Supplementary Information for

Genome-wide association studies identify novel loci associated with age at menarche and age at natural menopause

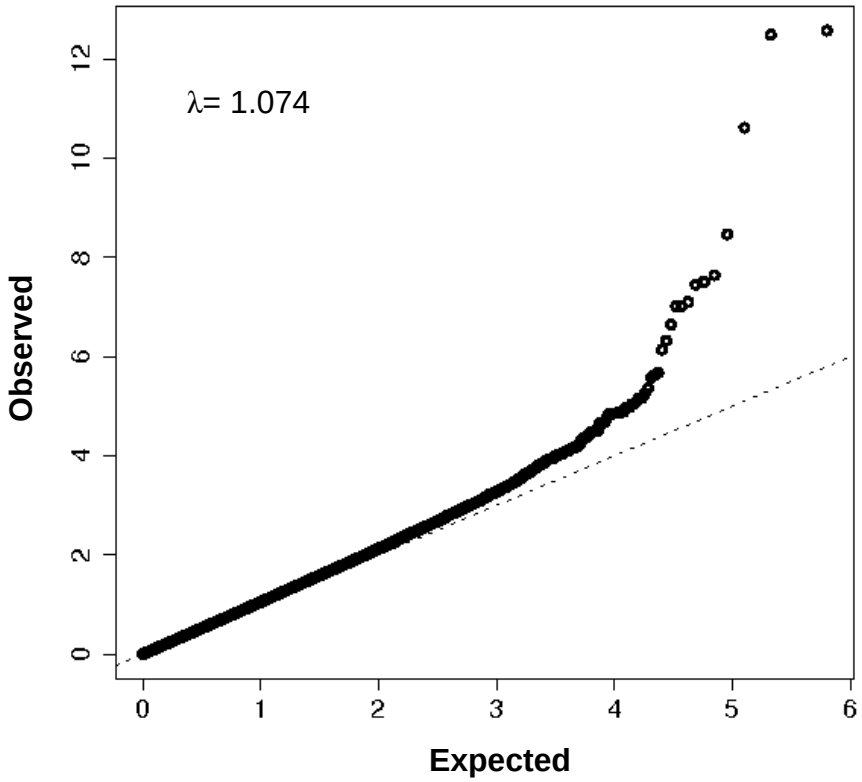
Chunyan He[1,2], Peter Kraft[1,2], Constance Chen[1,2], Julie E. Buring[2–4], Guillaume Paré[3–5], Susan E. Hankinson[2,6], Stephen J. Chanock[7], Paul M. Ridker[2–5], David J. Hunter[1,2,6–8] & Daniel I. Chasman[3–5]

[1]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA. [2]Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA. [3]Donald W. Reynolds Center for Cardiovascular Research, Brigham and Women's Hospital, Harvard Medical School, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA. [4]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA. [5]Center for Cardiovascular Disease Prevention, Brigham and Women's Hospital, Harvard Medical School, 900 Commonwealth Avenue East, Boston, Massachusetts 02115, USA. [6]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, Massachusetts 02115, USA. [7]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, 6120 Executive Boulevard, Bethesda, MD 20892, USA. [8]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, **7** Cambridge Center, Cambridge, Massachusetts 02142, USA. Correspondence should be addressed to C.H. (che@hsph.harvard.edu).

