nature genetics

Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies

To identify loci for age at menarche, we performed a meta-analysis of 32 genome-wide association studies in 87,802 women of European descent, with replication in up to 14,731 women. In addition to the known loci at *LIN28B* ($P = 5.4 \times 10^{-60}$) and 9q31.2 ($P = 2.2 \times 10^{-33}$), we identified 30 new menarche loci (all $P < 5 \times 10^{-8}$) and found suggestive evidence for a further 10 loci ($P < 1.9 \times 10^{-6}$). The new loci included four previously associated with body mass index (in or near *FTO*, *SEC16B*, *TRA2B* and *TMEM18*), three in or near other genes implicated in energy homeostasis (*BSX*, *CRTC1* and *MCHR2*) and three in or near genes implicated in hormonal regulation (*INHBA*, *PCSK2* and *RXRG*). Ingenuity and gene-set enrichment pathway analyses identified coenzyme A and fatty acid biosynthesis as biological processes related to menarche timing.

Menarche, the onset of first menstruation in girls, indicates the attainment of reproductive capacity and is a widely used marker of pubertal timing. Age at menarche varies widely between girls and is highly dependent on nutritional status[1]. Early menarche is associated with several adverse health outcomes, including breast cancer[2], endometrial cancer[3], obesity[4], type 2 diabetes[5] and cardiovascular disease[6], as well as shorter adult stature[4]. Studies of twins and extended families, although largely performed in populations free of nutritional deprivation, estimate that around 50% of the variance in menarche timing is attributable to genetic factors in such settings[7].

Recently, common variants in *LIN28B* were associated with age at menarche in four independent genome-wide association studies (GWAS)[8–11]. *LIN28B* is a human homolog of *lin-28* in *Caenorhabditis elegans*, which controls the rate of progression from larval stages to adult cuticle formation, indicating the possible conservation of specific micro-RNA regulatory mechanisms involved in developmental timing[9]. A second menarche locus was identified in an intergenic region at 9q31.2[8,10]. These two loci together explained only 0.6% of the variance in age at menarche[8]. We anticipated that a much larger GWAS would substantially increase the yield of loci associated with age at menarche.

Here we report a much expanded meta-analysis of GWAS for age at menarche. By combining data from the previous studies[8–11], plus several further studies to form the ReproGen Consortium, we identified at least 30 previously unidentified loci associated with age at menarche at genome-wide significance levels. Our findings show a close link between the genetic regulation of energy homeostasis and pubertal timing and suggest the presence of other diverse pathways.

RESULTS

Genome-wide association for age at menarche

This expanded GWAS includes data from 32 cohorts of European ancestry ($N = 87{,}802$). In most studies, age at menarche was determined by self recall, and the mean age at menarche in individual studies ranged from 12.4 to 13.6 years, excluding individuals with menarche <9 years and >17 years (Online Methods, **Supplementary Table 1** and **Supplementary Note**). Genome-wide SNP genotyping was performed using a variety of different platforms (**Supplementary Table 2** and **Supplementary Note**). Therefore, after applying standard quality control measures, we imputed the genotypes for ~2.5 million autosomal SNPs in the HapMap European CEU sample using Build 35 or 36 to allow inverse variance meta-analysis of additive genetic association results from each study. We also meta-analyzed results from X-chromosome SNPs in studies which had this data available ($N = 52{,}781$). Test statistics from each cohort were adjusted using genomic control to avoid inflation of results due to population stratification.

There was strong deviation from the uniform distribution of *P* values expected under the null hypothesis (**Supplementary Fig. 1**). This deviation was attenuated, but persisted, following removal of those signals associated with the two previously identified loci. In total, 945 SNPs representing 45 loci ($r^2 < 0.05$ based on HapMap in a 750-kb region) were associated with age at menarche at genome-wide significance levels ($P < 5 \times 10^{-8}$) (**Fig. 1** and **Supplementary Fig. 2**). None of these loci were located on the X chromosome. These 45 loci included three apparent second signals (defined as two genome-wide significant SNPs in low linkage disequilibrium (LD) ($r^2 < 0.05$) in the same 750-kb region) at 2q33.1, 6q21 and 14q32.2. The second signal at 6q21 (rs314279) had a low minor allele frequency (MAF = 6%) and was not present in many studies. We therefore genotyped this SNP *de novo* in the InCHIANTI cohort and found it was in LD with the top chromosome 6 signal (rs7759938, $r^2 = 0.3$). In HapMap, the r^2 between the two chromosome 6 SNPs was 0.015, but the *D* was 1.0. To verify the independence of additional loci, we performed a conditional analysis and a meta-analysis of all 32 studies using the top SNPs at all the 42 genome-wide significant regions as covariates (in addition to birth year). In these conditional analyses, the possible second signals on chromosomes 2 and 14 showed strong but not genome-wide significant associations with age at menarche ($P < 7.1 \times 10^{-6}$), suggestive of, but not confirming, second independent signals in these two regions (**Fig. 1** and **Supplementary Table 3**).

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

Received 21 May; accepted 19 October; published online 21 November 2010; doi:10.1038/ng.714

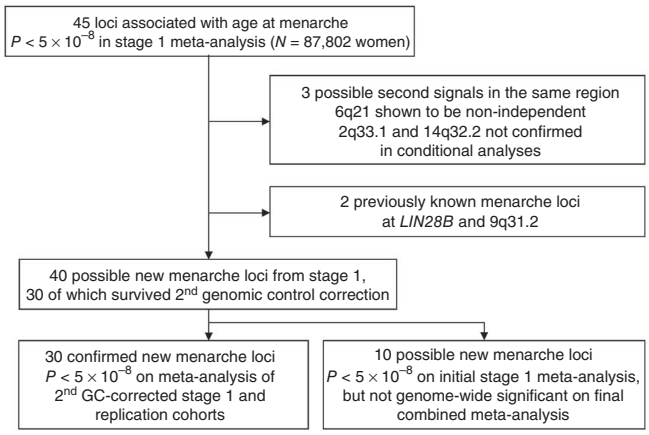

Figure 1 Flow diagram of the discovery and confirmation of new loci for age at menarche. GC, genomic control.

The two most significant loci for age at menarche confirmed the previously reported associations at *LIN28B* (rs7759938, $P = 1.6 \times 10^{-58}$) and 9q31.2 (rs2090409, $P = 4.4 \times 10^{-33}$) (**Table 1** and **Supplementary Fig. 3**). In addition, there were genome-wide significant signals for a further 40 possible previously unidentified loci, of which 30 survived a second more stringent correction for the overall genomic control in the stage 1 cohorts ($\lambda = 1.173$) (**Table 1**, **Fig. 1** and **Supplementary Fig. 3**).

Replication studies

We sought confirmation of the 40 possible new menarche loci in up to 14,731 women from 16 additional studies with *in silico* GWAS data and new genotyping data from one cohort (**Supplementary Tables 4** and **5**). This replication sample was substantially smaller than our stage 1 sample and was therefore underpowered to confirm individual SNP associations (**Supplementary Fig. 4**). Nonetheless, 37 of the 40 possible newly associated loci showed directionally consistent associations in both stages (**Table 1**; binomial sign test $P = 9.7 \times 10^{-9}$). A combined meta-analysis of the more stringent second genomic control–corrected stage 1 results and replication cohorts gave confirmatory evidence for 30 new menarche loci, leaving 10 unconfirmed possible menarche loci (**Table 1** and **Fig. 1**).

Based on the combined stage 1 and replication results, the estimated magnitudes of per-allele effects for the new menarche loci ranged from 4.5 to 2.1 weeks per allele (**Table 1**) and had an inverse relationship with MAF (**Supplementary Fig. 5**). Among the four largest *in silico* replication cohorts (each comprising >800 women), the variance in age at menarche explained by all 42 known, confirmed and possible new menarche loci ranged from 3.6% to 6.1% (**Supplementary Table 6**).

Candidate genes at new loci

The strongest new menarche signal was for rs1079866 (3.9 weeks per minor allele; 95% CI 2.9–5.0, $P = 5.5 \times 10^{-14}$) located approximately 250 kb downstream of *INHBA*, which encodes the protein subunit Inhibin beta A. Heterodimers of Inhibin beta A and the Inhibin alpha subunit form the female reproductive hormone Inhibin A[12]. Inhibin A, produced by granulosa cells in the ovary, increases dramatically during pubertal development in girls[13,14] and is involved in negative feedback regulation by inhibiting production of follicle stimulating hormone by the pituitary and secretion of gonadotrophin releasing hormone from the hypothalamus[15]. Conversely, homodimers of Inhibin beta A form the hormone Activin A, which stimulates pituitary follicle stimulating hormone production and also exhibits a wide range of biological activities, including the regulation of cellular proliferation and differentiation[16].

The second strongest new signal was for rs466639 ($P = 1.3 \times 10^{-13}$); this SNP is intronic in *RXRG*, which encodes retinoid X receptor gamma, a nuclear receptor that forms dimers with the receptors for retinoic acid, thyroid hormone and vitamin D, increasing both DNA binding and transcriptional function on their respective response elements[17].

Four new loci for menarche were previously identified by GWAS for adult body mass index (BMI)[18–20]: rs9939609 (in or near *FTO*, $P = 3.1 \times 10^{-8}$), rs633715 (*SEC16B*, $P = 2.1 \times 10^{-8}$), rs2002675 (*TRA2B* and *ETV5*, $P = 1.2 \times 10^{-9}$) and rs2947411 (*TMEM18*, $P = 1.7 \times 10^{-8}$). Apart from rs2002675, these menarche signals were either identical to or in tight LD ($r^2 > 0.9$) with those BMI loci, and in all cases, the BMI-increasing allele was associated with earlier menarche. Variants at these four loci have also been associated with childhood BMI[18–20], and these findings support a likely causal effect of childhood BMI on earlier pubertal timing.

