hum Genet. 2010; 128:515–27. [PubMed: 20734064]
10. He LN, et al. Association study of the oestrogen signalling pathway genes in relation to age at natural menopause. J Genet. 2007; 86:269–76. [PubMed: 18305346]
11. Weel AE, et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. J Clin Endocrinol Metab. 1999; 84:3146–50. [PubMed: 10487678]
12. Tempfer CB, et al. Polymorphisms associated with thrombophilia and vascular homeostasis and the timing of menarche and menopause in 728 white women. Menopause. 2005; 12:325–30. [PubMed: 15879922]
13. van Asselt KM, et al. Factor V Leiden mutation accelerates the onset of natural menopause. Menopause. 2003; 10:477–81. [PubMed: 14501610]
14. He C, et al. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat Genet. 2009; 41:724–8. [PubMed: 19448621]
15. Stolk L, et al. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. Nat Genet. 2009; 41:645–7. [PubMed: 19448619]
16. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073–81. [PubMed: 19561590]
17. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]
18. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–7. [PubMed: 20634204]
19. Marini F, Wood RD. A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. J Biol Chem. 2002; 277:8716–23. [PubMed: 11751861]
20. Pullirsch D, et al. The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development. 2010; 137:935–43. [PubMed: 20150277]
21. Copeland WC, Longley MJ. DNA polymerase gamma in mitochondrial DNA replication and repair. ScientificWorldJournal. 2003; 3:34–44. [PubMed: 12806118]
22. Tschopp J, Martinon F, Burns K. NALPs: a novel protein family involved in inflammation. Nat Rev Mol Cell Biol. 2003; 4:95–104. [PubMed: 12563287]
23. Segre AV, et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. 2010; 6
24. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]
25. Bishop DK, Park D, Xu L, Kleckner N. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell. 1992; 69:439–56. [PubMed: 1581960]
26. Qin Y, et al. NOBOX homeobox mutation causes premature ovarian failure. Am J Hum Genet. 2007; 81:576–81. [PubMed: 17701902]
27. Rajkovic A, Pangas SA, Ballow D, Suzumori N, Matzuk MM. NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. Science. 2004; 305:1157–9. [PubMed: 15326356]

28. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010; 363:166–76. [PubMed: 20647212]
29. Johnson FB, Sinclair DA, Guarente L. Molecular biology of aging. Cell. 1999; 96:291–302. [PubMed: 9988222]
30. Lee CK, Allison DB, Brand J, Weindruch R, Prolla TA. Transcriptional profiles associated with aging and middle age-onset caloric restriction in mouse hearts. Proc Natl Acad Sci U S A. 2002; 99:14988–93. [PubMed: 12419851]
31. Hamatani T, et al. Age-associated alteration of gene expression patterns in mouse oocytes. Hum Mol Genet. 2004; 13:2263–78. [PubMed: 15317747]
32. Promislow DE. DNA repair and the evolution of longevity: a critical analysis. J Theor Biol. 1994; 170:291–300. [PubMed: 7996857]
33. Du X, et al. Telomere shortening exposes functions for the mouse Werner and Bloom syndrome genes. Mol Cell Biol. 2004; 24:8437–46. [PubMed: 15367665]
34. Dorland M, van Kooij RJ, te Velde ER. General ageing and ovarian ageing. Maturitas. 1998; 30:113–8. [PubMed: 9871905]
35. Shiratori A, et al. Assignment of the 49-kDa (PRIM1) and 58-kDa (PRIM2A and PRIM2B) subunit genes of the human DNA primase to chromosome bands 1q44 and 6p11.1-p12. Genomics. 1995; 28:350–3. [PubMed: 8530050]
36. Pagnamenta AT, et al. Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. Hum Reprod. 2006; 21:2467–73. [PubMed: 16595552]
37. Trifunovic A, et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature. 2004; 429:417–23. [PubMed: 15164064]
38. Wong JC, Buchwald M. Disease model: Fanconi anemia. Trends Mol Med. 2002; 8:139–42. [PubMed: 11879775]
39. Nebel A, et al. A functional EXO1 promoter variant is associated with prolonged life expectancy in centenarians. Mech Ageing Dev. 2009; 130:691–9. [PubMed: 19698732]
40. Wei K, et al. Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. Genes Dev. 2003; 17:603–14. [PubMed: 12629043]
41. Tye BK. Insights into DNA replication from the third domain of life. Proc Natl Acad Sci U S A. 2000; 97:2399–401. [PubMed: 10716976]
42. Epstein CJ, Martin GM, Schultz AL, Motulsky AG. Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. Medicine (Baltimore). 1966; 45:177–221. [PubMed: 5327241]
43. Lee SJ, Lee SH, Ha NC, Park BJ. Estrogen prevents senescence through induction of WRN, Werner syndrome protein. Horm Res Paediatr. 2010; 74:33–40. [PubMed: 20395656]
44. Gao Y, et al. DNA repair gene polymorphisms and tobacco smoking in the risk for colorectal adenomas. Carcinogenesis. 32:882–7. [PubMed: 21504893]
45. Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. Clin Endocrinol (Oxf). 2008; 68:499–509. [PubMed: 17970776]
46. Hoek A, Schoemaker J, Drexhage HA. Premature ovarian failure and ovarian autoimmunity. Endocr Rev. 1997; 18:107–34. [PubMed: 9034788]
47. Robb L, et al. Infertility in female mice lacking the receptor for interleukin 11 is due to a defective uterine response to implantation. Nat Med. 1998; 4:303–8. [PubMed: 9500603]
48. Northup J, Griffis K, Hawkins J, Lockhart L, Velagaleti G. Unusual pseudo dicentric, psu dic (1;19)(q10;q13.42), in a female with premature ovarian failure. Fertil Steril. 2007; 87:697, e5–8. [PubMed: 17140574]
49. Tong ZB, Nelson LM. A mouse gene encoding an oocyte antigen associated with autoimmune premature ovarian failure. Endocrinology. 1999; 140:3720–6. [PubMed: 10433232]
50. Anasti JN, et al. Karyotypically normal spontaneous premature ovarian failure: evaluation of association with the class II major histocompatibility complex. J Clin Endocrinol Metab. 1994; 78:722–3. [PubMed: 8126148]

