nature genetics

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

To newly identify loci for age at natural menopause, we carried out a meta-analysis of 22 genome-wide association studies (GWAS) in 38,968 women of European descent, with replication in up to 14,435 women. In addition to four known loci, we identified 13 loci newly associated with age at natural menopause (at $P < 5 \times 10^{-8}$). Candidate genes located at these newly associated loci include genes implicated in DNA repair (*EXO1*, *HELQ*, *UIMC1*, *FAM175A*, *FANCI*, *TLK1*, *POLG* and *PRIM1*) and immune function (*IL11*, *NLRP11* and *PRRC2A* (also known as *BAT2*)). Gene-set enrichment pathway analyses using the full GWAS data set identified exoDNase, NF-κB signaling and mitochondrial dysfunction as biological processes related to timing of menopause.

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 owing 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 greater meiotic nondisjunction[2]. Oocyte quality may be controlled at the time germ cells are formed during fetal life, but it 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 unclear 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% (refs. 5–8). Thus far most genetic association studies regarding age at menopause have focused on candidate genes[9] from the estrogen pathway[10,11] or vascular components[12,13]. Recently, two GWAS have newly identified five loci associated with age at natural menopause on chromosomes 5, 6, 13, 19 and 20 (refs. 14,15). These loci, however, explained <1.5% of the phenotypic variation of age at natural menopause, suggesting that additional loci of small effect will probably 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[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 aged 40–60 (**Supplementary Tables 1** and **2**), we identified 20 regions with SNPs meeting the genome-wide significance criterion $P < 5 \times 10^{-8}$ (**Fig. 1**). Four of these loci confirmed earlier reports of associations on chromosomes 5, 6, 19 and 20 (refs. 14,15; regions 5b, 6a, 19a and 20, respectively, in **Table 1**) and 16 loci were previously unidentified. We did not confirm one reported association on chromosome 13 (13q34, rs7333181, $P = 0.12$). The overall genomic inflation factor was 1.03 (**Fig. 1**, inset; SNP with lowest P value from each region, **Table 1**). There was no between-study effect heterogeneity across discovery studies ($P > 0.05/20 = 0.0025$) for the 20 SNP associations presented. Within the Framingham Heart Study group, we tested for differences in effect size for the 20 SNPs in retrospectively and prospectively collected menopause age, and found no significant differences (data not shown). The effect sizes ranged from 0.17 years (8.7 weeks) to nearly 1 year (50.5 weeks) per each copy of the minor allele. We computed the effect sizes for dichotomized age at natural menopause in women from the Women's Genome Health Study (WGHS). For early menopause, we compared women with age at menopause <45 ($N = 745$) to those with age at menopause >45. For late menopause, we compared women with age at menopause >54 ($N = 1,632$) to those with age at menopause <54. 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 >400 kb but <1 Mb from the top SNP in another region on the same chromosome. The top SNP in each of these primary regions had low linkage disequilibrium (LD; $r^2 < 0.5$) with the top SNP in the nearby region. To determine whether these associations were independent, we carried out 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 and rs12461110, respectively), the effect estimates in the conditional analysis were unchanged compared

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

Received 21 July 2011; accepted 2 December 2011; published online 22 January 2012; doi:10.1038/ng.1051

with in the discovery analysis (differences of 0.3% and 4%, respectively), and the *P* values were genome-wide significant. However, for region 2c, the effect size was ~12.5% lower in the conditional analysis than in the initial analysis, and the SNP *P* value was no longer genome-wide significant ($P = 9.8 \times 10^{-7}$; **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 we did not pursue replication of rs7606918.

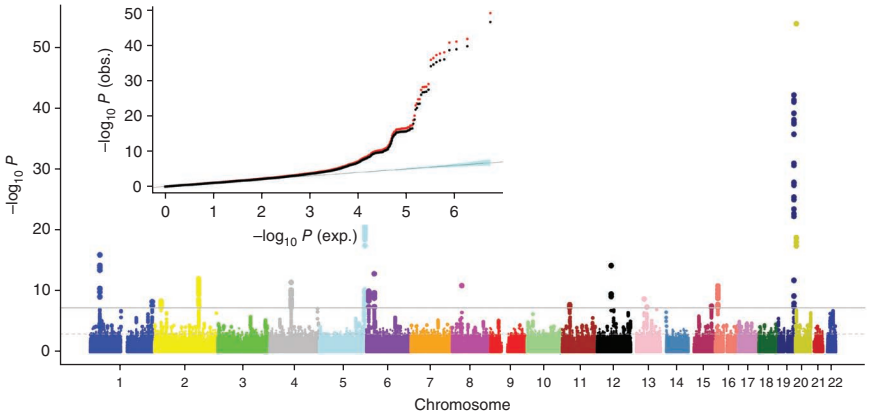

Figure 1 Discovery GWAS results. Manhattan plot of discovery meta-analysis. Inset, quantile-quantile plot of discovery primary analysis (red) and double genomic control–adjusted primary analysis (black). Obs., observed; exp., expected.

Replication

We used 21 studies contributing 14,435 women for replication of the 19 SNPs that defined the independent genome-wide significant regions from stage 1. We defined age at natural menopause using the same criteria as in the discovery studies (**Supplementary Table 1**). Of these studies, 17 ($n = 6,639$) 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**; effect sizes and *P* values for replication and combined meta-analysis of discovery and replication samples, **Table 1**). There was no evidence for effect heterogeneity among the replication studies (**Table 1**). We also 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 lead to significantly different effect size estimates than would have been obtained from direct genotyping. Of the 19 SNPs, 17 were 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 have been reported previously; thus our analysis identified 13 regions newly associated with age at natural menopause on the basis of 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 varied 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, $n = 11,379$) to explore whether substantial SNP-SNP 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 or near newly identified 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 have been predicted to have damaging protein function by SIFT[16], and one by PolyPhen2 (ref. 17). Using dbSNP and LocusZoom[18], we identified the genes underlying the newly identified top regions. We used SCAN (see URLs) to identify all genes with SNPs that are in 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 whether the SNPs are associated with expression (**Table 2**).

The strongest new signal was on chromosome 4 (region 4, rs4693089; $P = 2.4 \times 10^{-19}$). The SNP is located in an intron of *HELQ*, which encodes the protein HEL308, a DNA-dependent ATPase and DNA helicase[19]. The second strongest new signal was on chromosome 12 (region 12, rs2277339; $P = 2.5 \times 10^{-19}$). This SNP is a nonsynonymous variant in exon 1 of *PRIM1*. The top SNP was significantly associated with expression of *PRIM1* in visual cortex, cerebellum and prefrontal cortex (**Table 2**).

