nature genetics

Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair

Menopause timing has a substantial impact on infertility and risk of disease, including breast cancer, but the underlying mechanisms are poorly understood. We report a dual strategy in ~70,000 women to identify common and low-frequency protein-coding variation associated with age at natural menopause (ANM). We identified 44 regions with common variants, including two regions harboring additional rare missense alleles of large effect. We found enrichment of signals in or near genes involved in delayed puberty, highlighting the first molecular links between the onset and end of reproductive lifespan. Pathway analyses identified major association with DNA damage response (DDR) genes, including the first common coding variant in *BRCA1* associated with any complex trait. Mendelian randomization analyses supported a causal effect of later ANM on breast cancer risk (~6% increase in risk per year; $P = 3 \times 10^{-14}$), likely mediated by prolonged sex hormone exposure rather than DDR mechanisms.

Younger age at natural (non-surgical) menopause (ANM) is associated with lower risk of breast cancer but higher risks of osteoporosis, cardiovascular disease and type 2 diabetes[1]. Early menopause also has a substantial impact on fertility. It is estimated that natural fertility ceases on average 10 years before menopause[2], which is becoming increasingly relevant as women in many populations are delaying childbearing. For example, the birth rate in UK women aged 30–34 years is now higher than for women whose age falls in any other half-decade range. ANM is on average 51 years in European-ancestry populations, with natural menopause before the age of 40 years, or primary ovarian insufficiency (POI), occurring in 1% of the population[3].

Previous genome-wide association studies (GWAS) identified 18 common genetic loci associated with ANM, implicating several plausible gene candidates across a number of molecular pathways[4,5]. Together, these reported variants explained <5% of the variation in ANM, as compared to the 21% explained by all common variants on GWAS arrays[4]. We therefore undertook a more comprehensive genetic analysis in a substantially larger sample of nearly 70,000 women, incorporating both common and, for the first time to our knowledge, low-frequency coding variants. We were able to triple the number of independent signals associated with ANM, including two low-frequency coding variants in previously unreported loci. Our findings provide new insights into the causal relationship between ANM and breast cancer and identify molecular overlaps between ANM and puberty timing.

RESULTS

GWAS HapMap 2 meta-analysis

In a combined analysis of up to 69,360 women of European ancestry (**Supplementary Table 1**), 1,208 SNPs, of a total of ~2.6 million, reached the genome-wide significance threshold ($P < 5 \times 10^{-8}$) for association with ANM. Considering these SNPs, we identified 54 independent signals located in 44 genomic regions using approximate conditional analysis implemented in GCTA (**Fig. 1**, **Table 1** and **Supplementary Tables 2** and **3**). Eight loci contained secondary signals: six loci each contained two signals, and two loci each contained three signals. Across the 54 identified signals, minor allele frequency (MAF) ranged from 7 to 49%, and effect size ranged from 0.07 to 0.88 years per allele with no significant heterogeneity between studies. All of the 18 previously reported independent signals for ANM[4,5] retained directionally concordant genome-wide significance (maximum $P = 3.7 \times 10^{-11}$). These 18 signals were also directionally concordant in a subsidiary meta-analysis of the studies that were not included in the previous publication (P-value range of 1×10^{-30} to 1×10^{-3}). The top 29,958 independent SNPs with association $P < 0.05$ explained 21% (standard error = 9.7%; $P = 0.01$) of the variance in ANM, with this proportion decreasing to 6% (standard error = 1.6%; $P = 6.3 \times 10^{-12}$) for the top 54 SNPs with $P < 5 \times 10^{-8}$ (**Supplementary Table 4**). This finding contrasts with an estimate of 2.6% for the 18 previously identified index SNPs.

We assessed functional enrichment for all SNP associations with ANM in regions containing active histone marks across ten physiological cell type groups using stratified LD score regression[6] (Online Methods and **Supplementary Table 5**). Only the 'kidney-related cell types' group showed significant enrichment ($P = 0.003$), which could reflect the mesonephric embryonic origin of ovarian parenchymal cells[7]. Analysis by functional annotation showed the strongest enrichment for variants located in coding regions as defined by the UCSC Genome Browser (**Supplementary Table 5**), with ~1.5% of SNPs explaining 24.8% of the trait heritability ($P = 4.6 \times 10^{-3}$). The heritable component increased to 55% (standard error =11%; $P = 2.9 \times 10^{-7}$) when a flanking 500-bp window was added to the coding regions, capturing ~6.5% of SNPs.

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

Received 17 February; accepted 2 September; published online 28 September 2015; doi:10.1038/ng.3412

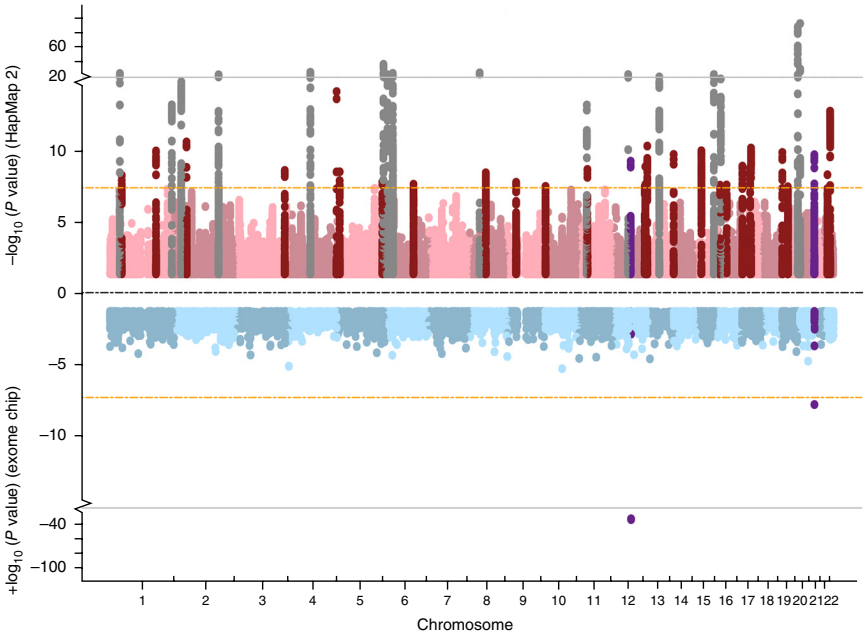

Figure 1 Miami plot of HapMap and exome SNP associations. Log-transformed *P* values are shown for association with ANM for SNPs from HapMap 2 (top; pink) and SNPs from the meta-analysis of exome chip data (bottom; blue). Previously known signals are shown in gray, and newly discovered signals are shown in red (HapMap 2) or purple (exome chip and HapMap 2). The yellow lines correspond to genome-wide significant levels in each direction; the gray lines indicate where the *y* axis has been truncated.

Exome array meta-analysis

To estimate the contribution of low-frequency coding variation to ANM, we performed a meta-analysis of up to 39,026 women genotyped on exome arrays (**Supplementary Table 6**). Only one signal, from two highly correlated ($r^2 = 0.73$, $D' = 1$) low-frequency missense variants in *HELB*, reached genome-wide significance in this discovery phase (**Fig. 1**, **Table 2** and **Supplementary Table 7**). Ten low-frequency (MAF <5%) nonsynonymous SNPs with association $P < 5 \times 10^{-4}$ were selected for follow-up in an independent sample of 10,157 women from the deCODE study that imputed rare variant genotypes. Directionally concordant effect estimates were observed for six of the eight variants considered (two of the ten variants failed quality control). The combined analysis identified missense alleles in *HELB* (rs75770066, MAF = 3.6%, effect (β) = 0.85 year/allele, $P = 1.2 \times 10^{-31}$) and *SLCO4A1* (rs140267842, MAF = 0.8%, $\beta = 0.79$, $P = 1.6 \times 10^{-8}$) as associated with ANM (**Table 2**, **Supplementary Fig. 1** and **Supplementary Table 7**).

DNA helicase B (encoded by *HELB*) is a DNA helicase that unwinds DNA during replication, transcription, repair and recombination. SLCO4A1 (solute carrier organic anion transporter family, member 4A1) transports organic anions such as thyroid hormones and estrone-3-sulfate. The exome array signals in *HELB* and *SLCO4A1* were located in ANM-associated loci newly identified by our parallel HapMap 2–based GWAS meta-analysis. At *HELB*, the association of the common index SNP, rs12371165, was fully explained by the associations at the two rare exome chip SNPs, which are in high linkage disequilibrium (LD) with each other ($r^2 = 0.73$, $D' = 1$) (**Fig. 2**). In contrast, the three independent signal SNPs identified through GCTA were not explained by the rare variant(s) (**Supplementary Table 8**). It thus appears that there are at least two non-redundant signals at this locus, and future fine-mapping experiments will be required to fully elucidate the number of independent causal variants. Functional studies have shown that substitution of aspartate by a nonpolar residue at amino acid 506 of DNA helicase B affects binding of the helicase to replication protein A (RPA)[8]. At *SLCO4A1*, all three variants (the common index SNP, the second signal from GCTA and the exome chip variant) appeared to reflect non-redundant signals, such that the association of each with ANM was unaffected by the presence of either of the other two (**Supplementary Table 8**).

ANM SNPs strongly enriched in DNA damage response pathways

Pathway analyses using MAGENTA and GRAIL indicated substantial enrichment of GWAS SNP associations in DDR pathways (**Supplementary Tables 9** and **10**). Seven of the ten ANM-associated pathways identified by MAGENTA at study-wise significance were involved in DDR, with the highest enrichment in the PANTHER-defined 'DNA repair pathway' ($P = 1 \times 10^{-6}$). After annotating likely causal genes at each locus, we found that 29 of the 44 GWAS-highlighted regions contained one or more DDR genes within 500 kb (**Table 1**). At 18 of these 29 regions, either the DDR candidate gene was the nearest gene or the signal was associated with the expression of a DDR gene at the locus.

The top SNP at GWAS signal 37 (**Table 1**) is highly correlated ($r^2 > 0.95$) with four common nonsynonymous variants in *BRCA1* (rs1799966, rs16942, rs16941 and rs799917), none of which is listed in the Human Gene Mutation Database (HGMD) as a known breast cancer susceptibility variant and all of which are listed as "not clinically important" by the Breast Cancer Information Core. In our exome array data, no low-frequency coding variants in *BRCA1* were associated with ANM ($P > 0.05$). Signal 37 is an expression quantitative trait locus (eQTL) for *BRCA1* in multiple tissues, including blood, skin, adipose and brain (**Supplementary Table 11**). There were 15 ANM signal genes that STRING analysis identified as having at least one direct link to *BRCA1* (**Supplementary Fig. 2** and **Supplementary Table 12**). Of these genes, there is experimental evidence that seven encode direct binding partners of BRCA1: *BRE* (signal 5), *MSH6* (signal 6), *POLR2H* (signal 8), *FAM175A* (signal 9), *UIMC1* (signal 13), *RAD51* (signal 30) and *CHEK2* (signal 43).

