Published in final edited form as:
Nature. 2014 October 2; 514(7520): 92–97. doi:10.1038/nature13545.

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

A full list of authors and affiliations appears at the end of the article.
[#] These authors contributed equally to this work.

Abstract

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 ($P<5\times10^{-8}$) 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 gamma-aminobutyric acid-B2 receptor signaling, among novel mechanisms that regulate pubertal timing in humans. Our findings suggest

[*]**Correspondence to:** John R.B. Perry (john.perry@mrc-epid.cam.ac.uk) and Joanne Murabito (Murabito@bu.edu) .

Author Contributions

Overall project management
JRBP, FD, CEE, PS, DJT, DFE, KS, JMM, KKO

Core analyses
JRBP, FD, CEE, PS, TF, DJT, DIC, TE

Individual study analysts
AAR, AD, AG, AJ, AT, AVS, BZA, BF, CEE, DFG, DIC, DJT, DLC, DLK, EA, EKW, EM, EMB, ET, FD, GM, GmM, IMN, JAV, JD, JH, JRBP, JT, JZ, KLL, KM, LLP, LMR, LMY, LS, MM, NF, NTs, PK, PS, RM, SK, SS, SSU, TC, TE, TF, TFo, THP, WQA, ZK

Individual study data management and generation
AAR, ACH, AD, ADC, AGU, AJO, AMS, AMu, AP, APo, BAO, CAH, DC, DIC, DJH, DK, DLw, DPK, DPS, DS, EAN, EP, EW, FA, FBH, FG, FR, GD, GE, GGW, HS, HW, ID, JC, JH, JPR, LF, LFr, LM, LMR, MEG, MJS, MJW, MKB, MMb, MP, MW, NA, NJT, NLP, PKM, QW, RH, SB, SC, SG, SL, SR, SSU, TE, US, UT, VS, WLM

Individual study PIs
AC, AGU, AH, AJO, AKD, AL, AM, AMD, AMm, AMu, AR, BB, BZA, BHRW, CB, CEP, CG, CH, CMv, DIB, DF, DFE, DJH, DL, DLw, DSP, DPS, DSs, EAS, EB, EEJd, EI, EW, EWD, FBH, FJC, GC, GD, GGG, GW, GW, GWM, HA, HAB, HB, HBe, HF, HN, HS, HV, ID, ILA, JAK, JB, JCC, JGE, JEB, JLH, JMC, JMM, JP, KC, KK, KKO, KP, KS, LC, LF, LJB, MCS, MG, MIM, MJ, MJE, MJH, MJS, MKS, MWB, MZ, NGM, NJW, PAF, PD, PDPP, PFM, PG, PH, PK, PMR, PN, PP, PPG, PR, PV, RJFL, RLM, RW, SB, SBm, SC, SEB, TBH, TDS, TIAS, UH, VG, VK, VS

Supplementary Information Supplementary information contains: Supplementary Tables and Figures, acknowledgements, author disclosures.

Data deposition statement
Plots of all 106 menarche loci and genome-wide summary level statistics are available at the ReproGen Consortium website: www.reprogen.org.

Europe PMC Funders Author Manuscripts

a genetic architecture involving at least hundreds of common variants in the coordinated timing of the pubertal transition.

Genome-wide array data were available on up to 132,989 women of European descent from 57 studies, and data on up to ~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] were available on an additional 49,427 women (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 ($P<5\times10^{-8}$) for association with age at menarche (Figure 1). Using GCTA[5], which approximates a conditional analysis adjusted for the effects of neighbouring SNPs (Extended Data Figure 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 www.reprogen.org). Of the 42 previously reported independent signals for age at menarche[4], all but one (rs2243803, *SLC14A2*, $P=2.3\times10^{-6}$) remained genome-wide significant in the expanded dataset.

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 $P_{Sign}=2.2\times10^{-15}$), 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% ($P<1\times10^{-20}$) of the variance in age at menarche, compared to 1.31% ($P=2.3\times10^{-14}$) 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%, $P=2.2\times10^{-6}$), indicating a highly polygenic architecture (Extended Data Figure 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 years 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 ($P_{Sign}=1.1\times10^{-13}$), 86/106 in girls only ($P_{Sign}=6.2\times10^{-11}$) 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/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

Europe PMC Funders Author Manuscripts

included 35,377 women with parental origins of alleles determined by a combination of genealogy and long-range phasing[6].