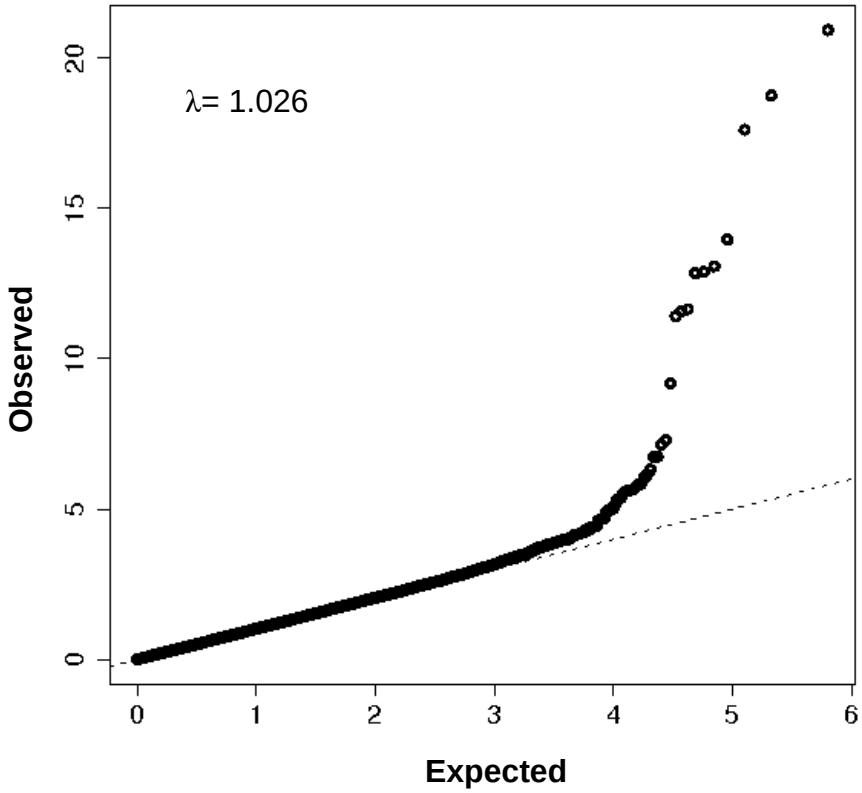
Supplementary Table 1 The means and standard deviations of age at menarche and age at natural menopause in the NHS, the WGHS and the combined sample

	NHS		WGHS		Combined	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Age at menarche	2,270	12.5 (1.39)	15,136	12.4 (1.44)	17,406	12.4 (1.43)
Age at natural menopause	1,344	50.8 (3.08)	7,768	50.6 (3.66)	9,112	50.6 (3.58)

