LETTER

doi:10.1038/nature13545

Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche

A list of authors and their affiliations appears at the end of the paper

Age at menarche is a marker of timing of puberty in females. It varies widely between individuals, is a heritable trait and is associated with risks for obesity, type 2 diabetes, cardiovascular disease, breast cancer and all-cause mortality[1]. Studies of rare human disorders of puberty and animal models point to a complex hypothalamic-pituitary-hormonal regulation[2,3], but the mechanisms that determine pubertal timing and underlie its links to disease risk remain unclear. Here, using genome-wide and custom-genotyping arrays in up to 182,416 women of European descent from 57 studies, we found robust evidence ($P < 5 \times 10^{-8}$) for 123 signals at 106 genomic loci associated with age at menarche. Many loci were associated with other pubertal traits in both sexes, and there was substantial overlap with genes implicated in body mass index and various diseases, including rare disorders of puberty. Menarche signals were enriched in imprinted regions, with three loci (*DLK1-WDR25*, *MKRN3-MAGEL2* and *KCNK9*) demonstrating parent-of-origin-specific associations concordant with known parental expression patterns. Pathway analyses implicated nuclear hormone receptors, particularly retinoic acid and γ-aminobutyric acid-B2 receptor signalling, among novel mechanisms that regulate pubertal timing in humans. Our findings suggest a genetic architecture involving at least hundreds of common variants in the coordinated timing of the pubertal transition.

Genome-wide array data were available from up to 132,989 women of European descent from 57 studies. In a further 49,427 women, data were available on up to approximately 25,000 single nucleotide polymorphisms (SNPs), or their proxy markers, that showed sub-genome-wide significant associations ($P < 0.0022$) with age at menarche in our previous genome-wide association study (GWAS)[4] (Supplementary Table 1). Association statistics for 2,441,815 autosomal SNPs that passed quality control measures (including minor allele frequency >1%) were combined across all studies by meta-analysis.

3,915 SNPs reached the genome-wide significance threshold ($P < 5 \times 10^{-8}$) for association with age at menarche (Fig. 1). Using GCTA[5], which approximates a conditional analysis adjusted for the effects of neighbouring SNPs (Extended Data Fig. 1 and Supplementary Table 2), we identified 123 independent signals for age at menarche at 106 genomic loci, including 11 loci containing multiple independent signals (Extended Data Tables 1–4; plots of all loci are available at http://www.reprogen.org). Of the 42 previously reported independent signals for age at menarche[4], all but one (gene *SLC14A2*, SNP variation rs2243803, $P = 2.3 \times 10^{-6}$) remained significant genome-wide in the expanded data set.

To estimate their overall contribution to the variation in age at menarche, we analysed an additional sample of 8,689 women. 104/123 signals showed directionally concordant associations or trends with menarche timing (binomial sign test $P_{Sign} = 2.2 \times 10^{-15}$), of which 35 showed nominal significance ($P_{Sign} < 0.05$) (Supplementary Table 3). In this independent sample, the top 123 SNPs together explained 2.71% ($P < 1 \times 10^{-20}$) of the variance in age at menarche, compared to 1.31% ($P = 2.3 \times 10^{-14}$) explained by the previously reported 42 SNPs. Consideration of further SNPs with lower levels of significance resulted in modest increases in the estimated variance explained with increasingly larger SNP sets, until we included all autosomal SNPs (15.8%, s.e. 3.6%,

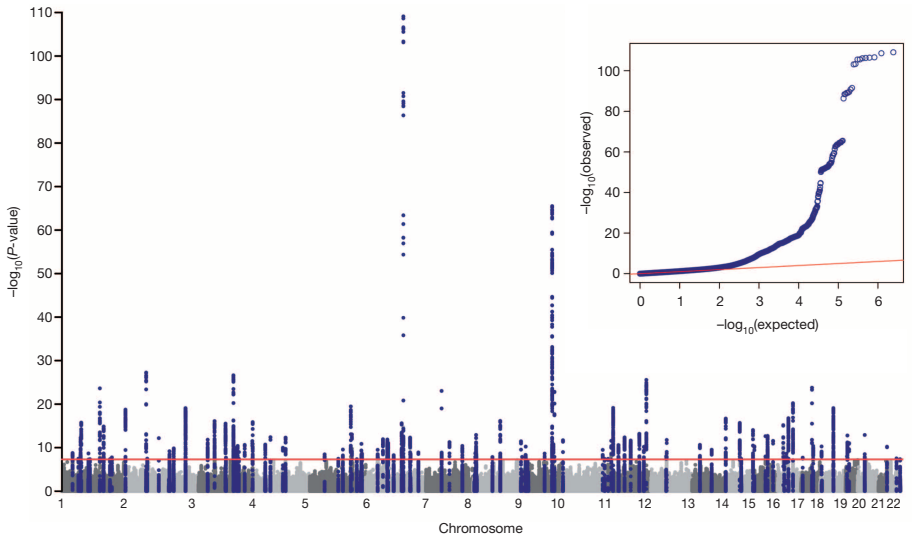


Figure 1 | Manhattan and quantile–quantile plot of the GWAS for age at menarche. Manhattan (main panel) and quantile–quantile (QQ) (embedded) plots illustrating results of the genome-wide association study (GWAS) meta-analysis for age at menarche in up to 182,416 women of European descent. The Manhattan plot presents the association $-\log_{10}$(*P*-values) for each genome-wide SNP (*y* axis) by chromosomal position (*x* axis). The red line indicates the threshold for genome-wide statistical significance ($P = 5 \times 10^{-8}$). Blue dots represent SNPs whose nearest gene is the same as that of the genome-wide significant signals. The QQ plot illustrates the deviation of association test statistics (blue dots) from the distribution expected under the null hypothesis (red line).

$P = 2.2 \times 10^{-6}$), indicating a highly polygenic architecture (Extended Data Fig. 2).

To test the relevance of menarche loci to the timing of related pubertal characteristics in both sexes, we examined their further associations with refined pubertal stage assessments in an overlapping subset of 10- to 12-year-old girls ($n = 6{,}147$). A further independent sample of 3,769 boys had similar assessments at ages 12 to 15 years. 90/106 menarche loci showed consistent directions of association with Tanner stage in boys and girls combined ($P_{Sign} = 1.1 \times 10^{-13}$), 86/106 in girls only ($P_{Sign} = 6.2 \times 10^{-11}$) and 72/106 in boys only ($P_{Sign} = 0.0001$), suggesting that the menarche loci are highly enriched for variants that regulate pubertal timing more generally (Supplementary Table 4).

Six independent signals were located in imprinted gene regions[6], which is an enrichment when compared to all published genome-wide-significant signals for any trait and/or disease[7] (6/123, 4.8% vs 75/4332, 1.7%; Fisher's exact test $P = 0.017$). Departure from Mendelian inheritance of pubertal timing has not been previously suspected, therefore we sought evidence for parent-of-origin-specific allelic associations in the deCODE Study, which included 35,377 women with parental origins of alleles determined by a combination of genealogy and long-range phasing[6].

Two independent signals (no. 85a and 85b; rs10144321 and rs7141210) lie on chromosome 14q32 harbouring the reciprocally imprinted genes *DLK1* and *MEG3*, which exhibit paternal-specific or maternal-specific expression, respectively, and may underlie the growth retardation and precocious puberty phenotype of maternal uniparental disomy-14[8]. In deCODE, for both signals the paternally inherited alleles were associated with age at menarche (rs10144321, $P_{pat} = 3.1 \times 10^{-5}$; rs7141210, $P_{pat} = 2.1 \times 10^{-4}$), but the maternally inherited alleles were not ($P_{mat} = 0.47$ and 0.12, respectively), and there was significant heterogeneity between paternal and maternal effect estimates (rs10144321, $P_{het} = 0.02$; rs7141210, $P_{het} = 2.2 \times 10^{-4}$) (Fig. 2; Supplementary Table 5). Notably, rs7141210 is reportedly a *cis*-acting methylation-quantitative trait locus (QTL) in adipose tissue[9] (Extended Data Table 5) and the menarche age-raising allele was also associated with lower transcript levels of *DLK1* (Supplementary Tables 6 and 7)[10], which encodes a transmembrane protein involved in adipogenesis and neurogenesis. In deCODE data, the maternally inherited rs7141210 allele was correlated with blood transcript levels of the maternally expressed genes *MEG3* ($P_{mat} < 5.6 \times 10^{-53}$), *MEG8* ($P_{mat} = 4.9 \times 10^{-41}$) and *MEG9* ($P_{mat} = 5.4 \times 10^{-5}$); however, lack of any correlation with the paternally inherited alleles ($P_{pat} = 0.18$, $P_{pat} = 0.87$ and $P_{pat} = 0.37$, respectively) suggests that these genes do not explain this paternal-specific menarche signal.

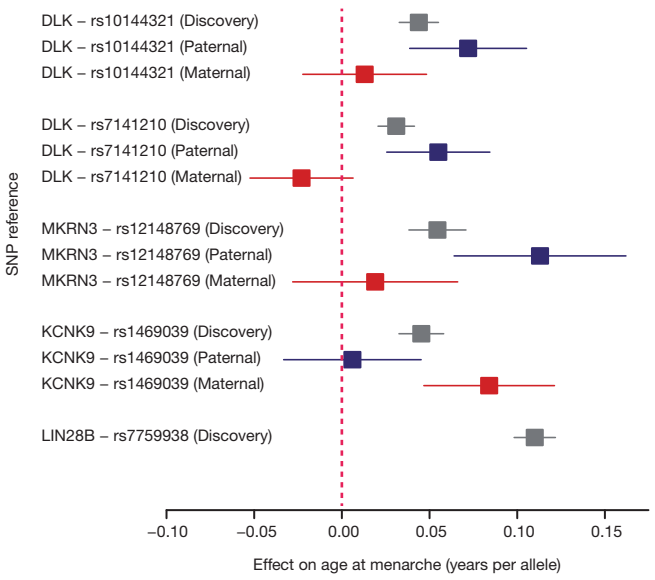


Figure 2 | Forest plot of parent-of-origin-specific allelic associations at three imprinted menarche loci. The forest plot illustrates the associations of variants in four independent genomic signals for age at menarche that are located in three imprinted gene regions. For each variant, squares (and error bars) indicate the estimated per-allele effect sizes on age at menarche in years (and 95% confidence intervals) from the standard additive models in the combined ReproGen meta-analysis (grey), and separately for the paternally inherited (blue) or maternally inherited allele (red) in up to 35,377 women from the deCODE study. The association for the menarche locus with the largest effect size at *LIN28B* is also shown for reference, illustrating the similar magnitude of effect size at the *MKRN3* locus when parent-of-origin is taken into account.

Signal no. 86 (rs12148769) lies in the imprinted critical region for Prader–Willi syndrome, which is caused by paternal-specific deletions of chromosome 15q11-13 and includes clinical features of hypogonadotropic hypogonadism and hypothalamic obesity[11]; conversely, a small proportion of cases have precocious puberty. For rs12148769, only the paternally inherited allele was associated with age at menarche ($P_{pat} = 2.4 \times 10^{-6}$), but the maternally inherited allele was not ($P_{mat} = 0.43$; $P_{het} = 5.6 \times 10^{-3}$) (Fig. 2). Recently, truncating mutations of *MAGEL2* affecting the paternal alleles were reported in Prader–Willi syndrome; all four reported cases had hypogonadism or delayed puberty[11], whereas paternally inherited deleterious mutations in *MKRN3* were found in patients with central precocious puberty[3]. It is as yet unclear which of these paternally expressed genes explains this menarche signal.

Signal no. 57 (rs1469039) is intronic in *KCNK9*, which shows maternal-specific expression in mouse and human brain[12]. Concordantly, only the maternally inherited allele was associated with age at menarche ($P_{mat} = 5.6 \times 10^{-6}$), but the paternally inherited allele was not ($P_{pat} = 0.76$; $P_{het} = 3.7 \times 10^{-3}$) (Fig. 2). The menarche age-increasing allele was associated with lower transcript levels of *KCNK9* in deCODE's blood expression data when maternally inherited ($P_{mat} = 0.003$), but not when paternally inherited ($P_{pat} = 0.31$). *KCNK9* encodes TASK-3, which belongs to a family of two-pore domain potassium channels that regulate neuronal resting membrane potential and firing frequency.

The two remaining signals located within imprinted regions (rs2137289 and rs947552) did not demonstrate either paternal- or maternal-specific association. We then systematically tested all 117 remaining independent menarche signals for parent-of-origin-specific associations with menarche timing and found only four (3.4%) with at least nominal associations ($P_{het} < 0.05$; Supplementary Table 5), which was proportionately fewer than signals at imprinted regions (4/6 (67.0%), Wilcoxon rank sum test $P = 0.009$).