Although many of the DDR genes highlighted are involved in homologous recombination for the repair of double-strand breaks, such as in the *BRCA1* pathway, other mechanisms of repair are also represented, for example, mismatch repair (*MSH5* and *MSH6*) and base-excision repair (*APEX1* and *PARP2*) (**Fig. 3**). Two genes act as DNA damage checkpoints (*CHEK2* and *BRSK1*), and others are involved in the cellular response to damage, having roles in activities such as cell cycle arrest, DNA replication, transcription control and apoptosis (**Fig. 3**). *CHEK2* is a well-known breast cancer–associated gene[9], but the ANM-associated signal is not in LD with the c.1100delC variant associated with breast cancer ($r^2 < 0.01$).

ANM SNPs enriched in known POI genes

In addition to DDR pathways, MAGENTA analyses also identified a fourfold enrichment of ANM GWAS SNP associations in or near a set of 31 genes reportedly associated with monogenic POI

Table 1 Association of 54 common HapMap 2 variants at 44 genomic loci with ANM

Region	Best SNP[a]	Signal SNP[b]	Chr.	Position (bp)[c]	Alleles[d]	EAF[e]	*n*	Univariate model[f]		Joint model[g]		Highlighted gene[h]
								Effect	*P*	Effect	*P*	
1*	rs4246511	rs4246511	1	39,152,972	C/T	0.71	69,116	–0.22 (0.02)	5.1×10^{-21}	–	–	*RHBDL2*$^{(B,N)}$, ***MYCBP***$^{(B)}$
2	rs12142240	rs12142240	1	46,519,888	T/C	0.68	69,356	–0.13 (0.02)	6.6×10^{-9}	–	–	***RAD54L***$^{(B,E)}$
3	rs1411478	rs1411478	1	179,228,905	A/G	0.41	68,680	–0.13 (0.02)	1.4×10^{-10}	–	–	*STX6*$^{(N,E)}$
4*	rs2236918	rs2236918	1	240,084,449	C/G	0.45	69,332	–0.15 (0.02)	8.3×10^{-14}	–	–	***EXO1***$^{(N,B,C)}$
5*	rs704795	rs704795	2	27,569,998	A/G	0.4	69,341	–0.16 (0.02)	2.1×10^{-15}	–	–	***BRE***$^{(B)}$, *GTF3C2*$^{(B,E)}$, *EIFB4*$^{(B)}$
6*	rs1800932	rs1800932	2	47,871,585	A/G	0.81	69,309	–0.17 (0.03)	3.2×10^{-11}	–	–	***MSH6***$^{(N,B,E)}$
7*	rs930036	rs930036	2	171,649,264	A/G	0.38	69,357	–0.19 (0.02)	3.1×10^{-19}	–	–	***TLK1***$^{(N,E,B)}$, *GAD1*$^{(B)}$
8	rs16858210	rs16858210	3	185,106,704	G/A	0.75	69,193	–0.14 (0.02)	3.1×10^{-9}	–	–	***PARL***$^{(B)}$, ***POLR2H***$^{(B)}$
9*	rs4693089	rs4693089	4	84,592,646	A/G	0.51	69,060	–0.20 (0.02)	9.2×10^{-23}	–	–	***HELQ***$^{(N,B)}$, ***FAM175A***$^{(B)}$
10	rs6856693	rs6856693	4	185,985,800	A/G	0.58	67,635	–0.16 (0.02)	9.8×10^{-15}	–	–	*ASCL1*$^{(N)}$, ***MLF1IP***$^{(B)}$
11	rs427394	rs427394	5	6,798,875	G/A	0.41	69,284	–0.13 (0.02)	3.8×10^{-9}	–	–	***PAPD7***$^{(N,B)}$
12	rs11738223	rs11738223	5	171,867,097	A/G	0.68	69,250	–0.12 (0.02)	2.0×10^{-8}	–	–	*SH3PXD2B*$^{(N)}$
13a*	rs365132	rs2241584	5	175,888,783	A/G	0.38	69,341	–0.14 (0.02)	1.5×10^{-11}	–0.14 (0.02)	3.2×10^{-11}	***UIMC1***$^{(B,E)}$
13b*	rs365132	rs365132	5	176,311,180	G/T	0.51	69,349	–0.24 (0.02)	1.4×10^{-33}	–0.24 (0.02)	7.9×10^{-33}	***UIMC1***$^{(N,B,E)}$
14a*	rs6899676	rs6899676	6	11,003,246	A/G	0.8	69,303	–0.23 (0.03)	2.2×10^{-19}	–0.21 (0.03)	6.2×10^{-16}	*SYCP2L*$^{(N,B)}$, *MAK*$^{(B)}$
14b*	rs6899676	rs9393800	6	11,059,723	G/A	0.27	69,124	–0.17 (0.02)	3.5×10^{-13}	–0.14 (0.02)	1.1×10^{-9}	*SYCP2L*$^{(N,B)}$, *MAK*$^{(B)}$
15a*	rs1046089	rs2230365	6	31,633,427	C/T	0.84	67,095	–0.17 (0.03)	7.6×10^{-10}	–0.16 (0.03)	2.7×10^{-8}	***MSH5***$^{(B)}$, *HLA*$^{(B)}$
15b*	rs1046089	rs707938	6	31,837,338	G/A	0.32	68,582	–0.17 (0.02)	7.2×10^{-15}	–0.16 (0.02)	2.3×10^{-13}	***MSH5***$^{(B,N,E)}$, *HLA*$^{(B)}$
16	rs12196873	rs12196873	6	111,704,751	A/C	0.85	69,313	–0.16 (0.03)	2.8×10^{-8}	–	–	***REV3L***$^{(B,C)}$
17*	rs2720044	rs2720044	8	38,099,744	A/C	0.84	63,917	–0.29 (0.03)	7.3×10^{-22}	–	–	*STAR*$^{(B)}$
18	rs10957156	rs10957156	8	61,791,955	A/G	0.76	69,341	–0.14 (0.02)	4.5×10^{-9}	–	–	***CHD7***$^{(N,B,E)}$
19	rs4879656	rs4879656	9	33,002,382	A/C	0.37	68,919	–0.12 (0.02)	2.0×10^{-8}	–	–	***APTX***$^{(N,B,E)}$
20	rs10905065	rs10905065	10	5,809,833	A/G	0.61	69,334	–0.11 (0.02)	3.9×10^{-8}	–	–	***FBXO18***$^{(B)}$
21a*	rs11031006	rs11031006	11	30,183,104	G/A	0.85	69,309	–0.22 (0.03)	8.5×10^{-14}	–0.25 (0.03)	4.0×10^{-17}	*FSHB*$^{(N,B)}$
21b*	rs11031006	rs6484478	11	30,263,016	G/A	0.74	69,099	–0.10 (0.02)	4.0×10^{-5}	–0.14 (0.02)	1.0×10^{-8}	*FSHB*$^{(B)}$
22	rs10734411	rs10734411	11	32,498,360	A/G	0.47	69,142	–0.12 (0.02)	2.6×10^{-9}	–	–	*EIF3M*$^{(N)}$
23*	rs2277339	rs2277339	12	55,432,336	G/T	0.1	67,603	–0.31 (0.03)	1.8×10^{-19}	–	–	***PRIM1***$^{(B,N,C,E)}$, *TAC3*$^{(B)}$
24a	rs12371165	rs3741604	12	64,982,677	T/C	0.52	69,100	–0.09 (0.02)	1.9×10^{-5}	–0.29 (0.03)	1.8×10^{-21}	***HELB***$^{(N,B,E,C)}$
24b	rs12371165	rs1183272	12	65,021,688	C/T	0.45	68,727	–0.07 (0.02)	7.3×10^{-4}	–0.31 (0.03)	3.0×10^{-24}	***HELB***$^{(B,N,C)}$
24c	rs12371165	rs7397861	12	65,100,733	G/C	0.64	69,095	–0.10 (0.02)	6.7×10^{-6}	–0.13 (0.02)	4.6×10^{-9}	***HELB***$^{(B,E,C)}$
25	rs551087	rs551087	12	119,693,576	G/A	0.29	69,001	–0.13 (0.02)	3.9×10^{-8}	–	–	*SPPL3*$^{(N)}$, *SRSF9*$^{(B)}$
26	rs1727326	rs1727326	12	122,166,039	C/G	0.15	68,870	–0.19 (0.03)	1.7×10^{-9}	–	–	***KNTC1***$^{(B)}$, *PITPNM2*$^{(N)}$
27	rs12824058	rs12824058	12	129,370,287	G/A	0.43	69,047	–0.14 (0.02)	6.1×10^{-11}	–	–	*PIWIL1*$^{(N)}$
28*	rs4886238	rs4886238	13	60,011,740	G/A	0.66	69,314	–0.18 (0.02)	2.5×10^{-16}	–	–	*TDRD3*$^{(B,N)}$
29	rs1713460	rs1713460	14	20,003,455	G/A	0.3	68,528	–0.14 (0.02)	2.4×10^{-10}	–	–	***APEX1***$^{(B)}$, ***PARP2***$^{(B)}$, *PNP*$^{(N,E)}$
30	rs9796	rs9796	15	39,058,739	T/A	0.46	69,317	–0.13 (0.02)	1.3×10^{-10}	–	–	***INO80***$^{(B,N,E)}$, ***RAD51***$^{(B)}$
31*	rs1054875	rs1054875	15	87,680,130	T/A	0.4	69,288	–0.19 (0.02)	1.7×10^{-19}	–	–	***POLG***$^{(B,N)}$, ***FANCI***$^{(B,C)}$
32	rs9039	rs9039	16	9,112,864	C/T	0.28	69,341	–0.12 (0.02)	3.3×10^{-8}	–	–	*C16orf72*$^{(N)}$, *ABAT*$^{(B)}$
33*	rs10852344	rs10852344	16	11,924,420	T/C	0.59	69,346	–0.16 (0.02)	1.3×10^{-15}	–	–	*GSPT1*$^{(N,C,E)}$, *BCAR4*$^{(B)}$
34	rs12599106	rs12599106	16	34,355,526	A/T	0.51	69,320	–0.12 (0.02)	3.1×10^{-8}	–	–	*UBE2MP1*$^{(N)}$
35	rs8070740	rs8070740	17	5,272,620	A/G	0.76	68,515	–0.15 (0.02)	1.5×10^{-9}	–	–	*RPAIN*$^{(N,E)}$
36	rs2941505	rs2941505	17	35,086,230	A/G	0.32	69,302	–0.13 (0.02)	1.9×10^{-9}	–	–	*STARD3*$^{(B)}$, *PGAP3*$^{(N,E)}$, *CDK12*$^{(B)}$
37	rs1799949	rs1799949	17	38,498,992	G/A	0.68	69,329	–0.14 (0.02)	8.4×10^{-11}	–	–	***BRCA1***$^{(N,E,B,C)}$
38	rs349306	rs349306	19	901,694	G/A	0.13	58,278	–0.23 (0.04)	1.7×10^{-10}	–	–	***POLR2E***$^{(B)}$, *KISS1R*$^{(B)}$
39	rs7259376	rs7259376	19	22,299,545	A/G	0.46	69,328	–0.11 (0.02)	4.2×10^{-8}	–	–	*ZNF729*$^{(N)}$
40a*	rs11668344	rs11668344	19	60,525,476	G/A	0.36	69,329	–0.41 (0.02)	5.5×10^{-85}	–0.41 (0.02)	4.2×10^{-84}	***BRSK1***$^{(B,E)}$, *NLRP11*$^{(N)}$, *U2AF2*$^{(B)}$
40b*	rs11668344	rs2547274	19	61,002,040	G/C	0.91	66,580	–0.28 (0.04)	3.4×10^{-13}	–0.22 (0.04)	2.7×10^{-8}	***BRSK1***$^{(B)}$, *NLRP11*$^{(N)}$, *U2AF2*$^{(B)}$
40c*	rs11668344	rs12461110	19	61,012,475	A/G	0.35	68,518	–0.17 (0.02)	7.6×10^{-16}	–0.15 (0.02)	5.0×10^{-12}	***BRSK1***$^{(B)}$, *NLRP11*$^{(N,C)}$, *U2AF2*$^{(B)}$
41a*	rs16991615	rs451417	20	5,889,999	A/C	0.12	65,420	–0.20 (0.03)	4.6×10^{-9}	–0.2 (0.03)	4.5×10^{-9}	***MCM8***$^{(N,C,B)}$
41b*	rs16991615	rs16991615	20	5,896,227	G/A	0.93	66,210	–0.88 (0.04)	1.6×10^{-89}	–0.88 (0.04)	4.4×10^{-89}	***MCM8***$^{(N,C,B)}$
42a	rs13040088	rs2236553	20	60,760,188	C/T	0.24	62,648	–0.16 (0.03)	6.1×10^{-10}	–0.16 (0.03)	4.4×10^{-10}	*SLCO4A1*$^{(N,C)}$, ***DIDO1***$^{(B,E)}$
42b	rs13040088	rs13040088	20	61,019,647	G/A	0.21	69,317	–0.16 (0.02)	2.4×10^{-10}	–0.16 (0.02)	1.9×10^{-10}	*SLCO4A1*$^{(C)}$, ***DIDO1***$^{(N,B,E)}$
43	rs5762534	rs5762534	22	26,963,571	T/C	0.84	69,322	–0.16 (0.03)	6.1×10^{-9}	–	–	***CHEK2***$^{(B)}$
44	rs763121	rs763121	22	37,209,886	G/A	0.36	66,632	–0.16 (0.02)	2.3×10^{-13}	–	–	***DMC1***$^{(B)}$, *DDX17*$^{(N,E,B)}$