a

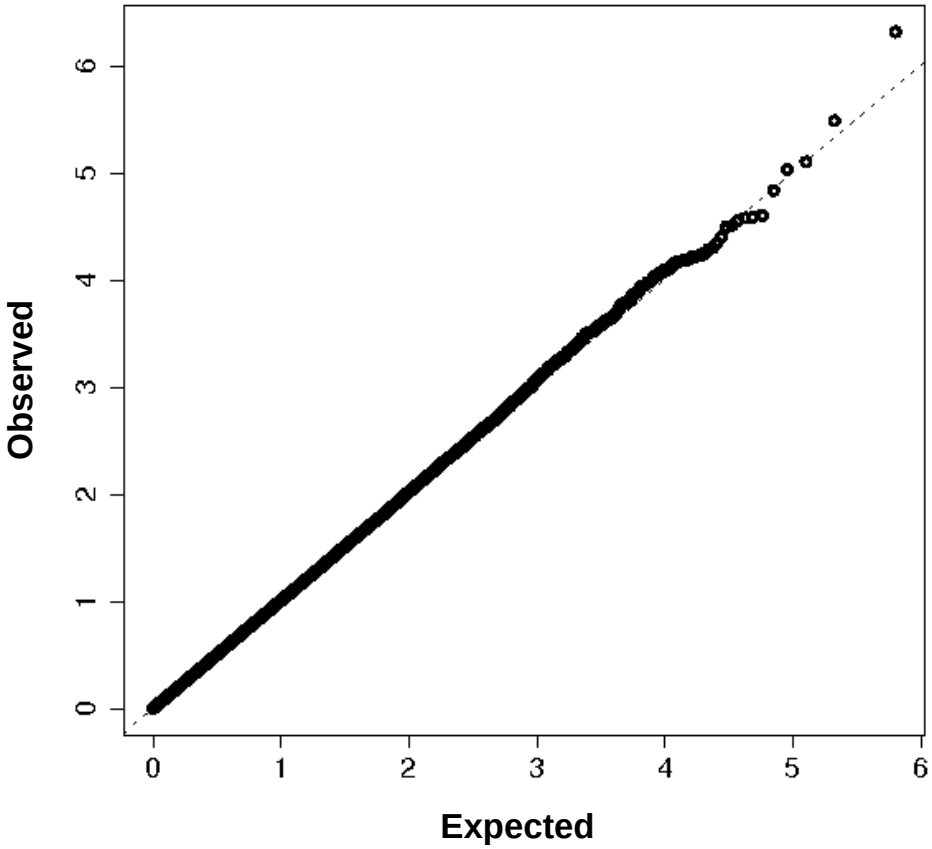

λ= 1.074
Observed
Expected
0
2
4
6
8
10
12
0
1
2
3
4
5
6

b

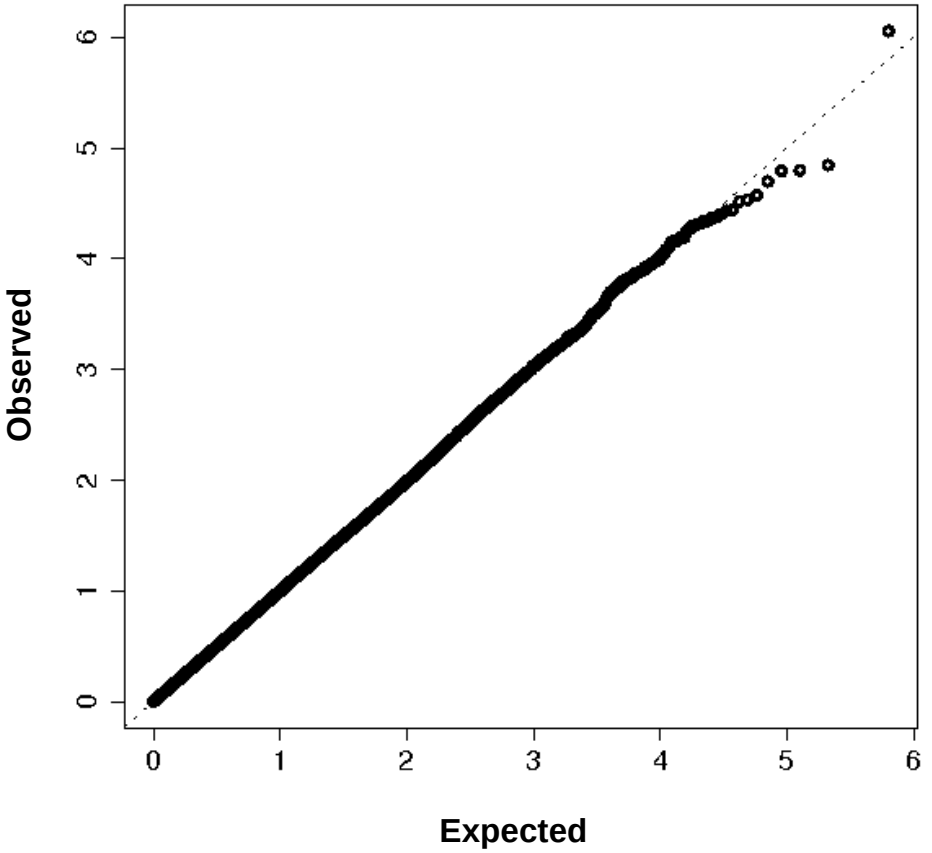

λ= 1.026
Observed
Expected
0
5
10
15
20
0
1
2
3
4
5
6

Supplementary Figure 1 Log quantile-quantile (QQ) P value plots for 317,759 single-SNP tests of association in joint analysis for age at menarche (**a**) and age at natural menopause (**b**), adjusting for the top principle components of genetic variation chosen for each study. Under the null hypothesis of no association at any locus, the points would be expected to follow the dashed line. Deviations from the dashed line correspond to loci that deviate from the null hypothesis. λ= mean (T^2) and T is the test statistic from the linear regression.

a



b



Supplementary Figure 2 Log quantile-quantile (QQ) P value for heterogeneity (Q statistic) plots for 317,759 single-SNP tests of association in joint analysis for age at menarche (**a**) and age at natural menopause (**b**), adjusting for the top principle components of genetic variation chosen for each study. Under the null hypothesis of no heterogeneity between the two studies at any locus, the points would be expected to follow the dashed line. Deviations from the dashed line correspond to loci that deviate from the null hypothesis.