Three new menarche loci were found in or near further genes implicated in the regulation of energy homeostasis and body weight in animal models: rs6589964 ($P = 1.9 \times 10^{-12}$) lies ~18 kb from *BSX*, rs10423674 ($P = 5.9 \times 10^{-9}$) is intronic in *CRTC1*, and rs4840046 ($P = 2.4 \times 10^{-8}$) lies ~160 kb from *MCHR2*. *BSX* encodes a DNA-binding protein and transcriptional activator. In mouse, *Bsx* is expressed specifically in the pineal gland, telencephalic septum, hypothalamic pre-mammillary body and arcuate nucleus and is necessary for postnatal growth, locomotory behavior, expression of the genes *Npy* and *Agrp*, and for the hyperphagic phenotype in leptin deficiency[21]. *CRTC1* encodes the CREB-regulated transcription coactivator 1, an activator of cellular gene expression. *Crtc1*$^{-/-}$ mice are hyperphagic, obese and infertile, and *Crtc1*$^{-/-}$ females have low circulating luteinizing hormone levels[22]. Leptin potentiates the effects of *Crtc1* transcriptional activity, and *Crtc1* overexpression in hypothalamic cells increases expression of Kisspeptin, which in turn activates secretion of the gonadotrophin releasing hormone. *MCHR2* encodes the melanin concentrating hormone receptor 2, an orphan G protein–coupled receptor which shows high affinity binding to the hypothalamic neuropeptide melanin-concentrating hormone (MCH), which regulates nutrient intake and energy homeostasis through *MCHR1*[23]. Furthermore, MCH directly inhibits gonadotrophin releasing hormone neurons and thereby links energy balance to reproduction[24].

rs852069 ($P = 3.3 \times 10^{-8}$) lies ~84 kb from *PCSK2*, which encodes proprotein convertase subtilisin/kexin type 2, an enzyme that cleaves latent precursor proteins, such as proinsulin and proopiomelanocortin, into their biologically active products. Although rare deleterious mutations and common variants in *PCSK1* are known to influence obesity risk, it is notable that *PCSK2* differs from *PCSK1* in that it additionally cleaves pro-luteinizing hormone-releasing hormone and could therefore have a more direct influence on the reproductive hormone axis.

Pathway analyses

Remaining new menarche loci were found in or near genes that are involved in a seemingly diverse range of biological functions (**Supplementary Table 7**). We used ingenuity pathway analysis (IPA) to identify potential biological pathways common to these identified loci. Based on direct interactions only, we identified two functional networks containing 16 and 11 genes, respectively, of those genes nearest to the new menarche loci (**Supplementary Fig. 6**). Network 1,

related to 'gene expression, cellular growth and proliferation, and cellular function and maintenance', covers a wide and nonspecific range of biological pathways. Functions in network 2 relate to 'lipid metabolism, small molecule biochemistry and molecular transport' (**Supplementary Table 8**). Central to network 2 are *RXRG* and several genes involved in fatty acid biosynthesis, including several fatty acid–binding proteins and *ACSL1*, which encodes an enzyme that converts free long-chain fatty acids into fatty acyl-CoA esters.

To identify potential further biological pathways that influence menarche timing, we used a gene set enrichment analysis (GSEA)

Table 1 Stage 1 and replication results for 42 known, confirmed or possible new loci for age at menarche

							Stage 1			Replication		Stage 1 and replication			
SNP	Nearest gene(s)	Distance from gene (kb)	Chr.	Position (Build 36)	MAF[a]	Alleles[b]	P_{het}[c]	P[d]	P[e 2-GC]	n	P[f]	β[g]	s.e.	Direction[h]	P[i]
Previous menarche loci															
rs7759938[j]	*LIN28B*	~26 kb	6	105,485,647	0.32	C/T	0.04	1.6×10^{-58}	4.3×10^{-50}	14,185	4.6×10^{-11}	6.4	0.4	+/+	5.4×10^{-60}
rs2090409	*TMEM38B*	~400 kb	9	108,006,909	0.31	A/C	0.05	4.4×10^{-33}	2.3×10^{-28}	14,708	2.7×10^{-6}	–4.7	0.4	–/–	2.2×10^{-33}
30 novel menarche loci															
rs1079866	*INHBA*	~250 kb	7	41,436,618	0.15	G/C	0.81	1.9×10^{-16}	2.7×10^{-14}	14,731	1.9×10^{-1}	3.9	0.5	+/+	5.5×10^{-14}
rs466639	*RXRG*	Intronic	1	163,661,506	0.13	T/C	0.80	7.8×10^{-15}	8.9×10^{-13}	14,279	3.1×10^{-2}	–4.2	0.6	–/–	1.3×10^{-13}
rs6438424	*3q13.32*	Intergenic	3	119,057,512	0.50	A/C	0.99	8.4×10^{-14}	4.6×10^{-12}	8,634	6.7×10^{-3}	–2.7	0.4	–/–	1.4×10^{-13}
rs1398217	*FUSSEL18*	Intronic	18	43,006,236	0.43	G/C	0.33	5.7×10^{-13}	2.5×10^{-11}	14,344	2.3×10^{-3}	–2.7	0.4	–/–	2.3×10^{-13}
rs12617311	*PLCL1*	~195 kb	2	199,340,810	0.32	A/G	0.90	2.6×10^{-13}	1.2×10^{-11}	14,007	1.1×10^{-2}	–3.0	0.4	–/–	6.0×10^{-13}
rs9635759	*CA10*	~94 kb	17	46,968,784	0.32	A/G	0.43	2.0×10^{-13}	1.5×10^{-11}	14,002	1.1×10^{-2}	3.0	0.4	+/+	7.3×10^{-13}
rs6589964	*BSX*	~18 kb	11	122,375,893	0.48	A/C	0.89	8.8×10^{-14}	4.3×10^{-12}	13,754	8.3×10^{-2}	–2.7	0.4	–/–	1.9×10^{-12}
rs10980926	*ZNF483*	Intronic	9	113,333,455	0.36	A/G	0.65	2.2×10^{-13}	9.2×10^{-12}	14,227	3.8×10^{-1}	2.5	0.4	+/+	4.2×10^{-11}
rs17268785	*CCDC85A*	Intronic	2	56,445,587	0.17	G/A	0.82	6.8×10^{-11}	2.0×10^{-9}	14,233	1.5×10^{-2}	3.2	0.5	+/+	9.7×10^{-11}
rs13187289	*PHF15*	~12 kb	5	133,877,076	0.20	G/C	0.99	2.0×10^{-10}	3.6×10^{-9}	14,303	1.4×10^{-2}	3.0	0.5	+/+	1.9×10^{-10}
rs7642134	*VGLL3*	~70 kb	3	86,999,572	0.38	A/G	0.65	2.3×10^{-9}	4.3×10^{-8}	14,205	2.1×10^{-3}	–2.4	0.4	–/–	3.5×10^{-10}
rs17188434	*NR4A2*	~84 kb	2	156,805,022	0.07	C/T	0.59	3.4×10^{-11}	9.1×10^{-10}	14,356	2.2×10^{-1}	–4.5	0.7	–/–	1.1×10^{-9}
rs2002675	*TRA2B, ETV5*	~4 kb, ~135 kb	3	187,112,262	0.42	G/A	0.94	3.9×10^{-9}	4.7×10^{-8}	14,334	6.6×10^{-3}	2.2	0.4	+/+	1.2×10^{-9}
rs7821178	*PXMP3*	~181 kb	8	78,256,392	0.34	A/C	0.38	6.7×10^{-10}	1.2×10^{-8}	14,151	8.0×10^{-2}	–2.4	0.4	–/–	3.0×10^{-9}
rs1659127	*MKL2*	~28 kb	16	14,295,806	0.34	A/G	0.19	3.0×10^{-9}	4.5×10^{-8}	14,021	2.5×10^{-2}	2.4	0.4	+/+	4.0×10^{-9}
rs10423674	*CRTC1*	Intronic	19	18,678,903	0.35	A/C	0.79	1.1×10^{-9}	1.7×10^{-8}	13,543	1.1×10^{-1}	2.3	0.4	+/+	5.9×10^{-9}
rs10899489	*GAB2*	Intronic	11	77,773,021	0.15	A/C	0.16	2.4×10^{-10}	4.7×10^{-9}	14,201	2.5×10^{-1}	3.1	0.5	+/+	8.1×10^{-9}
rs6575793	*BEGAIN*	Intronic	14	100,101,970	0.42	C/T	0.51	1.7×10^{-10}	3.7×10^{-9}	13,899	4.6×10^{-1}	2.3	0.4	+/+	1.2×10^{-8}
rs4929923	*TRIM66*	3′UTR	11	8,595,776	0.36	T/C	0.99	2.4×10^{-8}	2.2×10^{-7}	8,510	1.6×10^{-2}	2.3	0.4	+/+	1.2×10^{-8}
rs6439371	*TMEM108, NPHP3*	~146 kb, ~170 kb	3	134,093,442	0.34	G/A	0.35	1.5×10^{-8}	1.6×10^{-7}	8,581	3.0×10^{-2}	2.3	0.4	+/+	1.3×10^{-8}
rs900145	*ARNTL*	~5 kb	11	13,250,481	0.30	C/T	0.35	7.7×10^{-9}	1.1×10^{-7}	8,649	$6.5 \times 10^{--2}$	2.3	0.4	+/+	1.6×10^{-8}
rs6762477	*RBM6*	Intronic	3	50,068,213	0.44	G/A	0.22	1.4×10^{-9}	2.4×10^{-8}	12,447	1.5×10^{-1}	2.5	0.4	+/+	1.6×10^{-8}
rs2947411	*TMEM18*	~53 kb	2	604,168	0.17	A/G	0.27	2.1×10^{-8}	2.6×10^{-7}	8,657	1.9×10^{-2}	2.8	0.5	+/+	1.7×10^{-8}
rs1361108	*C6orf173, TRMT11*	~98 kb, ~407 kb	6	126,809,293	0.46	T/C	0.76	2.6×10^{-9}	3.0×10^{-8}	14,126	6.0×10^{-2}	–2.1	0.4	–/–	1.7×10^{-8}
rs1364063	*NFAT5*	~10 kb	16	68,146,073	0.43	C/T	0.05	4.4×10^{-8}	4.8×10^{-7}	8,669	7.1×10^{-3}	2.1	0.4	+/+	1.8×10^{-8}
rs633715	*SEC16B*	~44 kb	1	176,119,203	0.20	C/T	0.45	1.5×10^{-9}	2.3×10^{-8}	14,274	1.9×10^{-1}	–2.6	0.5	–/–	2.1×10^{-8}
rs4840086	*PRDM13, MCHR2*	~145 kb, ~160 kb	6	100,315,159	0.42	G/A	0.98	8.2×10^{-9}	1.2×10^{-7}	8,669	7.5×10^{-2}	–2.1	0.4	–/–	2.4×10^{-8}
rs7617480	*KLHDC8B*	Intronic	3	49,185,736	0.22	A/C	0.64	1.8×10^{-9}	2.7×10^{-8}	14,341	2.4×10^{-1}	2.4	0.4	+/+	2.8×10^{-8}
rs9939609	*FTO*	Intronic	16	52,378,028	0.40	A/T	0.17	3.3×10^{-11}	1.1×10^{-9}	8,665	5.3×10^{-1}	–2.1	0.4	–/+	3.1×10^{-8}
rs852069	*PCSK2*	~84 kb	20	17,070,593	0.37	A/G	0.47	1.1×10^{-9}	2.0×10^{-8}	14,306	3.3×10^{-1}	–2.1	0.4	–/–	3.3×10^{-8}
10 possible menarche loci[k]															
rs757647	*KDM3B*	Intronic	5	137,735,214	0.22	A/G	0.23	1.4×10^{-9}	2.0×10^{-8}	14,326	4.4×10^{-1}	–2.4	0.4	–/–	5.4×10^{-8}
rs9555810	*C13orf16, ARHGEF7*	~185 kb, ~223 kb	13	110,979,438	0.28	G/C	0.68	6.7×10^{-10}	1.4×10^{-8}	14,266	4.9×10^{-1}	2.3	0.4	+/+	5.6×10^{-8}
rs16938437	*PHF21A*	Intronic	11	46,009,151	0.09	T/C	0.32	1.4×10^{-9}	2.2×10^{-8}	14,330	3.8×10^{-1}	–3.7	0.7	–/–	5.9×10^{-8}
rs2687729	*EEFSEC*	Intronic	3	129,377,916	0.27	G/A	0.36	1.0×10^{-8}	1.4×10^{-7}	8,669	3.2×10^{-1}	2.3	0.4	+/+	1.3×10^{-7}
rs1862471	*OLFM2*	Intronic	19	9,861,322	0.47	G/C	0.17	4.6×10^{-10}	8.3×10^{-9}	13,470	9.4×10^{-1}	2.0	0.4	+/–	1.5×10^{-7}
rs12472911	*LRP1B*	Intronic	2	141,944,979	0.20	C/T	0.65	3.9×10^{-8}	3.9×10^{-7}	8,585	1.4×10^{-1}	2.5	0.5	+/+	1.5×10^{-7}
rs3914188	*ECE2*	3′ UTR	3	185,492,742	0.27	G/C	0.54	2.3×10^{-9}	3.2×10^{-8}	14,085	7.9×10^{-1}	–2.2	0.4	–/–	2.6×10^{-7}
rs2243803	*SLC14A2*	~238 kb	18	41,210,670	0.40	A/T	0.89	2.8×10^{-8}	3.3×10^{-7}	8,659	3.9×10^{-1}	2.0	0.4	+/+	3.4×10^{-7}
rs3743266	*RORA*	3′ UTR	15	58,568,805	0.32	C/T	0.24	2.6×10^{-8}	2.9×10^{-7}	8,666	7.8×10^{-1}	–2.0	0.4	–/–	8.0×10^{-7}
rs7359257	*IQCH*	Intronic	15	65,489,961	0.45	A/C	0.82	3.9×10^{-9}	4.7×10^{-8}	14,303	6.0×10^{-1}	1.7	0.4	+/–	1.9×10^{-6}