[a]Best regional SNP selected by clumping based on 1-Mb distance. [b]Lead independent SNP(s) in region selected through approximate conditional analysis. [c]Position in Build 36 of the reference genome. [d]Effect allele/other allele. [e]Effect allele frequency. [f]Univariate test statistics reported from the primary meta-analysis (no conditional analysis). [g]Test statistics derived from the joint model for regions containing more than one statistically independent SNP. [h]Highlighted gene in the region selected on the basis of the following criteria: N, nearest; B, biological candidate; E, eQTL effect; C, nonsynonymous SNP in high LD. Genes categorized as being in DDR pathways are shown in bold. An asterisk denotes a region previously described at genome-wide significance. Chr., chromosome.

Table 2 Results of the exome chip meta-analyses

SNP	Band	Gene	Amino acid change	Minor/common allele	Analysis	MAF (%)	Effect (SE) of minor allele in years	*P*	*n*	Heterogeneity *P*
rs75770066	12q14.3	*HELB*	p.Asp506Gly	G/A	Discovery	3.6	0.91 (0.08)	**1.79×10^{-32}**	39,026	
					Replication	1.7	0.32 (0.24)	0.171	10,157	
					Combined	3.4	0.85 (0.07)	**1.17×10^{-31}**	49,183	0.050
rs148126992	12q14.3	*HELB*	p.Glu522Asp	C/G	Discovery	2.5	1.03 (0.09)	**2.96×10^{-30}**	38,707	
					Replication	0.1	2.16 (1.75)	0.216	10,157	
					Combined	2.5	1.04 (0.09)	**1.69×10^{-30}**	48,864	0.116
rs140267842	20q13.33	*SLCO4A1*	p.Val263Ile	A/G	Discovery	0.8	0.80 (0.16)	5.58×10^{-7}	39,026	
					Replication	1.2	0.73 (0.28)	8.60×10^{-3}	10,157	
					Combined	0.9	0.79 (0.14)	**1.60×10^{-8}**	49,183	0.241

Amino acid change is from the amino acid coded by the common allele to the amino acid coded by the minor allele. Significant *P* values are shown in bold. SE, standard error.

(**Supplementary Tables 13** and **14**). Four of our genome-wide significant hits were located in or near reported POI genes. Autosomal recessive mutations in *MCM8* cause primary amenorrhea, hypothyroidism and hypergonadotropic hypogonadism[10]. Recessive mutations in *EIF2B4* (signal 5) cause ovarioleukodystrophy with vanishing white matter syndrome[11]. *POLG* (signal 31) mutations have been linked to POI in isolation or in association with other neurological conditions[12]. Mutations in *MSH5* (signals 15a and 15b) have been associated with various human diseases, including POI[13]. In addition, *TDRD3* (signal 28) is a primary binding partner of *FMR1*, in which triplet-repeat premutations are a risk factor for POI[14]. We saw no significant enrichment of ANM signals in our wider panel of ovarian function genes (**Supplementary Tables 13** and **15**).

Genetic correlation of ANM with other traits and diseases

We searched the GRASP database[15] and the National Human Genome Research Institute (NHGRI) GWAS catalog for pleiotropy between ANM signals and proxies ($r^2 > 0.5$) and GWAS-identified signals for other traits (**Supplementary Table 16**). The top overlapping signals were for liver enzymes, lipids, urate, height and fasting glucose ($P \leq 1 \times 10^{-10}$ for association of the ANM SNP or its proxy with the second trait). We found no overlap with any autoimmune traits and only a very weak link with a cancer (upper airway tract cancer, $P = 1 \times 10^{-8}$). To test the relationship between ANM and other health outcomes more broadly, we performed cross-trait LD Score regression to estimate genetic correlation, using data from 53 published GWAS meta-analyses (**Supplementary Table 17**). Adult obesity ranked highest in this analysis, with a negative trait correlation ($r_g = -0.15$; $P = 0.0004$), and there was supporting evidence from other growth-related and anthropometric traits, including age at menarche ($r_g = 0.14$; $P = 0.003$), body mass index (BMI; $r_g = -0.13$; $P = 0.003$), BMI in women but not men ($P = 0.002$ versus 0.17), waist circumference in women but not men ($P = 0.009$ versus 0.29) and waist-hip ratio (WHR) in men but not women ($P = 0.03$ versus 0.27). Other nominally significant associations included high-density lipoprotein (HDL) levels ($r_g = 0.14$; $P = 0.02$) and current or former smoking status ($r_g = 0.20$; $P = 0.04$), both of which are supported by epidemiological observations[16].

To elucidate the causal directions between these traits, we performed bidirectional Mendelian randomization analyses on ANM with both age at menarche and BMI. We were unable to resolve the causal direction with BMI (BMI to ANM: $P_{score} = 0.668$, **Supplementary Table 18**; ANM to BMI: $P_{binomial} = 0.683$, **Supplementary Table 19**). However, the 123 SNPs reported to be associated with age at menarche collectively predicted ANM in the expected direction ($P_{score} = 0.0005$; **Supplementary Table 20**), but the ANM SNP score was not associated with age at menarche ($P_{score} = 0.571$; **Supplementary Table 21**). We further explored the nature of this shared genetic architecture by

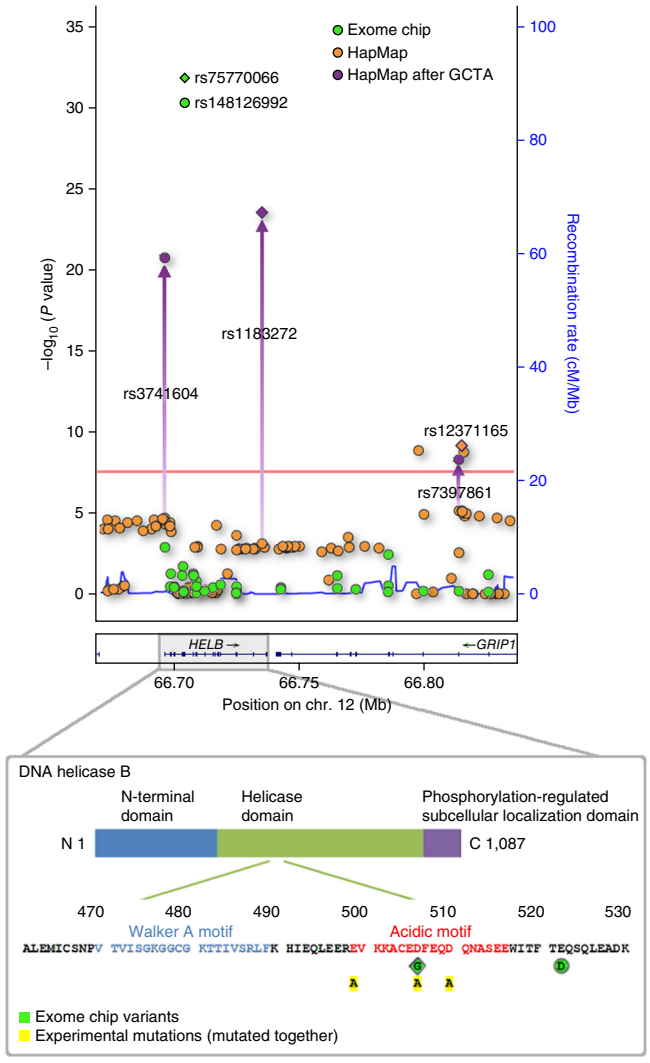

Figure 2 Multiple signals at *HELB* and relationship to DNA helicase B protein sequence. Positions are given in Build 37 coordinates of the reference genome. The top signal from the exome chip analysis maps to an acidic motif of DNA helicase B and results in the replacement of an acidic aspartate residue by a nonpolar glycine residue. Concurrent alteration of three acidic amino acids, (including the aspartate residue identified by the exome chip analysis) to nonpolar residues has been shown to reduce RPA binding[8].