Supplementary Table 2 Means of age at menarche and age at natural menopause by genotype for the genome-wide statistically significant SNPs identified in the combined sample of the NHS and the WGHS

		Major, Minor	Common Homozygote		Heterozygote		Variant Homozygote	
SNP	Total N	Alleles	N	Mean	N	Mean	N	Mean
Age at Menarche								
rs314277	17,366	C, A	12,768	12.4	4,265	12.6	333	12.7
rs314263	17,390	T, C	8,180	12.4	7,506	12.5	1,704	12.5
rs369065	17,338	T, C	7,770	12.4	7,741	12.5	1,827	12.5
rs7861820	17,404	T, C	4,706	12.5	8,654	12.5	4,044	12.4
rs314280	17,301	C, T	5,303	12.4	8,500	12.5	3,498	12.5
rs4946651	17,325	G, A	5,302	12.4	8,553	12.5	3,470	12.5
rs12684013	17,398	C, T	9,338	12.5	6,751	12.4	1,309	12.4
rs4452860	17,397	A, G	9,051	12.5	6,916	12.4	1,430	12.4
rs7028916	17,372	C, A	9,075	12.5	6,880	12.4	1,417	12.4
rs314262	16,816	T, C	4,937	12.4	8,492	12.5	3,387	12.5
Age at Natural Menopause								
rs16991615	9,111	G, A	8,089	50.6	989	51.7	33	52.3
rs1172822	9,096	C, T	3,715	51.0	4,115	50.6	1,266	49.8
rs2384687	9,087	T, C	3,422	51.0	4,233	50.6	1,432	49.9
rs897798	9,084	A, G	2,636	51.0	4,432	50.7	2,016	50.1
rs365132	8,807	G, T	2,356	50.3	4,210	50.7	2,241	51.1
rs7718874	9,111	A, G	2,354	50.3	4,509	50.7	2,248	51.1
rs402511	9,084	C, T	2,354	50.3	4,480	50.7	2,250	51.1
rs7246479	9,102	G, T	2,475	50.2	4,442	50.8	2,185	51.0
rs1551562	9,011	A, G	5,368	50.9	3,117	50.5	526	49.6
rs691141	9,088	C, T	2,742	50.4	4,443	50.7	1,903	51.2
rs12611091	8,882	T, C	2,396	50.3	4,418	50.7	2,068	51.0
rs2153157	9,046	C, T	2,373	50.4	4,448	50.7	2,225	50.9
rs2278493	9,100	G, A	4,023	50.9	4,070	50.6	1,007	50.3

Supplementary Table 3 Estimates of the square of the correlation coefficient (r^2) and D' for pairwise comparison of the identified genome-wide statistically significant SNPs for age at menarche and age at natural menopause within each chromosome region

Region	Chr.	SNP1	SNP2	R^2	D'
6q21	6	rs314262	rs314263	0.54	1.00
	6	rs314262	rs314277	0.20	0.99
	6	rs314262	rs314280	1.00	1.00
	6	rs314262	rs369065	0.48	0.90
	6	rs314262	rs4946651	0.99	1.00
	6	rs314263	rs314277	0.35	0.97
	6	rs314263	rs314280	0.54	1.00
	6	rs314263	rs369065	0.90	0.99
	6	rs314263	rs4946651	0.55	1.00
	6	rs314277	rs314280	0.20	0.99
	6	rs314277	rs369065	0.34	0.99
	6	rs314277	rs4946651	0.20	0.98
	6	rs314280	rs369065	0.49	0.91
	6	rs314280	rs4946651	0.99	1.00
	6	rs369065	rs4946651	0.48	0.90
9q31.2	9	rs12684013	rs4452860	0.92	0.99
	9	rs12684013	rs7028916	0.92	0.99
	9	rs12684013	rs7861820	0.38	0.98
	9	rs4452860	rs7028916	1.00	1.00
	9	rs4452860	rs7861820	0.42	1.00
	9	rs7028916	rs7861820	0.42	1.00
5q35.2	5	rs2278493	rs365132	0.46	0.97
	5	rs2278493	rs3762974	0.68	0.89
	5	rs2278493	rs402511	0.46	0.97
	5	rs2278493	rs691141	0.40	0.98
	5	rs2278493	rs7718874	0.46	0.97
	5	rs365132	rs3762974	0.57	1.00
	5	rs365132	rs402511	1.00	1.00
	5	rs365132	rs691141	0.83	1.00
	5	rs365132	rs7718874	1.00	1.00
	5	rs3762974	rs402511	0.57	1.00
	5	rs3762974	rs691141	0.48	1.00
	5	rs3762974	rs7718874	0.57	1.00
	5	rs402511	rs691141	0.83	1.00
	5	rs402511	rs7718874	1.00	1.00
	5	rs691141	rs7718874	0.83	1.00
19q13.42	19	rs1172822	rs12611091	0.36	0.82
	19	rs1172822	rs1551562	0.54	0.99
	19	rs1172822	rs2384687	0.81	0.94
	19	rs1172822	rs7246479	0.54	0.99
	19	rs1172822	rs897798	0.64	0.99
	19	rs12611091	rs1551562	0.20	0.83
	19	rs12611091	rs2384687	0.27	0.68
	19	rs12611091	rs7246479	0.50	0.71
	19	rs12611091	rs897798	0.54	0.81
	19	rs1551562	rs2384687	0.42	0.91
	19	rs1551562	rs7246479	0.30	0.99
	19	rs1551562	rs897798	0.35	0.99
	19	rs2384687	rs7246479	0.58	0.99
	19	rs2384687	rs897798	0.52	0.86
	19	rs7246479	rs897798	0.82	0.99