Several other previously unidentified signals are located in introns of genes for which mouse models exist. These were region 8 in *ASH2L* (rs2517388; $P = 9.3 \times 10^{-15}$), region 15 in *POLG* (rs2307449; $P = 3.6 \times 10^{-13}$) and region 1b in *EXO1* (rs1635501; $P = 8.5 \times 10^{-10}$). *ASH2L* encodes 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, which is 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; $P = 1.5 \times 10^{-11}$) 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 gland–expressed hormone that stimulates maturation of follicles. Region 19a (rs11668344; $P = 1.5 \times 10^{-59}$) 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 previously unknown regions are nonsynonymous coding variants. Region 6b, rs1046089 ($P = 1.6 \times 10^{-16}$), is in exon 22 of *PRRC2A* and was associated with expression of several transcripts in the human leukocyte antigen (HLA) region in several tissues (**Table 2**). Region 19b, rs12461110 ($P = 8.7 \times 10^{-10}$) is in exon 5 of *NLRP11*. *PRRC2A* encodes HLA-B associated transcript 2 and has several microsatellite repeats. *NLRP11* encodes the nucleotide-binding domain and leucine-rich repeat–containing (NLR) family pyrin domain–containing 11 protein, which is implicated in the activation of proinflammatory caspases[22].

Of the remaining five new regions, the top SNPs for regions 1a, 2a, 2b and 13b are located in introns. These were rs4246511 in *RHBDL2* (0.24 years per minor allele, $P = 9.1 \times 10^{-17}$), which is thought to function as an intramembrane serine protease; rs2303369 in *FNDC4*, which encodes fibronectin type III domain–containing 4 ($P = 2.3 \times 10^{-12}$);

Table 1 Discovery and replication results

Region number	Most significant SNP	Chr.	Location (bp)	Minor/major	Analysis	MAF	Effect per minor allele (years)	s.e.m.	*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	0.288
					Replication	0.293	0.133	0.052	0.01	6.9	0.680
					Combined		0.240	0.029	$\mathbf{9.08 \times 10^{-17}}$	12.5	0.012
1b	rs1635501	1	240107398	C/T	Discovery	0.478	–0.188	0.032	4.90×10^{-9}	9.8	0.080
					Replication	0.456	–0.110	0.048	0.023	5.7	0.016
					Combined		–0.164	0.027	$\mathbf{8.46 \times 10^{-10}}$	8.5	0.178
2a	rs2303369	2	27568920	T/C	Discovery	0.388	–0.174	0.030	3.80×10^{-9}	9.0	0.639
					Replication	0.391	–0.179	0.047	**0.000138**	9.3	0.542
					Combined		–0.175	0.025	$\mathbf{2.25 \times 10^{-12}}$	9.1	0.927
2b	rs10183486	2	171699217	T/C	Discovery	0.366	–0.219	0.031	7.88×10^{-13}	11.4	0.430
					Replication	0.362	–0.141	0.047	0.003	7.3	0.727
					Combined		–0.196	0.026	$\mathbf{2.21 \times 10^{-14}}$	10.2	0.169
2c[a]	rs7606918	2	172603695	G/A	Discovery	0.161	–0.228	0.041	2.89×10^{-8}	11.8	0.374
					Replication not attempted						
4	rs4693089	4	84592646	G/A	Discovery	0.486	0.209	0.030	3.28×10^{-12}	10.9	0.336
					Replication	0.492	0.273	0.047	$\mathbf{6.69 \times 10^{-9}}$	14.2	0.298
					Combined		0.228	0.025	$\mathbf{2.38 \times 10^{-19}}$	11.8	0.254
5a[b]	rs890835	5	175888877	A/C	Discovery	0.112	0.266	0.047	1.17×10^{-8}	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^{-6}	9.2	0.0004
5b	rs365132	5	176311180	T/G	Discovery	0.490	0.275	0.029	1.90×10^{-21}	14.3	0.115
					Replication	0.494	0.319	0.046	$\mathbf{4.26 \times 10^{-12}}$	16.6	0.728
					Combined		0.287	0.025	$\mathbf{9.11 \times 10^{-32}}$	14.9	0.416
6a	rs2153157	6	11005474	A/G	Discovery	0.492	0.184	0.028	9.40×10^{-11}	9.5	0.858
					Replication	0.500	0.116	0.046	0.012	6.0	0.329
					Combined		0.165	0.024	$\mathbf{7.76 \times 10^{-12}}$	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	$\mathbf{1.91 \times 10^{-4}}$	9.4	0.732
					Combined		–0.213	0.026	$\mathbf{1.63 \times 10^{-16}}$	11.1	0.427
8	rs2517388	8	38096889	G/T	Discovery	0.174	0.274	0.040	1.13×10^{-11}	14.2	0.670
					Replication	0.189	0.234	0.062	$\mathbf{1.52 \times 10^{-4}}$	12.2	0.708
					Combined		0.262	0.034	$\mathbf{9.31 \times 10^{-15}}$	13.6	0.591
11	rs12294104	11	30339475	T/C	Discovery	0.172	0.226	0.040	1.63×10^{-8}	11.8	0.721
					Replication	0.180	0.223	0.060	$\mathbf{2.18 \times 10^{-4}}$	11.6	0.239
					Combined		0.225	0.033	$\mathbf{1.46 \times 10^{-11}}$	11.7	0.970
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	$\mathbf{6.89 \times 10^{-6}}$	18.0	0.765
					Combined		–0.380	0.042	$\mathbf{2.47 \times 10^{-19}}$	19.7	0.610
13a	rs3736830	13	49204222	G/C	Discovery	0.157	–0.243	0.040	1.75×10^{-9}	12.6	0.859
					Replication	0.165	–0.033	0.062	0.594	1.7	0.905
					Combined		–0.180	0.034	9.41×10^{-8}	9.4	0.005
13b	rs4886238	13	60011740	A/G	Discovery	0.334	0.172	0.031	3.76×10^{-8}	8.9	0.974
					Replication	0.333	0.166	0.049	$\mathbf{6.09 \times 10^{-4}}$	8.6	0.953
					Combined		0.170	0.026	$\mathbf{9.53 \times 10^{-11}}$	8.9	0.919
15	rs2307449	15	87664932	G/T	Discovery	0.405	–0.167	0.030	2.59×10^{-8}	8.7	0.442
					Replication	0.387	–0.225	0.047	$\mathbf{1.61 \times 10^{-6}}$	11.7	0.328
					Combined		–0.184	0.025	$\mathbf{3.56 \times 10^{-13}}$	9.6	0.294
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	$\mathbf{1.01 \times 10^{-11}}$	8.7	0.054
19a	rs11668344	19	60525476	G/A	Discovery	0.363	–0.416	0.030	5.94×10^{-43}	21.6	0.112
					Replication	0.360	–0.415	0.048	$\mathbf{2.65 \times 10^{-18}}$	21.6	0.517
					Combined		–0.416	0.026	$\mathbf{1.45 \times 10^{-59}}$	21.6	0.987
19b[c]	rs12461110	19	61012475	A/G	Discovery	0.356	–0.174	0.030	9.49×10^{-9}	9.1	0.835
					Replication	0.344	–0.117	0.049	0.018	6.1	0.542
					Combined		–0.158	0.026	$\mathbf{8.74 \times 10^{-10}}$	8.2	0.320
20	rs16991615	20	5896227	A/G	Discovery	0.069	0.971	0.062	1.16×10^{-54}	50.5	0.356
					Replication	0.070	0.896	0.096	$\mathbf{7.90 \times 10^{-21}}$	46.6	0.088
					Combined		0.948	0.052	$\mathbf{1.42 \times 10^{-73}}$	49.3	0.509

Chr., chromosome. MAF, minor allele frequency. Heterogeneity, *P* values for heterogeneity among discovery studies and replication studies, and comparing all discovery to all replication studies. *P* value, replication *P* values that meet the criterion $P < 0.05/19 \approx 0.026$ are in bold. Combined analysis *P* values that reached genome-wide significance are in bold. [a]Conditional analysis: beta (SE): –0.199 (0.041); $P = 9.8 \times 10^{-7}$. [b]Conditional analysis beta (SE): 0.267 (0.046); $P = 6.5 \times 10^{-9}$. [c]Conditional analysis: beta (SE): –0.168 (0.031); $P = 3.8 \times 10^{-8}$.