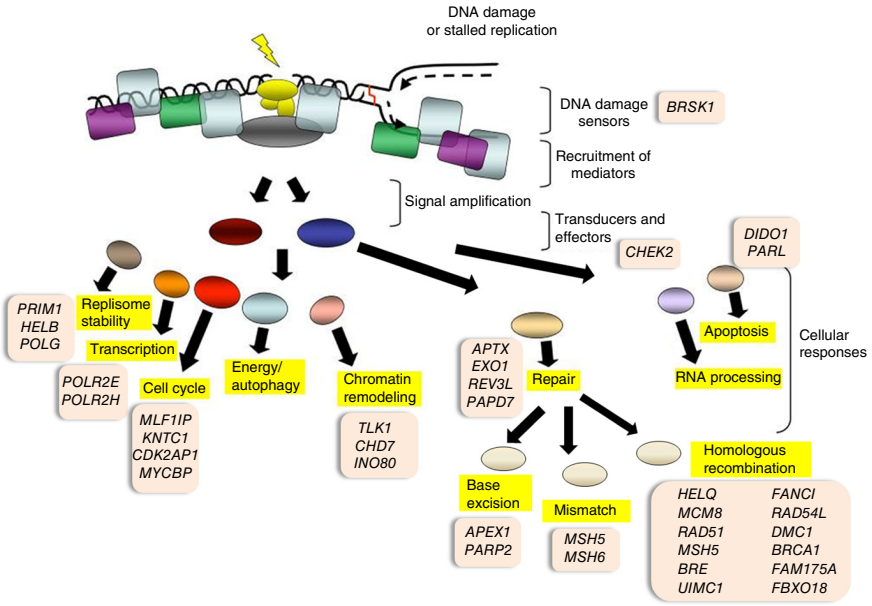

Figure 3 Classification of the genes identified as being involved in DDR pathways at genetic loci associated with ANM. The figure was adapted with permission from ref. 39.

testing for enrichment of all ANM-associated SNPs in or near genes implicated in monogenic or polygenic puberty timing[17]. Significant enrichment was found with the monogenic set (P = 0.01), underscored by the presence of ANM-associated SNPs in or near five genes reportedly causal for hypogonadotropic hypogonadism (*KISS1R*, *TAC3*, *CHD7*, *SOX10* and *FGFR1*) (**Supplementary Table 22**).

ANM variants demonstrate a causal link with breast cancer

Given the overwhelming enrichment of DDR genes among ANM-associated loci and the known epidemiological associations between ANM and breast cancer risk[18], we tested the causal relationship between ANM and breast cancer using a Mendelian randomization approach[19].

Across the 56 ANM-associated SNPs (54 from HapMap 2 and 2 from exome chip), there was a positive correlation between the effect sizes in ANM and the effect sizes for risk (log-transformed odds ratio) of breast cancer (in 46,347 breast cancer cases and 41,736 controls from the Breast Cancer Association Consortium (BCAC); $r = 0.67$, $P = 2.25 \times 10^{-8}$). A polygenic risk score comprising ANM-increasing alleles at the 56 SNPs, weighted by the size of their effects on ANM, was positively associated with breast cancer risk: each genetically predicted 1-year increase in ANM was associated with odds ratio (OR) = 1.064 higher risk of breast cancer (confidence interval (CI) = 1.050–1.081; $P = 2.78 \times 10^{-14}$; **Supplementary Fig. 3**). The size of this effect is larger than that reported by the largest pooled analysis of observational epidemiological studies (OR = 1.030, CI = 1.026–1.034)[18]. All of the women in the GWAS from the BCAC study were also included in the Mendelian randomization study (n = 14,884; ~14% of the total Mendelian randomization study). To confirm that this overlap did not bias our results, we conducted two analyses. First, a sensitivity analysis tested the effect on breast cancer of the 18 previously identified ANM-associated SNPs, which were identified from a meta-analysis that did not include BCAC cases, and a similar effect estimate was observed (OR = 1.062, CI = 1.033–1.101; $P = 1.58 \times 10^{-7}$). Second, the reverse analysis tested 63 SNPs with independent robust associations with breast cancer[20] and found no association between these breast cancer signals and ANM ($P_{score} > 0.05$), which reduces the likelihood of case ascertainment bias in our discovery meta-analysis (**Supplementary Table 23**).

Stratified analyses identified significantly larger effect estimates for the ANM risk score in estrogen receptor (ER)-positive versus ER-negative breast cancer cases (OR = 1.07, CI = 1.05–1.10, $P = 1.73 \times 10^{-12}$ versus OR = 1.03, CI = 1.00–1.07, P = 0.043; P = 0.0086 for the case-only analysis) and women aged ≥55 versus ≤45 years (OR = 1.06, CI = 1.04–1.10, $P = 2.23 \times 10^{-7}$ versus OR = 1.00, CI = 0.97–1.05, P = 0.95; case-only $P = 2.30 \times 10^{-5}$). Consideration of DDR-linked SNPs versus those not related to DDR in the polygenic risk score also produced discordant effect estimates (OR = 1.05, CI = 1.03–1.08, $P = 1.06 \times 10^{-7}$ versus OR = 1.12, CI = 1.06–1.21, $P = 7.84 \times 10^{-10}$; heterogeneity P = 0.01), a difference that was further reinforced in age-stratified analyses (**Supplementary Fig. 3** and **Supplementary Table 24**).

Furthermore, lack of association between ANM risk scores and risk of prostate cancer in men (in 25,074 cases and 24,272 controls; P = 0.36; **Supplementary Table 25**) provides no evidence to support an effect of ANM-related DDR mechanisms on other cancer risks. We therefore surmise that ANM genetic variants influence breast cancer risk primarily through variation in menopause timing.

DISCUSSION

Our study represents a greatly expanded genetic discovery effort for ANM, both in terms of increased sample size and breadth of variation tested. By more than doubling the GWAS sample size, we have increased the number of loci robustly associated with the trait by threefold. In addition, we assessed the role of low-frequency protein-coding variation using exome genotyping arrays. This approach identified the first such variants of large effect for ANM, implicating both *HELB* and *SLCO4A1* in the etiology of reproductive aging. Both of these regions contain common variants we identified in parallel, producing 'synthetic associations' at the *HELB* locus[21].

Our analyses suggest a far more substantial role for DDR processes in ovarian aging than originally estimated. Both manual assessment and formal computational approaches identified an overwhelming excess of DDR genes mapping to the 44 GWAS loci, possibly explaining up to approximately two-thirds of the associations. Despite the limitations of our GWAS approach in definitively mapping SNPs to genes, 19 of the 44 loci contained signal SNPs where plausible DDR candidates were either the closest gene or were linked via altered expression levels to the associated variant. This level of enrichment is comparable to that observed in GWAS meta-analyses of several cancers[22,23].

A notable inclusion in our list of DDR annotated genes was *BRCA1*, which was the nearest gene, was linked as an eQTL and contained multiple nonsynonymous SNPs in high LD with the lead index SNP. Although rare loss-of-function *BRCA1* alleles are well studied in the context of cancer predisposition, coding variants in this gene are generally regarded as neutral and have not previously been mapped to any complex trait or disease, including breast cancer. Titus *et al.* have shown that *BRCA1* expression decreases in human ovaries with age and that reduced *Brca1* expression in mouse models leads to reduced

ovarian reserve[24]. These findings are consistent with our data, where the ANM-lowering allele reduces *BRCA1* expression in blood. BRCA1 directly inhibits the transcriptional activation function of ERα, and thus *BRCA1* variants could also affect ANM through altered estrogen signalling[25]. Of the 34 DDR genes highlighted in **Table 1**, 15 have experimental links to *BRCA1*, three of which form part of the BRCA1-A complex: *BRE* (BRCC45), *FAM175A* (abraxas) and *UIMC1* (RAP80). Dispensable for the major tumor-suppressive role of BRCA1 in promoting DNA double-strand break repair by homologous recombination, the BRCA1-A complex components RAP80 and abraxas are actually involved in counteracting this activity, restricting BRCA1-dependent homologous recombination to appropriate levels[26]. Similarly, the DNA helicase FBH1 (*FBXO18*; signal 20) negatively regulates homologous recombination[27,28]. Although homologous recombination is essential for cell viability, such anti-recombinase activities are also important for maintaining genome stability, and failure of this regulation is associated with inappropriate recombination events and the accumulation of toxic recombination intermediates, DNA repair activities associated with driving translocations, loss of heterozygosity and chromosomal abnormalities[29].

Double-strand break repair is an important response to metabolic and environmental damage to DNA but is also a key process in meiosis for resolving recombination events. Aberrant meiotic recombination is known to cause meiotic arrest and affect the viability of oocytes. Menopause occurs when the number of oocytes in the ovary falls below a threshold number (approximately 1,000), and thus processes that affect the size of the oocyte pool will affect the timing of menopause. Recent studies have shown that recessive mutations in both *MCM8* and *MCM9* result in genomic instability, caused by a deficiency in double-strand break repair, which has a devastating effect on the oocyte pool, causing POI[10,30]. *MCM8* is one of the genes highlighted in our study (signal 41), and a further 12 are also involved in homologous recombination–mediated repair, including two that are specific for meiotic repair (*MSH5* and *DMC1* (DNA meiotic recombinase 1)). Thus, double-strand break repair, during recombination at meiosis, appears to be a major mechanism by which oocyte numbers are regulated, thus determining depletion of the oocyte pool and ANM.