Three menarche signals were in genes encoding JmjC-domain-containing lysine-specific demethylases (enrichment $P = 0.006$ for all genes in this family); signal no. 1 (rs2274465) is intronic in *KDM4A*, signal no. 37 (rs17171818) is intronic in *KDM3B*, and signal no. 59b (rs913588) is a missense variant in *KDM4C*. Notably, *KDM3B*, *KDM4A* and *KDM4C* all encode activating demethylases for lysine 9 on histone H3, which was recently identified as the chromatin methylation target that mediates the remarkable long-range regulatory effects of *IPW*, a paternally expressed long noncoding RNA in the imprinted Prader–Willi syndrome region on chromosome 15q11-13, on maternally expressed genes at the imprinted *DLK1-MEG3* locus on chromosome 14q32[13]. Examination of sub-genome-wide signals showed another potential locus intronic in *KDM4B* (rs11085110, $P = 2.3 \times 10^{-6}$). Pubertal onset in female mice is reportedly triggered by DNA methylation of the Polycomb group silencing complex of genes (including *CBX7* near signal no. 105), leading to enrichment of activating lysine modifications on histone H3[14]. Specific histone demethylases could potentially regulate cross-links between imprinted regions to influence pubertal timing.

Menarche signals also tended to be enriched in or near genes that underlie rare Mendelian disorders of puberty (enrichment $P = 0.05$)[2,3]. As well as rs12148769 near *MKRN3*, signals were found near *LEPR-LEPROT* (signal no. 2; rs10789181), which encodes the leptin receptor, and immediately upstream of *TACR3* (signal no. 32; rs3733631), which encodes the receptor for neurokinin B. A further variant approximately 10 kilobases (kb) from *GNRH1* approached genome-wide significance (rs1506869,

$P = 1.8 \times 10^{-6}$) and was also associated with *GNRH1* expression in adipose tissue ($P = 3.7 \times 10^{-5}$). Signals no. 34 (rs17086188) and 103 (rs852069) lie near *PCSK1* and *PCSK2*, respectively, indicating a common function of the type 1 and 2 prohormone convertases in pubertal regulation. Signals in or near several further genes with relevance to pituitary development/function included: signal no. 20 (rs7642134) near *POU1F1*, signal no. 39 (rs9647570) within *TENM2*, and signal no. 42 (rs2479724) near *FRS3*. Furthermore, signals no. 71 (rs7103411) and no. 92 (rs1129700) are *cis*-expression QTLs (eQTLs) for *LGR4* and *TBX6*, respectively, both of which encode enhancers for the pituitary development factor *SOX2*. Signals no. 52 (rs6964833 intronic in *GTF2I*) and no. 104 (rs2836950 intronic in *BRWD1*) were found in critical regions for complex conditions that include abnormal reproductive phenotypes, Williams–Beuren syndrome (early puberty)[15] and Down syndrome (hypogonadism in boys), respectively[16].

Including signals described above, we identified 29 menarche signals in or near genes with possible roles in hormonal functions (Fig. 3, Supplementary Table 8), many more than the three signals we described previously (*INHBA*, *PCSK2* and *RXRG*)[4]. Two signals were found in or near genes related to steroidogenesis. Signal 35 (rs251130) was a *cis*-eQTL for *STARD4*, which encodes a StAR-related lipid transfer protein involved in the regulation of intra-cellular cholesterol trafficking. Signal no. 9 (rs6427782) is near *NR5A2*, which encodes a nuclear receptor with key roles in steroidogenesis and oestrogen-dependent cell proliferation.

We observed that SNPs in or near a custom list of genes that encode nuclear hormone receptors, co-activators or co-repressors were enriched for associations with menarche timing (enrichment $P = 6 \times 10^{-5}$). Individually, nine genome-wide significant signals mapped to within 500 kb of these genes, including those encoding the nuclear receptors for oestrogen, progesterone, thyroid hormone and 1,25-dihydroxyvitamin D3. Several nuclear hormone receptors are involved in retinoic acid signalling. SNPs in or near *RXRG* and *RORA* reached genome-wide significance, and three other genes contained sub-genome-wide signals (*RXRA* (rs2520094, $P = 4 \times 10^{-7}$), *RORB* (rs4237264, $P = 9.4 \times 10^{-6}$), *RXRB* (rs241438, $P = 7.1 \times 10^{-5}$)). Two other genome-wide significant signals mapped to genes with roles in retinoic acid function (no. 67 *CTBP2* and no. 101 *RDH8*). The active metabolites of vitamin A, all-*trans*-retinoic acid and 9-*cis*-retinoic acid, have differential effects on gonadotropin-releasing hormone (GnRH) expression and secretion[17]. Other possible mechanisms linking retinoic acid signalling to pubertal timing include inhibition of embryonic GnRH neuron migration, and enhancement of steroidogenesis and gonadotropin secretion[18]. The relevance of our findings to observations of low circulating vitamin A levels and use of dietary vitamin A in delayed puberty[19] are yet unclear.

To identify other mechanisms that regulate pubertal timing, we tested all SNPs genome-wide for collective enrichment across any biological pathway defined in publicly available databases. The top ranked pathway reaching study-wise significance (false discovery rate = 0.009) was gamma-aminobutyric acid ($GABA_B$) receptor II signalling (Extended Data Table 6); each of the nine genes in this pathway contained a SNP with sub-genome-wide significant association with menarche (Extended Data Table 7). Notably, $GABA_B$ receptor activation inhibits hypothalamic GnRH secretion in animal models[20].

Regarding the relevance of our findings to other traits, we confirmed[4] and extended the overlap between genome-wide significant loci for menarche and adult body mass index (BMI)[21]. At all nine loci (in or near *FTO*, *SEC16B*, *TMEM18*, *NEGR1*, *TNNI3K*, *GNPDA2*, *BDNF*, *BCDIN3D* and *GPRC5B*) the menarche age-raising allele was also associated with lower adult BMI (Supplementary Table 9). Three menarche signals overlapped known loci for adult height[22]. The menarche age-raising alleles at signals no. 47c (rs7759938, *LIN28B*) and no. 83 (rs1254337, *SIX6*) were also associated with taller adult height, which is directionally concordant with epidemiological observations. Conversely, the menarche age-raising allele at signal no. 48 (rs4895808, *CENPW-NCOA7*) was associated with shorter adult height (Supplementary Table 9).

Further menarche signals overlapped reported GWAS loci for other traits, but in each case at only a single locus, therefore possibly reflecting small-scale pleiotropy rather than a broader shared genetic aetiology. Signal no. 26 (rs900400) was a *cis*-eQTL for *LEKR1*, and is the same lead SNP associated with birth weight[23]. The menarche age-raising allele was also associated with higher birth weight, directionally concordant with epidemiological observations[24]. Signal no. 48 (rs4895808, a *cis*-eQTL for *CENPW*) is in linkage disequilibrium (LD) ($r^2 = 0.90$) with the lead SNP for the autoimmune disorder type 1 diabetes, rs9388489[25], which also showed robust association with menarche timing ($P = 6.49 \times 10^{-12}$). Signal no. 41 (rs16896742) is near *HLA-A*, which encodes the class I, A major histocompatibility complex, and is a known locus for various immunity or inflammation-related traits[7]. Signal no. 50 (rs6933660) is near *ESR1*, which encodes the oestrogen receptor, a known locus for breast cancer[26] and bone mineral density[27]. Notably, the menarche age-raising allele at rs6933660 was associated with higher femoral neck bone mineral density ($P = 6 \times 10^{-5}$)[27], which is directionally discordant with the epidemiological association[28]. Signal no. 70 (rs11022756) is intronic in *ARNTL*, a known locus for circulating plasminogen activator inhibitor type 1 (PAI-1) levels[29]; the reported lead SNP (rs6486122) for PAI-1[29] also showed robust association with menarche timing ($P = 9.3 \times 10^{-10}$).

Our findings indicate both BMI-related and BMI-independent mechanisms that could underlie the epidemiological associations between

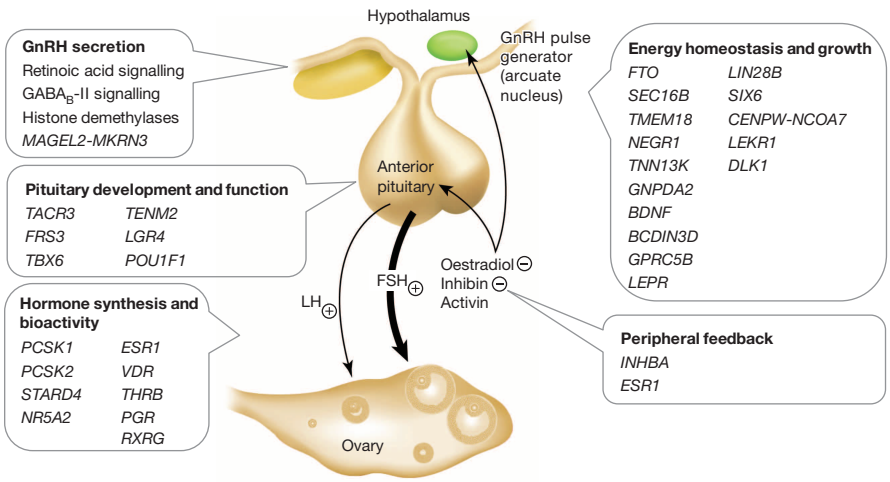


Figure 3 | Schematic diagram indicating possible roles in the hypothalamic-pituitary-ovarian axis of several of the implicated genes and biological mechanisms for menarche timing.

early menarche and higher risks of adult disease[1]. These include actions of *LIN28B* on insulin sensitivity through the mTOR pathway, $GABA_B$ receptor signalling on inhibition of oxidative stress-related β-cell apoptosis, and *SIRT3* (mitochondrial sirtuin 3), which could link early life nutrition to metabolism and ageing. Finally, only few parent-of-origin-specific allelic associations at imprinted loci have been described for complex traits[6]. Our findings implicate differential pubertal timing, a trait with putative selection advantages[30], as a potential additional target for the evolution of genomic imprinting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2013; accepted 30 May 2014.
Published online 23 July 2014.

1. Prentice, P. & Viner, R. M. Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. *Int. J. Obes.* **37,** 1036–1043 (2013).
2. Silveira, L. F. G. & Latronico, A. C. Approach to the patient with hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* **98,** 1781–1788 (2013).
3. Abreu, A. P. *et al.* Central precocious puberty caused by mutations in the imprinted gene *MKRN3*. *N. Engl. J. Med.* **368,** 2467–2475 (2013).
4. Elks, C. E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genet.* **42,** 1077–1085 (2010).
5. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genet.* **44,** 369–375 (2012).
6. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462,** 868–874 (2009).
7. Hindorff, L. A. *et al.* A catalog of published genome-wide association studies. Available at http://www.genome.gov/gwastudies. (Accessed, 1 November 2013).
8. Temple, I. K., Shrubb, V., Lever, M., Bullman, H. & Mackay, D. J. G. 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).
9. Grundberg, E. *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* **93,** 876–890 (2013), corrected **93,** 1158 (2013).
10. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genet.* **45,** 1238–1243 (2013).
11. Schaaf, C. P. *et al.* Truncating mutations of *MAGEL2* cause Prader-Willi phenotypes and autism. *Nature Genet.* **45,** 1405–1408 (2013).
12. Ruf, N. *et al.* Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the *KCNK9* potassium channel gene in mouse and human. *Hum. Mol. Genet.* **16,** 2591–2599 (2007).
13. Stelzer, Y., Sagi, I., Yanuka, O., Eiges, R. & Benvenisty, N. The noncoding RNA *IPW* regulates the imprinted *DLK1-DIO3* locus in an induced pluripotent stem cell model of Prader-Willi syndrome. *Nature Genet.* **46,** 551–557 (2014).
14. Lomniczi, A. *et al.* Epigenetic control of female puberty. *Nature Neurosci.* **16,** 281–289 (2013).
15. Partsch, C.-J. *et al.* Central precocious puberty in girls with Williams syndrome. *J. Pediatr.* **141,** 441–444 (2002).
16. Grinspon, R. P. *et al.* Early onset of primary hypogonadism revealed by serum anti-Müllerian hormone determination during infancy and childhood in trisomy 21. *Int. J. Androl.* **34,** e487–e498 (2011).
17. Cho, S. *et al.* 9-cis-Retinoic acid represses transcription of the gonadotropin-releasing hormone (GnRH) gene via proximal promoter region that is distinct from all-trans-retinoic acid response element. *Brain Res. Mol. Brain Res.* **87,** 214–222 (2001).
18. Nagl, F. *et al.* Retinoic acid-induced nNOS expression depends on a novel PI3K/Akt/DAX1 pathway in human TGW-nu-I neuroblastoma cells. *Am. J. Physiol. Cell Physiol.* **297,** C1146–C1156 (2009).
19. Zadik, Z., Sinai, T., Zung, A. & Reifen, R. Vitamin A and iron supplementation is as efficient as hormonal therapy in constitutionally delayed children. *Clin. Endocrinol.* **60,** 682–687 (2004).
20. Constantin, S. *et al.* GnRH neuron firing and response to GABA *in vitro* depend on acute brain slice thickness and orientation. *Endocrinology* **153,** 3758–3769 (2012).
21. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42,** 937–948 (2010).
22. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
23. Horikoshi, M. *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature Genet.* **45,** 76–82 (2013).
24. D'Aloisio, A. A., DeRoo, L. A., Baird, D. D., Weinberg, C. R. & Sandler, D. P. Prenatal and infant exposures and age at menarche. *Epidemiology* **24,** 277–284 (2013).
25. Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genet.* **41,** 703–707 (2009).
26. Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genet.* **41,** 324–328 (2009).
27. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44,** 491–501 (2012).
28. Parker, S. E. *et al.* Menarche, menopause, years of menstruation, and the incidence of osteoporosis: the influence of prenatal exposure to diethylstilbestrol. *J. Clin. Endocrinol. Metab.* **99,** 594–601 (2014).
29. Huang, J. *et al.* Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood* **120,** 4873–4881 (2012).
30. Migliano, A. B., Vinicius, L. & Lahr, M. M. Life history trade-offs explain the evolution of human pygmies. *Proc. Natl Acad. Sci. USA* **104,** 20216–20219 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements A full list of acknowledgements can be found in the Supplementary Information.