Table 2 Characteristics of top SNP in each region

Region	SNP ID	Chr.	Location (bp)	Gene	Feature	Other ref. genes <60 kb from SNP	Genes with SNPs in LD (r^2 > 0.5) with SNP (gene symbol:r^2)	eQTL
1a	rs4246511	1	39152972	*RHBDL2*	Intron	*MYCBP*, *GJA10*, *RRAGC*	*MYCBP*:0.678; *RHBDL2*:0.678; *RRAGC*:0.678; GJA9:0.678; *LOC100130627*:0.678	
1b	rs1635501	1	240107398	*EXO1*	Intron		*LOC100131576*:0.678; *WDR64*:0.923; *LOC100133057*:1; EXO1:1	
2a	rs2303369	2	27568920	*FNDC4*	Intron	*GCKR*, *KRTCAP3*, *IFT172*, *NRBP1*	*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 *IFT172* in lymph, adipose and blood; r^2 > 0.9 with multiple eSNPs for *KRTCAP3* in lymph and CD4; r^2 > 0.9 with eSNP for *SNX17* in PFC
2b	rs10183486	2	171699217	*TLK1*	Intron		*GORASP2*:0.929; TLK1:0.964	r^2 > 0.9 with eSNPs for *TLK1* in LCL and PFC
4	rs4693089	4	84592646	*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	175888877	*RNF44*	Intron	*UBXD8*, *PCLKC*	*SNCB*:0.318; *RNF44*:1; *FAF2*:1; *PCDH24*:1	
5b	rs365132	5	176311180	*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 *ZNF346* in PFC; for *ZNF346* in VC
6a	rs2153157	6	11005474	*SYCP2L*	Intron	*GCM2*	*MAK*:0.602; *GCM2*:0.602; *SYCP2L*:0.602	eSNP for *SYCP2L* in monocytes
6b	rs1046089[a]	6	31710946	*PRRC2A*	Missense	*BAT3*, *LST1*, *C6orf47*, *APOM*, *AIF1*, *NCR3*, *LY6G5C*, *LTB*, *BAT5*, *CSNK2B*, *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	38096889	*ASH2L*	Intron	*BAG4*, *EIF4EBP1*, *LSM1*, *STAR*	*STAR*:0.831; *ASH2L*:0.831; *LSM1*:0.831	
11	rs12294104	11	30339475	–		*MPPED2*, *C11orf46*	*FSHB*:0.92; *C11orf46*:1	r^2 > 0.9 with eSNP for *C11orf46* in lymph
12	rs2277339[b]	12	55432336	*PRIM1*	Missense	*HSD17B6*, *NACA*	–	eSNP for *PRIM1* in VC, CR and PFC
13a	rs3736830	13	49204222	*KPNA3*	Intron	*EBPL*	*KPNA3*:0.734; *EBPL*:0.734; *ARL11*:0.623; *LOC100131941*:0.623	
13b	rs4886238	13	60011740	*TDRD3*	Intron		*TDRD3*:1; *LOC390407*:0.731	r^2 > 0.8 with eSNP in adipose for *TDRD3*
15	rs2307449	15	87664932	*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	11924420	–		*TNFRSF17*, *RUNDC2A*, *GSPT1*	*TNFRSF17*:0.662; *GSPT1*:1; *COX6CP1*:1; *RSL1D1*:1; *ZC3H7A*:0.701; *RUNDC2A*:0.662; *LOC729978*:1	
19a	rs11668344	19	60525476	*TMEM150B*	Intron	*BRSK1*, *HSPBP1*, *COX6B2*, *LOC284417*, *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 *MGC2705* in adipose and blood
19b	rs12461110[c]	19	61012475	*NLRP11*	Missense	*NLRP4*	*NLRP4*:0.514; *NLRP11*:0.514; *RFPL4A*:0.514; *LOC646663*:0.514; *LOC729974*:0.514	
20	rs16991615[d]	20	5896227	*MCM8*	Missense	*CRLS1*, *CHGB*, *TRMT6*	–	

CD4, $CD4^+$ lymphocyte cells; PFC, prefrontal cortex cells; LCL, lymphoblastoid cell line; CR, cerebellum; VC, visual cortex cells.
[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).

rs10183486 in *TLK1* ($P = 2.2 \times 10^{-14}$), a nuclear serine-threonine kinase that is potentially involved in the regulation of chromatin assembly; and rs4886238 in *TDRD3* ($P = 9.5 \times 10^{-11}$). *TDRD3* is a binding partner for *FMR1*, which has been associated with primary ovarian insufficiency (POI). The top SNP in the final newly identified region, 16, is within 60 kb of three genes, *TNFRSF17*, *GSPT1* and *RUNDC2A*. It is in LD ($r^2 > 0.5$) with SNPs in these three genes and four others (rs10852344; $P = 1.0 \times 10^{-11}$; **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. Although all three approaches (ingenuity pathway analysis (IPA, see URLs), MAGENTA[23] and GRAIL[24]) are based on published data, thus linking the gene products of our top hits to each other in functional pathways, each uses a substantially different methodology and uses 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. On the basis of the genes physically nearest the 17 loci, we identified four major functional networks applying direct interactions only (**Supplementary Table 4**). Network 1, related to lipid metabolism, molecular transport and small molecule biochemistry, contained 14 of the genes nearest the menopause loci ($P = 1 \times 10^{-30}$). Central to this network is the *HNF4A* gene, which has a role in diabetes. Network 2, containing 12 of the input genes, relates to cell cycle, cell death and cancer ($P = 1 \times 10^{-24}$). The *ESR1* gene is central in this network, suggesting that genes in this network influence or are influenced by estrogen signaling. Network 3 is also partially related to cell death, and includes TNF and NF-κB ($P = 1 \times 10^{-19}$). Network 4 relates to infection mechanism, DNA replication, recombination and repair and gene expression ($P = 1 \times 10^{-12}$). Notably, several of the input genes included in network 1 (*EXO1* and *HELQ*) and network 2 (*UIMC1*, *FANCI* and *TLK1*) are also involved in DNA repair mechanisms.