In this study, however, the repair mechanisms highlighted are not confined to homologous recombination; mismatch repair and base-excision repair are also implicated, as well as mitotic repair and repair checkpoints. Thus, it appears that the mechanisms are not confined to repair of meiotic crossovers, but more general mechanisms are also involved. Seven million oogonia are produced during fetal development by mitosis. Inefficient repair of DNA damage during these mitotic events could result in apoptosis and thus a reduction in the initial oocyte pool. Loss of oocytes throughout female life predominantly occurs by atresia rather than ovulation. It is likely that oocytes are particularly sensitive to DNA damage because of the prolonged state of cell cycle arrest, lasting up to 50–60 years. Aberrant repair throughout life could affect the rate of atresia and thus ANM.

Several of the genes highlighted in our study are robust cancer predisposition genes, for example, *BRCA1*, *CHEK2* and *MSH6*. Additionally, *BCAR4* and *STARD3* have also been linked with breast cancer predisposition. However, common susceptibility variants have not been mapped to any of these genes through GWAS approaches for any cancer (NHGRI GWAS catalog). Patients with known pathogenic mutations in *BRCA1* predisposing to breast cancer have been reported to have lower ANM[31], although other studies have not replicated these findings[32].

We found that carrying higher numbers of ANM-increasing variants was associated with increased breast cancer risk. This association is consistent with (and indeed slightly larger than) the observed epidemiological association. Our Mendelian randomization approach indicates a causal relationship between ANM and breast cancer risk, with prolonged estrogen and/or progesterone exposure likely to be the mechanism[33]. Consistent with this proposed mechanism, the effect size was greater for ER-positive than ER-negative breast cancer.

At first sight, this observation might appear paradoxical given the enrichment of DDR genes associated with menopause. However, we noted that the association between ANM variants and breast cancer risk was weaker for variants in or near DDR genes than those in the non-DDR set. This raises the possibility that the DDR variants that reduce menopausal age do modestly increase breast cancer risk, but this increase in risk is counterbalanced by the larger effect due to altered hormonal exposure. Alternatively, it is possible that variants in the non-DDR set may have a residual effect on breast cancer risk through hormonal or other mechanisms or that both mechanisms could have a role (**Supplementary Fig. 4**). *BRCA1* mutations are known to be risk factors for prostate cancer[34], and yet we found no association with prostate cancer predisposition for the ANM variants, supporting the hypothesis that the association with breast cancer is mediated via menopause and not a direct effect of the DDR variants. That the effect of the ANM polygenic risk score on breast cancer risk was larger than that predicted from observational studies might indicate measurement error in the reporting of age at menopause or residual negative confounding in epidemiological studies; in either case, the Mendelian randomization analysis performed here using the polygenic risk score as an instrumental variable can give a more accurate estimate of the effect of age at menopause on breast cancer risk. Such measurement error would also be present in studies in the ANM GWAS from which the polygenic risk score weights were derived; hence, the 'true' effect of later menopause on breast cancer risk may actually be larger even than the ~6% increase in risk/year predicted here.

Our findings provide new evidence for a neural influence on the timing of ovarian follicular ageing. Until now, it has been thought that hypothalamic or pituitary activity in relation to menopause is simply secondary to the loss of feedback inhibition from ovarian hormones[35]. We identified five ANM loci containing genes reported to be causal for hypogonadotropic hypogonadism. Monogenic disruption of three of these genes (*CHD7*, *FGFR1* and *SOX10*) is a cause of Kallmann syndrome, characterized by anosmic hypogonadotropic hypogonadism due to failure of embryonic migration of gonadotropin-releasing hormone (GnRH)-secreting neurons from the olfactory bulb to the hypothalamus[36]. In addition, *KISS1R* (*GPR54*) encodes the receptor for kisspeptin, a key hypothalamic activator of the reproductive hormone axis, and *TAC3* encodes neurokinin B, which is highly expressed in hypothalamic neurons that also express kisspeptin and promotes the pulse frequency of luteinizing hormone (LH) secretion from the pituitary. A possible central influence on ovarian aging is also supported by the ANM locus in or near *FSHB* (which is reportedly also associated with circulating follicle-stimulating hormone (FSH) levels). Alternatively, recent studies have identified expression of TAC3, KISS1R and kisspeptin in ovarian granulosa cells[37], suggesting peripheral actions of these neuropeptides and their receptors[38]. Indeed, GPR54 haploinsufficiency in mice leads to progressive oocyte and follicle loss without affecting gonadotropin secretion[38]. Regardless of site of action, our findings indicate several mechanisms that could link the regulation of puberty to ANM and therefore influence both the start and end of the female reproductive lifespan.

In summary, our findings suggest a surprisingly narrow range of biological pathways governing ANM, highlighting a substantial role for DDR pathways in the etiology of ovarian ageing. We demonstrate the usefulness of genetics in informing epidemiological

observations, identifying shared biological pathways linking puberty timing, breast cancer and reproductive aging.

URLs. UK Office for National Statistics publications, http://www.ons.gov.uk/ons/publications/; Human Gene Mutation Database (HGMD), http://www.hgmd.cf.ac.uk/; Breast Cancer Information Core, http://research.nhgri.nih.gov/bic/; National Human Genome Research Institute (NHGRI) GWAS catalog, http://www.genome.gov/gwastudies/; SNP info file SNPInfo_HumanExome-12v1_rev5.tsv.txt, http://www.chargeconsortium.com/main/exomechip; STRING program, http://string-db.org/; SNAP, http://www.broadinstitute.org/mpg/snap/; ldsc, https://github.com/bulik/ldsc;1000 Genomes Project Consortium, http://www.1000genomes.org/; ReproGen Consortium, http://www.reprogen.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

For full acknowledgments, see the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

All authors reviewed the original and revised manuscripts. Statistical analysis: F.R.D., K.S.R., D.J.T., K.L.L., N.P., D.I.C., L.S., H.K.F., P.S., B.B.-S., T.E., A.D.J., C.E.E., N.F., C. He, E. Altmaier, J.A.B., L.L.F., J.E.H., S.E.J., M.F.K., P.F.M., T.N., E.P., A. Robino, L.M.R., U.M.S., J.A.S., A.T., M.T., D. Vuckovic, J.Y., W. Zhao, E. Albrecht, N.A., T.C., J.-J.H., M.M., A.V.S., T. Tanaka, J.R.B.P. Sample collection, genotyping and phenotyping: G.R.A., I.L.A., H.A.-C., A.C.A., V.A., A.M.A., C. Barbieri, M.W.B., A.B.-F., J.B., L.B., S.J.B., C. Blomqvist, E.B., N.V.B., S.E.B., M.K.B., A.-L.B.-D., T.S.B., H. Brauch, H. Brenner, T.B., B.B., A. Campbell, H.C., S.J.C., J.R.C., Y.-D.I.C., G.C.-T., F.J.C., A.D.C., A. Cox, K.C., H.D., I.D.V., E.W.D., J.D., P.D., T.D., I.d.-S.-S., A.M.D., J.D.E., P.A.F., J.D.F., J.F., D.F.-J., I.G., M.E.G., M.G.-C., G.G. Giles, G.G. Girotto, M.S.G., A.G.-N., M.O.G., M.L.G., D.F.G., P.G., X.G., C.A.H., P.H., U.H., B.E.H., L.J.H., A.H., G.H., M.J.H., J.L.H., F.B.H., J.H., K.H., D.J.H., A.J., M.K., D.K., J.A.K., I.K., C.K., V.-M.K., J.K., V.K., D.L., C.L., J. Li, X.L., S.L., Y.L., J. Luan, J. Lubinski, R.M., A. Mannermaa, J. Manz, S.M., J. Marten, N.G.M., C.M., A. Meindl, K.M., E.M., L.M., R.L.M., M.M.-N., M.N., B.M.N., H.N., P.N., A.B.N., B.G.N., J.E.O., S.P., P.P., U.P., A. Petersmann, J.P., P.D.P.P., N.N.P., A. Pirie, G.P., O.P., D.P., B.M.P., K.P., P.R., L.J.R., F.R., I.R., A. Rudolph, D.R., C.F.S., S.S., E.J.S., D. Schlessinger, M.K.S., F.S., R.K.S., M.J.S., R.A.S., C.M.S., J.S., R.S., M.C.S., D. Stöckl, K. Strauch, A.S., K.D.T., U.T., A.E.T., I.T., T. Truong, L.T., S.T.T., D. Vozzi, Q.W., M.W., G.W., J.F.W., R.W., B.B.H.R.W., A.F.W., D.Y., T.Z., W. Zheng, M.Z. Individual study principal investigators: S.B., D.I.B., J.E.B., L.F., G.W.M., V.G., T.D.S., C.M.v.D., B.Z.A., M.C., L.C., D.F.E., P.P.G., C.G., T.B.H., C. Hayward, S.L.R.K., P.K., B.M., A. Metspalu, A.C.M., A.P.R., P.M.R., J.I.R., D.T., A.G.U., S.U., H.V., N.J.W., D.R.W., L.M.Y.-A., A.L.P., K. Stefansson, J.A.V., K.K.O., J.C.-C., J.M.M., A. Murray. Working group: F.R.D., K.S.R., D.J.T., K.L.L., N.P., D.I.C., L.S., H.K.F., P.S., B.B.-S., T.E., A.D.J., C.E.E., N.F., C. He, A.L.P., K. Stefansson, J.A.V., K.K.O., J.C.-C., J.M.M., J.R.B.P., A. Murray.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hartge, P. Genetics of reproductive lifespan. *Nat. Genet.* **41**, 637–638 (2009).
2. Lambalk, C.B., van Disseldorp, J., de Koning, C.H. & Broekmans, F.J. Testing ovarian reserve to predict age at menopause. *Maturitas* **63**, 280–291 (2009).
3. te Velde, E.R. & Pearson, P.L. The variability of female reproductive ageing. *Hum. Reprod. Update* **8**, 141–154 (2002).
4. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat. Genet.* **44**, 260–268 (2012).
5. Perry, J.R. *et al.* DNA mismatch repair gene *MSH6* implicated in determining age at natural menopause. *Hum. Mol. Genet.* **23**, 2490–2497 (2014).
6. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* doi:10.1038/ng.3404 (28 September 2015).
7. Oktem, O. & Oktay, K. The ovary: anatomy and function throughout human life. *Ann. NY Acad. Sci.* **1127**, 1–9 (2008).
8. Guler, G.D. *et al.* Human DNA helicase B (HDHB) binds to replication protein A and facilitates cellular recovery from replication stress. *J. Biol. Chem.* **287**, 6469–6481 (2012).
9. Weischer, M., Bojesen, S.E., Ellervik, C., Tybjaerg-Hansen, A. & Nordestgaard, B.G. *CHEK2**1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J. Clin. Oncol.* **26**, 542–548 (2008).
10. AlAsiri, S. *et al.* Exome sequencing reveals *MCM8* mutation underlies ovarian failure and chromosomal instability. *J. Clin. Invest.* **125**, 258–262 (2015).
11. Fogli, A. *et al.* Ovarian failure related to eukaryotic initiation factor 2B mutations. *Am. J. Hum. Genet.* **72**, 1544–1550 (2003).
12. Trifunovic, A. *et al.* Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417–423 (2004).
13. Mandon-Pépin, B. *et al.* Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur. J. Endocrinol.* **158**, 107–115 (2008).
14. Linder, B. *et al.* Tdrd3 is a novel stress granule–associated protein interacting with the Fragile-X syndrome protein FMRP. *Hum. Mol. Genet.* **17**, 3236–3246 (2008).
15. Eicher, J.D. *et al.* GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–D804 (2015).
16. Morris, D.H. *et al.* Body mass index, exercise, and other lifestyle factors in relation to age at natural menopause: analyses from the Breakthrough Generations study. *Am. J. Epidemiol.* **175**, 998–1005 (2012).
17. Perry, J.R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
18. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* **13**, 1141–1151 (2012).
19. Vimaleswaran, K.S. *et al.* Association of vitamin D status with arterial blood pressure and hypertension risk: a Mendelian randomisation study. *Lancet Diabetes Endocrinol.* **2**, 719–729 (2014).
20. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
21. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
22. Monteiro, A.N. & Freedman, M.L. Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. *J. Intern. Med.* **274**, 414–424 (2013).
23. Ghoussaini, M., Pharoah, P.D. & Easton, D.F. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *Am. J. Pathol.* **183**, 1038–1051 (2013).
24. Titus, S. *et al.* Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci. Transl. Med.* **5**, 172ra21 (2013).
25. Fan, S. *et al.* BRCA1 inhibition of estrogen receptor signaling in transfected cells. *Science* **284**, 1354–1356 (1999).
26. Hu, Y. *et al.* RAP80-directed tuning of BRCA1 homologous recombination function at ionizing radiation–induced nuclear foci. *Genes Dev.* **25**, 685–700 (2011).
27. Tsutsui, Y. *et al.* Multiple regulation of Rad51-mediated homologous recombination by fission yeast Fbh1. *PLoS Genet.* **10**, e1004542 (2014).
28. Simandlova, J. *et al.* FBH1 helicase disrupts RAD51 filaments *in vitro* and modulates homologous recombination in mammalian cells. *J. Biol. Chem.* **288**, 34168–34180 (2013).
29. Chapman, J.R., Taylor, M.R. & Boulton, S.J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* **47**, 497–510 (2012).
30. Wood-Trageser, M.A. *et al. MCM9* mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am. J. Hum. Genet.* **95**, 754–762 (2014).
31. Oktay, K., Kim, J.Y., Barad, D. & Babayev, S.N. Association of *BRCA1* mutations with occult primary ovarian insufficiency: a possible explanation for the link between infertility and breast/ovarian cancer risks. *J. Clin. Oncol.* **28**, 240–244 (2010).
32. Collins, I.M. *et al.* Do *BRCA1* and *BRCA2* mutation carriers have earlier natural menopause than their noncarrier relatives? Results from the Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. *J. Clin. Oncol.* **31**, 3920–3925 (2013).
33. Manson, J.E. *et al.* Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *J. Am. Med. Assoc.* **310**, 1353–1368 (2013).
34. Levy-Lahad, E. & Friedman, E. Cancer risks among *BRCA1* and *BRCA2* mutation carriers. *Br. J. Cancer* **96**, 11–15 (2007).
35. Rance, N.E. Menopause and the human hypothalamus: evidence for the role of kisspeptin/neurokinin B neurons in the regulation of estrogen negative feedback. *Peptides* **30**, 111–122 (2009).
36. Silveira, L.F. & Latronico, A.C. Approach to the patient with hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* **98**, 1781–1788 (2013).
37. García-Ortega, J. *et al.* Expression of neurokinin B/NK3 receptor and kisspeptin/KISS1 receptor in human granulosa cells. *Hum. Reprod.* **29**, 2736–2746 (2014).
38. Gaytan, F. *et al.* Kisspeptin receptor haplo-insufficiency causes premature ovarian failure despite preserved gonadotropin secretion. *Endocrinology* **155**, 3088–3097 (2014).
39. Jackson, S.P. & Bartke, J. The DNA damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).