UTR, untranslated region.
[a]Minor allele frequency. [b]Minor/major allele. [c]*P* value for effect heterogeneity between studies. [d]*P* value from stage 1 meta-analysis with genomic control applied to individual studies (up to 87,802 women from 32 studies). [e]*P* value from stage 1 meta-analysis with additional adjustment for overall genomic control. [f]*P* value from *in silico* replication studies (up to 14,731 women). [g]Per allele change in age at menarche (weeks) obtained from a meta-analysis of stage 1 and replication cohorts. [h]Direction of minor allele association with age at menarche in stage 1/replication cohorts. [i]*P* value from meta-analysis of stage 1 (second genomic-control–corrected estimates) and replication cohorts. [j]rs314276 was used as a proxy in the ALSPAC replication sample. [k]These loci reached genome-wide significance in stage 1 but not in the final analysis with second genomic-control correction and combination with replication cohorts.

Table 2 Associations between known obesity-related SNPs and age at menarche

Nearby gene	SNP[a]	Chr.	Obesity phenotype	Menarche β (weeks per allele)	Menarche s.e.	Menarche *P*	Obesity-susceptibility allele	Menarche-decreasing allele
FTO	rs9939609	16q12	BMI	2.5	0.4	3.3×10^{-11}	A	A
SEC16B	rs10913469	1q25	BMI	2.6	0.5	2.4×10^{-8}	C	C
GNPDA2	rs10938397	4p13	BMI	2.1	0.4	8.7×10^{-8}	G	G
NEGR1	rs2815752	1p31	BMI	1.9	0.4	5.9×10^{-7}	A	A
TMEM18	rs6548238	2p25	BMI	2.7	0.5	7.1×10^{-7}	C	C
FAIM2	rs7138803	12q13	BMI	1.8	0.4	1.7×10^{-6}	A	A
BDNF	rs4923461	11p14	BMI	1.7	0.5	3.1×10^{-4}	A	A
KCTD15	rs11084753	19q13	BMI	1.4	0.4	5.9×10^{-4}	G	G
TRA2B, ETV5	rs7647305	3q27	BMI	1.2	0.5	9.0×10^{-3}	C	C
TFAP2B	rs987237	6p12	WHR	1.6	0.5	7.8×10^{-4}	G	G
MSRA	rs7826222	8p23	WHR	1.8	0.8	2.4×10^{-2}	G	G

BMI, body mass index; WHR, waist-hip ratio.
[a]Selected SNPs at each locus are those published for association with BMI, WHR or obesity (rather than those with the strongest signal for age at menarche). SNPs listed are those with a significant association ($P < 0.05$) with age at menarche. A full version of this table including SNPs related to adiposity traits but not reaching significance for menarche can be found in **Supplementary Table 13**.

approach in meta-analysis gene-set enrichment of variant associations (MAGENTA), in which each gene in the genome is assigned an adjusted score that represents its association with age at menarche, and predefined pathways are tested for enrichment of multiple associations (Online Methods). The most significant pathway ($P = 4.9 \times 10^{-3}$) was the biosynthesis of coenzyme A, which is a carrier of acyl groups and is necessary for pyruvate oxidation and fatty acid synthesis and oxidation (**Supplementary Table 9**).

Functional SNP and structural assessment

We explored the potentially functional impacts of our new menarche loci in order to identify their likely genetic mechanisms. In addition, by particularly focusing on those groups of SNPs that have been identified as functional, we aimed to identify possible further menarche loci which did not reach genome-wide significance in our primary meta-analysis.

Copy number variation. Using data from a recent genomic map of copy number variation (CNV)[25], we established that none of the 42 known, confirmed or possible new menarche loci were related to CNVs. Next, we explored the 1,052 CNV-tagging SNPs for association with age at menarche in our GWAS sample. Only one tag SNP was associated with age at menarche after Bonferroni correction (rs3101336, $P = 3 \times 10^{-7}$; **Supplementary Fig. 7**). This SNP tags a CNV near the NEGR1 gene locus, which has been previously associated with body mass index[20].

Non-synonymous SNPs. None of the 42 known, confirmed or possible new menarche variants were amino acid changing. However, two were in strong LD ($r^2 \geq 0.8$) with non-synonymous variants. rs1862471 (intronic in OLFM2 at 19p13.2) is in LD ($r^2 = 0.8$) with rs2303100, which encodes an arginine to glutamine residue change in OLFM2. Second, rs4929923 (in the 3 untranslated region of TRIM66 at 11p15.4) is in LD ($r^2 = 0.92$) with rs11042023, which encodes a histidine to arginine residue change in TRIM66.