We used 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 (false discovery rate (FDR) < 0.05), exoDNase ($P = 0.0005$), NF-κB signaling ($P = 0.0006$) 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 genes near our top SNPs. Genes are considered related if they share informative words. GRAIL scores for genes associated with three of the replicating genome-wide significant SNPs were significant, *EXO1*, *FKBPL* and *BRSK1*. When we applied this method 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* and *REV3L*. *BRSK1* was also borderline significant with GRAIL FDR = 0.06.

Candidate genes

Within the discovery GWAS, 18,327 SNPs were within 60 kb 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* and *RFPL4A*) were significantly associated with age at natural menopause. 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 $P = 1.6 \times 10^{-7}$ ($P = 0.0009$ corrected for candidate gene SNP analyses); age at natural menopause was lower by ~0.18 years per copy of the minor allele. *DMC1* encodes a protein that is essential for meiotic homologous recombination and is regulated by *NOBOX*, mutations in which can cause 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 1 Mb of a previously published genome-wide significant SNP, but most of the colocalized 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 glycemic 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 two-stage GWAS, we confirmed four established menopause loci and identified and replicated 13 loci newly associated with age at natural menopause. Of these 17 hits, all but two are intronic or exonic to known genes. For associated SNPs in GWAS, on average 40% are intergenic, whereas only 2% of our hits are intergenic. Furthermore, we found twice the nonsynonymous top hits typically observed in GWAS (24% versus 12%; ref. 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 show that longevity and DNA repair function are generally related[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 progress toward menopause owing to follicle depletion[34].

In biological pathway 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* and *PRIM1*; **Supplementary Table 4**). The protein encoded by *UIMC1* physically interacts with BRCA1 and estrogen receptor α and is thought to recruit BRCA1 to DNA damage sites and to initiate checkpoint control in the G2/M phase of the cell cycle. PRIM1 (primase) is involved in DNA replication by synthesizing RNA primers for Okazaki fragments during discontinuous replication[35]. A mutation in *POLG* can segregate with POI[36]. *Polg* knock-in mice show lower lifespan, premature aging

and lower fertility compared with wild type[37]. *FANCI*, another gene at the same locus adjacent to *POLG*, is a member of the Fanconi anemia complementation group. Fanconi anemia is a recessive disorder characterized by cytogenetic instability and defective DNA repair. Fanconi anemia patients experience irregular menstruation with menopause occurring around age 30 (ref. 38). The functional polymorphism correlated to our top hit in *EXO1* is associated with longevity in female centenarians[39]. Male and female *Exo1* knockout mice are sterile because the gene is essential for male and female meiosis[40]. In addition to the GWAS regions in or near genes associated with early menopause, we investigated a panel of candidate genes identified before 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. With altered DNA repair mechanisms, damage could accumulate, rendering poor-quality oocytes for selection. In contrast, the number of damaged follicles may increase with aging, leading to a greater rate of follicle loss through atresia. The top hit in this study, a nonsynonymous 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 prereplication 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, but with subgenome-wide levels of significance for association with age of natural menopause. These 12 candidates (**Supplementary Table 5**) included the gene encoding Werner helicase (WRN), mutations in which cause Werner syndrome, a classic progeria with advanced aging phenotype and ovarian aging[42]. Estrogen can enhance WRN expression, preventing cell senescence, suggesting that WRN is involved in menopause[43]. The identification of DNA repair as one of the biological pathways involved in menopause may also explain 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, is associated with colorectal adenomas in smokers only[44]. Additional studies are needed to determine whether smoking status modifies the association between age at natural menopause and polymorphisms in DNA repair genes, as has been observed for various cancers.

Pathway-based analysis indicated that genes related to autoimmune 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 autoimmunity can have a role[45]. 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 interleukin 11 (IL-11) receptor α chain. Female mice with null mutations in *Il11ra* are infertile owing to defective uterine decidualization, the process necessary for successful embryo implantation[47]. *NLRP11* (region 19b) is a member of the NLRP family of genes, which have important roles in the innate immune system and reproductive system. Several NLRP genes show an oocyte-specific expression pattern[46], whereas *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 *PRRC2A* (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 *PRRC2A* SNPs in GWAS, 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 show 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 a candidate for a proinflammatory component to oocyte depletion that affects menopause age. Indeed, the enrichment of genes involved in NF-κB signaling (*TNF*, *TNFRSF17* and *CSNK2B*) in biological pathway 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 identified (DNA repair and inflammation) may contribute to variation in age at natural menopause.

Three of the 17 regions can be linked to hormonal regulation, an additional route to follicle pool exhaustion. The top SNP in region 11 (rs12294104) is in high LD with SNPs in *FSHB* ($r^2 = 0.92$, **Table 2**), which limits the rate of production of FSH, a key pituitary gland–expressed hormone that stimulates maturation of follicles. FSH-deficient female mice are infertile[53]. Transgenic mice that overexpress FSH show premature infertility owing to postimplantation reduction of embryo-fetal survival[54]. FSH concentrations rise in women approaching menopause; this might be related to a decrease in growing follicles[55]. Mutations in *FSHB* cause hypogonadism and primary amenorrhea in women[56] and lead to greater FSH concentrations and infertility in males compared with wild type[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*, which encodes 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 a probable 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, *BCAR4*, which encodes 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 have a role in hormonal stimulation in the ovary. In breast cancer treatment, tumors highly expressing *BCAR4* are more resistant to tamoxifen treatment[61], reinforcing the role of *BCAR4* in transduction of hormonal signals.

In summary, our findings demonstrate the role of genes that regulate DNA repair and immune function, and genes affecting neuroendocrine pathways of ovarian function in regulating age at menopause, indicating that the process of aging is involved in both somatic and germ line aging.

We expect that several additional common variants with small effects on age at natural menopause are yet to be identified, and that many of them are in genes in pathways identified in this study. Sequencing and exome chip studies to determine whether low-frequency and rare variants of large effect also contribute to age at natural menopause are underway or being planned in many of the cohorts involved in this GWAS. A collaboration of several consortia is examining the contribution of common genetic variants to age at natural menopause in African-American women, and could allow researchers to determine whether the genetic variation that affects age at natural menopause in African-American women is the same or substantially different from that for women of primarily European descent. We are now conducting a study of women with POI to determine whether variants associated with age at natural menopause