Author Contributions Overall project management: J.R.B.P., F.D., C.E.E., P.S., D.J.T., D.F.E., K.S., J.M.M. and K.K.O. Core analyses: J.R.B.P., F.D., C.E.E., P.S., T.F., D.J.T., D.I.C. and T.E. Individual study analysts: A.A.R., A.D., A.G., A.J., A.T., A.V.S., B.Z.A., B.F., C.E.E., D.F.G., D.I.C., D.J.T., D.L.C., D.L.K., E.A., E.K.W., E.M., E.M.B., E.T., F.D., G.M., G.McMahon, I.M.N., J.A.V., J.D., J.H., J.R.B.P., J.T., J.Z., K.L.L., K.M., L.L.P., L.M.R., L.M.Y., L.S., M.M., N.F., N.Ts., P.K., P.S., R.M., S.K., S.S., S.S.U., T.C., T.E., T.F., T.Fo., T.H.P., W.Q.A. and Z.K. Individual study data management and generation: A.A.R., A.C.H., A.D., A.D.C., A.G.U., A.J.O., A.M.S., A.Mu., A.P., A.Po., B.A.O., C.A.H., D.C., D.I.C., D.J.H., D.K., D.Lw., D.P.K., D.P.S., D.S., E.A.N., E.P., E.W., F.A., F.B.H., F.G., F.R., G.D., G.E., G.G.W., H.S., H.W., I.D., J.C., J.H., J.P.R., L.F., L.Fr., L.M., L.M.R., M.E.G., M.J.S., M.J.W., M.K.B., M.Melbye, M.P., M.W., N.A., N.J.T., N.L.P., P.K.M., Q.W., R.H., S.B., S.C., S.G., S.L., S.R., S.S.U., T.E., U.S., U.T., V.S. and W.L.M. Individual study principal investigators: A.C., A.G.U., A.H., A.J.O., A.K.D., A.L., A.M., A.M.D., A.Mannermaa, A.Mu., A.R., B.B., B.Z.A., B.H.R.W., C.B., C.E.P., C.G., C.H., C.van Duijn, D.I.B., D.F., D.F.E., D.J.H., D.L., D.Lw., D.S.P., D.P.S., D.Schlessinger, E.A.S., E.B., E.E.J.d.G., E.I., E.W., E.W.D., F.B.H., F.J.C., G.C., G.D., G.G.G., G.Wa., G.Wi., G.W.M., H.A., H.A.B., H.B., H.Be., H.F., H.N., H.S., H.V., I.D., I.L.A., J.A.K., J.B., J.C.C., J.G.E., J.E.B., J.L.H., J.M.C., J.M.M., J.P., K.C., K.K., K.K.O., K.P., K.S., L.C., L.F., L.J.B., M.C.S., M.G., M.I.M., M.J., M.J.E., M.J.H., M.J.S., M.K.S., M.W.B., M.Z., N.G.M., N.J.W., P.A.F., P.D., P.D.P.P., P.F.M., P.G., P.H., P.K., P.M.R., P.N., P.P., P.P.G., P.R., P.V., R.J.F.L., R.L.M., R.W., S.B., S.Bergmann, S.C., S.E.B., T.B.H., T.D.S., T.I.A.S., U.H., V.G., V.K. and V.S.

Author Information Plots of all 106 menarche loci and genome-wide summary level statistics are available at the ReproGen Consortium website: http://www.reprogen.org. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R.B.P. (john.perry@mrc-epid.cam.ac.uk) and J.M. (Murabito@bu.edu).

John R. B. Perry[1,2,3,4]*, Felix Day[1]*, Cathy E. Elks[1]*, Patrick Sulem[5]*, Deborah J. Thompson[6], Teresa Ferreira[3], Chunyan He[7,8], Daniel I. Chasman[9,10], Tõnu Esko[11,12,13,14], Gudmar Thorleifsson[5], Eva Albrecht[15], Wei Q. Ang[16], Tanguy Corre[17,18], Diana L. Cousminer[19], Bjarke Feenstra[20], Nora Franceschini[21], Andrea Ganna[22], Andrew D. Johnson[23], Sanela Kjellqvist[24], Kathryn L. Lunetta[23,25], George McMahon[26,27], Ilja M. Nolte[28], Lavinia Paternoster[26], Eleonora Porcu[29,30], Albert V. Smith[31,32], Lisette Stolk[33,34], Alexander Teumer[35], Natalia Tšernikova[11,36], Emmi Tikkanen[19,37], Sheila Ulivi[38], Erin K. Wagner[7,8], Najaf Amin[39], Laura J. Bierut[40], Enda M. Byrne[41,42], Jouke-Jan Hottenga[43], Daniel L. Koller[44], Massimo Mangino[4], Tune H. Pers[12,13,45,46], Laura M. Yerges-Armstrong[47], Jing Hua Zhao[1], Irene L. Andrulis[48,49], Hoda Anton-Culver[50], Femke Atsma[51], Stefania Bandinelli[52,53], Matthias W. Beckmann[54], Javier Benitez[55,56], Carl Blomqvist[57], Stig E. Bojesen[58,59], Manjeet K. Bolla[6], Bernardo Bonanni[60], Hiltrud Brauch[61,62], Hermann Brenner[63,64], Julie E. Buring[9,10], Jenny Chang-Claude[65], Stephen Chanock[66], Jinhui Chen[67,68], Georgia Chenevix-Trench[69], J. Margriet Collée[70], Fergus J. Couch[71], David Couper[72], Andrea D. Coviello[73], Angela Cox[74], Kamila Czene[22], Adamo Pio D'adamo[38,75], George Davey Smith[26,27], Immaculata De Vivo[76,77], Ellen W. Demerath[78], Joe Dennis[6], Peter Devilee[79], Aida K. Dieffenbach[63,64], Alison M. Dunning[80], Gudny Eiriksdottir[31], Johan G. Eriksson[81,82,83,84], Peter A. Fasching[54], Luigi Ferrucci[85], Dieter Flesch-Janys[86], Henrik Flyger[87], Tatiana Foroud[44], Lude Franke[88], Melissa E. Garcia[89], Montserrat García-Closas[90,91], Frank Geller[20], Eco E. J. de Geus[43,92], Graham G. Giles[93,94], Daniel F. Gudbjartsson[5,95], Vilmundur Gudnason[31,32], Pascal Guénel[96,97], Suiqun Guo[98], Per Hall[22], Ute Hamann[99], Robin Haring[100], Catharina A. Hartman[101], Andrew C. Heath[102], Albert Hofman[103], Maartje J. Hooning[104], John L. Hopper[94], Frank B. Hu[76,77,105], David J. Hunter[13,76,77], David Karasik[10,106], Douglas P. Kiel[106,107], Julia A. Knight[108,109], Veli-Matti Kosma[110,111], Zoltan Kutalik[17,18], Sandra Lai[29], Diether Lambrechts[112,113], Annika Lindblom[114], Reedik Mägi[11], Patrik K. Magnusson[22], Arto Mannermaa[110,111], Nicholas G. Martin[69], Gisli Masson[5], Patrick F. McArdle[47], Wendy L. McArdle[27], Mads Melbye[20,115], Kyriaki Michailidou[6], Evelin Mihailov[11,36], Lili Milani[11], Roger L. Milne[93,94], Heli Nevanlinna[116], Patrick Neven[117], Ellen A. Nohr[118], Albertine J. Oldehinkel[119], Ben A. Oostra[39], Aarno Palotie[19,120,121,122], Munro Peacock[123], Nancy L. Pedersen[22], Paolo Peterlongo[124], Julian Peto[125], Paul D. P. Pharoah[80], Dirkje S. Postma[126], Anneli Pouta[81,127], Katri Pylkäs[128], Paolo Radice[129], Susan Ring[26,27], Fernando Rivadeneira[33,34,103], Antonietta Robino[38,75], Lynda M. Rose[9], Anja Rudolph[65], Veikko Salomaa[81], Serena Sanna[29], David Schlessinger[130], Marjanka K. Schmidt[131], Mellissa C. Southey[132], Ulla Sovio[133,134], Meir J. Stampfer[76,77,105], Doris Stöckl[135,136], Anna M. Storniolo[123], Nicholas J. Timpson[26,27], Jonathan Tyrer[80], Jenny A. Visser[33], Peter Vollenweider[137], Henry Völzke[138,139], Gerard Waeber[137], Melanie Waldenberger[140], Henri Wallaschofski[100,139], Qin Wang[6], Gonneke Willemsen[43], Robert Winqvist[128], Bruce H. R. Wolffenbuttel[141], Margaret J. Wright[142], Australian Ovarian Cancer Study†, The GENICA Network†, kConFab†, The LifeLines Cohort

Study†, The InterAct Consortium†, Early Growth Genetics (EGG) Consortium†, Dorret I. Boomsma[43], Michael J. Econs[44,123], Kay-Tee Khaw[143], Ruth J. F. Loos[1,144], Mark I. McCarthy[3,145,146], Grant W. Montgomery[142], John P. Rice[40], Elizabeth A. Streeten[47,147], Unnur Thorsteinsdottir[5,95], Cornelia M. van Duijn[34,39,148], Behrooz Z. Alizadeh[28], Sven Bergmann[17,18], Eric Boerwinkle[149], Heather A. Boyd[20], Laura Crisponi[29], Paolo Gasparini[38,75], Christian Gieger[15], Tamara B. Harris[89], Erik Ingelsson[150], Marjo-Riitta Järvelin[133,151,152,153,154], Peter Kraft[76,155], Debbie Lawlor[26,27], Andres Metspalu[11,36], Craig E. Pennell[16], Paul M. Ridker[9,10], Harold Snieder[28], Thorkild I. A. Sørensen[156,157], Tim D. Spector[4], David P. Strachan[158], André G. Uitterlinden[33,34,103], Nicholas J. Wareham[1], Elisabeth Widen[19], Marek Zygmunt[159], Anna Murray[2], Douglas F. Easton[6], Kari Stefansson[5,95]*, Joanne M. Murabito[23,160]* & Ken K. Ong[1,161]*