To identify possible further menarche loci, we then explored the set of 12,062 non-synonymous SNPs for association with age at menarche in our GWAS sample. Outside of the already associated regions, three non-synonymous SNPs were associated with age at menarche after correction for multiple testing (the Bonferroni threshold for 12,062 independent tests was $P < 4.1 \times 10^{-6}$). These non-synonymous SNPs were rs1254319 in *C14orf39* ($P = 1.9 \times 10^{-7}$), rs7653652 in *C3orf38* ($P = 1.4 \times 10^{-6}$) and rs913588 in *JMJD2C* ($P = 3.3 \times 10^{-6}$).

Expression QTLs. Three of the forty-two known, confirmed or possible new menarche variants were highly significantly cis associated with mRNA expression ($P < 1 \times 10^{-6}$ for mRNA transcript abundance) based on publicly available data from lymphoblastoid cell lines of 400 children (mRNA by SNP Browser). These transcripts were in GAB2 (associated with rs10899489), RBM6 (rs6762477) and NARG2 (rs3743266) (**Supplementary Table 10**). As these genomic loci included a number of genes (**Supplementary Fig. 3**), these specific transcript associations inform the likely functional gene at each locus.

Table 3 Associations between known height SNPs and age at menarche

Gene	SNP[a]	Chr.	Position	Menarche β (weeks per allele)	Menarche s.e.	Menarche *P*	Height-increasing allele	Menarche-increasing allele
LIN28B	rs314277	6	105,514,355	6.9	0.6	2.1×10^{-35}	A	A
PXMP3	rs7846385	8	78,322,734	2.5	0.4	1.9×10^{-9}	C	T
C6orf173	rs4549631	6	127,008,001	1.8	0.4	4.9×10^{-7}	C	T
SCMH1	rs6686842	1	41,303,458	–1.1	0.4	3.3×10^{-3}	T	C
Histone cluster 1	rs10946808	6	26,341,366	1.1	0.4	6.4×10^{-3}	A	A
NOG	rs4794665	17	52,205,328	–0.9	0.4	1.1×10^{-2}	A	G
HMGA2	rs1042725	12	64,644,614	–0.8	0.4	2.0×10^{-2}	C	C
TBX2	rs757608	17	56,852,059	–0.9	0.4	2.2×10^{-2}	A	G
HLA Class III	rs2844479	6	31,680,935	–0.9	0.4	2.4×10^{-2}	A	C
ZBTB38	rs6440003	3	142,576,899	0.8	0.4	3.5×10^{-2}	A	A
CABLES1	rs4800148	18	18,978,326	–1.0	0.5	3.7×10^{-2}	A	G

$\chi^2 = 7.02$, $P = 0.008$ for 11 out of 44 height-associated SNPs also associated with age at menarche (at $P < 0.05$) compared to the 2.2 expected by chance. However, seven height-increasing SNPs are associated with earlier menarche and four are associated with later menarche. Menarche *P* values are derived from our stage 1 meta-analysis of 32 studies with genomic control applied to individual studies.
[a]Selected SNPs at each locus are those published for association with height (rather than those with the strongest signal for age at menarche). SNPs listed are those with a significant association ($P < 0.05$) with age at menarche. A full version of this table including SNPs associated with adult height but not reaching significance for menarche can be found in **Supplementary Table 14**.

Given the likely close biological interaction between the regulation of age at menarche and adiposity, we hypothesized that adipose tissue expressed SNPs (eSNPs) might show a preponderance of associations with age at menarche. Of the 5,184 adipose eSNPs identified in the Icelandic Family Adipose cohort[26], 23 were significantly associated with age at menarche after correction for multiple testing (using a $1/n$ P value threshold for 5,184 independent tests ($P < 1.9 \times 10^{-4}$)) (**Supplementary Table 11**). Of these adipose eSNPs, rs10835211 (menarche $P = 9.4 \times 10^{-6}$) is near *BDNF*, which is a BMI locus and is implicated in eating behavior and body weight regulation[27,28]. rs7160413 (menarche $P = 2.2 \times 10^{-5}$) is near *DLK1*, a gene implicated in early onset puberty[29]. rs133934508 (menarche $P = 3.6 \times 10^{-5}$) is associated with expression of *PITX1*, which encodes a pituitary transcriptional regulator[30].

Candidate gene assessment

Candidate gene studies for age at menarche have largely focused on genes involved in sex steroid-hormone biosynthesis and metabolism, highlighted through animal models or human cases with extreme delayed puberty or hypogonadotrophic hypogonadism[31]. We examined 8,770 SNPs in 16 candidate genes[31–33] and their surrounding regions (±300 kb) for association with age at menarche in our GWAS meta-analysis sample (**Supplementary Table 12**). SNPs in the regions of *TAC3R* (top hit, rs17034046, $P = 3.4 \times 10^{-7}$, ~19 kb upstream of *TAC3R*) and *ESR1* (top hit, rs9383922, $P = 2.2 \times 10^{-6}$, 110 kb upstream of *ESR1*) were significantly associated with age at menarche after correction for multiple testing (the Bonferroni threshold for 8,770 independent tests was $P < 5.7 \times 10^{-6}$). Rare deleterious mutations in *TAC3R*, encoding a receptor for Neurokinin B, and in its ligand *TAC3* have been found in families affected by hypogonadotropic hypogonadism and pubertal failure[31]. *ESR1* encodes an estrogen receptor that is essential for sexual development and reproductive function, and polymorphisms in *ESR1* have previously been nominally associated with age at menarche[33].

Overlapping heritability of body size and menarche timing

Family studies have suggested a substantial coinheritance of the timing of puberty and BMI[34], and this is supported by our finding of four established BMI variants among our new menarche loci. We therefore systematically assessed whether established loci for adiposity-related traits (BMI, waist-hip ratio (WHR) and obesity) and adult height were also associated with age at menarche. Nine of the twelve BMI loci and two of the four WHR loci tested were associated with age at menarche (**Table 2** and **Supplementary Table 13**). In all cases, the BMI- or WHR-increasing allele was associated with earlier menarche, which is consistent with the direction of association in epidemiological studies[35]. Eleven of the forty-four adult height loci were associated with age at menarche (**Table 3** and **Supplementary Table 14**). However, for seven of these loci, the adult height-increasing allele was associated with earlier menarche, which is in the opposite direction to the association in individual-level epidemiological studies[35].

We then assessed the relevance of our new menarche loci to adult BMI and height by exploring *in silico* data from the GIANT consortium. Nine of the forty-two menarche loci were associated with adult BMI (at $P < 0.05$; $N = 32,530$); in all cases, the allele associated with higher BMI was associated with earlier menarche (**Supplementary Table 15**). Eighteen of the menarche loci were associated with adult height (at $P < 0.05$; $N \sim 130,000$); although for three of these loci, the direction of effect was opposite to that predicted from epidemiological studies (**Supplementary Table 16**). Despite these joint associations with body size, in Avon Longitudinal Study of Parents and Children (ALSPAC) mothers, the combined influence of the menarche loci on age at menarche appeared to be completely unattenuated following adjustment for adult height and BMI (**Supplementary Table 17**), suggesting that in general, these menarche loci have direct effects on age at menarche. However, we acknowledge that further large studies with childhood growth data are needed to establish the causal directions of effect of these loci.

DISCUSSION

In a large GWAS meta-analysis comprising over 87,000 women, we identified 30 new loci for the timing of menarche and provide evidence for a further ten possible new loci. These loci were in or near genes associated with cellular development, body weight regulation, hormonal regulation and a wide variety of other biological functions. Previous studies comprising up to 17,510 women had detected only one or two genome-wide significant signals[8–11]. We now show that those earlier signals at *LIN28B* and 9q31.2 represented the 'low-hanging fruit' with particularly large effect sizes relative to their MAF (**Supplementary Fig. 5**). The list of functions of those genes nearest to the menarche loci (**Supplementary Table 7**) and the results of pathway analyses indicate a wide diversity of biological processes that regulate the timing of female pubertal maturation.

Among the confirmed new menarche loci were several loci implicated in body weight regulation, including four loci with established associations with BMI (in or near *FTO*, *SEC16B*, *TRA2B* and *TMEM18*). Furthermore, our systematic analysis of established BMI-related SNPs showed that the majority of alleles related to higher BMI and WHR also showed at least nominal associations with earlier menarche (**Table 2**). It is noteworthy that three new menarche loci are in or near genes implicated in energy homeostasis in animal models (*BSX*, *CRTC1* and *MCHR2*). In the GIANT consortium data, we did not detect any associations between these loci and adult BMI, however the *BSX* and *MCHR2* loci were nominally associated with adult height. In order to robustly investigate whether menarche loci have pleiotropic effects on growth or whether the association with menarche timing is driven through increased adiposity, measures of body fatness before menarche or even before the onset of puberty would be required but were unavailable in most studies. Further functional studies of these new menarche loci may also help to clarify the biological mechanisms linking these traits. In addition to influencing the timing of pubertal initiation, sufficient adiposity is also required for the maintenance of normal hypothalamic-pituitary-gonadal function through signaling by adipocytokines such as leptin[36]. Our pathway analyses highlighted coenzyme A and fatty acid biosynthesis as biological pathways related to menarche timing. Hypothalamic levels of long-chain fatty acyl coenzyme As have been shown to regulate rodent feeding behavior and glucose homeostasis[37], and genetic variants in this pathway could therefore potentially alter central nutrient sensing.

Earlier age at menarche is related to shorter adult stature in large epidemiological studies[35]. We found that several adult height-increasing alleles were also associated with age at menarche (**Table 3**), but at different loci, these alleles were associated with either earlier or later menarche. These paradoxical associations suggest a complex interplay between growth and pubertal timing. Earlier menarche is associated with taller, rather than shorter, childhood height, and there are likely separate causal effects of rapid linear growth on earlier puberty and of earlier pubertal maturation on earlier growth plate fusion and cessation of growth.