Supplementary Table 4 Hapmap SNPs in the specified region that are in high LD (pairwise correlation $r^2 \geq 0.8$) with the identified genome-wide statistically significant SNPs for age at menarche and age at natural menopause in the study based on HapMap phase I+II dataset

Cytological Location	Specified Region (bp)	Genome-wide significant SNP in the study	HapMap SNPs
Age at menarche			
6q21	Chr6:105445489..105596453	rs314277 [a]	-
		rs314263 [a]	-
		rs369065 [a]	rs314268 [a], rs314273 [a], rs7759938 [b], rs9391253 [b], rs395962 [a], rs167539 [a], rs2095812 [c], rs314276 [a]
		rs314280 [a]	-
		rs4946651 [b]	rs1744206 [a], rs314266 [a], rs314270 [a], rs314289 [a], rs364663 [a]
		rs314262 [a]	rs11156429 [b]
9q31.2	Chr9:105924867..106110634	rs7861820 [b]	rs1516890 [b]
		rs12684013 [b]	rs997068 [b], rs2676451 [b], rs7862517 [b], rs13299139 [b], rs2676450 [b], rs13296266 [b], rs13302191 [b], rs4273906 [b]
		rs4452860 [b]	rs13300395 [b], rs2008393 [b], rs7020666 [b], rs12686569 [b], rs10978431 [b], rs2222133 [b], rs1516882 [b], rs1516883 [b], rs7020240 [b], rs12684029 [b], rs12684045 [b], rs10978430 [b], rs9299121 [b], rs12682937 [b], rs7047281 [b], rs7032974 [b], rs10453225 [b], rs1516881 [b], rs12352703 [b]
		rs7028916 [b]	rs7865332 [b], rs7865367 [b], rs13290190 [b], rs12684048 [b], rs10118978 [b], rs2138628 [b], rs11999547 [b], rs7866767 [b], rs13294242 [b], rs7862972 [b], rs1516887 [b], rs7867013 [b], rs2090409 [b], rs10156597 [b]
Age at natural Menopause			
5q35.2	Chr5:176205976..176412209	rs365132 [d]	-
		rs7718874 [a]	rs1700490 [d]
		rs402511 [e]	rs251844 [a], rs601923 [a], rs149307 [a], rs2456181 [a], rs2292256 [a], rs353467 [b], rs353468 [b], rs353478 [a], rs10068703 [a], rs6861925 [a], rs2292255 [a], rs353471 [a]
		rs691141 [a]	rs4490607 [a], rs353496 [a], rs4976665 [a], rs353495 [a], rs3923879 [a], rs251843 [a], rs353479 [a], rs2962842 [a], rs2962844 [a], rs2940521 [a], rs353494 [a], rs187114 [a], rs547798 [e]
		rs2278493 [a]	rs4976662 [a], rs6875296 [d], rs2278492 [d]
6p24.2	Chr6:10958298..11046090	rs2153157 [a]	-
19q13.42	Chr19:60429993..60571684	rs1172822 [a]	rs11668344 [a], rs11668309 [a], rs4806660 [a]
		rs2384687 [a]	rs10425848 [a], rs1109368 [a], rs4806661 [a]
		rs897798 [a]	rs7252864 [a]
		rs7246479 [f]	rs8113016 [a], rs10411773 [a], rs10412726 [a], rs734518 [a]
		rs1551562 [a]	-
		rs12611091 [a]	rs2607336 [a], rs1172818 [a]
20p12.3	chr20:5851323..5944887	rs16991615 [f]	-