[1]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. [2]University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [4]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK. [5]deCODE Genetics, Reykjavik IS-101, Iceland. [6]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [7]Department of Epidemiology, Indiana University Richard M Fairbanks School of Public Health, Indianapolis, Indiana 46202, USA. [8]Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, Indiana 46202, USA. [9]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. [10]Harvard Medical School, Boston, Massachusetts 02115, USA. [11]Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia. [12]Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [13]Broad Institute of the Massachusetts Institute of Technology and Harvard University, 140 Cambridge, Massachusetts 02142, USA. [14]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [15]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany. [16]School of Women's and Infants' Health, The University of Western Australia, WA-6009, Australia. [17]Department of Medical Genetics, University of Lausanne, CH-1005 Lausanne, Switzerland. [18]Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland. [19]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, FI-00014, Finland. [20]Department of Epidemiology Research, Statens Serum Institut, DK-2300 Copenhagen, Denmark. [21]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27599-7400, USA. [22]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 17177 Stockholm, Sweden. [23]NHLBI's and Boston University's Framingham Heart Study, Framingham, Massachusetts 01702-5827, USA. [24]Science for Life Laboratory, Karolinska Institutet, Stockholm, Box 1031, 17121 Solna, Sweden. [25]Boston University School of Public Health, Department of Biostatistics, Boston, Massachusetts 02118, USA. [26]MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK. [27]School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. [28]Department of Epidemiology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands. [29]Institute of Genetics and Biomedical Research, National Research Council, Cagliari, 09042 Sardinia, Italy. [30]University of Sassari, Department of Biomedical Sciences, 07100 Sassari, Italy. [31]Icelandic Heart Association, IS-201 Kopavogur, Iceland. [32]University of Iceland, IS-101 Reykjavik, Iceland. [33]Department of Internal Medicine, Erasmus MC, 3015 GE Rotterdam, the Netherlands. [34]Netherlands Consortium on Health Aging and National Genomics Initiative, 2300 RC Leiden, the Netherlands. [35]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, D-17475 Greifswald, Germany. [36]Department of Biotechnology, University of Tartu, 51010 Tartu, Estonia. [37]Hjelt Institute, University of Helsinki, FI-00014, Finland. [38]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", 34137 Trieste, Italy. [39]Genetic Epidemiology Unit Department of Epidemiology, Erasmus MC, 3015 GE, Rotterdam, the Netherlands. [40]Department of Psychiatry, Washington University, St Louis, Missouri 63110, USA. [41]The University of Queensland, Queensland Brain Institute, St Lucia, Queensland 4072, Australia. [42]QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006, Australia. [43]Department of Biological Psychology, VU University Amsterdam, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. [44]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202-3082, USA. [45]Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [46]Center for Biological Sequence Analysis, Department of Systems Biology, Technical 142 University of Denmark, DK-2800 Lyngby, Denmark. [47]Program in Personalized and Genomic Medicine, and Department of Medicine, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [48]Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [49]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [50]Department of Epidemiology, University of California Irvine, Irvine, California 92697-7550, USA. [51]Sanquin Research, 6525 GA Nijmegen, The Netherlands. [52]Tuscany Regional Health Agency, Florence, Italy, I.O.T. and Department of Medical and Surgical Critical Care, University of Florence, 50134 Florence, Italy. [53]Geriatric Unit, Azienda Sanitaria di Firenze, 50122 Florence, Italy. [54]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, D-91054 Erlangen, Germany. [55]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), E-28029 Madrid, Spain. [56]Centro de Investigación en Red de Enfermedades Raras (CIBERER), E-46010 Valencia, Spain. [57]Department of Oncology, University of Helsinki and Helsinki University Central Hospital, FI-00100 Helsinki, Finland. [58]Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, DK-2100 Copenhagen, Denmark. [59]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, DK-2100 Copenhagen, Denmark. [60]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), 20139 Milan, Italy. [61]DrMargarete Fischer-Bosch-Institute of Clinical Pharmacology, D-70376 Stuttgart, Germany. [62]University of Tübingen, D-72074 Tübingen, Germany. [63]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. [64]German Cancer Consortium (DKTK), D-69120 Heidelberg, Germany. [65]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. [66]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. [67]Departments of Anatomy and Neurological Surgery, Indiana University school of Medicine, Indianapolis, Indiana 46202, USA. [68]Stark Neuroscience Research Center, Indiana University school of Medicine, Indianapolis, Indiana 46202, USA. [69]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006 Australia. [70]Department of Clinical Genetics, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands. [71]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55905, USA. [72]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599-7420, USA. [73]Boston University School of Medicine, Department of Medicine, Sections of Preventive Medicine and Endocrinology, Boston, Massachusetts 02118, USA. [74]Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield S10 2RX, UK. [75]Department of Clinical Medical Sciences, Surgical and Health, University of Trieste, 34149 Trieste, Italy. [76]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [77]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [78]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455, USA. [79]Department of Human Genetics & Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. [80]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge CB1 8RN, UK. [81]National Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland. [82]Department of General Practice and Primary health Care, University of Helsinki, FI-00014 Helsinki, Finland. [83]Helsinki University Central Hospital, Unit of General Practice, FI-00029 HUS Helsinki, Finland. [84]Folkhalsan Research Centre, FI-00290 Helsinki, Finland. [85]Longitudinal Studies Section, Clinical Research Branch, Gerontology Research Center, National Institute on Aging, Baltimore, Maryland 20892, USA. [86]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, D-20246 Hamburg, Germany. [87]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. [88]Department of Genetics, University of Groningen, University Medical Centre Groningen, P.O. Box 72, 9700 AB Groningen, The Netherlands. [89]National Insitute on Aging, National Institutes of Health, Baltimore, Maryland 20892, USA. [90]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. [91]Breakthrough Breast Cancer Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK. [92]EMGO + Institute for Health and Care Research, VU University Medical Centre, Van der Boechorststraat 7, 1081 Bt, Amsterdam, The Netherlands. [93]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria 3004, Australia. [94]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria 3010, Australia. [95]Faculty of Medicine, University of Iceland, IS-101 Reykjavik, Iceland. [96]Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, F-94807 Villejuif, France. [97]University Paris-Sud, UMRS 1018, F-94807 Villejuif, France. [98]Department of Obstetrics and Gynecology, Southern Medical University, 510515 Guangzhou, China. [99]Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), D-69120 Heidelberg, Germany. [100]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany. [101]Department of Psychiatry, University of Groningen, University Medical Center Groningen, P.O. Box 72, 9700 AB Groningen, The Netherlands. [102]Washington University, Department of Psychiatry, St Louis, Missouri 63110, USA. [103]Department of Epidemiology, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, the Netherlands. [104]Department of Medical Oncology, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. [105]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [106]Hebrew SeniorLife Institute for Aging Research, Boston, Massachusetts 02131, USA. [107]Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02115, USA. [108]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [109]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5T 3M7, Canada. [110]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland. [111]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, P.O. Box 100, FI-70029 Kuopio, Finland. [112]Vesalius Research Center (VRC), VIB, 3000 Leuven, Belgium. [113]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, 3000 Leuven, Belgium. [114]Department of Molecular Medicine and Surgery, Karolinska Institutet, SE-171 77 Stockholm, Sweden. [115]Department of Medicine, Stanford School of Medicine, Stanford, California 94305-5101, USA. [116]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, P.O. Box 100, FI-00029 HUS Helsinki, Finland. [117]KULeuven (University of Leuven), Department of Oncology, Multidisciplinary Breast Center, University Hospitals Leuven, 3000 Leuven, Belgium. [118]Research Unit of Obstetrics & Gynecology, Institute of Clinical Research, University of Southern Denmark, DK-5000 Odense C, Denmark. [119]Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands. [120]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [121]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [122]Psychiatric & Neurodevelopmental Genetics

Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [123]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [124]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, 20139 Milan, Italy. [125]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. [126]University Groningen, University Medical Center Groningen, Department Pulmonary Medicine and Tuberculosis, GRIAC Research Institute, P.O. Box 30.001, NL-9700 RB Groningen, The Netherlands. [127]Department of Obstetrics and Gynecology, Oulu University Hospital, P.O. Box 10, FI-90029 OYS Oulu, Finland. [128]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu University Hospital/NordLab Oulu, P.O. Box 3000, FI-90014 Oulu, Finland. [129]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), 20133 Milan, Italy. [130]National Institute on Aging, Intramural Research Program, Baltimore, Maryland 21224-6825, USA. [131]Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Postbus 90203, 1006 BE Amsterdam, The Netherlands. [132]Department of Pathology, The University of Melbourne, Melbourne, Victoria 3010, Australia. [133]Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK. [134]Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge CB2 0SW, UK. [135]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, D-8576 Neuherberg, Germany. [136]Department of Obstetrics and Gynaecology, Campus Grosshadern, Ludwig-Maximilians-University, D-81377 Munich, Germany. [137]Department of Internal Medicine, Lausanne University Hospital, CH-1015 Lausanne, Switzerland. [138]Institute for Community Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany. [139]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, D-17475 Greifswald, Germany. [140]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-8576 Neuherberg, Germany. [141]Department of Endocrinology, University of Groningen, University Medical Centre Groningen, P.O. Box 72, 9700 AB Groningen, The Netherlands. [142]Queensland Insitute of Medical Research, Brisbane, Queensland 4029, Australia. [143]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge CB2 0QQ, UK. [144]Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, 1 Gustave L Levy Place, Box 1003, New York, New York 10029, USA. [145]NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford OX3 7LE, UK. [146]Oxford Centre for Diabetes, Endocrinology, & Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. [147]Geriatric Research and Education Clinical Center (GRECC) - Veterans Administration Medical Center, Baltimore, Maryland 21201, USA. [148]Centre of Medical Systems Biology, PO Box 9600, 2300 RC Leiden, the Netherlands. [149]Human Genetics Center and Divof Epidemiology, University of Houston, P.O. Box 20186, Texas 77025 USA. [150]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Box 256, 751 05 Uppsala, Sweden. [151]Institute of Health Sciences, University of Oulu, P.O. Box 5000, FI-90014 Oulu, Finland. [152]Biocenter Oulu, University of Oulu, P.O. Box 5000, Aapistie 5A, FI-90014 Oulu, Finland. [153]Department of Children and Young People and Families, National Institute for Health and Welfare, Aapistie 1, Box 310, FI-90101 Oulu, Finland. [154]Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O. Box 20, FI-90220 Oulu, 90029 OYS, Finland. [155]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [156]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200, Denmark. [157]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, The Capital Region, Copenhagen, DK-2000 Frederiksberg, Denmark. [158]Division of Population Health Sciences and Education, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK. [159]Department of Obstetrics and Gynecology, University Medicine Greifswald, D-17475 Greifswald, Germany. [160]Boston University School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, Massachusetts 02118, USA. [161]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK.

*These authors contributed equally to this work.
†Lists of participants and their affiliations appear in Supplementary Information.

METHODS

GWAS meta-analysis. We performed an expanded GWAS meta-analysis for self-reported age at menarche in up to 182,416 women of European descent from 58 studies (Supplementary Table 1). All participants provided written informed consent and the studies were approved by the respective Local Research Ethics committees or Institutional Review Boards. Consistent with our previous analysis protocol[4], women who reported their age at menarche as <9 years or >17 years were excluded from the analysis; birth year was included as the only covariate to allow for the secular trends in menarche timing. Genome-wide SNP array data were available on up to 132,989 women from 57 studies. Each study imputed genotype data based on HapMap Phase II CEU build 35 or 36. Data on an additional 49,427 women from the Breast Cancer Association Consortium (BCAC) were generated on the Illumina iSelect "iCOGS" array[31]. This array included up to ~25,000 SNPs, or their proxy markers, that showed sub-genome-wide associations ($P < 0.0022$) with age at menarche in our earlier GWAS[4]. SNPs were excluded from individual study data sets if they were poorly imputed or were rare (minor allele frequency < 1%). Test statistics for each study were adjusted using study-specific genomic control inflation factors and where appropriate individual studies performed additional adjustments for relatedness (Supplementary Table 1). Association statistics for each of the 2,441,815 autosomal SNPs that passed QC in at least half of the studies were combined across studies in a fixed effects inverse-variance meta-analysis implemented in METAL[32].

On meta-analysis, 3,915 SNPs reached the genome-wide significance threshold ($P < 5 \times 10^{-8}$) for association with age at menarche (Fig. 1). The overall GC inflation factor was 1.266, consistent with an expected high yield of true positive findings in large-scale GWAS meta-analysis of highly polygenic traits[33].

Selection of independent signals. Given the genome-wide results of the meta-analysis, SNPs showing evidence for association at genome-wide significant *P*-values were selected and clumped based on a physical (kb) threshold <1 megabase. The lead SNPs of the 105 clumps formed constitute the list of SNPs independently associated with age at menarche (Extended Data Tables 1-4).

To augment this list we performed approximate conditional analysis using GCTA software[34], where the LD between variants was estimated from the Northern Finland Birth Cohort (NFBC66) consisting of 5,402 individuals of European ancestry with GWAS data imputed using CEU haplotypes from Hapmap Phase II. Assuming that the LD correlations between SNPs more than 10 Mb away or on different chromosomes are zero, we performed the GCTA model selection to select SNPs independently associated with age at menarche at genome-wide significant *P*-values. This software selected as independently associated with age at menarche 115 SNPs at 98 loci, 11 of which had two or more signals of association (six loci contained two signals, four loci contained three signals, and one locus contained four signals). Plots of all 106 loci are available at http://www.reprogen.org. SNPs with A/T or C/G alleles were excluded from this analysis to prevent strand issues leading to false-positive results.