within the normal range of age 40–60 also contribute to disease conditions related to the early-menopause phenotype.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to the study participants and staff from all cohorts involved in this study. Extended acknowledgments per cohort are in **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Individual study design and management were done by D.I.C., C. He, E.M.B., P.K., J.S.B., M. Boban, E.B., H.C., S.J.C., U.d.F., I.J.D., G.V.Z.D., S.E., J.G.E., L. Ferrucci, A.R.F., P.G., C.J.M.W.G., C.G., D.E.G., P.H., S.E.H., A.H., E.I., S.L.R.K., D.A.L., P.K.E.M., M. Marongiu, N.G.M., V.M., N.C.O.-M., G. Paré, A.N.P., N.L.P., P.H.M.P., O.P., B.M.P., K.R., I.R., J.M.S., R.P.S., V.J.M.P., V.G., G.E., T.B.H., L.J.L.,Y.V.S., A.T., Y.T.v.d.S., P.M.V., G. Waeber, H.E.W., J.F.W., B.H.R.W., L.M.L., D.I.B., J.E.B., L.C., F.B.H., D.J.H., G.W.M., B.A.O., P.M.R., D.S., T.D.S., K. Stefansson, E.A.S., M.U., C.M.v.D., H.V., S.B., A. Salumets and A. Metspalu. Data collection was done by B.Z.A., S.B., L.B.L., J.S.B., M. Boban, A.B., H.C., P.dA., U.dF., I.J.D., G.V.Z.D., S.E., J.G.E., L. Ferreli, L. Ferrucci, A.R.F., D.E.G., P.H., S.E.H., A.C.H., A.H., A.C.J.W.J., I.K., S.L., D.A.L., P.K.E.M., N.G.M., I.M.K., A.B.N., N.L.P., P.H.M.P., O.P., B.M.P., K.R., I.R., A. Scuteri, S.N.S., J.M.S., M.G.S., U.S., B.T., L.T., S.U., V.J.M.P., V.G., G.E., Y.T.v.d.S., C.H.v.G., J.M.V., G. Waeber, H.E.W., A.F.W., T.Z., L.Z., M.C.Z., M.Z., L.M.L., A.M.A., J.E.B., E.A.S., A.G.U., J.M.M., P.v.N. and A. Metspalu. Genotyping was done by J.K., L.S., T.E., E.B., F.J.B., S.J.C., P.dA., G.D., P.D., C. Hayward, A.C.H., J. Liu, B.C.J.M.F., S.E.M., V.M., P.N., D.R.N., N.C.O.-M., A.S.P., A.N.P., B.M.P., J.I.R., A. Singleton, K. Stirrups, A.T., K.T., Y.T.v.d.S., M.V., E.W., T.Z., L.M.L., F.B.H., G.W.M., B.A.O. and U.T. Genotype preparation was done by J.R.B.P., M. Barbalic, N.F., E.P., S.-Y.S., W.V.Z., L.B.L., P.dA., G.D., C.J.M.W.G., C.G., C. Hayward, E.I., S.L.R.K., P.K.E.M., I.M.K., C.M., S.E.M., P.N., D.R.N., N.C.O.-M., A.P., G. Pistis, E.R., C.S., J.A.S., H.S., N.S., Y.V.S., A.T., M.V., E.W., T.Z., L.M.L., F.B.H., D.J.H., S.S., K.L.L. and M.H. Phenotype preparation was done by J.R.B.P., L.S., C. He, M. Mangino, M. Barbalic, L.B., E.M.B., F.E., N.F., D.F.G., J.-J.H., P.K., G.Z., W.V.Z., L.B.L., M. Boban, E.J.C.d.G., I.J.D., G.V.Z.D., M.G.S., S.E., J.G.E., C.H.v.G., L. Ferreli, K.F., M.H., C. Hayward, E.I., A.C.J.W.J., S.L.R.K., I.K., J. Lahti, S.L., T.L., D.A.L., L.M.L., P.K.E.M., I.M.K., C.M., P.H.M.P., G. Pistis, O.P., E.R., C.S., A. Scuteri, J.A.S., B.T., S.U., R.M.v.D., V.G., G.E., T.A., Y.T.v.d.S., H.W., G. Willemsen, B.H.R.W., A.F.W., M.C.Z., L.M.L., A.M.A., E.W.D., A. Metspalu, K.L.L. and J.M.M. Analysis plan development was done by L.S., D.I.C., C. He, E.M.B., P.K., T.C., P.G., D.K., D.P.K., N.G.M., D.T., D.J.H., G.W.M., K.L.L., J.M.M. and A. Murray. Analysis plan review was done by L.S., D.I.C., C. He, A.V.S., A.D.C., N.G., D.K., D.P.K., B.M., A.B.N., B.M.P., A.M.A., D.J.H., K.L.L., J.M.M. and A. Murray. Study data analysis was done by L.S., D.I.C., C. He, M. Mangino, M. Barbalic, L.B., E.M.B., F.E., T.E., N.F., D.F.G., J.-J.H., P.K., P.F.M., E.P., S.-Y.S., A.V.S., S.v.W., G.Z., W.V.Z., E.A., C.C., M.C.C., T.C., N.G., T.H., C. Hayward, Z.K., J. Lahti, D.A.L., L.M.L., D.M., N.C.O.M., G. Paré, G. Pistis, A.S.P., V.E., E.R., J.A.S., M.G.S., S.U., Y.T.vdS., M.V., L.M.Y.-A., L.Z., S.S. and U.T. Review and interpretation of analyses were done by J.R.B.P., L.S., D.I.C., C. He, P.S., M. Barbalic, E.M.B., N.F., P.K., P.F.M., A.V.S., B.Z.A., S.J.C., A.D.C., I.J.D., S.E.H., E.I., D.K., S.L.R.K., D.P.K., L.M.L., P.K.E.M., N.G.M., G. Paré, A.S.P., J.A.S., H.S., L.M.Y.-A., J.E.B., E.W.D., F.B.H., D.J.H., G.W.M., P.M.R., C.M.v.D., H.V., K.L.L., J.M.M., P.v.N. and A. Murray. Meta-analyses were done by L.S., P.S. and K.L.L. Pathway and other analyses were done by J.R.B.P., D.I.C., C. He, A.D.J., M. Mangino, G. Paré and J.A.V. Menopause study was designed by J.R.B.P., L.S., D.I.C., C. He, D.E.G., P.H.M.P., Y.T.v.d.S., C.H.v.G., A. Metspalu, K.L.L., J.M.M. and A. Murray. Manuscript was prepared by J.R.B.P., L.S., D.I.C., C. He, T.E., N.F., A.D.C., D.K., D.P.K., E.W.D., A. Metspalu, K.L.L., J.M.M., A. Murray and J.A.V. Manuscript was reviewed by J.R.B.P., L.S., D.I.C., C. He, N.F., P.K., P.F.M., A.V.S., B.Z.A., T.A., L.B.L., J.S.B., M. Boban, F.J.B., H.C., S.J.C., C.C., M.C.C., A.D.C., G.D., U.d.F., I.J.D., G.E., B.C.J.M.F., M.E.G., N.G., D.E.G., P.H., S.E.H., C. Hayward, E.I., D.K., D.P.K., S.L.R.K., J.A.S., I.K., Z.K., T.L., T.E., A. Salumets, A. Metspalu, J.S.E.L., J. Liu, L.M.L., Y.V.L., P.K.E.M., I.M.K., B.M., V.M., P.N., A.B.N., N.C.O.-M., G. Paré, A.N.P., N.L.P., P.H.M.P., O.P., B.M.P., J.I.R., I.R., H.S., J.M.S., R.P.S., A.T., R.M.v.D., Y.T.v.d.S., C.H.v.G., P.M.V., M.V., G. Waeber, J.F.W., B.H.R.W., A.F.W., L.M.Y.-A., T.Z., L.Z., M.C.Z., V.J.M.P., L.M.L., A.M.A., D.I.B., J.E.B., E.W.D., V.G., T.B.H., F.B.H., D.J.H., L.J.L., P.M.R., T.D.S., E.A.S., H.V., K.L.L., J.M.M., A. Murray, P.v.N. and J.A.V. Consortium was overseen by K.L.L., J.M.M., A. Murray and J.A.V.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Burger, H.G. The menopausal transition. *Baillieres Clin. Obstet. Gynaecol.* **10**, 347–359 (1996).
2. Broekmans, F.J., Soules, M.R. & Fauser, B.C. Ovarian aging: mechanisms and clinical consequences. *Endocr. Rev.* **30**, 465–493 (2009).
3. te Velde, E.R. & Pearson, P.L. The variability of female reproductive ageing. *Hum. Reprod. Update* **8**, 141–154 (2002).
4. te Velde, E.R., Dorland, M. & Broekmans, F.J. Age at menopause as a marker of reproductive ageing. *Maturitas* **30**, 119–125 (1998).
5. Murabito, J.M., Yang, Q., Fox, C., Wilson, P.W. & Cupples, L.A. Heritability of age at natural menopause in the Framingham Heart Study. *J. Clin. Endocrinol. Metab.* **90**, 3427–3430 (2005).
6. Snieder, H., MacGregor, A.J. & Spector, T.D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J. Clin. Endocrinol. Metab.* **83**, 1875–1880 (1998).
7. van Asselt, K.M. *et al.* Heritability of menopausal age in mothers and daughters. *Fertil. Steril.* **82**, 1348–1351 (2004).
8. Vink, J. & Boomsma, D.I. Modeling age at menopause. *Fertil. Steril.* **83**, 1068 (2005).
9. He, C. *et al.* A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum. Genet.* **128**, 515–527 (2010).
10. He, L.N. *et al.* Association study of the oestrogen signalling pathway genes in relation to age at natural menopause. *J. Genet.* **86**, 269–276 (2007).
11. Weel, A.E. *et al.* Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. *J. Clin. Endocrinol. Metab.* **84**, 3146–3150 (1999).
12. Tempfer, C.B. *et al.* Polymorphisms associated with thrombophilia and vascular homeostasis and the timing of menarche and menopause in 728 white women. *Menopause* **12**, 325–330 (2005).
13. van Asselt, K.M. *et al.* Factor V Leiden mutation accelerates the onset of natural menopause. *Menopause* **10**, 477–481 (2003).
14. He, C. *et al.* Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.* **41**, 724–728 (2009).
15. Stolk, L. *et al.* Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat. Genet.* **41**, 645–647 (2009).
16. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
17. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
18. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
19. Marini, F. & Wood, R.D. A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. *J. Biol. Chem.* **277**, 8716–8723 (2002).
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 X inactivation. *Development* **137**, 935–943 (2010).
21. Copeland, W.C. & Longley, M.J. DNA polymerase gamma in mitochondrial DNA replication and repair. *ScientificWorldJournal* **3**, 34–44 (2003).
22. Tschopp, J., Martinon, F. & Burns, K. NALPs: a novel protein family involved in inflammation. *Nat. Rev. Mol. Cell Biol.* **4**, 95–104 (2003).
23. Segrè, A.V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** pii, e1001058 (2010).
24. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
25. Bishop, D.K., 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* **69**, 439–456 (1992).
26. Qin, Y. *et al.* NOBOX homeobox mutation causes premature ovarian failure. *Am. J. Hum. Genet.* **81**, 576–581 (2007).
27. Rajkovic, A., Pangas, S.A., Ballow, D., Suzumori, N. & Matzuk, M.M. NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. *Science* **305**, 1157–1159 (2004).
28. Manolio, T.A. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166–176 (2010).
29. Johnson, F.B., Sinclair, D.A. & Guarente, L. Molecular biology of aging. *Cell* **96**, 291–302 (1999).
30. Lee, C.K., Allison, D.B., Brand, J., Weindruch, R. & Prolla, T.A. Transcriptional profiles associated with aging and middle age-onset caloric restriction in mouse hearts. *Proc. Natl. Acad. Sci. USA* **99**, 14988–14993 (2002).
31. Hamatani, T. *et al.* Age-associated alteration of gene expression patterns in mouse oocytes. *Hum. Mol. Genet.* **13**, 2263–2278 (2004).
32. Promislow, D.E. DNA repair and the evolution of longevity: a critical analysis. *J. Theor. Biol.* **170**, 291–300 (1994).
33. Du, X. *et al.* Telomere shortening exposes functions for the mouse Werner and Bloom syndrome genes. *Mol. Cell. Biol.* **24**, 8437–8446 (2004).