Although our pathway analyses strongly identified potential new biological pathways involved in pubertal timing, we acknowledge that the ability to assign putative functions to these menarche loci is substantially limited by the lack of identification of the causal variant at each locus. Many of the strongest associated SNPs were located

hundreds of kilobases distant to the nearest gene, and some menarche loci contained several plausible genes. Indeed, none of the top signals represented non-synonymous SNPs and only two SNPs were in LD with such variants (in *OLFM2* and *TRIM66*). Use of eQTLs helped to identify the likely causal genes (*GAB2*, *RBM6* and *NARG2*) at three menarche loci that spanned multiple genes. However, much future work will be required to identify the causal variants and implicated genes related to these menarche loci.

Despite the large size of our meta-analysis and the substantial increase in the number of menarche loci, these together explained between 3.6%–6.1% of the variance in age at menarche, equivalent to 7.2%–12.2% of its heritability. The majority of menarche loci had estimated effect sizes of between 2 and 3 weeks per allele. Assuming the presence of many true menarche SNPs with an effect size of 2 weeks per allele, even our large meta-analysis would only have had sufficient power to detect half of those SNPs with a MAF of 50% and only one in ten of those SNPs with MAF of 10% (**Supplementary Fig. 8**).

We corrected for population stratification by applying the genomic control method[38] to each of the individual study results. When we applied a more stringent second correction for the overall genomic control inflation factor across all 32 studies, 10 of the 40 possible new menarche variants fell below genome-wide significance (**Fig. 1** and **Table 1**). However, our subsequent finding of confirmatory evidence ($P < 0.05$) even in our limited replication studies for four of these ten variants (in or near *TRIM66*, *TMEM108*, *TMEM18* and *NFAT5*) suggests that the second correction for genomic control is likely to be overconservative.

Our identification of strong associations with SNPs near the candidate genes *TAC3R* and *ESR1* supports the likely presence of further menarche loci which did not meet the genome-wide significance threshold. Systematic assessment of functional genetic variants identified several further putative menarche loci. rs3101336, which tags a CNV near the BMI locus *NEGR1*, showed strong, but not genome-wide significant, association with age at menarche ($P = 3 \times 10^{-7}$). Exploration of adipose tissue eQTLs also identified further putative menarche loci related to genes implicated in eating behavior (*BDNF*), precocious puberty (*DLK1*) and pituitary function (*PITX1*). It has been suggested that lower levels of statistical significance may be applied to variants with prior biological candidacy, however this must be balanced against the desire to avoid false positives, and we suggest that these putative menarche loci require confirmation in further studies.

Notably all of the top menarche variants had MAF ≥ 7%. Although it has been suggested that low-frequency variants have larger effects than common variants[39], our study was clearly underpowered to detect low-frequency variants (MAF < 5%) with modest effect sizes. It is also possible that rare variants are not well captured using genome-wide chips. Future imputation using deep sequencing data from the 1000 Genomes Project may identify additional low frequency hits as well as refine the location of possible functional variants.

In the majority of studies contributing to this report, age at menarche was recalled several years later and often to the nearest completed whole year. Although recalled age at menarche is a valid measure[40] and is unlikely to show systematic bias by genotype, any nondifferential error would lead to reduced statistical power. Menarche indicates the completion of puberty in females, and it is unclear whether our new menarche loci also influence timing of other pubertal phenotypes. The known menarche locus in *LIN28B* was shown to also influence the onset of breast development in girls, the timing of pubic hair development and voice breaking in boys[9] and the timing of the pubertal growth spurt in both boys and girls[41]. Although our new menarche loci might also regulate such wider pubertal processes, it is plausible that some (for example, *INHBA*) might have sex-specific effects. Our study was restricted to cohorts of European ancestry and our results are therefore not generalized to other groups. African-American girls tend to show earlier pubertal maturation compared to girls of European ancestry[42], and genetic studies in such populations might reveal different menarche loci.

In summary, we identified at least 30 new loci for age at menarche. Our findings demonstrate the role of genes which regulate energy homeostasis and hormone pathways and illustrate the complexity of the regulation of the timing of puberty.

URLs. KBiosciences, http://www.kbioscience.co.uk; MACH, http://www.sph.umich.edu/csg/abecasis/MaCH/; METAL, www.sph.umich.edu/csg/abecasis/metal; mRNA by SNP Browser, http://www.sph.umich.edu/csg/liang/asthma/; MAGENTA, http://www.broadinstitute.org/mpg/magenta/; PANTHER, http://www.pantherdb.org/.

METHODS

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

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

ACKNOWLEDGMENTS

Academy of Finland (Finnish Centre of Excellence in Complex Disease Genetics 129680, 120315, 129287, 129494); Affymetrix (N02-HL-6-4278); Agency of Science, Technology and Research of Singapore (A*STAR); Althingi (the Icelandic Parliament); American Heart Association (0855082E); Amgen; Augustinus Foundation; Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498, 619667); Australian Research Council (A7960034, A79801419, A79906588, DP0212016, DP0343921, DP0770096); Baltimore Veterans Administration Medical Center Geriatrics Research; Canadian Institutes of Health Research (grant ID 166067); Cancer Research United Kingdom; the Cariplo Foundation; Center for Disease Control and Prevention (USA), Centre for Medical Systems Biology (CMSB); the Chief Scientist Office of the Scottish Government; Clinical Nutrition Research Unit of Maryland (P30 DK072488); Danish National Research Foundation; Danish Pharmacists' Fund; the Egmont Foundation; Erasmus Medical Center; Erasmus University; Estonian Government (SF0180142s08); the European Commision (212111 BBMRI, 205419 ECOGENE, 201413 ENGAGE, HEALTH-F2-2008-201865 - GEFOS, HEALTH-F2-2008-35627, HEALTH-F4-2007-201413, HEALTH-F4-2007-201550 HYPERGENES, 245536 OPENGENE, TREAT-OA, GenomEUtwin Project QLG2-CT-2002-01254, EU/QLRT-2001-01254; ERC-230374); European Union framework program 6 EUROSPAN project (LSHG-CT-2006-018947, LSHM-CT-2003-503041); European Regional Development Fund; Faculty of Biology and Medicine of Lausanne, Switzerland; Fondazione Compagnia di San Paolo Health Ministry, Framingham Heart Study (N01-HC-25195); GENEVA Coordinating Center (U01HG004446); German Federal Ministry of Education and Research; German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823); Giorgi-Cavaglieri Foundation; GlaxoSmithKline; Health Fund of the Danish Health Insurance Societies; Helmholtz Zentrum München-German Research Center for Environmental Health; Hjartavernd (the Icelandic Heart Association); Italian Ministry of Health (ICS110.1/RF97.71, RF-FSR-2007-647201); Juvenile Diabetes Research Foundation International (JDRF); Leenaards Foundation; March of Dimes Birth Defects Foundation; the Medical Research Council (G0000934, G0500539, G0701863); National Cancer Institute (CA40356, CA98233, P01CA087969, P01CA055075, CA047988, CA63464, CA54281, CA136792, CA089392, CA104021); Munich Center of Health Sciences (MC Health, LMUinnovativ); National Health and Medical Research Council of Australia (572613 and 003209); National Heart, Lung, and Blood Institute (HL 043851, HL087679, HL69757, N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367, R01HL086694, RC2 HL102419); National Human Genome Research Institute (NHGRI); National Institute of Aging (263 MD 9164, 263 MD 821336, N.1-AG-1-1, N.1-AG-1-2111, N01-AG-5-0002, AG-16592, Genetics of Reproductive Life Period and Health Outcomes - R21AG032598); National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Institute on Aging ((NIAMS/NIA) R01 AR/AG 41398); National Institute of Child Health and Human Development (HD-061437); National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, R01DK058845, U01 DK062418); National Institute on Alcohol Abuse and Alcoholism (NIAAA); National Institute of Allergy and Infectious Diseases (NIAID); National Institute on

Drug Abuse (NIDA); National Institutes of Health (AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, HHSN268200782096C, M01 RR-00750, MH66206, N01-AG-12100, N01-AG-1-2109, P01-AG-18397, R01-088119, R01-DA013423, RFAHG006033, U01DE018993, U01DE018903, U01HG004422, U01HG004446, U01HL72515, U01 HL84756, U01HG004399, U01HG004402, U01HG004415, U01HG004423, U01HG004436, U01HG004438, U01HG004402, U01HG004446, U01HG004726, U01HG004728, U01HG004729, U01HG004735, U01HG004738, U01HG04424, U10AA008401, U54RR025204-01, UL1RR025005, UL1 RR025774, Z01CP010200, HHSN268200625226C); the National Institute for Health Research Cambridge Biomedical Research Centre; Netherlands Organization for Scientific Research (904-61-193, 575-25-006, 480-04-004, 56-464-14192, NWO 480-05-003, 175.010.2005.011, 911-03-012, 014-93-015); Netherlands Genomic Initiative (NGI, 050-060-810), National Institute on Aging (NIA) Intramural Research Program; and Laboratory of Neurogenetics; US National Institutes of Health Genes, Environment and Health Initiative; Republic of Croatia Ministry of Science, Education and Sports (108-1080315-0302); Robert Dawson Evans Endowment; the Royal Society, Royal Swedish Academy of Science; State of Bavaria, Germany; Susan G. Komen Breast Cancer Foundation; Swedish Foundation for Strategic Research (SSF); Swedish Heart-Lung Foundation; Swedish Ministry of Higher Education and Research; Swedish Research Council; Swedish Society of Medicine; Swiss National Science Foundation (Ref: 33CSCO-122661, 3100AO-116323/1); the University of Bristol, University of Maryland General Clinical Research Center (M01 RR 16500); the Wellcome Trust (068545/Z/02, 076467/Z/05/Z, 077016/Z/05/Z, 89061/Z/09/Z, strategic award 079895); Western Australian DNA Bank; Western Australian Genetic Epidemiology Resource (see **Supplementary Note** for extended acknowledgments).