To summarize the information obtained from the single-SNP and GCTA analyses, the 105 SNPs selected from the uni-variate analysis and the 115 SNPs selected from the GCTA model selection analysis were combined into a single list of signals independently associated with age at menarche (Supplementary Table 2), using the following selection process (Extended Data Fig. 1). For loci with no evidence of allelic heterogeneity, if the uni-variate signal was genome-wide significant, the lead uni-variate SNP was selected (94 independent association signals follow this criterion); otherwise the lead GCTA SNP was selected instead (one independent signal). For loci where evidence for allelic heterogeneity was found, all signals identified in the GCTA joint model were selected if GCTA selected the uni-variate index SNP (21 independent signals at 8 loci) or a very good proxy ($r^2 > 0.8$) (7 independent signals at 3 loci). When instead GCTA selected a SNP independent from the uni-variate index SNP, both the lead uni-variate SNP and all signals identified in the GCTA joint model were selected (0 independent signals).

To determine likely causal genes at each locus, we used a combination of criteria. The gene nearest to each top SNP was selected by default. This gene was replaced or added to if the top SNP was (in high LD with) an expression quantitative-trait locus (eQTL) or a non-synonymous variant in another gene, or if there was an alternative neighbouring biological candidate gene. 31/123 signals mapped as eQTLs in data from Westra *et al.* (E)[10], five were annotated as non-synonymous functional (F), 60 as biological candidates (C), and four mapped to gene deserts (nearest gene >500 kb) (Supplementary Tables 6–8). We also used publicly available whole blood and adipose tissue methylation-QTL data to map 9/123 signals to *cis*-acting changes in methylation level (Extended Data Table 5)[9].

Follow up in the EPIC-InterAct study. We used an independent sample of 8689 women from the EPIC-InterAct study[35] to follow up our menarche signals. To test associations between each identified SNP and age at menarche with correction for cryptic relatedness, we ran a linear mixed model association test implemented in GCTA[34] (--mlma-loco option), adjusting for birth year, disease status and research centre. Given the relatively small sample size compared to our discovery set, directional consistency with results from the discovery-meta analysis was assessed using a binomial sign test. Variance explained by menarche loci was estimated using restricted maximum likelihood analysis in GCTA[34]. In addition to the 123 confirmed menarche loci, variance explained in subsets of menarche loci below the genome-wide significance thresholds was also assessed.

eQTL analyses. In order to estimate the potential downstream regulatory effects of age at menarche associated variants, we used publicly available blood eQTL data (downloadable from http://genenetwork.nl/bloodeqtlbrowser/) from a recently published paper by Westra *et al.*[10]. Westra *et al.* conducted *cis*-eQTL mapping by testing, for a large set of genes, all SNPs (HapMap2 panel) within 250 kb of the transcription start site of the gene for association with total RNA expression level of the gene. The publicly available data contain, for each gene, a list of all SNPs that were found to be significantly associated with gene expression using a false discovery rate (FDR) of 5%. For a detailed description of the quality control measures applied to the original data, see Westra *et al.*[10]. Their meta-analysis was based on a pooled sample of 5,311 individuals from 7 population-based cohorts with gene expression levels measured from full blood. We used the software tool SNAP (http://www.broadinstitute.org/mpg/snap/) to identify variants in close linkage disequilibrium ($r^2 \geq 0.8$) with the trait associated variants. All eQTL effects at FDR 5% and also lists of the strongest SNP effect for all the significant genes are shown in Supplementary Table 7.

Index SNPs (or highly correlated proxies) were also interrogated against a collected database of eQTL results from a range of tissues. Blood cell related eQTL studies included fresh lymphocytes[36], fresh leukocytes[37], leukocyte samples in individuals with Coeliac disease[38], whole blood samples[39–43], lymphoblastoid cell lines (LCL) derived from asthmatic children[44,45], HapMap LCL from 3 populations[46], a separate study on HapMap CEU LCL[47], additional LCL population samples[48–50] (and Mangravite *et al.* (unpublished)), $CD19^+$ B cells[51], primary PHA-stimulated T cells[48], $CD4^+$ T cells[52], peripheral blood monocytes[51,53,54], $CD11^+$ dendritic cells before and after *Mycobacterium tuberculosis* infection[55]. Micro-RNA QTLs[56] and DNase-I QTLs[57] were also queried for LCL. Non-blood cell tissue eQTLs searched included omental and subcutaneous adipose[39,50,58], stomach[58], endometrial carcinomas[59], ER+ and ER− breast cancer tumour cells[60], brain cortex[53,61,62], pre-frontal cortex[63,64], frontal cortex[65], temporal cortex[62,65], pons[65], cerebellum[62,65], 3 additional large studies of brain regions including prefrontal cortex, visual cortex and cerebellum, respectively[66], liver[58,67–70], osteoblasts[71], intestine[72], lung[73], skin[50,74] and primary fibroblasts[48]. Micro-RNA QTLs were also queried for gluteal and abdominal adipose[75]. Only results that reach study-wise significance thresholds in their respective data sets were included (Supplementary Table 6). Expression data was also available on adipose tissue and whole blood samples from deCODE where parent-of-origin-specific analyses were possible.

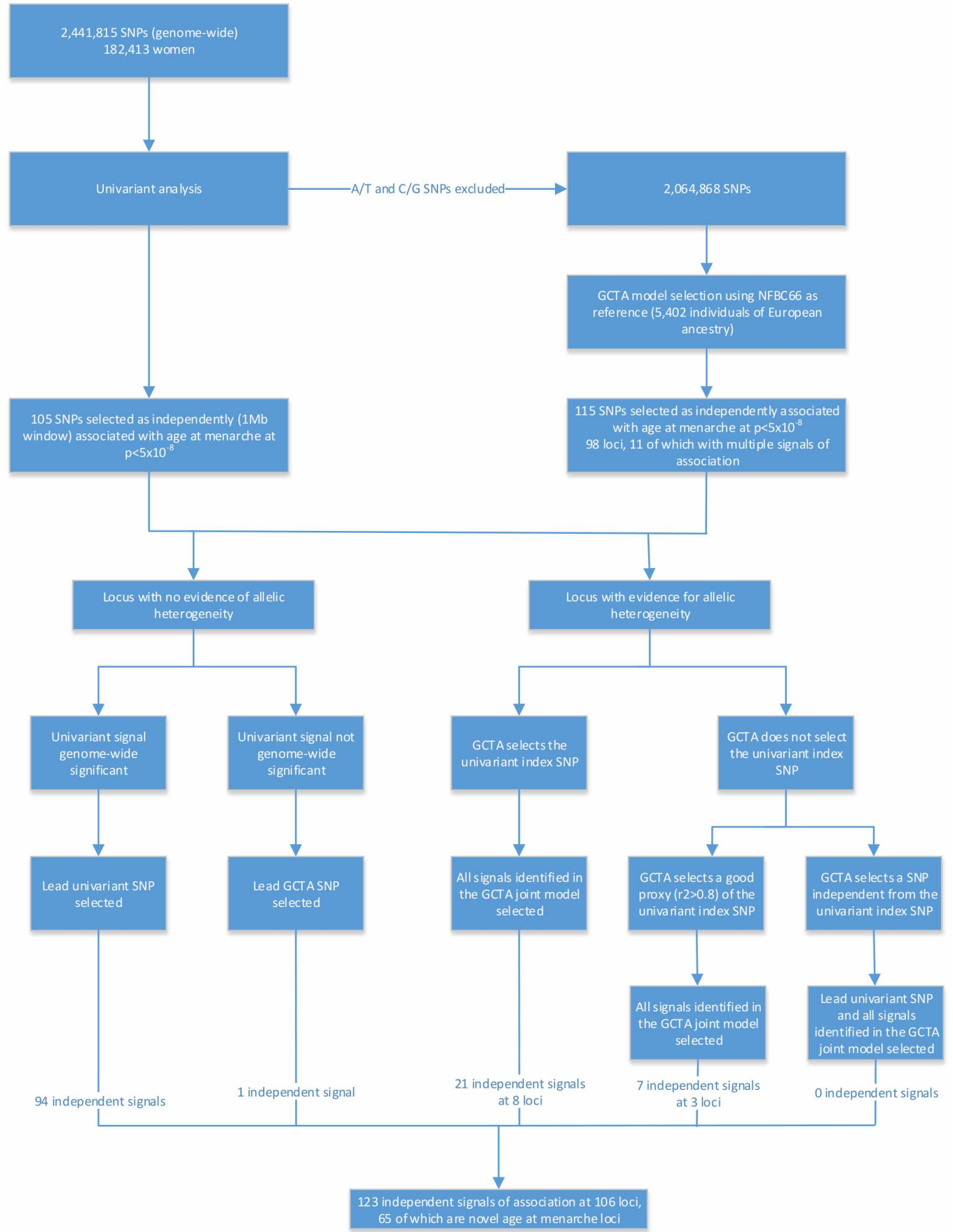
Parent-of-origin-specific associations. Evidence for parent-of-origin-specific allelic associations at imprinted loci was sought in the deCODE Study, which included 35,377 women with parental origins of alleles determined by a combination of genealogy and long-range phasing as previously described[6]. Briefly, using SNP chip data in each proband, genome-wide, long range phasing was applied to overlapping tiles, each 6 centimorgan (cM) in length, with 3 cM overlap between consecutive tiles. For each tile, the parental origins of the two phased haplotypes were determined regardless of whether the parents of the proband were chip-typed. Using the Icelandic genealogy database, for each of the two haplotypes of a proband, a search was performed to identify, among those individuals also known to carry the same haplotype, the closest relative on each of the paternal and maternal sides. Results for the two haplotypes were combined into a robust single-tile score reflecting the relative likelihood of the two possible parental origin assignments. Haplotypes from consecutive tiles were then stitched together based on sharing at the overlapping region. For haplotypes derived by stitching, a contig-score for parental origin was computed by summing the individual single-tile scores. Similarly, parent-of-origin-specific allelic associations at imprinted loci were also sought in the deCODE blood cells and adipose tissue expression data sets.

Pathway analyses. Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) was used to explore pathway-based associations in the full GWAS data set. MAGENTA implements a gene set enrichment analysis (GSEA) based approach, as previously described[76]. Briefly, each gene in the genome is mapped to a single index SNP with the lowest *P*-value within a 110 kb upstream, 40 kb downstream window. This *P*-value, representing a gene score, is then corrected for confounding factors such as gene size, SNP density and LD-related properties in a regression model. Genes within the HLA-region were excluded from analysis due to difficulties in accounting for gene density and LD patterns. Each mapped gene in the genome is then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, is calculated. This observed statistic is then compared to 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA *P*-value for each pathway. Significance was determined when an individual pathway reached a false discovery rate (FDR) < 0.05 in either analysis. In total, 2529 pathways from Gene Ontology, PANTHER, KEGG

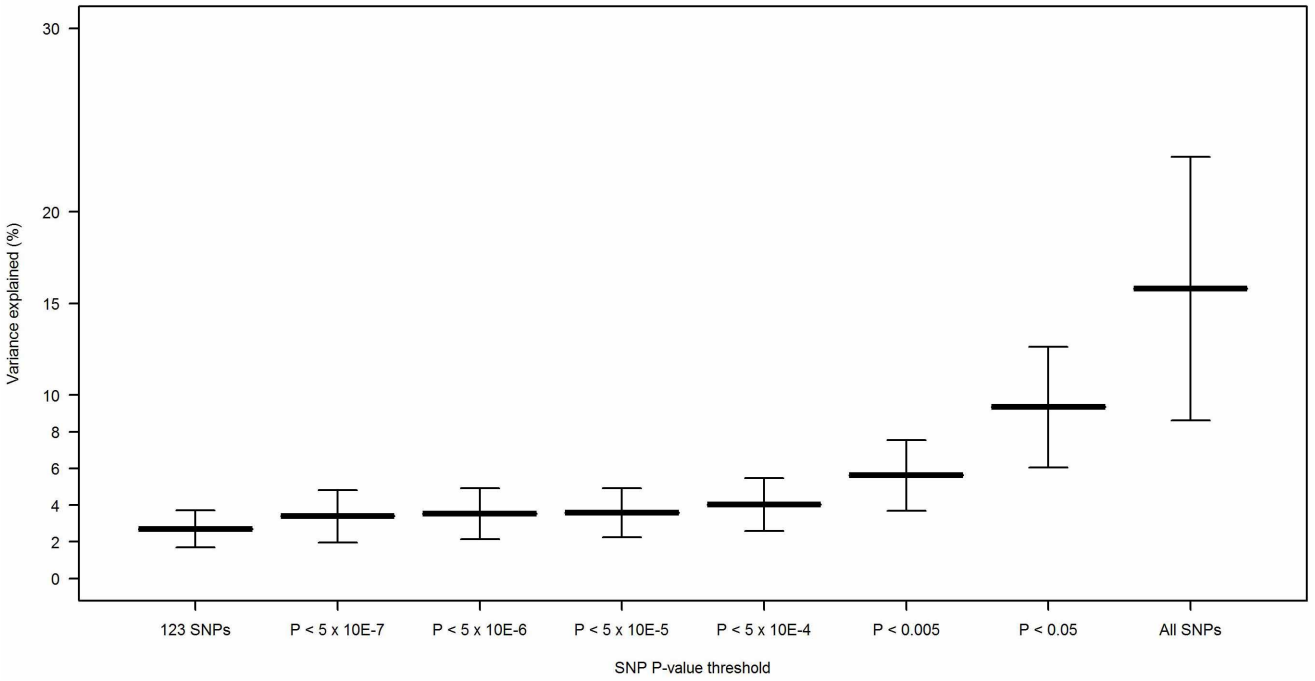
and Ingenuity were tested for enrichment of multiple modest associations with age at menarche. MAGENTA software was also used for enrichment testing of custom gene sets.