51. Jaroudi KA, Arora M, Sheth KV, Sieck UV, Willemsen WN. Human leukocyte antigen typing and associated abnormalities in premature ovarian failure. Hum Reprod. 1994; 9:2006–9. [PubMed: 7868665]
52. Gasser S, Raulet DH. The DNA damage response arouses the immune system. Cancer Res. 2006; 66:3959–62. [PubMed: 16618710]
53. Kumar TR, Wang Y, Lu N, Matzuk MM. Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. Nat Genet. 1997; 15:201–4. [PubMed: 9020850]
54. McTavish KJ, et al. Rising follicle-stimulating hormone levels with age accelerate female reproductive failure. Endocrinology. 2007; 148:4432–9. [PubMed: 17540727]
55. Lambalk CB, De Koning CH, Braat DD. The endocrinology of dizygotic twinning in the human. Mol Cell Endocrinol. 1998; 145:97–102. [PubMed: 9922105]
56. Kottler ML, et al. A new FSHbeta mutation in a 29-year-old woman with primary amenorrhea and isolated FSH deficiency: functional characterization and ovarian response to human recombinant FSH. Eur J Endocrinol. 2010; 162:633–41. [PubMed: 19966036]
57. Grigorova M, et al. Increased Prevalance of the −211 T allele of follicle stimulating hormone (FSH) beta subunit promoter polymorphism and lower serum FSH in infertile men. J Clin Endocrinol Metab. 2010; 95:100–8. [PubMed: 19897680]
58. Hoogendoorn B, et al. Functional analysis of human promoter polymorphisms. Hum Mol Genet. 2003; 12:2249–54. [PubMed: 12915441]
59. Kaku U, et al. Ovarian histological findings in an adult patient with the steroidogenic acute regulatory protein (StAR) deficiency reveal the impairment of steroidogenesis by lipoid deposition. Endocr J. 2008; 55:1043–9. [PubMed: 18724044]
60. Pisarska MD, Bae J, Klein C, Hsueh AJ. Forkhead l2 is expressed in the ovary and represses the promoter activity of the steroidogenic acute regulatory gene. Endocrinology. 2004; 145:3424–33. [PubMed: 15059956]
61. Godinho M, Meijer D, Setyono-Han B, Dorssers LC, Agthoven TV. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. J Cell Physiol. 2011; 226:1741–9. [PubMed: 21506106]
62. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26:2190–1. [PubMed: 20616382]
63. Goring HH, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet. 2007; 39:1208–16. [PubMed: 17873875]
64. Idaghdour Y, et al. Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet. 2010; 42:62–7. [PubMed: 19966804]
65. Heap GA, et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC Med Genomics. 2009; 2:1. [PubMed: 19128478]
66. Dixon AL, et al. A genome-wide association study of global gene expression. Nat Genet. 2007; 39:1202–7. [PubMed: 17873877]
67. Stranger BE, et al. Population genomics of human gene expression. Nat Genet. 2007; 39:1217–24. [PubMed: 17873874]
68. Kwan T, et al. Genome-wide analysis of transcript isoform variation in humans. Nat Genet. 2008; 40:225–31. [PubMed: 18193047]
69. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–50. [PubMed: 19644074]
70. Heinzen EL, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol. 2008; 6:e1. [PubMed: 19222302]
71. Zeller T, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One. 2010; 5:e10693. [PubMed: 20502693]
72. Murphy A, et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Hum Mol Genet. 2010; 19:4745–57. [PubMed: 20833654]
73. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–8. [PubMed: 18344981]

74. Webster JA, et al. Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet. 2009; 84:445–58. [PubMed: 19361613]

75. Gibbs JR, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 2010; 6:e1000952. [PubMed: 20485568]

76. Liu C, et al. Whole-genome association mapping of gene expression in the human prefrontal cortex. Mol Psychiatry. 2010; 15:779–84. [PubMed: 20351726]

77. Schadt EE, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008; 6:e107. [PubMed: 18462017]

78. Grundberg E, et al. Population genomics in a disease targeted primary cell model. Genome Res. 2009; 19:1942–52. [PubMed: 19654370]

79. Dabney, A.; Storey, JD.; Warnes, GR. R package version 1.22.0. edn. 2010. qvalue: Q-value estimation for false discovery rate control.

80. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005; 95:221–7. [PubMed: 16077740]