AUTHOR CONTRIBUTIONS

Writing group: D.I.B., H.A.B., D.I.C., D.L.C., L.C., E.W.D., M.J.E., C.E.E., T.E., B.F., N.F., F.G., D.F.G., C. He, T.B.H., D.J.H., D.L.K., K.K.L., M. Mangino, M. Marongiu, J.M.M., A. Metspalu, A. Murray, K.K.O., J.R.B.P., L.S., E.A.S., P.S., U.T., A.G.U., C.M.v.D., S.H.v.W., J.A.V., E.W., G.Z.

Analysis working group: C.E.E., A. Murray, K.K.O., J.R.B.P., P.S.

Secondary analyses: J.A.V., J.R.B.P., A.D.J., D.L., A.S.P., V.E., A.V. Segrè, The GIANT Consortium.

Oversight of contributing cohorts: D.I.B., H.A.B., L.J.B., L.C., C.M.v.D., D.F.E., M.J.E., U.d.F., P.G., A.H., D.J.H., P.H., T.B.H., K.-T.K. M. Marongiu, A. Metspalu, A. Murray, G.W.M., J.M.M., S.S.M., V.M., C.E.P., P.P., I.R., P.M.R., E.A.S., D.P.S., G.D.S., K.S., N.J.S., T.D.S., D.T., A.G.U., A.F.W., D.M.W., E.W., H.E.W., J.F.W., N.J.W.

Analytical lead of contributing studies: E.A., E.M.B., D.I.C., D.L.C., T.C., C.E.E., T.E., B.F., N.F., D.F.G., C. Hayward, C. He, J.-J.H., E.I., D.L.K., Z.K., K.L.L., P.L., R.J.F.L., M. Mangino, P.F.M., P.K.E.M., K.N., E.P., J.R.B.P., A.V. Smith, E.N.S., L.S., P.S., S.S., S.U., J.A.V., N.M.W., S.H.v.W., J.H.Z., L.Z.

Genotyping and phenotyping of contributing studies: H.A., N.A., P. d'Adamo, T.A., G.S.B., H.B., I.B., E.B., F.B., J.E.B., S. Bandinelli, S. Bergmann, A.D.C., D.C., H.C., M.C.C., S.J.C., W.C., A.D., R.M.v.D., P.D., G.E., J.E., A.C.H., S.E.H., A.R.F., C.G., L. Ferreli, L. Ferrucci, T.F., F.G., E.J.C.d.G., M.G., V.G., D.G.H., F.B.H., T.I., M.-R.J., D.K., D.P.K., I.K., P.K., T.O.K., J.L., J.S.E.L., L.J.L., S.L., J.B.J.v.M., J.C.M., J.M.M., M. Melbye, N.G.M., W.L.M., A.R.N., F.N., M.N., M.A.N., P.N., K.K.O., B.A.O., A. Palotie, A. Pouta, A.N.P., C.E.P., G.P., L.P., L.J.P., M.P., N.L.P., O.P., F.R., J.P.R., S.M.R., T.R., A.S., A.R.S., C.S., D.S., N.S., S.R.S., U.S., V.S., E.T., K.T., L.T., M.-L.T., J.T., M.U., P.V., G. Waeber, G. Willemsen, M.N.W., L.Y., G.Z., W.V.Z.

COMPETING FINANCIAL INTERESTS

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

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Parent, A.S. *et al.* The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24**, 668–693 (2003).
2. Kvale, G. Reproductive factors in breast cancer epidemiology. *Acta Oncol.* **31**, 187–194 (1992).
3. Purdie, D.M. & Green, A.C. Epidemiology of endometrial cancer. *Best Pract. Res. Clin. Obstet. Gynaecol.* **15**, 341–354 (2001).
4. Ong, K.K. *et al.* Earlier mother's age at menarche predicts rapid infancy growth and childhood obesity. *PLoS Med.* **4**, e132 (2007).
5. He, C. *et al.* Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. *Am. J. Epidemiol.* **171**, 334–344 (2010).
6. Lakshman, R. *et al.* Early age at menarche associated with cardiovascular disease and mortality. *J. Clin. Endocrinol. Metab.* **94**, 4953–4960 (2009).
7. Towne, B. *et al.* Heritability of age at menarche in girls from the Fels Longitudinal Study. *Am. J. Phys. Anthropol.* **128**, 210–219 (2005).
8. He, C. *et al.* Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.* **41**, 724–728 (2009).
9. Ong, K.K. *et al.* Genetic variation in *LIN28B* is associated with the timing of puberty. *Nat. Genet.* **41**, 729–733 (2009).
10. Perry, J.R. *et al.* Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.* **41**, 648–650 (2009).
11. Sulem, P. *et al.* Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat. Genet.* **41**, 734–738 (2009).
12. Raivio, T. & Dunkel, L. Inhibins in childhood and puberty. *Best Pract. Res. Clin. Endocrinol. Metab.* **16**, 43–52 (2002).
13. Crofton, P.M. *et al.* Changes in dimeric inhibin A and B during normal early puberty in boys and girls. *Clin. Endocrinol.* **46**, 109–114 (1997).
14. Sehested, A. *et al.* Serum inhibin A and inhibin B in healthy prepubertal, pubertal, and adolescent girls and adult women: relation to age, stage of puberty, menstrual cycle, follicle-stimulating hormone, luteinizing hormone, and estradiol levels. *J. Clin. Endocrinol. Metab.* **85**, 1634–1640 (2000).
15. Burger, H.G. Evidence for a negative feedback role of inhibin in follicle stimulating hormone regulation in women. *Hum. Reprod.* **8**(Suppl 2), 129–132 (1993).
16. Sulzbacher, S., Schroeder, I.S., Truong, T.T. & Wobus, A.M. Activin A-induced differentiation of embryonic stem cells into endoderm and pancreatic progenitors—the influence of differentiation factors and culture conditions. *Stem Cell Rev.* **5**, 159–173 (2009).
17. Dolle, P. Developmental expression of retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7**, e006 (2009).
18. Frayling, T.M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
19. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
20. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
21. Sakkou, M. *et al.* A role for brain-specific homeobox factor Bsx in the control of hyperphagia and locomotory behavior. *Cell Metab.* **5**, 450–463 (2007).
22. Altarejos, J.Y. *et al.* The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat. Med.* **14**, 1112–1117 (2008).
23. Pissios, P., Bradley, R.L. & Maratos-Flier, E. Expanding the scales: The multiple roles of MCH in regulating energy balance and other biological functions. *Endocr. Rev.* **27**, 606–620 (2006).
24. Wu, M., Dumalska, I., Morozova, E., van den Pol, A. & Alreja, M. Melanin-concentrating hormone directly inhibits GnRH neurons and blocks kisspeptin activation, linking energy balance to reproduction. *Proc. Natl. Acad. Sci. USA* **106**, 17217–17222 (2009).
25. Conrad, D.F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2009).
26. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
27. Kernie, S.G., Liebl, D.J. & Parada, L.F. *BDNF* regulates eating behavior and locomotor activity in mice. *EMBO J.* **19**, 1290–1300 (2000).
28. Xu, B. *et al.* Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. *Nat. Neurosci.* **6**, 736–742 (2003).
29. Temple, I.K., Shrubb, V., Lever, M., Bullman, H. & Mackay, D.J. Isolated imprinting mutation of the *DLK1/GTL2* locus associated with a clinical presentation of maternal uniparental disomy of chromosome 14. *J. Med. Genet.* **44**, 637–640 (2007).
30. Drouin, J., Lamolet, B., Lamonerie, T., Lanctot, C. & Tremblay, J.J. The PTX family of homeodomain transcription factors during pituitary developments. *Mol. Cell. Endocrinol.* **140**, 31–36 (1998).
31. Topaloglu, A.K. *et al. TAC3* and *TACR3* mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat. Genet.* **41**, 354–358 (2009).
32. Gajdos, Z.K. *et al.* Association studies of common variants in 10 hypogonadotropic hypogonadism genes with age at menarche. *J. Clin. Endocrinol. Metab.* **93**, 4290–4298 (2008).
33. Stavrou, I., Zois, C., Ioannidis, J.P. & Tsatsoulis, A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. *Hum. Reprod.* **17**, 1101–1105 (2002).
34. Kaprio, J. *et al.* Common genetic influences on BMI and age at menarche. *Hum. Biol.* **67**, 739–753 (1995).
35. Onland-Moret, N.C. *et al.* Age at menarche in relation to adult height: the EPIC study. *Am. J. Epidemiol.* **162**, 623–632 (2005).
36. Welt, C.K. *et al.* Recombinant human leptin in women with hypothalamic amenorrhea. *N. Engl. J. Med.* **351**, 987–997 (2004).
37. Pocai, A. *et al.* Restoration of hypothalamic lipid sensing normalizes energy and glucose homeostasis in overfed rats. *J. Clin. Invest.* **116**, 1081–1091 (2006).
38. Devlin, B., Bacanu, S.A. & Roeder, K. Genomic Control to the extreme. *Nat. Genet.* **36**, 1129–1130 author reply 1131 (2004).
39. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
40. 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**, 672–679 (2002).
41. Widén, E. *et al.* Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am. J. Hum. Genet.* **86**, 773–782 (2010).
42. Herman-Giddens, M.E. *et al.* Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics* **99**, 505–512 (1997).