Relevance of menarche loci to other traits. We assessed the relevance of identified menarche loci to other traits by comparing SNPs significantly associated with age at menarche with published GWAS findings or by using publicly available data from the Genetic Investigation of Anthropometric Traits (GIANT) consortium[21,22] and the GEnetic Factors for OS (GEFOS) consortium[27]. In addition, we requested look-ups up the 123 menarche SNPs for association with puberty timing assessed by Tanner staging in the Early Growth Genetics (EGG) consortium[77].

31. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genet.* **45,** 353–361 (2013).
32. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
33. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2011).
34. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).
35. The InterAct Consortium Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54,** 2272–2282 (2011).
36. Göring, H. H. H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genet.* **39,** 1208–1216 (2007).
37. Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nature Genet.* **42,** 62–67 (2010).
38. Heap, G. A. *et al.* Complex nature of SNP genotype effects on gene expression in primary human leucocytes. *BMC Med. Genomics* **2,** 1 (2009).
39. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452,** 423–428 (2008).
40. Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7,** e1002197 (2011).
41. Mehta, D. *et al.* Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet.* **21,** 48–54 (2013).
42. Maeda, T. *et al.* The correlation between clinical laboratory data and telomeric status of male patients with metabolic disorders and no clinical history of vascular events. *Aging Male* **14,** 21–26 (2011).
43. Sasayama, D. *et al.* Identification of single nucleotide polymorphisms regulating peripheral blood mRNA expression with genome-wide significance: an eQTL study in the Japanese population. *PLoS ONE* **8,** e54967 (2013).
44. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39,** 1202–1207 (2007).
45. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* **23,** 716–726 (2013).
46. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39,** 1217–1224 (2007).
47. Kwan, T. *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nature Genet.* **40,** 225–231 (2008).
48. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325,** 1246–1250 (2009).
49. Cusanovich, D. A. *et al.* The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. *Hum. Mol. Genet.* **21,** 2111–2123 (2012).
50. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nature Genet.* **44,** 1084–1089 (2012).
51. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genet.* **44,** 502–510 (2012).
52. Murphy, A. *et al.* Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum. Mol. Genet.* **19,** 4745–4757 (2010).
53. Heinzen, E. L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6,** e1 (2008).
54. Zeller, T. *et al.* Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5,** e10693 (2010).
55. Barreiro, L. B. *et al.* Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl Acad. Sci. USA* **109,** 1204–1209 (2012).
56. Huang, R. S. *et al.* Population differences in microRNA expression and biological implications. *RNA Biol.* **8,** 692–701 (2011).
57. Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482,** 390–394 (2012).
58. Greenawalt, D. M. *et al.* A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* **21,** 1008–1016 (2011).
59. Kompass, K. S. & Witte, J. S. Co-regulatory expression quantitative trait loci mapping: method and application to endometrial cancer. *BMC Med. Genomics* **4,** 6 (2011).
60. Li, Q. *et al.* Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152,** 633–641 (2013).
61. Webster, J. A. *et al.* Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* **84,** 445–458 (2009).
62. Zou, F. *et al.* Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* **8,** e1002707 (2012).
63. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478,** 519–523 (2011).
64. Liu, C. *et al.* Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol. Psychiatry* **15,** 779–784 (2010).
65. Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6,** e1000952 (2010).
66. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153,** 707–720 (2013).
67. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6,** e107 (2008).
68. Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7,** e1002078 (2011).
69. 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).
70. Schröder, A. *et al.* Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. *Pharmacogenomics J.* **13,** 12–20 (2013).
71. Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res.* **19,** 1942–1952 (2009).
72. Kabakchiev, B. & Silverberg, M. S. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* **144,** 1488–1496e3 (2013).
73. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8,** e1003029 (2012).
74. Ding, J. *et al.* Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in *cis*-eQTL signals. *Am. J. Hum. Genet.* **87,** 779–789 (2010).
75. Rantalainen, M. *et al.* MicroRNA expression in abdominal and gluteal adipose tissue is associated with mRNA expression levels and partly genetically driven. *PLoS ONE* **6,** e27338 (2011).
76. Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
77. Cousminer, D. L. *et al.* Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum. Mol. Genet.*; Epub ahead of print (2014).



Extended Data Figure 1 | Flow chart illustrating the selection criteria used to identify independent signals for age at menarche.



Extended Data Figure 2 | Estimates of genetic variance explained. Variance in age at menarche in the EPIC-InterAct replication sample ($N = 8,689$) explained by combined sets of SNPs defined by their strength of association in the discovery set.

Extended Data Table 1 | Details of the 123 independent signals for menarche timing at 106 genomic loci—signals no. 1 to 30

Locus	SNP	Location[1]	Novel (r-sq)[2]	N	Alleles / Freq[3]	Uni-variate Model[4]		Joint Model[5]		Gene[6]
						Beta (se)	P	Beta (se)	P	
1	rs2274465	1-43894144	Yes	179348	c/g/0.66	0.03 (0.005)	1.7E-09	n/a	n/a	*KDM4A*[NC], *PTPRF*[EC]
2	rs10789181	1-65589155	Yes	177560	a/g/0.39	0.03 (0.005)	3.5E-08	n/a	n/a	*LEPR*[C]
3	rs3101336	1-72523773	Yes	182404	t/c/0.4	0.04 (0.005)	5.2E-13	n/a	n/a	*NEGR1*[NC]
4	rs7514705	1-74779308	Yes	179631	c/t/0.56	0.04 (0.005)	1.8E-16	n/a	n/a	*TNNI3K*[N], *TYW3*[E]
5	rs11165924	1-98148036	Yes	174006	a/g/0.69	0.03 (0.006)	2.2E-09	n/a	n/a	*DPYD*[N]
6	rs11578152	1-102349609	Yes	179433	g/a/0.44	0.03 (0.005)	4.5E-08	n/a	n/a	*OLFM3*[N]
7	rs466639	1-163661506	No (Same)	179432	c/t/0.87	0.08 (0.007)	2.4E-24	n/a	n/a	*RXRG*[NC]
8	rs543874	1-176156103	No (0.91)	179613	a/g/0.8	0.05 (0.006)	1.4E-15	n/a	n/a	*SEC16B*[N]
9	rs6427782	1-198064962	Yes	175785	a/g/0.51	0.03 (0.005)	4.6E-08	n/a	n/a	*NR5A2*[NC]
10	rs951366	1-203951975	Yes	179567	t/c/0.6	0.03 (0.005)	1.7E-08	n/a	n/a	*NUCKS1*[NE], *RAB7L1*[E]
11	rs2947411	2-604168	No (Same)	179608	a/g/0.17	0.06 (0.007)	1.8E-19	n/a	n/a	*TMEM18*[NC]
12	rs6747380	2-56441253	No (1.0)	182377	a/g/0.17	0.07 (0.007)	5.6E-28	n/a	n/a	*CCDC85A*[N]
13	rs268067	2-59734549	Yes	179406	a/g/0.8	0.04 (0.006)	3.3E-08	n/a	n/a	*BCL11A*[N~800kb]
14	rs6758290	2-105231258	Yes	167496	t/c/0.5	0.04 (0.005)	6.6E-13	n/a	n/a	*GPR45*[N]
15	rs12472911	2-141944979	No (Same)	182269	c/t/0.2	0.04 (0.006)	6.7E-10	n/a	n/a	*LRP1B*[N]
16a	rs17236969	2-156460705	Yes	162496	t/c/0.14	0.05 (0.008)	2.6E-09	0.05 (0.008)	1.0E-08	*NR4A2*[NC]
16b	rs4369815	2-156835210	No (Same)	174922	t/g/0.93	0.06 (0.01)	1.5E-10	0.06 (0.01)	5.5E-10	*NR4A2*[NC]
17a	rs1400974	2-199346935	No (0.78)	179605	a/g/0.64	0.05 (0.005)	8.3E-20	0.04 (0.005)	3.0E-17	*SATB2*[N]
17b	rs17233066	2-199352283	No (0.22)	168273	c/t/0.93	0.09 (0.014)	6.1E-11	0.08 (0.014)	1.8E-09	*SATB2*[N]
17c	rs17266097	2-199983454	Yes	179181	t/c/0.42	0.04 (0.005)	3.3E-18	0.04 (0.005)	2.4E-16	*SATB2*[N]
18	rs6770162	3-24686017	Yes	179304	a/g/0.51	0.04 (0.005)	1.5E-12	n/a	n/a	*THRB*[NC]
19a	rs7647973	3-49485935	No (0.74)	179667	a/g/0.26	0.05 (0.006)	1.3E-16	0.05 (0.006)	2.4E-16	*WDR6*[EC], *UBA7*[C]
19b	rs6762477	3-50068213	No (Same)	138679	g/a/0.44	0.04 (0.006)	7.8E-12	0.04 (0.006)	2.2E-11	*WDR6*[EC], *UBA7*[C]
20	rs7642134	3-86999572	No (Same)	182263	g/a/0.61	0.04 (0.005)	3.0E-16	n/a	n/a	*POU1F1*[C] *(PIT1)*
21	rs9849248	3-88323964	Yes	179654	c/t/0.15	0.04 (0.007)	1.9E-08	n/a	n/a	*ZNF654*[NEF], *HTR1F*[C]
22	rs11715566	3-119045126	No (0.97)	179637	t/c/0.5	0.05 (0.005)	2.4E-27	n/a	n/a	*IGSF11*[N~1Mb]
23	rs2687729	3-129377916	No (Same)	179617	g/a/0.27	0.04 (0.006)	1.0E-10	n/a	n/a	*EEFSEC*[NE]
24	rs2600959	3-134098154	No (0.97)	174583	a/g/0.34	0.04 (0.005)	4.1E-11	n/a	n/a	*ACAD11*[E]
25	rs13067731	3-138472681	Yes	179330	t/c/0.16	0.04 (0.007)	1.0E-09	n/a	n/a	*IL20RB*[N]
26	rs900400	3-158281469	Yes	179649	t/c/0.61	0.03 (0.005)	2.3E-11	n/a	n/a	*LEKR1*[NE], *CCNL1*[C]
27	rs939317	3-185528493	No (0.8)	179622	g/a/0.74	0.04 (0.006)	3.0E-12	n/a	n/a	*EIF4G1*[N]
28	rs16860328	3-187118379	No (0.93)	179646	g/a/0.42	0.04 (0.005)	1.4E-16	n/a	n/a	*TRA2B*[N], *IGF2BP2*[C]
29	rs1038903	4-28361152	Yes	179610	t/c/0.73	0.04 (0.006)	2.0E-11	n/a	n/a	*PCDH7*[N~2Mb]
30	rs10938397	4-44877284	Yes	179167	a/g/0.57	0.04 (0.005)	4.0E-13	n/a	n/a	*GNPDA2*[N]

[1]All positions mapped to Hapmap build 36.
[2]Novel indicates previously unidentified loci. If the locus was established, r-sq refers to the linkage disequilibrium between the reported SNP and the previous signal. Some regions with known associations and no prior evidence for allelic heterogeneity now have multiple independent signals.
[3]Alleles/freq refers to the menarche age-increasing allele (from the uni-variate SNP discovery), and the decreasing allele/increasing allele frequencies from meta-analysis study estimates.
[4]Uni-variate models included only one SNP per model.
[5]Joint models were performed using GCTA software. These models approximate conditional analysis; that is, the effect estimates are adjusted for the effects of other neighbouring SNPs.
[6]Gene refers to the consensus gene(s) reported at that locus mapped using 4 approaches: N, nearest; C, biological candidate; F, 1000 Genomes missense variant in high LD ($r^2 > 0.8$); E, gene expression linked by eQTL. See Supplementary Tables 5, 7 and 8 for more information.