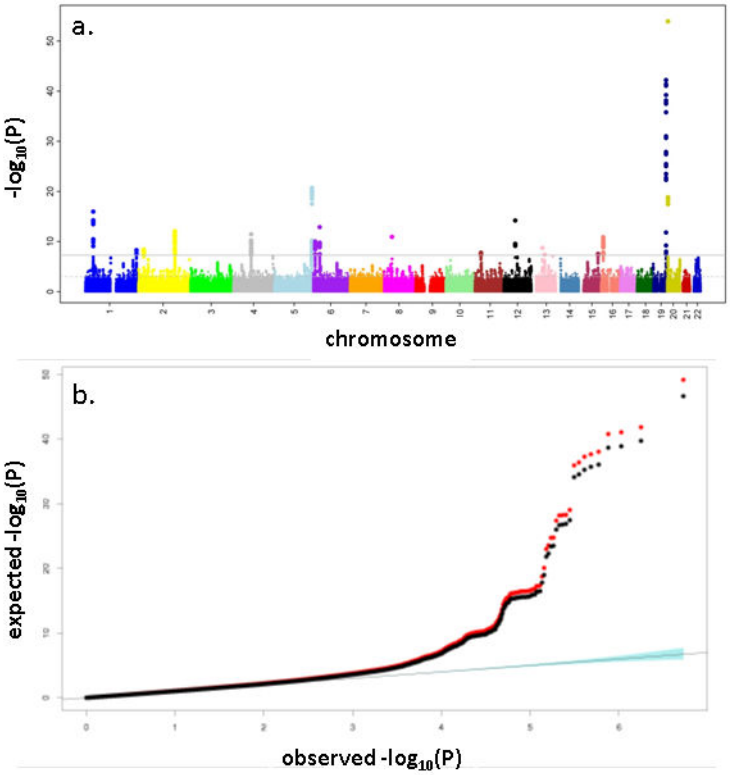

Figure 1.

a. Manhattan plot of discovery meta analysis. **b.** quantile-quantile plot of discovery primary analysis (red), and double genomic control adjusted primary analysis (black)

Table 1

Discovery and replication results

Region number	Most Significant SNP	Chr	Location (bp)	Minor/Major	Analysis	MAF	Effect per minor allele (years)	SE	P	Absolute Effect per minor allele (weeks)	Heterogeneity P-value
1a	rs4246511	1	39152972	t/c	Discovery	0.271	0.289	0.035	1.02×10^{-16}	15	0.288
					Replication	0.293	0.133	0.052	0.01	6.9	0.68
					Combined		0.24	0.029	***9.08 × 10 −17***	12.5	0.012
1b	rs1635501	1	240107398	c/t	Discovery	0.478	−0.188	0.032	4.90×10^{-09}	9.8	0.08
					Replication	0.456	−0.11	0.048	0.023	5.7	0.016
					Combined		−0.164	0.027	***8.46 × 10 −10***	8.5	0.178
2a	rs2303369	2	27568920	t/c	Discovery	0.388	−0.174	0.03	3.80×10^{-09}	9	0.639
					Replication	0.391	−0.179	0.047	***0.000138***	9.3	0.542
					Combined		−0.175	0.025	***2.25 × 10 −12***	9.1	0.927
2b	rs10183486	2	171699217	t/c	Discovery	0.366	−0.219	0.031	7.88×10^{-13}	11.4	0.43
					Replication	0.362	−0.141	0.047	0.003	7.3	0.727
					Combined		−0.196	0.026	***2.21 × 10 −14***	10.2	0.169
2c[a]	rs7606918	2	172603695	g/a	Discovery	0.161	−0.228	0.041	2.89×10^{-08}	11.8	0.374
					Replication not attempted						
4	rs4693089	4	84592646	g/a	Discovery	0.486	0.209	0.03	3.28×10^{-12}	10.9	0.336
					Replication	0.492	0.273	0.047	***6.69 × 10 −09***	14.2	0.298
					Combined		0.228	0.025	***2.38 × 10 −19***	11.8	0.254
5a[b]	rs890835	5	175888877	a/c	Discovery	0.112	0.266	0.047	1.17×10^{-08}	13.8	0.003
					Replication	0.115	−0.037	0.072	0.613	1.9	0.486
					Combined		0.177	0.039	6.10×10^{-06}	9.2	0.0004
5b	rs365132	5	176311180	t/g	Discovery	0.49	0.275	0.029	1.90×10^{-21}	14.3	0.115
					Replication	0.494	0.319	0.046	***4.26 × 10 −12***	16.6	0.728
					Combined		0.287	0.025	***9.11 × 10 −32***	14.9	0.416