Cathy E Elks[1,106], John R B Perry[2,106], Patrick Sulem[3,106], Daniel I Chasman[4,5], Nora Franceschini[6], Chunyan He[7,8], Kathryn L Lunetta[9,10], Jenny A Visser[11], Enda M Byrne[12,13], Diana L Cousminer[14], Daniel F Gudbjartsson[3], Tõnu Esko[15–17], Bjarke Feenstra[18], Jouke-Jan Hottenga[19], Daniel L Koller[20], Zoltán Kutalik[21,22], Peng Lin[23], Massimo Mangino[24], Mara Marongiu[25], Patrick F McArdle[26], Albert V Smith[27,28], Lisette Stolk[11,29], Sophie H van Wingerden[30], Jing Hua Zhao[1], Eva Albrecht[31], Tanguy Corre[32], Erik Ingelsson[33], Caroline Hayward[34], Patrik K E Magnusson[33], Erin N Smith[35–37], Shelia Ulivi[38], Nicole M Warrington[39], Lina Zgaga[40], Helen Alavere[15], Najaf Amin[30], Thor Aspelund[27,28], Stefania Bandinelli[41], Inês Barroso[42], Gerald S Berenson[43], Sven Bergmann[21,22], Hannah Blackburn[42], Eric Boerwinkle[44], Julie E Buring[4,5,45], Fabio Busonero[25], Harry Campbell[40], Stephen J Chanock[46], Wei Chen[43], Marilyn C Cornelis[47], David Couper[48], Andrea D Coviello[49], Pio d'Adamo[38], Ulf de Faire[50], Eco J C de Geus[19], Panos Deloukas[42], Angela Döring[31], George Davey Smith[51], Douglas F Easton[52,53], Gudny Eiriksdottir[27], Valur Emilsson[54], Johan Eriksson[55–58], Luigi Ferrucci[59], Aaron R Folsom[60], Tatiana Foroud[20], Melissa Garcia[61], Paolo Gasparini[38], Frank Geller[18], Christian Gieger[31], The GIANT Consortium[62], Vilmundur Gudnason[27,28], Per Hall[33], Susan E Hankinson[45,63], Liana Ferreli[25], Andrew C Heath[64], Dena G Hernandez[65], Albert Hofman[29,66], Frank B Hu[45,47,63], Thomas Illig[31], Marjo-Riitta Järvelin[67], Andrew D Johnson[9,68], David Karasik[69], Kay-Tee Khaw[70], Douglas P Kiel[69], Tuomas O Kilpeläinen[1], Ivana Kolcic[71], Peter Kraft[45,47,63], Lenore J Launer[61], Joop S E Laven[72], Shengxu Li[1], Jianjun Liu[73], Daniel Levy[9,68,74], Nicholas G Martin[75], Wendy L McArdle[76], Mads Melbye[18], Vincent Mooser[77], Jeffrey C Murray[78], Sarah S Murray[35–37], Michael A Nalls[79], Pau Navarro[34], Mari Nelis[15–17], Andrew R Ness[80], Kate Northstone[76], Ben A Oostra[30,81], Munro Peacock[82], Lyle J Palmer[39], Aarno Palotie[14,42,83], Guillaume Paré[4,5,84], Alex N Parker[85], Nancy L Pedersen[33], Leena Peltonen[14,42,55,83,105], Craig E Pennell[86], Paul Pharoah[52,53], Ozren Polasek[71,87], Andrew S Plump[88], Anneli Pouta[55], Eleonora Porcu[25], Thorunn Rafnar[3], John P Rice[23], Susan M Ring[76], Fernando Rivadeneira[11,29,66], Igor Rudan[40,89], Cinzia Sala[32], Veikko Salomaa[55], Serena Sanna[25], David Schlessinger[90], Nicholas J Schork[35–37], Angelo Scuteri[25,91], Ayellet V Segrè[83,92], Alan R Shuldiner[26,93], Nicole Soranzo[24,42], Ulla Sovio[67], Sathanur R Srinivasan[43], David P Strachan[94], Mar-Liis Tammesoo[15], Emmi Tikkanen[14,55], Daniela Toniolo[32], Kim Tsui[85], Laufey Tryggvadottir[95], Jonathon Tyrer[52,53], Manuela Uda[25], Rob M van Dam[47,96,97], Joyce B J van Meurs[11,29], Peter Vollenweider[98], Gerard Waeber[98], Nicholas J Wareham[1], Dawn M Waterworth[77], Michael N Weedon[2], H Erich Wichmann[31,99,100], Gonneke Willemsen[19], James F Wilson[40], Alan F Wright[34], Lauren Young[85], Guangju Zhai[24], Wei Vivian Zhuang[10], Laura J Bierut[23], Dorret I Boomsma[19], Heather A Boyd[18], Laura Crisponi[25], Ellen W Demerath[60], Cornelia M van Duijn[29,30], Michael J Econs[20,82], Tamara B Harris[61], David J Hunter[45–47,63], Ruth J F Loos[1], Andres Metspalu[15–17], Grant W Montgomery[101], Paul M Ridker[4,5,45,102], Tim D Spector[24], Elizabeth A Streeten[26], Kari Stefansson[3,103], Unnur Thorsteinsdottir[3,103], André G Uitterlinden[11,29,66], Elisabeth Widen[14], Joanne M Murabito[9,49,107], Ken K Ong[1,104,107] & Anna Murray[2,107]

[1]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [2]Genetics of Complex Traits, Peninsula Medical School, University of Exeter, UK. [3]deCODE Genetics, Reykjavik, Iceland. [4]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [5]Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [7]Department of Public Health, Indiana University School of Medicine, Indiana, USA. [8]Melvin and Bren Simon Cancer Center, Indiana University, Indiana, USA. [9]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [10]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [11]Department of Internal Medicine, Erasmus Medical Center (MC), Rotterdam, The Netherlands. [12]Queensland Statistical Genetics, Queensland Institute of Medical Research, Brisbane, Australia. [13]The University of Queensland, Brisbane, Australia. [14]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland. [15]Estonian Genome Center, University of Tartu, Tartu, Estonia. [16]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [17]Genotyping Core Facility, Estonian Biocenter, Tartu, Estonia. [18]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [19]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. [20]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [21]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [22]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [23]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [24]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [25]Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy. [26]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [27]Icelandic Heart Association, Kopavogur, Iceland. [28]University of Iceland, Reykjavik, Iceland. [29]Netherlands Consortium of Healthy Aging, Rotterdam, The Netherlands. [30]Genetic-Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [31]Institute of Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany. [32]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy. [33]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [34]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK. [35]Scripps Genomic Medicine, La Jolla, California, USA. [36]The Scripps Translational Science Institute, La Jolla, California, USA. [37]The Scripps Research Institute, La Jolla, California, USA. [38]Medical Genetics, Department of Reproductive Sciences and Development, University of Trieste, Trieste, Italy. [39]Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Australia. [40]Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, Scotland. [41]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy. [42]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [43]Tulane University, New Orleans, Louisiana, USA. [44]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [45]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [46]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA. [47]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts,

USA. [48]Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [49]Sections of General Internal Medicine, Preventive Medicine and Endocrinology, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [50]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [51]MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol, UK. [52]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [53]Department of Oncology, Strangeways Research Laboratories, University of Cambridge, Cambridge, UK. [54]Molecular Programming and Research Informatics (MPRI), Merck and Co., Inc., Rahway, New Jersey, USA. [55]National Institute for Health and Welfare, Helsinki, Finland. [56]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [57]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland. [58]Folkhalsan Research Centre, Helsinki, Finland. [59]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA. [60]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA. [61]Laboratory of Epidemiology, Demography and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [62]A full list of members is provided in the **Supplementary Note**. [63]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA. [64]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [65]Laboratory of Neurogenetics, National Institute of Aging, Bethesda, Maryland, USA. [66]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [67]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [68]National Heart, Lung, and Blood Institute (NHLBI) Center for Population Studies, Bethesda, Maryland, USA. [69]Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, USA. [70]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK. [71]Medical School, University of Zagreb, Zagreb, Croatia. [72]Department of Obstetrics and Gynaecology, Erasmus, MC, Rotterdam, The Netherlands. [73]Human Genetics, Genome Institute of Singapore, Singapore. [74]Division of Cardiology, Boston University School of Medicine, Boston, Massachusetts, USA. [75]Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [76]Avon Longitudinal Study of Parents and Children (ALSPAC), Department of Social Medicine, University of Bristol, Bristol, UK. [77]Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania, USA. [78]Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA. [79]Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [80]Department of Oral and Dental Science, University of Bristol, Bristol, UK. [81]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. [82]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA. [83]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [84]Genetic and Molecular Epidemiology Laboratory, McMaster University, Hamilton, Ontario, Canada. [85]Amgen, Cambridge, Massachusetts, USA. [86]School of Women's and Infants' Health, The University of Western Australia, Crawley, Australia. [87]Gen-Info Ltd., Zagreb, Croatia. [88]Cardiovascular Disease, Merck Research Laboratory, Rahway, New Jersey, USA. [89]Croatian Centre for Global Health, University of Split Medical School, Split, Croatia. [90]Gerontology Research Center, National Institute on Aging, Baltimore, Maryland, USA. [91]Unità Operativa Geriatria-Istituto Nazionale Ricovero e Cura per Anziani (INRCA), IRCCS, Rome, Italy. [92]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [93]Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland, USA. [94]Division of Community Health Sciences, St. George's, University of London, London, UK. [95]Icelandic Cancer Registry, Reykjavik, Iceland. [96]Department of Epidemiology and Public Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [97]Department of Medicine, National University of Singapore, Singapore. [98]Department of Internal Medicine, BH-10 Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland. [99]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [100]Klinikum Grosshadern, Munich, Germany. [101]Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [102]Division of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [103]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [104]Department of Paediatrics, University of Cambridge, Cambridge, UK. [105]Deceased. [106]These authors contributed equally to this work. [107]These senior authors jointly supervised this work. Correspondence should be addressed to A. Murray (anna.murray@pms.ac.uk), K.K.O. (ken.ong@mrc-epid.cam.ac.uk) or J.M.M. (murabito@bu.edu).