Extended Data Table 2 | Details of the 123 independent signals for menarche timing at 106 genomic loci—signals no. 31 to 58

Locus	SNP	Location[1]	Novel (r-sq)[2]	N	Alleles / Freq[3]	Uni-variate Model[4]		Joint Model[5]		Gene[6]
						Beta (se)	P	Beta (se)	P	
31	rs13135934	4-95426711	Yes	178661	c/g/0.4	0.03 (0.005)	1.1E-10	n/a	n/a	*SMARCAD1*[NEF]
32	rs3733631	4-104860552	Yes	179623	c/g/0.15	0.05 (0.007)	4.8E-13	n/a	n/a	*TACR3*[NC]
33	rs1532331	5-43152587	Yes	179201	g/t/0.32	0.03 (0.005)	3.5E-09	n/a	n/a	*ZNF131*[NEC], *GHR*[C]
34	rs17086188	5-95871610	Yes	176967	a/g/0.94	0.07 (0.013)	3.6E-08	n/a	n/a	*PCSK1*[NC]
35	rs251130	5-110887696	Yes	179429	g/a/0.73	0.04 (0.006)	2.8E-10	n/a	n/a	*STARD4*[NEC]
36	rs13179411	5-133928412	No (0.53)	179579	t/g/0.17	0.06 (0.007)	3.4E-20	n/a	n/a	*PHF15*[N], *TCF7*[E]
37	rs17171818	5-137752902	No (1.0)	182224	c/t/0.77	0.04 (0.006)	8.9E-14	n/a	n/a	*KDM3B*[NC], *BRD8*[C]
38	rs7701886	5-153527602	Yes	179664	a/g/0.58	0.03 (0.005)	4.5E-08	n/a	n/a	*GALNT10*[N]
39	rs9647570	5-167302841	Yes	179600	g/t/0.14	0.05 (0.007)	1.4E-11	n/a	n/a	*TENM2*[NC]
40	rs6555855	5-168682315	Yes	179462	g/a/0.23	0.04 (0.006)	2.4E-09	n/a	n/a	*SLIT3*[N]
41	rs16896742	6-30030719	Yes	171665	g/a/0.38	0.04 (0.006)	3.2E-10	n/a	n/a	*HLA-A*[N]
42	rs2479724	6-41998960	Yes	179630	t/c/0.45	0.03 (0.005)	1.2E-12	n/a	n/a	*BYSL*[NE], *FRS3*[C]
43	rs988913	6-54864267	Yes	182407	c/t/0.66	0.04 (0.005)	1.4E-12	n/a	n/a	*FAM83B*[N], *HCRTR2*[C]
44	rs9475752	6-56888700	Yes	178646	c/t/0.81	0.04 (0.006)	8.3E-12	n/a	n/a	*DST*[N], *BEND6*[E]
45	rs9447700	6-77224806	Yes	179648	c/t/0.69	0.03 (0.005)	5.6E-09	n/a	n/a	*IMPG1*[N]
46a	rs9321659	6-100222813	Yes	182356	a/g/0.13	0.06 (0.008)	2.5E-16	0.06 (0.008)	2.9E-16	*SIM1*[C], *MCHR2*[C]
46b	rs4840086	6-100315159	No (Same)	179666	a/g/0.58	0.04 (0.005)	9.2E-14	0.04 (0.005)	4.3E-13	*SIM1*[C], *MCHR2*[C]
46c	rs13196561	6-100866891	Yes	182278	c/a/0.78	0.04 (0.006)	8.4E-12	0.06 (0.006)	3.4E-20	*SIM1*[NC], *MCHR2*[C]
46d	rs239198	6-101240798	Yes	179496	t/c/0.46	0.03 (0.005)	2.5E-08	0.04 (0.005)	3.1E-15	*SIM1*[C], *ASCC3*[NEF]
47a	rs4946632	6-105207901	Yes	132973	c/t/0.1	0.01 (0.01)	0.14	-0.07 (0.01)	3.1E-12	*LIN28B*[C]
47b	rs2153127	6-105455237	Yes	182110	t/c/0.52	0.08 (0.005)	5.5E-59	0.03 (0.006)	2.1E-09	*LIN28B*[EC]
47c	rs7759938	6-105485647	No (Same)	179557	c/t/0.32	0.12 (0.005)	7.8E-110	0.11 (0.006)	1.2E-69	*LIN28B*[NC]
48	rs4895808	6-126823127	No (1.0)	179655	c/t/0.54	0.03 (0.005)	4.8E-13	n/a	n/a	*CENPW*[NE], *NCOA7*[C]
49	rs6938574	6-128432673	Yes	178428	t/c/0.16	0.04 (0.007)	2.4E-09	n/a	n/a	*PTPRK*[N]
50	rs6933660	6-151845447	Yes	182379	c/a/0.69	0.03 (0.005)	1.3E-09	n/a	n/a	*ESR1*[C]
51	rs1079866	7-41436618	No (Same)	172036	g/c/0.15	0.07 (0.007)	9.3E-24	n/a	n/a	*INHBA*[NC]
52	rs6964833	7-73739845	Yes	171484	t/c/0.75	0.04 (0.006)	5.3E-12	n/a	n/a	*GTF2I*[NC]
53	rs11767400	7-121947978	Yes	179658	a/c/0.3	0.04 (0.006)	4.1E-11	n/a	n/a	*CADPS2*[N]
54a	rs2688325	8-3754618	Yes	182244	t/c/0.29	0.03 (0.006)	2.1E-09	0.03 (0.006)	9.7E-10	*CSMD1*[N]
54b	rs7828501	8-4547489	Yes	179434	g/a/0.45	0.04 (0.005)	1.2E-13	0.04 (0.005)	2.8E-15	*CSMD1*[N]
54c	rs7463166	8-4821198	Yes	179542	a/g/0.63	0.03 (0.005)	1.3E-08	0.03 (0.005)	5.9E-09	*CSMD1*[N]
55	rs16918254	8-53931766	Yes	179635	a/g/0.92	0.05 (0.009)	1.4E-08	n/a	n/a	*NPBWR1*[NC]
56	rs7821178	8-78256392	No (Same)	179533	c/a/0.65	0.04 (0.005)	7.3E-17	n/a	n/a	*PEX2*[N]
57	rs1469039	8-140720961	Yes	174755	a/g/0.19	0.05 (0.007)	3.5E-12	n/a	n/a	*KCNK9*[N]
58	rs4875053	8-144944399	Yes	136628	g/c/0.44	0.03 (0.006)	1.3E-08	n/a	n/a	*SCRIB*[N], *PARP10*[E]

[1]All positions mapped to Hapmap build 36.
[2]Novel indicates previously unidentified loci. If the locus was established, r-sq refers to the linkage disequilibrium between the reported SNP and the previous signal. Some regions with known associations and no prior evidence for allelic heterogeneity now have multiple independent signals.
[3]Alleles/freq refers to the menarche age-increasing allele (from the uni-variate SNP discovery), and the decreasing allele/increasing allele frequencies from meta-analysis study estimates.
[4]Uni-variate models included only one SNP per model.
[5]Joint models were performed using GCTA software. These models approximate conditional analysis; that is, the effect estimates are adjusted for the effects of other neighbouring SNPs.
[6]Gene refers to the consensus gene(s) reported at that locus mapped using 4 approaches: N, nearest; C, biological candidate; F, 1000 Genomes missense variant in high LD ($r^2 > 0.8$); E, gene expression linked by eQTL. See Supplementary Tables 5, 7 and 8 for more information.

Extended Data Table 3 | Details of the 123 independent signals for menarche timing at 106 genomic loci—signals no. 59 to 87

						Uni-variate Model[4]		Joint Model[5]		
Locus	SNP	Location[1]	Novel (r-sq)[2]	N	Alleles / Freq[3]	Beta (se)	P	Beta (se)	P	Gene[6]
59a	rs7037266	9-6932940	Yes	179488	a/c/0.37	0.03 (0.005)	4.7E-09	0.03 (0.005)	3.5E-09	*KDM4C*[NC]
59b	rs913588	9-7164673	Yes	182403	g/a/0.49	0.03 (0.005)	5.8E-11	0.03 (0.005)	3.8E-11	*KDM4C*[NFC]
60	rs7865468	9-10264080	Yes	179418	a/g/0.7	0.03 (0.005)	1.3E-07	0.03 (0.005)	1.9E-08	*PTPRD*[N]
61	rs7853970	9-85905386	Yes	169702	t/c/0.47	0.03 (0.005)	2.3E-09	n/a	n/a	*RMI1*[N], *NTRK2*[C]
62a	rs10816359	9-107797491	Yes	169277	t/g/0.86	0.04 (0.008)	1.6E-08	0.05 (0.008)	1.2E-12	*TMEM38B*[N]
62b	rs10453225	9-107960041	No (0.73)	179631	g/t/0.68	0.09 (0.005)	5.8E-66	0.07 (0.006)	3.5E-33	*TMEM38B*[N]
62c	rs10739221	9-108100651	No (0.42)	179624	c/t/0.77	0.08 (0.006)	3.9E-41	0.05 (0.007)	1.9E-11	*TMEM38B*[N]
63	rs11792861	9-110849116	Yes	179618	a/c/0.7	0.04 (0.005)	1.7E-11	n/a	n/a	*TMEM245*[NE]
64a	rs10980854	9-113090178	Yes	181999	a/g/0.06	0.06 (0.011)	1.3E-08	0.06 (0.011)	4.3E-09	*ZNF483 / OR2K2*[N]
64b	rs10980921	9-113319733	No (0.12)	172160	c/t/0.09	0.09 (0.009)	1.7E-23	0.09 (0.009)	4.3E-23	*ZNF483 / OR2K2*[N]
65	rs1874984	10-1721871	Yes	179112	c/g/0.47	0.04 (0.005)	1.9E-12	n/a	n/a	*ADARB2*[N]
66	rs12571664	10-121698919	Yes	179629	t/c/0.79	0.04 (0.006)	3.3E-10	n/a	n/a	*SEC23IP*[NE]
67	rs1915146	10-126836204	Yes	182401	g/a/0.4	0.03 (0.005)	3.7E-08	n/a	n/a	*CTBP2*[NC]
68	rs7104764	11-219977	Yes	179664	g/a/0.25	0.03 (0.006)	3.7E-08	n/a	n/a	*SIRT3*[NEC]
69	rs4929947	11-8596570	No (1.0)	179331	g/c/0.36	0.04 (0.005)	2.6E-12	n/a	n/a	*TRIM66*[NEF]
70	rs11022756	11-13272015	No (0.88)	179401	a/c/0.29	0.05 (0.006)	7.4E-20	n/a	n/a	*ARNTL*[N], *PTH*[C]
71	rs7103411	11-27656701	Yes	179656	c/t/0.21	0.04 (0.006)	2.6E-11	n/a	n/a	*BDNF*[NC], *LGR4*[C]
72	rs16918636	11-29080758	Yes	182237	t/c/0.79	0.03 (0.006)	3.2E-08	n/a	n/a	*FSHB*[CN~1Mb]
73	rs4756059	11-46107195	No (0.65)	179478	t/c/0.92	0.07 (0.01)	4.5E-13	n/a	n/a	*PHF21A*[N]
74	rs2063730	11-77726172	No (0.75)	179293	c/a/0.18	0.05 (0.007)	2.3E-12	n/a	n/a	*GAB2*[N], *THRSP*[C]
75	rs10895140	11-100941931	Yes	179647	g/a/0.66	0.04 (0.005)	6.7E-14	n/a	n/a	*TRPC6*[N], *PGR*[C]
76	rs11215400	11-114557845	Yes	179376	c/a/0.27	0.04 (0.006)	6.8E-11	n/a	n/a	*CADM1*[N]
77	rs1461503	11-122350285	No (0.34)	179603	c/a/0.57	0.05 (0.005)	2.7E-26	n/a	n/a	*BSX*[NC]
78	rs7955374	12-46166416	Yes	179419	t/c/0.13	0.04 (0.008)	9.5E-09	n/a	n/a	*VDR*[C]
79	rs7138803	12-48533735	Yes	174834	g/a/0.62	0.04 (0.005)	1.7E-12	n/a	n/a	*BCDIN3D*[N]
80	rs6563739	13-39137785	Yes	179667	g/t/0.34	0.03 (0.005)	2.3E-11	n/a	n/a	*COG6*[NE]
81	rs1324913	13-73533589	Yes	182393	g/t/0.65	0.03 (0.005)	3.1E-10	n/a	n/a	*KLF12*[N]
82	rs9560113	13-110981349	No (1.0)	179359	g/a/0.28	0.05 (0.006)	2.1E-17	n/a	n/a	*TEX29*
83	rs1254337	14-59990278	Yes	179658	t/a/0.31	0.04 (0.005)	2.1E-16	n/a	n/a	*SIX6*[N]
84	rs1958560	14-65106548	Yes	179655	a/g/0.59	0.03 (0.005)	3.7E-08	n/a	n/a	*FUT8*[NE]
85a	rs10144321	14-99952158	Yes	179595	a/g/0.75	0.04 (0.006)	9.0E-15	0.04 (0.006)	1.1E-14	*DLK1*[C], *WDR25*[E]
85b	rs7141210	14-100252223	Yes	172034	t/c/0.34	0.03 (0.005)	5.8E-09	0.03 (0.005)	4.1E-09	*DLK1*[NEC]
86	rs12148769	15-21703187	Yes	182411	g/a/0.9	0.05 (0.008)	5.2E-11	n/a	n/a	*MKRN3*[C], *MAGEL2*[C]
87	rs3743266	15-58568805	No (Same)	182389	t/c/0.68	0.04 (0.005)	2.4E-13	n/a	n/a	*RORA*[NC]