Region number	Most Significant SNP	Chr	Location (bp)	Minor/Major	Analysis	MAF	Effect per minor allele (years)	SE	P	Absolute Effect per minor allele (weeks)	Heterogeneity P-value
6a	rs2153157	6	11005474	a/g	Discovery	0.492	0.184	0.028	9.40×10^{-11}	9.5	0.858
					Replication	0.5	0.116	0.046	0.012	6	0.329
					Combined		0.165	0.024	***7.76 × 10 −12***	8.6	0.211
6b	rs1046089	6	31710946	a/g	Discovery	0.353	−0.226	0.031	1.31×10^{-13}	11.8	0.426
					Replication	0.358	−0.181	0.049	***1.91 × 10 −04***	9.4	0.732
					Combined		−0.213	0.026	***1.63 × 10 −16***	11.1	0.427
8	rs2517388	8	38096889	g/t	Discovery	0.174	0.274	0.04	1.13×10^{-11}	14.2	0.67
					Replication	0.189	0.234	0.062	***1.52 × 10 −04***	12.2	0.708
					Combined		0.262	0.034	***9.31 × 10 −15***	13.6	0.591
11	rs12294104	11	30339475	t/c	Discovery	0.172	0.226	0.04	1.63×10^{-08}	11.8	0.721
					Replication	0.18	0.223	0.06	***2.18 × 10 −04***	11.6	0.239
					Combined		0.225	0.033	***1.46 × 10 −11***	11.7	0.97
12	rs2277339	12	55432336	g/t	Discovery	0.102	−0.394	0.051	5.99×10^{-15}	20.5	0.088
					Replication	0.105	−0.347	0.077	***6.89 × 10 −06***	18	0.765
					Combined		−0.38	0.042	***2.47 × 10 −19***	19.7	0.61
13a	rs3736830	13	49204222	g/c	Discovery	0.157	−0.243	0.04	1.75×10^{-09}	12.6	0.859
					Replication	0.165	−0.033	0.062	0.594	1.7	0.905
					Combined		−0.18	0.034	9.41×10^{-08}	9.4	0.005
13b	rs4886238	13	60011740	a/g	Discovery	0.334	0.172	0.031	3.76×10^{-08}	8.9	0.974
					Replication	0.333	0.166	0.049	***6.09 × 10 −04***	8.6	0.953
					Combined		0.17	0.026	***9.53 × 10 −11***	8.9	0.919
15	rs2307449	15	87664932	g/t	Discovery	0.405	−0.167	0.03	2.59×10^{-08}	8.7	0.442
					Replication	0.387	−0.225	0.047	***1.61 × 10 −06***	11.7	0.328
					Combined		−0.184	0.025	***3.56 × 10 −13***	9.6	0.294

Region number	Most Significant SNP	Chr	Location (bp)	Minor/Major	Analysis	MAF	Effect per minor allele (years)	SE	P	Absolute Effect per minor allele (weeks)	Heterogeneity P-value
16	rs10852344	16	11924420	c/t	Discovery	0.415	0.198	0.029	1.28×10^{-11}	10.3	0.014
					Replication	0.426	0.093	0.046	0.042	4.8	0.599
					Combined		0.168	0.025	***1.01 × 10 −11***	8.7	0.054
19a	rs11668344	19	60525476	g/a	Discovery	0.363	−0.416	0.03	5.94×10^{-43}	21.6	0.112
					Replication	0.36	−0.415	0.048	***2.65 × 10 −18***	21.6	0.517
					Combined		−0.416	0.026	***1.45 × 10 −59***	21.6	0.987
19b[c]	rs12461110	19	61012475	a/g	Discovery	0.356	−0.174	0.03	9.49×10^{-09}	9.1	0.835
					Replication	0.344	−0.117	0.049	0.018	6.1	0.542
					Combined		−0.158	0.026	***8.74 × 10 −10***	8.2	0.32
20	rs16991615	20	5896227	a/g	Discovery	0.069	0.971	0.062	1.16×10^{-54}	50.5	0.356
					Replication	0.07	0.896	0.096	***7.90 × 10 −21***	46.6	0.088
					Combined		0.948	0.052	***1.42 × 10 −73***	49.3	0.509

MAF: minor allele frequency; Heterogeneity: p-values for heterogeneity among discovery studies, replication studies, and comparing all discovery to all replication studies. P- value: Replication p-values that meet the criterion P<0.05/19 are indicated in bold italics. Combined analysis p-values that reached genome-wide significance are indicated in bold italics.