Felix R Day[1,178], Katherine S Ruth[2,178], Deborah J Thompson[3,178], Kathryn L Lunetta[4,5], Natalia Pervjakova[6,7], Daniel I Chasman[8,9], Lisette Stolk[10,11], Hilary K Finucane[12,13], Patrick Sulem[14], Brendan Bulik-Sullivan[15–17], Tõnu Esko[6,18–20], Andrew D Johnson[5], Cathy E Elks[1], Nora Franceschini[21], Chunyan He[22,23], Elisabeth Altmaier[24–26], Jennifer A Brody[27], Lude L Franke[28], Jennifer E Huffman[5,29], Margaux F Keller[30], Patrick F McArdle[31], Teresa Nutile[32], Eleonora Porcu[33–35], Antonietta Robino[36], Lynda M Rose[8], Ursula M Schick[37], Jennifer A Smith[38], Alexander Teumer[39], Michela Traglia[40], Dragana Vuckovic[36,41], Jie Yao[42], Wei Zhao[38], Eva Albrecht[25], Najaf Amin[43], Tanguy Corre[44,45], Jouke-Jan Hottenga[46], Massimo Mangino[47,48], Albert V Smith[49,50], Toshiko Tanaka[51], Gonçalo R Abecasis[35], Irene L Andrulis[52,53], Hoda Anton-Culver[54], Antonis C Antoniou[3], Volker Arndt[55], Alice M Arnold[56], Caterina Barbieri[36,40], Matthias W Beckmann[57], Alicia Beeghly-Fadiel[58], Javier Benitez[59,60], Leslie Bernstein[61], Suzette J Bielinski[62], Carl Blomqvist[63], Eric Boerwinkle[64,65], Natalia V Bogdanova[66], Stig E Bojesen[67,68], Manjeet K Bolla[3], Anne-Lise Borresen-Dale[69,70], Thibaud S Boutin[29], Hiltrud Brauch[71–73], Hermann Brenner[55,73,74], Thomas Brüning[75], Barbara Burwinkel[76,77], Archie Campbell[78], Harry Campbell[79], Stephen J Chanock[80], J Ross Chapman[81], Yii-Der Ida Chen[42], Georgia Chenevix-Trench[82], Fergus J Couch[83], Andrea D Coviello[84,85], Angela Cox[86], Kamila Czene[87], Hatef Darabi[87], Immaculata De Vivo[12,88], Ellen W Demerath[89], Joe Dennis[3], Peter Devilee[90,91], Thilo Dörk[92], Isabel dos-Santos-Silva[93], Alison M Dunning[94], John D Eicher[5], Peter A Fasching[57,95], Jessica D Faul[96], Jonine Figueroa[97], Dieter Flesch-Janys[98,99], Ilaria Gandin[36,41], Melissa E Garcia[100], Montserrat García-Closas[101,102], Graham G Giles[103,104], Giorgia G Girotto[41], Mark S Goldberg[105,106], Anna González-Neira[59], Mark O Goodarzi[107], Megan L Grove[64], Daniel F Gudbjartsson[14,108], Pascal Guénel[109,110], Xiuqing Guo[42], Christopher A Haiman[111], Per Hall[87], Ute Hamann[112], Brian E Henderson[111], Lynne J Hocking[113], Albert Hofman[43], Georg Homuth[114], Maartje J Hooning[115], John L Hopper[103], Frank B Hu[12,88,116], Jinyan Huang[117], Keith Humphreys[87], David J Hunter[12,20,88,116], Anna Jakubowska[118], Samuel E Jones[2], Maria Kabisch[112], David Karasik[9,119], Julia A Knight[120,121], Ivana Kolcic[122], Charles Kooperberg[37], Veli-Matti Kosma[123–125], Jennifer Kriebel[24,26,126], Vessela Kristensen[69,70,127], Diether Lambrechts[128,129], Claudia Langenberg[1], Jingmei Li[87], Xin Li[12], Sara Lindström[12], Yongmei Liu[130], Jian'an Luan[1], Jan Lubinski[118], Reedik Mägi[6], Arto Mannermaa[123–125], Judith Manz[24,26], Sara Margolin[131], Jonathan Marten[29], Nicholas G Martin[132], Corrado Masciullo[40], Alfons Meindl[133], Kyriaki Michailidou[3], Evelin Mihailov[6], Lili Milani[6], Roger L Milne[103,104], Martina Müller-Nurasyid[25,134,135], Michael Nalls[136], Benjamin M Neale[15–17], Heli Nevanlinna[137], Patrick Neven[138], Anne B Newman[139–141], Børge G Nordestgaard[67,68], Janet E Olson[62], Sandosh Padmanabhan[142], Paolo Peterlongo[143], Ulrike Peters[37], Astrid Petersmann[144], Julian Peto[93], Paul D P Pharoah[3,94], Nicola N Pirastu[36,41], Ailith Pirie[3], Giorgio Pistis[33–35], Ozren Polasek[122], David Porteous[78], Bruce M Psaty[27,145–147], Katri Pylkäs[148,149], Paolo Radice[150], Leslie J Raffel[151,152], Fernando Rivadeneira[10,11,43], Igor Rudan[79], Anja Rudolph[153], Daniela Ruggiero[32], Cinzia F Sala[40], Serena Sanna[33], Elinor J Sawyer[154], David Schlessinger[155], Marjanka K Schmidt[156], Frank Schmidt[114], Rita K Schmutzler[157–159], Minouk J Schoemaker[101], Robert A Scott[1], Caroline M Seynaeve[115], Jacques Simard[160], Rossella Sorice[32], Melissa C Southey[161], Doris Stöckl[26], Konstantin Strauch[25,162], Anthony Swerdlow[101,163], Kent D Taylor[42], Unnur Thorsteinsdottir[14,50], Amanda E Toland[164], Ian Tomlinson[81,165], Thérèse Truong[109,110], Laufey Tryggvadottir[166], Stephen T Turner[167], Diego Vozzi[36], Qin Wang[3], Melissa Wellons[168], Gonneke Willemsen[46], James F Wilson[29,79], Robert Winqvist[148,149], Bruce B H R Wolffenbuttel[169,170], Alan F Wright[29], Drakoulis Yannoukakos[171], Tatijana Zemunik[122], Wei Zheng[58], Marek Zygmunt[172], Sven Bergmann[44,45], Dorret I Boomsma[46], Julie E Buring[8,9], Luigi Ferrucci[51], Grant W Montgomery[132], Vilmundur Gudnason[49,50], Tim D Spector[47], Cornelia M van Duijn[43], Behrooz Z Alizadeh[173], Marina Ciullo[32], Laura Crisponi[33], Douglas F Easton[3,94], Paolo P Gasparini[36,41], Christian Gieger[24–26], Tamara B Harris[100], Caroline Hayward[29], Sharon L R Kardia[38], Peter Kraft[12,174], Barbara McKnight[56], Andres Metspalu[6], Alanna C Morrison[64], Alex P Reiner[37,145], Paul M Ridker[8,9], Jerome I Rotter[42], Daniela Toniolo[40], André G Uitterlinden[10,11,43], Sheila Ulivi[36], Henry Völzke[39], Nicholas J Wareham[1], David R Weir[96], Laura M Yerges-Armstrong[31], The PRACTICAL Consortium[175], kConFab Investigators[175], AOCS Investigators[175], Generation Scotland[175], EPIC-InterAct Consortium[175], LifeLines Cohort Study[175], Alkes L Price[12], Kari Stefansson[14,50], Jenny A Visser[10], Ken K Ong[1,176], Jenny Chang-Claude[153], Joanne M Murabito[5,177,179], John R B Perry[1,179] & Anna Murray[2,179]