34. Dorland, M., van Kooij, R.J. & te Velde, E.R. General ageing and ovarian ageing. *Maturitas* **30**, 113–118 (1998).
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* **28**, 350–353 (1995).
36. Pagnamenta, A.T. *et al.* Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. *Hum. Reprod.* **21**, 2467–2473 (2006).
37. Trifunovic, A. *et al.* Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417–423 (2004).
38. Wong, J.C. & Buchwald, M. Disease model: Fanconi anemia. *Trends Mol. Med.* **8**, 139–142 (2002).
39. Nebel, A. *et al.* A functional EXO1 promoter variant is associated with prolonged life expectancy in centenarians. *Mech. Ageing Dev.* **130**, 691–699 (2009).
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.* **17**, 603–614 (2003).
41. Tye, B.K. Insights into DNA replication from the third domain of life. *Proc. Natl. Acad. Sci. USA* **97**, 2399–2401 (2000).
42. Epstein, C.J., Martin, G.M., Schultz, A.L. & Motulsky, A.G. Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. *Medicine (Baltimore)* **45**, 177–221 (1966).
43. Lee, S.J., Lee, S.H., Ha, N.C. & Park, B.J. Estrogen prevents senescence through induction of WRN, Werner syndrome protein. *Horm. Res. Paediatr.* **74**, 33–40 (2010).
44. Gao, Y. *et al.* DNA repair gene polymorphisms and tobacco smoking in the risk for colorectal adenomas. *Carcinogenesis* **32**, 882–887 (2011).
45. Welt, C.K. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clin. Endocrinol.* **68**, 499–509 (2008).
46. Hoek, A., Schoemaker, J. & Drexhage, H.A. Premature ovarian failure and ovarian autoimmunity. *Endocr. Rev.* **18**, 107–134 (1997).
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.* **4**, 303–308 (1998).
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.* **87** 697 e5–e8 (2007).
49. Tong, Z.B. & Nelson, L.M. A mouse gene encoding an oocyte antigen associated with autoimmune premature ovarian failure. *Endocrinology* **140**, 3720–3726 (1999).
50. Anasti, J.N. *et al.* Karyotypically normal spontaneous premature ovarian failure: evaluation of association with the class II major histocompatibility complex. *J. Clin. Endocrinol. Metab.* **78**, 722–723 (1994).
51. Jaroudi, K.A., Arora, M., Sheth, K.V., Sieck, U.V. & Willemsen, W.N. Human leukocyte antigen typing and associated abnormalities in premature ovarian failure. *Hum. Reprod.* **9**, 2006–2009 (1994).
52. Gasser, S. & Raulet, D.H. The DNA damage response arouses the immune system. *Cancer Res.* **66**, 3959–3962 (2006).
53. Kumar, T.R., Wang, Y., Lu, N. & Matzuk, M.M. Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. *Nat. Genet.* **15**, 201–204 (1997).
54. McTavish, K.J. *et al.* Rising follicle-stimulating hormone levels with age accelerate female reproductive failure. *Endocrinology* **148**, 4432–4439 (2007).
55. Lambalk, C.B., De Koning, C.H. & Braat, D.D. The endocrinology of dizygotic twinning in the human. *Mol. Cell. Endocrinol.* **145**, 97–102 (1998).
56. Kottler, M.L. *et al.* A new FSHβ 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.* **162**, 633–641 (2010).
57. Grigorova, M. *et al.* Increased prevalance of the -211 T allele of follicle stimulating hormone (FSH) β subunit promoter polymorphism and lower serum FSH in infertile men. *J. Clin. Endocrinol. Metab.* **95**, 100–108 (2010).
58. Hoogendoorn, B. *et al.* Functional analysis of human promoter polymorphisms. *Hum. Mol. Genet.* **12**, 2249–2254 (2003).
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.* **55**, 1043–1049 (2008).
60. Pisarska, M.D., Bae, J., Klein, C. & Hsueh, A.J. Forkhead l2 is expressed in the ovary and represses the promoter activity of the steroidogenic acute regulatory gene. *Endocrinology* **145**, 3424–3433 (2004).
61. Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L.C. & Agthoven, T.V. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell. Physiol.* **226**, 1741–1749 (2011).