[a] Intronic; [b] Unkown; [c] 3'-UTR; [d] Exonic, synonymous; [e] 5'-UTR; [f] Exonic, non-synonymous.

Supplementary Methods

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

The NHS nested breast cancer case-control study was derived from the 32,826 women in the blood subcohort who were free of diagnosed breast cancer at blood collection and followed for incidence disease until June 1, 2004. Breast cancer follow-up in the NHS was conducted by personal mailings and searches of the National Death Index. Controls were women not diagnosed with breast cancer during follow-up, and were one-to-one matched to cases based on age at diagnosis, blood collection variables (time of day, season, and year of blood collection, as well as recent (<3 months) use of postmenopausal hormones), ethnicity (all cases and controls are self-reported Caucasians), and menopausal status (all cases were postmenopausal at diagnosis). The 2,287 NHS participants included in the present analysis were from this nested breast cancer case-control study and were self-described Caucasians with genotype data available from the National Cancer Institute's Cancer Genetic Marker of Susceptibility (CGEMS) project[1].

The WGHS cohort comprises 28,345 American women, who are participants in the ongoing Women's Health Study (WHS) who had no prior history of cardiovascular disease, cancer, or other major chronic illness, and who provided a blood sample at baseline with consent for analyses linking blood-derived observations with baseline risk factor profiles and incident disease events. These women were 45 years of age or older at baseline in 1993. Details of the rationale, design, and methodology of the WGHS and the WHS are described elsewhere[2,3]. While genotyping of the cohort is ongoing, there were 15,151 Caucasian WGHS participants with genotype data available for the present analysis.

Validity of phenotype measurement. Age at menarche and age at natural menopause were self-reported, and age at menarche was determined retrospectively. Thus, both traits might be measured with error, although this error is likely to be random with respect to genotype. Age at menarche and age at natural menopause are associated with breast cancer and other endpoints in both the NHS and the WHS with magnitudes and directions of association that are consistent with what has been observed in other published studies, attesting to the validity of the measurements. In addition, Colditz *et al.* found a correlation of 0.82 for reported ages at natural menopause between two questionnaires in the NHS[4]. Must *et al.* observed a correlation of 0.79 between contemporarily documented and recalled age at menarche after 30 years[5]. Moreover, random measurement error, if present, would be expected to attenuate associations in our study, rather than elicit false positive associations.

Population stratification. In the NHS GWA study, population stratification was analyzed based on approximately 10,000 unlinked markers[1]. After excluding admixed (>15%) individuals using STRUCTURE[6], population stratification was estimated using the four largest principal components of genetic variation that were calculated using the EIGENSTRAT software[1,7] and chosen on the basis of significant ($p<0.05$) Tracy-Wisdom tests[8].

In the WGHS GWA study, a principal component analysis using 1,443 ancestry informative SNPs was performed using PLINK [9] in order to confirm self-reported ancestry. Subjects (0.3%) were removed from analysis if they did not cluster with other Caucasians. Furthermore, to rule out residual stratification among Caucasians only, we performed a principal component analysis with EIGENSTRAT software [7] using 124,931 SNPs chosen to have pair-wise linkage disequilibrium lower than r2 = 0.4. The first ten components were then used as covariates for adjustment of age at menarche or natural menopause.

References

1. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
2. Ridker, P.M. et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54**, 249-55 (2008).
3. Ridker, P.M. et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med* **352**, 1293-304 (2005).
4. Colditz, G.A. et al. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* **126**, 319-25 (1987).
5. Must, A. et al. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am J Epidemiol* **155**, 672-9 (2002).
6. Pritchard, J.K. & Rosenberg, N.A. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**, 220-8 (1999).
7. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
8. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
9. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).