[a] Conditional analysis: beta(SE): −0.199(0.041); p-value:9.8×10^{-07}

[b] Conditional analysis beta(SE): 0.267(0.046); p-value:6.5×10^{-09}

[c] Conditional analysis: beta(SE): −0.168(0.031); p-value:3.8×10^{-08}

Table 2

Characteristics of the top SNP in each region.

Region	SNPID	Chr	Location (bp)	gene	feature	Other refgenes<60kb from SNP	Genes with SNPs in LD (r^2>0.5) with SNP (gene:r^2)	eQTL
1a	rs4246511	1	39,152,972	RHBDL2	intron	MYCBP;GJA10;RRAGC	MYCBP:0.678; RHBDL2:0.678; RRAGC: 0.678; GJA9:0.678; LOC100130627:0.678	
1b	rs1635501	1	240,107,398	EXO1	intron		LOC100131576:0.678; WDR64:0.923; LOC100133057:1; EXO1:1	
2a	rs2303369	2	27,568,920	FNDC4	intron	GCKR;KRTCAP3;IFT172;N RBP1	FTHL3P:0.967; GCKR:1; GTF3C2:0.967; MPV17:0.967; PPM1G:0.967; UCN: 0.841; EIF2B4:0.967; SNX17:0.967; IFT172:1; NRBP1:0.967; TRIM54:0.841; FNDC4:1; C2orf16:0.513; ZNF513:0.967; KRTCAP3:1; DNAJC5G:0.841; LOC100130981:0.901	r^2>0.9 with multiple eSNPs for IFT172n lymph, adipose, and blood; r2>0.9 with multiple eSNPs for KRTCAP3 in lymph and CD4; r2>0.9 with eSNP for SNX17 in PFC
2b	rs10183486	2	171,699,217	TLK1	intron		GORASP2:0.929; TLK1:0.964	r^2>0.9 with eSNPs for TLK1 in LCL and PFC
4	rs4693089	4	84,592,646	HELQ	intron	MRPS18C;FAM175A	MRPS18C:1; FAM175A:1; AGPAT9:1; HEL308:1; OK/SW-CL.36:1	r^2>0.80 with eSNPs for MRPS18C in Lymph and PFC; r^2>0.80 for AGPAT9 in Fibroblast
5a	rs890835	5	175,888,877	RNF44	intron	UBXD8;PCLKC	SNCB:0.318; RNF44:1; FAF2:1; PCDH24:1	
5b	rs365132	5	176,311,180	UIMC1	coding-synonymous	HK3	FGFR4:0.871; HK3:0.967; ZNF346:0.966; UIMC1:0.967; UNC5A:0.932	eSNP for Hs.484258 in Lymph; for FGFR4 and