Lisette Stolk[1,2,117], John R B Perry[3,4,117], 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[15–17], Nora Franceschini[18], Daniel F Gudbjartsson[10], Jouke-Jan Hottenga[19], Peter Kraft[20,21], Patrick 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 J C de Geus[19,47], Ian J Deary[45,48], George V Z Dedoussis[49], Panagiotis Deloukas[24], Shah Ebrahim[50], Gudny Eiriksdottir[25], Valur Emilsson[25], Johan G Eriksson[51–55], Bart C J M 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 J W Janssens[12], Andrew D Johnson[42], David Karasik[42,64], Sharon L R 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 S E Laven[70], Debbie A Lawlor[71], Jianjun Liu[72], Lorna M Lopez[45,48], Yvonne V Louwers[70], Patrik K E 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 H M Peeters[59,82], Giorgio Pistis[40], Andrew S Plump[83], Ozren Polasek[35], Victor J M 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[39,99,100], Yvonne T van der Schouw[59], Carla H van Gils[59], Peter van Nierop[101], Jacqueline M Vink[19], Peter M Visscher[48,102], Marlies Voorhuis[36,59], Gérard Waeber[103], Henri Wallaschofski[104], H Erich Wichmann[105–107], Elisabeth Widen[77], Colette J M Wijnands-van Gent[108], Gonneke Willemsen[19], James F Wilson[37], Bruce H R Wolffenbuttel[91,109], Alan F Wright[62], Laura M Yerges-Armstrong[22],

Tatijana Zemunik[35], Lina Zgaga[37,110], M Carola Zillikens[1], Marek Zygmunt[111], The LifeLines Cohort Study[91], Alice M Arnold[74], Dorret I Boomsma[19,47], Julie E Buring[5,6,112], Laura Crisponi[23], Ellen W Demerath[57], Vilmundur Gudnason[25,26], Tamara B Harris[58], Frank B Hu[20,39,61], David J Hunter[20,21,39,61], Lenore J Launer[58], Andres Metspalu[15–17,88], Grant W Montgomery[13], Ben A Oostra[113], Paul M Ridker[5,6,112,114], Serena Sanna[23], David Schlessinger[115], 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[116], Anna Murray[3,117], Joanne M Murabito[41,42,117], Jenny A Visser[1,117] & Kathryn L Lunetta[28,42,117]

[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, 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, Massachusetts, USA. [6]Harvard Medical School, Boston, Massachusetts, 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 MC, Rotterdam, The Netherlands. [13]Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [14]Interfakultäres Institut für Genomforschung, Universität Greifswald, 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]Department of 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, Cambridge, Massachusetts, 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, Newfoundland and Labrador, 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]Croatian Science Foundation, Zagreb, Croatia. [33]Department of Medical Genetics, University of Lausanne, 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, US National Institutes of Health (NIH), 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, Massachusetts, USA. [42]National Heart, Lung and Blood Institute Framingham Heart Study, Framingham, Massachusetts, USA. [43]Institute for Maternal and Child Health, Istituto di Ricovero e Cura a Carattere Scientifico Materno-infantile Burlo Garofolo Trieste, 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 for Health and Care Research, VU Medical Center, Amsterdam, The Netherlands. [48]Centre for Cognitive Aging and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK. [49]Department of Nutrition and Dietetics, Harokopio University, Athens, Greece. [50]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [51]National Institute for Health and Welfare, Helsinki, Finland. [52]Department of General Practice and Primary Health Care, University of Helsinki, 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 (NIA), 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, NIA, NIH, Bethesda, Maryland, 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]Medical Research Council (MRC) Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. [63]Washington University St. Louis, St. Louis, Missouri, USA. [64]Hebrew SeniorLife Institute for Aging Research, Harvard Medical School, Boston, Massachusetts, USA. [65]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA. [66]Diagnostic GP Laboratory Eindhoven, Eindhoven, The Netherlands. [67]Swiss Institute of Bioinformatics, Lausanne, 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 and Gynaecology, Erasmus MC, Rotterdam, The Netherlands. [71]MRC Centre for Causal Analysis in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, 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, Washington, 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, University of Helsinki, Helsinki, Finland. [78]Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland. [79]Genetic and Molecular Epidemiology Laboratory, McMaster University, Hamilton, Ontario, Canada. [80]Amgen, Cambridge, Massachusetts, USA. [81]Foundation Medicine, Cambridge, Massachusetts, USA. [82]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK. [83]Cardiovascular Diseases, Merck Research Laboratory, Rahway, New Jersey, 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, Washington, USA. [86]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [87]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [88]Competence Centre on Reproductive Medicine and Biology, Tartu, Estonia. [89]Lab Cardiovascular Sciences, NIA, NIH, Baltimore, Maryland, USA. [90]Laboratory of Neurogenetics, NIA, Bethesda, Maryland, USA. [91]LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [92]Geriatric Medicine Unit, University of Edinburgh, Edinburgh, UK. [93]Cardiovascular Genetics Research Unit, Université Henri Poincaré, Nancy, France. [94]Department of Epidemiology, Emory University, Atlanta, Georgia, USA. [95]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, Consiglio Nazionale delle Ricerche, Pavia, Italy. [98]Icelandic Cancer Registry, Reykjavik, Iceland. [99]Saw Swee Hock School of Public Health, National University of Singapore, Singapore. [100]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [101]Municipal Health Service Brabant-Zuidoost, Helmond, The Netherlands. [102]Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [103]Department of Internal Medicine, CHUV, University Hospital, Lausanne, Switzerland. [104]Institute for Clinical Chemistry and Laboratory Medicine, University of Greifswald, Greifswald, Germany. [105]Institute of Epidemiology I, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany. [106]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [107]Klinikum Grosshadern, Munich, Germany. [108]POZOB Veldhoven, Veldhoven, The Netherlands. [109]Department of Endocrinology, University Medical Center Groningen, University of Groningen, The Netherlands. [110]Andrija Stampar School of Public Health, Medical School, University of Zagreb, Zagreb, Croatia. [111]Klinik für Gynäkologie und Geburtshilfe, Universität Greifswald, Greifswald, Germany. [112]Harvard School of Public Health, Boston, Massachusetts, USA. [113]Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands. [114]Division of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [115]NIA, Intramural Research Program, Baltimore, Maryland, USA. [116]Institut für Community Medicine, Universität Greifswald, Greifswald, Germany. [117]These authors contributed equally to this work. Correspondence should be addressed to A.M. (anna.murray@pms.ac.uk) or K.L.L. (klunetta@bu.edu).