[1]Medical Research Council (MRC) Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. [2]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [3]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [4]Department of Biostatistics, Boston University School of Public Health,

Boston, Massachusetts, USA. [5]National Heart, Lung, and Blood Institute and Boston University's Framingham Heart Study, Framingham, Massachusetts, USA. [6]Estonian Genome Center, University of Tartu, Tartu, Estonia. [7]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [8]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [9]Harvard Medical School, Boston, Massachusetts, USA. [10]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [11]Netherlands Consortium on Health Aging and National Genomics Initiative, Leiden, the Netherlands. [12]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [13]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [14]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [15]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [16]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [17]Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [18]Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA. [19]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [20]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [21]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA. [22]Department of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA. [23]Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, Indiana, USA. [24]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [25]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [26]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [27]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [28]Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. [29]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [30]Merck Pharmaceuticals, Boston, Massachusetts, USA. [31]Program in Personalized Medicine, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [32]Institute of Genetics and Biophysics, National Research Council, Naples, Italy. [33]Institute of Genetics and Biomedical Research, National Research Council, Cagliari, Italy. [34]Department of Biomedical Sciences, University of Sassari, Sassari, Italy. [35]Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [36]Institute for Maternal and Child Health, Scientific Institute for Research, Hospitalisation and Health Care 'Burlo Garofolo', Trieste, Italy. [37]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [38]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. [39]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [40]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy. [41]Department of Clinical Medical Sciences, Surgical and Health, University of Trieste, Trieste, Italy. [42]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, California, USA. [43]Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [44]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [45]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [46]Department of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands. [47]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [48]National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK. [49]Icelandic Heart Association, Kopavogur, Iceland. [50]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [51]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland, USA. [52]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. [53]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [54]Department of Epidemiology, University of California–Irvine, Irvine, California, USA. [55]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [56]Department of Biostatistics, University of Washington, Seattle, Washington, USA. [57]Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany. [58]Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [59]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [60]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain. [61]Beckman Research Institute of City of Hope, Duarte, California, USA. [62]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [63]Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [64]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. [65]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [66]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. [67]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [68]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark. [69]Department of Genetics, Institute for Cancer Research, Radiumhospitalet, Oslo University Hospital, Oslo, Norway. [70]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [71]Dr. Margarete Fischer Bosch Institute of Clinical Pharmacology, Stuttgart, Germany. [72]Institute of Clinical Pharmacology, University of Tübingen, Tübingen, Germany. [73]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [74]Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [75]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance Institute of Ruhr University Bochum (IPA), Bochum, Germany. [76]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [77]Molecular Biology of Breast Cancer, Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany. [78]Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [79]Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [80]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. [81]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [82]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [83]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. [84]Section of Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [85]Section of Endocrinology, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [86]Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK. [87]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [88]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [89]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, USA. [90]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [91]Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands. [92]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. [93]Non-Communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, UK. [94]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [95]Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [96]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA. [97]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA. [98]Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Hamburg, Germany. [99]Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. [100]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, USA. [101]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. [102]Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. [103]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. [104]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. [105]Department of Medicine, McGill University, Montreal, Quebec, Canada. [106]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montreal, Quebec, Canada. [107]Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, California, USA. [108]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [109]Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France. [110]University Paris–Sud, UMRS 1018, Villejuif, France. [111]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [112]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. [113]Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Aberdeen, UK. [114]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [115]Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, the Netherlands. [116]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [117]State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [118]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. [119]Hebrew SeniorLife Institute for Aging Research, Boston, Massachusetts, USA. [120]Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. [121]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [122]Faculty of Medicine, University of Split, Split, Croatia. [123]Cancer Center, Kuopio University Hospital, Kuopio, Finland. [124]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. [125]Imaging Center,

Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. [126]German Center for Diabetes Research, Neuherberg, Germany. [127]Department of Clinical Molecular Biology, Oslo University Hospital, University of Oslo, Oslo, Norway. [128]Vesalius Research Center (VRC), VIB, Leuven, Belgium. [129]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium. [130]Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [131]Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. [132]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [133]Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany. [134]Department of Medicine I, Ludwig Maximilians University Munich, Munich, Germany. [135]German Center for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany. [136]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, USA. [137]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [138]Department of Oncology, University Hospitals Leuven, Leuven, Belgium. [139]Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [140]Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [141]Department of Clinical and Translational Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [142]British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [143]Fondazione Istituto FIRC di Oncologia Molecolare (IFOM), Milan, Italy. [144]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [145]Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA. [146]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [147]Department of Health Services, University of Washington, Seattle, Washington, USA. [148]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry, University of Oulu, Oulu, Finland. [149]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland. [150]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. [151]Medical Genetics Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [152]UCLA Clinical and Translational Science Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [153]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [154]Research Oncology, Guy's Hospital, King's College London, London, UK. [155]National Institute on Aging, Intramural Research Program, Baltimore, Maryland, USA. [156]Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. [157]Division of Molecular Gyneco-Oncology, Department of Gynecology and Obstetrics, University Hospital of Cologne, Cologne, Germany. [158]Center of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany. [159]Center for Integrated Oncology, University Hospital of Cologne, Cologne, Germany. [160]Centre Hospitalier Universitaire de Québec Research Center, Laval University, Quebec City, Quebec, Canada. [161]Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia. [162]Chair of Genetic Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, *Ludwig* Maximilians Universität, Munich, Germany. [163]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. [164]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA. [165]NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK. [166]Icelandic Cancer Registry, Reykjavik, Iceland. [167]Division of Nephrology and Hypertension, Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA. [168]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [169]Department of Endocrinology, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. [170]LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [171]Molecular Diagnostics Laboratory, Institute of Radioisotopes and Radiodiagnostic Products, National Centre for Scientific Research 'Demokritos', Athens, Greece. [172]Department of Obstetrics and Gynecology, University Medicine Greifswald, Greifswald, Germany. [173]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [174]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [175]A full list of members and affiliations appears in the **Supplementary Note**. [176]Department of Paediatrics, University of Cambridge, Cambridge, UK. [177]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [178]These authors contributed equally to this work. [179]These authors jointly supervised this work. Correspondence should be addressed to J.R.B.P. (john.perry@mrc-epid.cam.ac.uk).

ONLINE METHODS

Menopause data collection. ANM was self-reported and defined as the age at last naturally occurring menstrual period followed by at least 12 consecutive months of amenorrhea. Recall bias or error in ANM reporting may have reduced our power to detect associations but would be unlikely to introduce systematic error. We assessed this issue in our previous meta-analysis and found no significant differences in effect estimates when considering retrospective versus prospective studies[4]. We included women with ANM who were 40–60 years of age in our analyses, excluding those with menopause induced by hysterectomy, bilateral ovariectomy, radiation or chemotherapy and those using hormone replacement therapy (HRT) before menopause (**Supplementary Table 1**). Within each of the studies included, each participant provided written informed consent and the study protocol was approved by the institutional review board at the parent institution.

GWAS. A total of 33 studies contributed genome-wide association data using self-reported ANM (**Supplementary Table 1**). One of the 33 studies was from BCAC, comprising 17 separate studies with menopause data; samples were genotyped using an Illumina iSelect array (iCOGS)[20]. There was a maximum total sample size of 69,360 individuals of European descent. Studies were asked to use the full imputed set of HapMap Phase 2 autosomal SNPs and to run an additive model including top principal components and study-specific covariates.

In some cases, studies submitted data using 1000 Genomes Projects–based imputation; in these cases, SNPs not included in the HapMap 2 set were removed. Once data were submitted, each study underwent quality control centrally according to standard quality control protocols implemented independently by two analysts. SNPs were filtered out if the MAF was less than 1% or the imputation quality metrics were low (imputation quality <0.4). Studies and SNPs passing quality control were combined using an inverse variance–weighted meta-analysis, implemented using METAL[40]. Again, this meta-analysis was run independently by two analysts who then separately ran PLINK clumping commands[41] to identify the most significant SNPs in associated regions (termed index SNPs), using only SNPs that had data from more than 50% of the studies. SNPs were considered genome-wide significant if $P < 5 \times 10^{-8}$ (P value of 0.05 Bonferroni corrected for 1 million tests). Comparisons were made to ensure concordance of the identified signals between the two independent analysts.