Region	SNPID	Chr	Location (bp)	gene	feature	Other refgenes<60kb from SNP	Genes with SNPs in LD (r^2>0.5) with SNP (gene:r^2)	eQTL
								ZNF346 in PFC; for Z ZNF346 in PFC; for Z ZNF346 in PFC; for Z ZNF346 in PFC; for Z
6a	rs2153157	6	11,005,474	SYCP2L	intron	GCM2	MAK:0.602; GCM2:0.602; SYCP2L:0.602	eSNP for SYCP2L in monocytes
6b	rs1046089[a]	6	31,710,946	BAT2	missense	BAT3;LST1;C6orf47;APOM; AIF1; NCR3; LY6G5C;LTB;BAT5;CSNK2 B;LY6G5B;BAT4;TNF	AIF1:0.963; CLIC1:0.682; CSNK2B: 0.963; HSPA1A:0.649; HSPA1B:0.587; HSPA1L:0.649; LTB:0.963; MSH5:0.682; NEU1:0.587; VARS:0.682; BAT2:0.963; BAT3:0.963; BAT4:0.963; BAT5:0.963; LST1:0.963; DDAH2:0.682; SNORD52:0.587; SNORD48:0.587; C6orf48:0.587; APOM:0.963; LSM2:0.649; C6orf47:0.963; LY6G5B: 0.963; LY6G6D:0.963; LY6G6E:0.963; SLC44A4:0.587; C6orf27:0.682; C6orf25:0.682; LY6G6C:0.961; LY6G5C: 0.963; NCR3:0.963; LY6G6F:0.963; C6orf26:0.682; SNORA38:0.963; LOC100130756:0.963	eSNP for LY6G5C in CD4 and Lymph; for HLA-DRB4 in monocytes; for C6orf10 in VC; for AIF1 in Lymph; for HLA-DQA1 in LCL
8	rs2517388	8	38,096,889	ASH2L	intron	BAG4;EIF4EBP1;LSM1;ST AR	STAR:0.831; ASH2L:0.831; LSM1:0.831	
11	rs12294104	11	30,339,475	--		MPPED2;C11orf46	FSHB:0.92; C11orf46:1	r^2>0.9 with eSNP for C11orf46 in Lymph
12	rs2277339[b]	12	55,432,336	PRIM1	missense	HSD17B6;NACA	-	eSNP for PRIM1 in VC, CR, PFC
13a	rs3736830	13	49,204,222	KPNA3	intron	EBPL	KPNA3:0.734; EBPL:0.734; ARL11:0.623; LOC100131941:0.623	
13b	rs4886238	13	60,011,740	TDRD3	intron		TDRD3:1; LOC390407:0.731	r^2>0.8 with eSNP in Adipose for TDRD3
15	rs2307449	15	87,664,932	POLG	intron	FANCI	POLG:0.965; RLBP1:0.898; ABHD2:0.898; FANCI:0.965; LOC728003:0.898; LOC100131654:0.683	r^2>0.8 with eSNP for RLBP1 in PFC
16	rs10852344	16	11,924,420	--		TNFRSF17;RUNDC2A;GSP T1	TNFRSF17:0.662; GSPT1:1; COX6CP1:1; RSL1D1:1; ZC3H7A:0.701; RUNDC2A:0.662; LOC729978:1	

Region	SNPID	Chr	Location (bp)	gene	feature	Other refgenes<60kb from SNP	Genes with SNPs in LD (r^2>0.5) with SNP (gene:r^2)	eQTL
19a	rs11668344	19	60,525,476	TMEM150B	intron	BRSK1;HSPBP1;COX6B2;L OC284417;IL11; SUV420H2	IL11:0.962; SAPS1:0.894; HSPBP1:0.962; BRSK1:0.962; SUV420H2:0.962; COX6B2:0.962; LOC284417:0.962; FAM71E2:0.962	r^2>0.8 with eSNP for MGC27005 in adipose and blood
19b	rs12461110[c]	19	61,012,475	NLRP11	missense	NLRP4	NLRP4:0.514; NLRP11:0.514; RFPL4A: 0.514; LOC646663:0.514; LOC729974:0.514	
20	rs16991615[d]	20	5,896,227	MCM8	missense	CRLS1;CHGB;TRMT6	-	

[a] Arg > His; Predicted to be Damaging (SIFT), Benign (Polyphen2)

[b] Asp > Ala; Predicted to be Damaging (SIFT), Probably damaging (Polyphen2)

[c] Pro > Leu; Predicted to be Damaging (SIFT), Benign (Polyphen2)

[d] Glu > Lys; Predicted to be Tolerated (SIFT), Benign (Polyphen2)