ONLINE METHODS

Discovery. Age at natural menopause was defined as age at last menstrual period that occurred naturally with at least 12 consecutive months of amenorrhea. This analysis included women with natural menopause aged 40–60. Women of self-reported non-European ancestry were excluded, as were women with menopause owing to hysterectomy and/or bilateral ovariectomy, chemotherapy or irradiation, if validated by medical records, and women using HRT before menopause. Most cohorts collected age at natural menopause retrospectively; in the Framingham Offspring, the Atherosclerosis Risk in Communities Study (ARIC), Nurses Health Study (NHS) and WGHS studies, some women became menopausal under study observation. Study-specific questions, mean age at menopause and age at interview are in **Supplementary Table 1**. Genotyping and imputation information for discovery cohorts are in **Supplementary Table 2**. Descriptions of each study are in **Supplementary Note**. For all studies participating in the meta-analyses, each participant provided written informed consent. The Institutional Review Board at the parent institution for each respective cohort approved the study protocol.

Replication. A total of 14,435 women from 21 studies meeting the same inclusion and exclusion criteria as the women in discovery analysis were included in replication analysis. The women had mean and s.d. of age at natural menopause similar to the discovery set (**Supplementary Table 1**). Genotyping and imputation methods for the *in silico* replication cohorts are in **Supplementary Table 2**. Genotyping information for the studies that genotyped the SNPs *de novo* is in **Supplementary Table 2**. Descriptions of each study are in **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 done with METAL using genomic control[62]. A SNP within a study was omitted if the minor allele frequency was <1% or imputation quality score was <0.2. The discovery meta-analysis included 2,551,160 autosomal SNPs and 38,968 samples.

Expression quantitative trait locus 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 (LCLs) derived from asthmatic children[66], HapMap LCLs from three populations[67], HapMap CEU LCLs[68], fibroblasts, T cells and LCLs derived from cord blood[69], peripheral blood monocytes[70,71], $CD4^+$ lymphocytes[72], adipose and blood samples[73], brain cortex[70,74], brain regions including prefrontal cortex, visual cortex and cerebellum (three large studies; V.E., unpublished data), cerebellum, frontal cortex, temporal cortex and caudal pons[75], 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. eQTL findings for replicated GWAS SNPs are summarized in **Table 2**.

Conditional analysis. On each chromosome, SNPs of the lowest *P* value that met genome-wide significance were identified. Genome-wide-significant SNPs >250,000 bp and <1 Mb apart that also had pairwise HapMap CEU LD values of $r^2 < 0.5$ were considered potentially independent regions. Potential independent regions that were within 1 Mb of a second region with a 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 retested for independent association.

Pathway analyses. 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 60 kb 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 molecular relationships between genes or gene products were graphically represented. Proteins are depicted as nodes in various shapes representing functional class of protein. Lines depict the biological relationships between nodes. To determine the probability that the analyzed gene would be found in a network from Ingenuity Pathways Knowledge Base owing 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, determined by 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 data set. MAGENTA implements a GSEA-based approach that has been described[23]. Briefly, each gene in the genome is mapped to a single index SNP of 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 owing 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 with 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA *P* value for each pathway. Significance was determined when an individual pathway reached FDR <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 would be expected by chance, suggesting 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 2006 edition of the database of genes and informative words. Separately, GRAIL was applied to a list of 66 SNPs, one from each locus that had at least one SNP meeting a FDR threshold of 0.05 from the QVALUE software in R[79]. For meta-analysis of age at natural menopause, the FDR < 0.05 threshold implied $P < 2.8 \times 10^{-5}$.

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 in which gene mutations affect ovarian function ($n = 37$), human studies of menopause or isolated POI ($n = 48$), syndromes including ovarian failure ($n = 4$) or genes expressed in the ovary or female germ cells ($n = 38$; **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 genome-wide-significant menopause signals to the online catalog of published GWAS (GWAS catalog; see URLs). All reported associations that reached $P < 5 \times 10^{-8}$ and were within 1 Mb of the menopause signal were considered. LD estimates between the SNP pairs were assessed using HapMap (CEU, release 27). Results are in **Supplementary Table 7**.

URLs. GWAS catalog, http://www.genome.gov/gwastudies/; SCAN, http://www.scandb.org/newinterface/about.html/; IPA, http://www.ingenuity.com/.

62. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
63. Göring, H.H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* **39**, 1208–1216 (2007).
64. Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat. Genet.* **42**, 62–67 (2010).
65. Heap, G.A. *et al.* Complex nature of SNP genotype effects on gene expression in primary human leucocytes. *BMC Med. Genomics* **2**, 1 (2009).
66. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
67. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
68. Kwan, T. *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.* **40**, 225–231 (2008).
69. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
70. Heinzen, E.L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6**, e1 (2008).
71. Zeller, T. *et al.* Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
72. Murphy, A. *et al.* Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum. Mol. Genet.* **19**, 4745–4757 (2010).
73. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
74. Webster, J.A. *et al.* Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* **84**, 445–458 (2009).
75. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
76. Liu, C. *et al.* Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol. Psychiatry* **15**, 779–784 (2010).
77. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
78. Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res.* **19**, 1942–1952 (2009).
79. Dabney, A., Storey, J.D. & Warnes, G.R. qvalue: Q-value estimation for false discovery rate control. R package version 1.22.0. (2010).
80. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).

doi:10.1038/ng.1051