Exome chip. Exome genotyping data were analyzed for 22 studies of European ancestry with questionnaire data on ANM (**Supplementary Table 6**). Genotype calling was performed using the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) joint calling protocol, including X-chromosome variants. Each contributing study carried out study-level analysis in the R package skatMeta or seqMeta using the skatCohort command, with the top genetic principal components included in the model and alleles coded according to a common reference file (SNPInfo_HumanExome-12v1_rev5.tsv.txt; see URLs)[42]. After data submission, two data analysts carried out checks to ensure the consistency of allele coding. We carried out a single-variant meta-analysis in METAL[40], with a total sample size of 39,026; associations were considered significant if $P < 5 \times 10^{-8}$. Variants were put forward for replication in the deCODE study (n = 10,157) if they were present in more than half of the studies in the discovery stage and had $P < 5 \times 10^{-5}$ (MAF <1%) or $P < 5 \times 10^{-4}$ (MAF 1–5%).

Selection of independent signals and conditional analysis. Independent signals (termed signal SNPs) for ANM were identified using approximate conditional analysis implemented in the GCTA software package[43]. LD between variants was estimated using three independently genotyped studies as reference panels: the Rotterdam Study I (n = 5,974) and two EPIC-InterAct data sets (n = 7,397 and 9,294); these comprised males and females of European ancestry with GWAS data imputed using CEU (European-ancestry) haplotypes from HapMap 2. We assumed zero correlation between SNPs more than 10 Mb apart or on different chromosomes. We considered signals to be independent if they were observed in at least two of the three LD reference panels and were located in a 10-Mb region that contained a genome-wide significant SNP according to univariate test statistics.

We assessed the independence of the exome array and HapMap 2 signals by performing formal conditional analyses in the Women's Genome Health Study (WGHS; n = 11,664). Regression was performed including all significant index SNPs in additive models, with the same study covariates as used in the primary analysis. LD computation in Haploview[44] used experimental genotypes where possible (the exome chip rare variants and the common variants rs3741604 and rs2236553) but HapMap 2–imputed genotypes for the other common variants (MaCH v1.0.16; all Rsq >0.99).

Gene identification. At each locus identified by the GWAS meta-analysis, we annotated the likely causative gene(s) (**Supplementary Table 3**), selecting genes that were identified by at least one of the gene prioritization or pathway programs (GRAIL or STRING), genes for which the top SNP or a proxy ($r^2 > 0.8$) was an eQTL in one of 108 tissues or genes in which the top SNP or a proxy ($r^2 > 0.8$) was a coding variant (**Supplementary Fig. 5** and **Supplementary Tables 9–12**, **26** and **27**). In case of overlap between the results of the GWAS and exome array analyses, the gene indicated by the exome array analysis was chosen. Further manual annotation was used to select additional likely candidates on the basis of known biology (for example, monogenic POI) or biology highlighted by hypothesis-free pathway testing (**Supplementary Table 15**). If no candidate was identified by these methods, the nearest gene was chosen.

GRAIL is a literature-based text mining program used to suggest the most likely causal gene at each locus[45], controlling for gene size and without any seed regions. GRAIL $P < 0.05$ was taken to indicate a suggested causal gene (**Supplementary Table 9**). All genes located within 500 kb of the top SNP at each locus were assessed using the STRING program, which was used to highlight any connectivity between genes in different regions (**Supplementary Table 12**).

Expression quantitative trait loci. Each independent SNP signal was assessed in over 100 separate eQTL data sets[46] (**Supplementary Table 11** and **Supplementary Note**). If an independent signal SNP was in high LD ($r^2 > 0.8$; using SNAP) with the most significant signal for an eQTL, then the eQTL-associated gene was highlighted as a potential causal candidate. The collected eQTL results met criteria of statistical thresholds for association with gene transcript levels as described in the original papers.

Pathway identification. We tested for signal enrichment across 2,580 predefined biological pathways in GO, KEGG, Ingenuity, Panther, Reactome and Biocarta using MAGENTA[47] with the full HapMap 2–imputed meta-analysis (**Supplementary Table 10**). Analysis was performed using the same default settings as described in our previous paper[4], with study-wise significance declared at false discovery rate (FDR) < 0.05. In addition to these predefined pathways, we also tested four custom pathways comprising genes involved in POI (n = 31), ovarian function (n = 130), monogenic disorders of puberty (n = 21) and age at menarche (n = 154) (**Supplementary Tables 13–15** and **22**).

Estimating variance explained by SNP sets. An estimate of the total variance explained by highlighted ANM-associated SNPs was calculated using REML (restricted maximum likelihood) implemented in GCTA[43]. Using individual-level data from the EPIC-InterAct cohort (n = 1,761), we calculated the attributable variance for the genome-wide significant SNPs at varying significance thresholds (5×10^{-7}, 5×10^{-6}, 5×10^{-5}, 5×10^{-4}, 0.005, 0.05 and all SNPs passing quality control) obtained from a repeated meta-analysis excluding EPIC-InterAct samples.

We used stratified LD score regression to quantify evidence of functional enrichment specific to groups of cell types[6]. We used the same baseline model as in Finucane *et al.*[6], which comprises 53 overlapping categories, including basic annotations such as coding, UTR, promoter and intronic regions, as well as annotations for several histone marks, DNase I hypersensitivity site (DHS) regions, ChromHMM predictions[48], regions that are conserved in mammals[49], super-enhancers[50] and FANTOM5 enhancers[51]. We evaluated enrichment for each of these categories, which are not specific to a particular cell type. We then took 230 cell type–specific annotations for four histone marks (H3K4me1, H3K4me3, H3K9ac (ref. 52) and H3K27ac (ref. 53); **Supplementary Table 5**)

doi:10.1038/ng.3412

and grouped them into ten cell type groups (adrenal/pancreas, central nervous system, cardiovascular, connective/bone, gastrointestinal, immune/hematopoietic, kidney, liver, skeletal muscle and other)[6]. We added each cell type group to the baseline model one at a time and measured the *P* value of the resulting LD Score regression coefficient of the cell type group using the –h2 flag in ldsc with LD Scores from 1000 Genomes Project Europeans. We ranked the cell type groups by whether the per-SNP heritability in the functional annotation was larger than the per-SNP heritability outside this annotation, controlling for the other annotations in the baseline model.

Breast and prostate cancer Mendelian randomization. To assess the association of the ANM SNPs with breast cancer risk, we used breast cancer cases (n = 46,347) and controls (n = 41,736) of European ancestry from 41 studies in BCAC, who had been genotyped using a custom Illumina Infinium array (iCOGS). After standard quality control exclusions (as described in ref. 20), genotypes were available for 199,961 SNPs. Further genotypes were imputed in a two-stage procedure using SHAPEIT and IMPUTEv2 (ref. 54) with the 1000 Genomes Project March 2012 release as the reference data set[55], giving ~11.6 million SNPs with imputation $r^2 > 0.3$ and MAF >0.005. The 4,747 breast cancer cases and 7,285 controls in the BCAC data set for which ANM information was available had also been included in the ANM GWAS analysis.

The genotypes or imputed genotype dosages for the 56 significant SNPs in **Tables 1** and **2** were used to construct a polygenic risk score for each breast cancer case and control, such that for the ith woman

$$\mathrm{PRS}_i = \sum_{j=1}^{56} \beta_j G_{ij}$$

where β_j is the ANM regression coefficient for the effect allele of the jth SNP (conditional β values were used for the correlated SNPs) and G_{ij} is the number of copies of the effect allele at the jth SNP carried by the ith woman (G_{ij} is between 0 and 2).

The association between the polygenic risk score and breast cancer was tested using unconditional logistic regression, adjusting for study and seven principal components (as estimated on the basis of a subset of 37,000 uncorrelated markers, including ~1,000 selected as ancestry-informative markers). The log-transformed odds ratio was scaled according to the effect size of a one-unit increase in polygenic risk score on ANM in control subjects, so as to obtain an estimated log-transformed odds ratio for a 1-year increase in genetically predicted ANM. Hence, the polygenic risk score can be thought of as an instrumental variable in a Mendelian randomization of ANM against breast cancer.

Additional analyses were conducted specifically for ER-positive (n = 27,026) or ER-negative (n = 7,401) cases and for participants with age at diagnosis (for cases) or interview (for controls) of ≤45 years (8,547 cases and 8,029 controls) or ≥55 years (24,841 cases and 20,410 controls) (as a surrogate for pre- or postmenopausal age at diagnosis, as ANM was not known for all participants), with heterogeneity evaluated in case-only analyses.

We also tested the association of ANM SNPs with prostate cancer risk, to determine whether any effect of the genetic variants was specific to breast cancer. Prostate cancer data were available from a similar sample size as for breast cancer, and there is known overlap in genetic risk for breast and prostate cancers. Individual-level data were not available for prostate cancer; we therefore assessed the impact of ANM using an approximated allele score comprising the 54 HapMap 2 GWAS SNPs derived from summary-level results[56]. The score was assessed using summary statistics from a recent prostate cancer meta-analysis, comprising 25,074 cases and 24,272 controls from 32 studies in the PRACTICAL Consortium[57], genotyped using the iCOGS array, with quality control and imputation carried out in the same way as for the BCAC iCOGS study.

Genetic correlation with additional traits. Cross-trait LD Score regression was used to estimate the genetic correlation between menopause timing and 54 individual traits from published studies, including anthropometric and metabolic traits[58]. We estimated genetic correlations with the method described in ref. 59 and the --rg flag in the ldsc software package, with LD Scores from 1000 Genomes Project Europeans and default settings. Briefly, this method regresses the product of effect size estimates for trait 1 and trait 2 for each SNP against LD Score. The product of the slope and a constant estimates the genetic correlation, and the intercept estimates the product of the number of overlapping samples and the correlation between phenotypes among the overlapping samples.

Bidirectional Mendelian randomization analyses on ANM with age at menarche and BMI were carried out using similar methods as described for prostate cancer, with a weighted allele score[56] generated from summary statistics. Information on the associations with age at menarche came from the most recent GWAS for the trait (n = 182,416 women from 57 studies)[17]. The BMI data were taken from the most recent analysis (n = 249,796 from 64 studies)[60]. Although it was possible to calculate a full allele score for the genome-wide significant BMI-associated SNPs in ANM analysis, this was not possible for the ANM-associated SNPs in BMI analysis; instead, a binomial test of the consistency of effect direction was used.

40. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
42. Zhou, J.J. *et al.* A comparative analysis of family-based and population-based association tests using whole genome sequence data. *BMC Proc.* **8**, S33 (2014).
43. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
44. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
45. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
46. Zhang, X. *et al.* Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. *BMC Genomics* **15**, 532 (2014).
47. 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**, e1001058 (2010).
48. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
49. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
50. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
51. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
52. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
53. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
54. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
55. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
56. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
57. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45**, 385–391 (2013).
58. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
59. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* doi:10.1038/ng.3406 (28 September 2015).
60. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).