[1]All positions mapped to Hapmap build 36.
[2]Novel indicates previously unidentified loci. If the locus was established, r-sq refers to the linkage disequilibrium between the reported SNP and the previous signal. Some regions with known associations and no prior evidence for allelic heterogeneity now have multiple independent signals.
[3]Alleles/freq refers to the menarche age-increasing allele (from the uni-variate SNP discovery), and the decreasing allele/increasing allele frequencies from meta-analysis study estimates.
[4]Uni-variate models included only one SNP per model.
[5]Joint models were performed using GCTA software. These models approximate conditional analysis; that is, the effect estimates are adjusted for the effects of other neighbouring SNPs.
[6]Gene refers to the consensus gene(s) reported at that locus mapped using 4 approaches: N, nearest; C, biological candidate; F, 1000 Genomes missense variant in high LD ($r^2 > 0.8$); E, gene expression linked by eQTL. See Supplementary Tables 5, 7 and 8 for more information.

Extended Data Table 4 | Details of the 123 independent signals for menarche timing at 106 genomic loci—signals no. 88 to 106

						Uni-variate Model[4]		Joint Model[5]		
Locus	SNP	Location[1]	Novel (r-sq)[2]	N	Alleles / Freq[3]	Beta (se)	P	Beta (se)	P	Gene[6]
88	rs8032675	15-65746518	No (0.39)	179630	t/c/0.4	0.04 (0.005)	2.1E-13	n/a	n/a	*MAP2K5*[N]
89	rs12915845	15-86843471	Yes	179535	c/t/0.58	0.03 (0.005)	2.7E-12	n/a	n/a	*DET1*[NE]
90	rs246185	16-14302933	Yes (0.84)	177773	c/t/0.33	0.04 (0.006)	6.8E-16	n/a	n/a	*MKL2*[N]
91	rs12446632	16-19842890	Yes	182401	a/g/0.13	0.04 (0.007)	1.3E-08	n/a	n/a	*GPRC5B*[NC]
92	rs1129700	16-29825535	Yes	181797	t/c/0.44	0.03 (0.005)	2.3E-09	n/a	n/a	*KCTD13*[N], *TBX6*[EC]
93	rs8050136	16-52373776	No (1.0)	182365	c/a/0.6	0.04 (0.005)	1.7E-17	n/a	n/a	*FTO*[NC]
94a	rs1364063	16-68146073	No (Same)	182393	c/t/0.43	0.05 (0.005)	6.2E-21	0.04 (0.005)	4.8E-18	*COG4*[C], *NFAT5*[N]
94b	rs929843	16-68603249	Yes	177329	a/c/0.23	0.04 (0.006)	1.2E-11	0.04 (0.006)	5.9E-09	*COG4*[C], *WWP2*[N]
95	rs7215990	17-5975555	Yes	170053	g/a/0.76	0.04 (0.006)	1.9E-08	n/a	n/a	*WSCD1*[NE] *ALOX15B*[E]
96	rs9635759	17-46968784	No (Same)	179649	a/g/0.32	0.05 (0.005)	1.7E-24	n/a	n/a	*CA10*[N]
97	rs244293	17-50585721	Yes	179560	g/a/0.6	0.03 (0.005)	4.2E-11	n/a	n/a	*STXBP4*[NE]
98	rs12607903	18-3807134	Yes	179171	c/t/0.3	0.04 (0.005)	5.4E-11	n/a	n/a	*DLGAP1*[N]
99	rs2137289	18-43006123	No (0.74)	178617	a/g/0.59	0.05 (0.005)	8.2E-20	n/a	n/a	*SKOR2*[N]
100	rs652260	19-7806562	Yes	182356	t/c/0.54	0.03 (0.005)	9.9E-09	n/a	n/a	*EVI5L*[N], *RETN*[C]
101	rs889122	19-9856867	No (0.33)	179397	g/t/0.72	0.04 (0.006)	1.6E-13	n/a	n/a	*OLFM2*[N], *RDH8*[C]
102	rs10423674	19-18678903	No (Same)	182377	a/c/0.34	0.04 (0.005)	9.2E-12	n/a	n/a	*CRTC1*[NC]
103	rs852069	20-17070593	No (Same)	182413	g/a/0.64	0.04 (0.005)	1.2E-13	n/a	n/a	*PCSK2*[NC]
104	rs2836950	21-39526299	Yes	178602	c/g/0.64	0.03 (0.005)	6.2E-11	n/a	n/a	*BRWD1*[NC]
105	rs13053505	22-37575564	Yes	177596	g/t/0.8	0.04 (0.007)	3.0E-08	n/a	n/a	*NPTXR*[NE], *CBX7*[C]
106	rs6009583	22-48063650	Yes	181839	c/t/0.74	0.03 (0.006)	4.6E-08	n/a	n/a	*C22orf34*[N]

[1]All positions mapped to Hapmap build 36.
[2]Novel indicates previously unidentified loci. If the locus was established, r-sq refers to the linkage disequilibrium between the reported SNP and the previous signal. Some regions with known associations and no prior evidence for allelic heterogeneity now have multiple independent signals.
[3]Alleles/freq refers to the menarche age-increasing allele (from the uni-variate SNP discovery), and the decreasing allele/increasing allele frequencies from meta-analysis study estimates.
[4]Uni-variate models included only one SNP per model.
[5]Joint models were performed using GCTA software. These models approximate conditional analysis; that is, the effect estimates are adjusted for the effects of other neighbouring SNPs.
[6]Gene refers to the consensus gene(s) reported at that locus mapped using 4 approaches: N, nearest; C, biological candidate; F, 1000 Genomes missense variant in high LD ($r^2 > 0.8$); E, gene expression linked by eQTL. See Supplementary Tables 5, 7 and 8 for more information.

Extended Data Table 5 | Methylation QTLs based on Illumina 450K whole blood and adipose methylome data in 648 twins

				Adipose tissue			Whole blood	
Locus	SNP	Consensus gene	Methylation probe[1,2]	Beta[3]	SE	P	Beta[3]	P
16b	rs4369815	*NR4A2* (N,C)	cg14912644	0.006	0.002	7.3E-04	-	-
33	rs1532331	*ZNF131* (N,E,C), *GHR* (C)	cg18254356	-0.01	0.003	4.4E-04	-	-
36	rs13179411	*PHF15* (N), *TCF7* (E)	cg00043364	-0.02	0.003	7.9E-11	-0.35	7.3E-03
64b	rs10980921	*ZNF483* / *OR2K2* (N)	cg01294431	0.01	0.002	1.1E-08	-	-
67	rs1915146	*CTBP2* (N,C)	cg17191109	0.01	0.001	6.9E-16	0.75	2.8E-18
83	rs1254337	*SIX6* (N)	cg00157572	-0.005	0.001	3.8E-05	-	-
85b	rs7141210	*DLK1* (N,E,C)	cg17008318	0.02	0.002	1.3E-18	-	-
100	rs652260	*EVI5L* (N), *RETN* (C)	cg06793867	-0.03	0.003	1.3E-23	-	-
100	rs652260	*EVI5L* (N), *RETN* (C)	cg14209047	0.01	0.002	2.4E-12	0.35	1.9E-04
100	rs652260	*EVI5L* (N), *RETN* (C)	cg15974673	-0.03	0.003	4.8E-27	-0.6	2.1E-11
102	rs10423674	*CRTC1* (N,C)	cg19861427	-0.007	0.002	1.4E-05	-	-

[1]Methylation-QTLs were derived for associations between genotypes and methylation in 648 adipose samples from the MuTHER study using a 1% FDR level, corresponding to $P < 8.6 \times 10^{-41}$. Significant methylation-QTLs were also tested for replication in whole blood in 200 individuals.
[2]Methylation data available from ref. 9.
[3]Methylation betas are presented per menarche-age-increasing allele.

Extended Data Table 6 | MAGENTA pathway analyses

			95th percentile enrichment cut-off			75th percentile enrichment cut-off		
Database	Gene set	Genes (mapped)[1]	P	FDR	Enrichment[2] Exp. (obs.)	P	FDR	Enrichment[2] Exp. (obs.)
Panther	$GABA_B$ receptor II signaling	9 (9)	8.00E-04	9.25E-03	0 (4)	9.70E-03	1.12E-01	2 (6)
Panther	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	5 (5)	6.00E-04	1.39E-02	0 (3)	1.39E-02	9.78E-02	1 (4)
GOTERM	Regulation of transcription	991 (844)	1.30E-05	2.65E-01	42 (69)	1.00E-06	7.00E-04	211 (271)
GOTERM	Transcription factor activity	947 (788)	4.51E-03	4.19E-01	39 (55)	2.40E-05	3.89E-02	197 (242)
BIOCARTA	ETC_PATHWAY	12 (9)	3.78E-01	5.59E-01	0 (1)	1.20E-03	4.23E-02	2 (7)
GOTERM	Chromatin assembly or disassembly	38 (31)	4.69E-01	9.05E-01	2 (2)	1.10E-05	1.15E-02	8 (19)
Panther	5HT3 type receptor mediated signaling	7 (5)	1.00E+00	9.27E-01	0 (0)	1.10E-03	1.65E-02	1 (5)
Custom	Nuclear hormone receptors	57 (55)	6.00E-05	6.00E-05	3 (11)	4.58E-03	9.60E-03	14 (23)
Custom	Lysine specific demethylases	24 (24)	5.60E-03	5.60E-03	1 (5)	1.24E-01	1.24E-01	6 (9)
Custom	Mendelian pubertal disorders[3]	20 (18)	5.30E-02	5.30E-02	1 (3)	1.38E-01	1.38E-01	5 (7)

Results are shown for database pathways and custom pathways that reached study-wise statistical significance (FDR <0.05).
[1]Genes denotes number of genes in pathway (number of genes successfully mapped by MAGENTA).
[2]Enrichment denotes expected number of genes at enrichment threshold (observed number of genes).
[3]Genes for Mendelian pubertal disorders, as described in refs 2 and 3.

Extended Data Table 7 | $GABA_B$ receptor II signalling pathway genes

Gene	Gene P	Gene size(kb)	Number of SNPs	Number of Recombination Hotspots	Best SNP	Best SNP p value
ADCY8	2.87E-03	260	489	9	rs4392877	6.83E-08
ADCY6	4.89E-03	23	92	3	rs2446999	8.70E-07
GABBR1	9.32E-03	31	405	2	rs1362126	1.33E-06
PRKAR2A	9.04E-03	97	59	2	rs11713694	1.99E-06
PRKAR2B	2.81E-01	117	209	4	rs2244846	1.17E-03
ADCY9	3.42E-01	154	309	7	rs879150	1.51E-03
GABBR2	5.51E-01	421	698	10	rs2485144	2.86E-03
ADCY1	6.08E-01	149	184	3	rs10951832	1.27E-02
ADCY5	7.13E-01	164	207	5	rs9880405	2.31E-02