ONLINE METHODS

Stage 1 GWAS populations. Thirty-two studies contributed to the stage 1 GWAS meta-analysis, comprising 87,802 women of European ancestry. The consortium was made up of populations from the Age, Gene/Environment Susceptibility Study (AGES, $n = 1849$), the Amish population (Amish, $n = 557$), the Atherosclerosis Risk in Communities study (ARIC, $n = 4247$), the British 1958 Birth Cohort (B58C-T1DGC and B58C-WTCCC, $n = 1584$), CoLaus ($n = 2797$), deCODE ($n = 15,864$), the Danish National Birth Cohort (DNBC, $n = 1748$), the Estonian Genome Center, University of Tartu (EGCUT, $n = 987$), the European Prospective Investigation into Cancer and Nutrition (EPIC-obesity cases and cohort, $n = 1840$), the Erasmus Rucphen Family Study (ERF, $n = 1103$), the Framingham Heart Study (FHS, $n = 3801$), the Helsinki Birth Cohort (HBCS, $n = 976$), the Health 2000 study (Health 2000 cases and controls, $n = 922$), InCHIANTI ($n = 597$), the Indiana University premenopausal Caucasian women peak BMD study (Indiana, $n = 1497$), the Nurse's Heath Studies (NHS, $n = 5360$), the Northern Finland Birth cohort (NFBC, $n = 2648$), the Netherlands Twin Register (NTR, $n = 1051$), the Queensland Institute of Medical Research (QIMR, $n = 3528$), the Rotterdam studies (RS1, RS2 and RS3, $n = 5406$), the Study of Addiction: Genetics and Environment (SAGE, $n = 1376$), the SardiNIA study ($n = 2158$), Twins UK I, II and III ($n = 3962$), and the Women's Genome Health Study (WGHS, $n = 22,028$). Full details can be found in the **Supplementary Note**. All studies were approved by local ethics committees and all participants provided written informed consent.

Phenotype measurement and inclusion criteria. Age at menarche recalled by the participant was recorded in each study. Specific questions asked can be found in **Supplementary Table 1**. Only women of European ancestry with a valid age at menarche between 9 and 17 years were included in this analysis, as this represents the normal physiological range. Information on birth year was also collected in each study.

Genotyping. The 32 stage 1 studies were genotyped using a variety of Affymetrix (6.0, GeneChip 500K, 250K, MIP50K and 10K) and Illumina (HumanHap 550K, 318K, HumanHap 300K, HumanHap 370K CNV, HumanHap610 quad, Human660W-Quad BeadChip, 6K and Human 1Mv1_C) genotyping arrays. Genotyping call rate cutoffs were at least 90%, and SNPs were filtered for those with a minor allele frequency of greater than 1%. More details on the filtering criteria for genotypes in each individual study can be found in **Supplementary Table 2**.

Genotype imputation. In order to increase genomic coverage and allow the evaluation of the same SNPs across as many study populations as possible, each study imputed genotype data based on the HapMap CEU Build 35 or 36. Algorithms were used to infer unobserved genotypes in a probabilistic manner in either MACH, IMPUTE[43], or software that was developed by the researchers. As a quality control measure, we excluded non-genotyped SNPs with an imputation quality less than 0.3 (for observed versus expected variance in MACH) or 0.4 (for IMPUTE's proper info statistic) from the meta-analysis.

Association testing. Each study performed genome-wide association testing for age at menarche across approximately 2.5 million SNPs based on linear regression under an additive genetic model. Analyses were adjusted for birth year in order to remove the effect of the temporal decline in age at menarche. Studies used PLINK, ProABEL, MACH2QTL, SNPTEST, R packages or MERLIN-fastassoc for the association testing. The results from individual studies were corrected by their respective genomic inflation factors (λ) (**Supplementary Table 1**) according to the genomic control method to correct for population stratification[38].

Meta-analysis. We used an inverse-variance meta-analysis to test the effects of each genetic variant on age at menarche across the 32 studies. Fixed effects models were used, although in the absence of significant heterogeneity, choice of model has little impact on the results. In order to correct for potential relatedness between two Icelandic cohorts (AGES and deCODE), the corrected association results for these cohorts were first meta-analyzed and the genomic-control method was reapplied to the results of the combined sample. These results were then meta-analyzed with the remaining 30 studies. We also displayed further results following a second correction for genomic control using the overall genomic inflation factor calculated from the meta-analysis of all 32 studies. All meta-analyses were conducted using the METAL software package. We considered P values $< 5 \times 10^{-8}$ to indicate genome-wide significance.

We also meta-analyzed results from X-chromosome SNPs in a subset of studies with this data available. This included seven imputed datasets and one directly genotyped dataset. Total sample size was ~60% of the autosomal meta-analysis ($N = 52,781$) and the same statistical model was tested.

Conditional analysis. In order to establish whether genome-wide significant SNPs with low LD in the same chromosomal region (defined as $r^2 < 0.05$ in a 750-kb region) were independent loci, we carried out a conditional analysis. Each study performed a genome-wide analysis for age at menarche using linear regression adjusting for the top signal at each of the 42 associated regions to determine whether potential second signals remained significant even after adjusting for these variants. Birth year was also included as a covariate. Results from each individual study were meta-analyzed to determine whether these potential second signals were truly independent (that is, if $P < 5.0 \times 10^{-8}$).

Replication studies. In order to confirm our possible new menarche loci, we tested our 42 top hits for *in silico* association with age at menarche in 8,669 women from 16 studies with GWAS data and which were not included in the first stage meta-analysis (**Supplementary Table 4**). In addition, new genotype data was generated for 30 of the 42 menarche loci and tested for association with age at menarche in up to 6,118 women from the Avon Longitudinal Study of Parents and Children (ALSPAC). Genotyping was performed by KBiosciences (Hoddesdon, UK) using their own unique system of fluorescence-based competitive allele-specific PCR (KASPar). As in stage 1, analyses were restricted to women reporting age at menarche between 9 and 17 years and adjustment was made for birth year. Mean age at menarche ranged from 12.4 to 13.5 years, consistent with studies in the stage 1 meta-analysis. Linear regression was used to test the association between each variant and age at menarche in an additive genetic model. These results were then meta-analyzed with genomic control–adjusted statistics from our stage 1 meta-analysis using inverse-variance fixed effects models.

In order to calculate the overall variance explained by these menarche loci in each of the replication cohorts, we calculated the r^2 value from a model including all 42 known, confirmed and possible new menarche variants and birth year and compared this to a model including birth year alone. We only included cohorts with >800 women in their full model analyses, as sample sizes smaller than this may give spurious results.

Pathway analysis. Ingenuity pathway analysis (IPA) Knowledge Base 8.5 (Ingenuity Systems) was used to explore the functional relationship between proteins encoded by the 42 known, confirmed and possible new menarche loci. The IPA Knowledge Base contains millions of findings curated from the literature. Genes or nearest genes to the 42 loci (**Table 1**) were entered into the Ingenuity database. These 'focus genes' were analyzed for direct interactions only. Networks were generated with a maximum size of 35 genes and shown as graphical representations of the molecular relationships between genes and gene products. Proteins are depicted as nodes in various shapes representing the functional class of the protein. The biological relationships between nodes are depicted by lines. To determine the probability of the analyzed gene to be found together in a network from Ingenuity Pathways Knowledge Base due to random chance alone, IPA applies a Fisher's exact test. The network score or P value represents the significance of the focus gene enrichment. There are 25 diseases and disorders categories and 32 molecular and cellular function categories in the IPA Knowledge Base. Enrichment of focus genes to these diseases and functional categories was also evaluated. The P value, based on a right-tailed Fisher's exact test, considers the number of identified focus genes and the total number of molecules known to be associated with these categories in the IPA knowledge database.

MAGENTA was used to explore pathway-based associations in the full GWAS dataset. MAGENTA implements a GSEA-based approach, the methodology of which has been previously described[44]. Briefly, each gene in the genome is mapped to a single index SNP with the lowest P value within a 110 kb upstream, 40 kb downstream window. This P value, representing a gene score, is then corrected for confounding factors such as gene size, SNP density and

doi:10.1038/ng.714

LD-related properties in a regression model. Genes within the HLA region were excluded from analysis due to difficulties in accounting for gene density and LD patterns. Each mapped gene in the genome is then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, is calculated. This observed statistic is then compared to 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA *P* value for each pathway. Significance was determined when an individual pathway reached a false discovery rate < 0.05 in either analysis (**Supplementary Table 9**). In total, 2,529 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with age at menarche.

eQTLs. We tested the association between 5,184 adipose tissue eSNPs identified in the Icelandic Family Adipose (IFA) cohort ($n = 673$) with age at menarche in our stage 1 meta-analysis sample. The IFA cohort dataset included the expression of 23,720 transcripts representing 84% of the 20,060 protein-coding genes annotated in the Ensembl database (v 33)[26].

43. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
44. 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).