Two independent signals (#85a-b; 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, $P_{pat}=3.1\times10^{-5}$; rs7141210, $P_{pat}=2.1\times10^{-4}$), 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, $P_{het}=2.2\times10^{-4}$) (Figure 2; Supplementary Table 5). Notably, rs7141210 is reportedly a *cis*-acting methylation-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* ($P_{mat}<5.6\times10^{-53}$), *MEG8* ($P_{mat}=4.9\times10^{-41}$) and *MEG9* ($P_{mat}=5.4\times10^{-5}$); 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.

Signal #86 (rs12148769) lies in the imprinted critical region for Prader Willi Syndrome (PWS), 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 ($P_{pat}=2.4\times10^{-6}$), but the maternally-inherited allele was not ($P_{mat}=0.43$; $P_{het}=5.6\times10^{-3}$) (Figure 2). Recently, truncating mutations of *MAGEL2* affecting the paternal alleles were reported in PWS; 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 #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 ($P_{mat}=5.6\times10^{-6}$), but the paternally-inherited allele was not ($P_{pat}=0.76$; $P_{het}=3.7\times10^{-3}$) (Figure 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 4 (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 #1 (rs2274465) is intronic in *KDM4A*, signal #37 (rs17171818) is intronic in *KDM3B*, and signal #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 PWS 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, $P=2.3\times10^{-6}$). Pubertal onset in female mice is reportedly triggered by DNA methylation of the Polycomb group silencing complex of genes (including *CBX7* near signal #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/near genes that underlie rare Mendelian disorders of puberty (enrichment P=0.05)[2,3]. As well as rs12148769 near to *MKRN3*, signals were found near *LEPR/LEPROT* (signal #2; rs10789181), which encodes the leptin receptor, and immediately upstream of *TACR3* (signal #32; rs3733631), which encodes the receptor for Neurokinin B. A further variant ~10 kb from *GNRH1* approached genome-wide significance (rs1506869, $P=1.8\times10^{-6}$) and was also associated with *GNRH1* expression in adipose tissue ($P=3.7\times10^{-5}$). Signals #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/near several further genes with relevance to pituitary development/function included: signal #20 (rs7642134) near *POU1F1*, signal #39 (rs9647570) within *TENM2*, and signal #42 (rs2479724) near *FRS3*. Furthermore, signals #71 (rs7103411) and #92 (rs1129700) are *cis*-eQTLs for *LGR4* and *TBX6*, respectively, both of which encode enhancers for the pituitary development factor *SOX2*. Signals #52 (rs6964833 intronic in *GTF2I*) and #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/near genes with possible roles in hormonal functions (Figure 3, Supplementary Table 8), many more than the three signals we described previously (*INHBA*, *PCSK2* and *RXRG*)[4]. Two signals were found in/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 #9 (rs6427782) is near *NR5A2*, which encodes a nuclear receptor with key roles in steroidogenesis and estrogen-dependent cell proliferation.

We observed that SNPs in/near a custom list of genes that encode nuclear hormone receptors, co-activators or co-repressors were enriched for associations with menarche

timing (enrichment $P=6\times10^{-5}$). 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 (RA) signaling. SNPs in/near *RXRG* and *RORA* reached genome-wide significance, and three other genes contained sub-genome-wide signals (*RXRA* [rs2520094, $P=4\times10^{-7}$], *RORB* [rs4237264, $P=9.4\times10^{-6}$], *RXRB* [rs241438, $P=7.1\times10^{-5}$]). Two other genome-wide significant signals mapped to genes with roles in RA function (#67 *CTBP2* and #101 *RDH8*). The active metabolites of vitamin A, all-*trans*-RA and 9-*cis*-RA, have differential effects on GnRH expression and secretion[17]. Other possible mechanisms linking RA signaling to pubertal timing include inhibition of embryonic GnRH neuron migration, and enhancement of steroidogenesis and gonadotrophin 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 (FDR=0.009) was gamma-aminobutyric acid ($GABA_B$) receptor II signaling (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 BMI[21]. At all nine loci (in/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 #47c (rs7759938, *LIN28B*) and #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 #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 #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 #48 (rs4895808, a *cis*-eQTL for *CENPW*) is in 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 ($P=6.49\times10^{-12}$). Signal #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 #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 ($P=6\times10^{-5}$)[27], which is directionally

discordant with the epidemiological association[28]. Signal #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=9.3×10^{-10}).

Our findings indicate both BMI-related and BMI-independent mechanisms that could underlie the epidemiological associations between early menarche and higher risks of adult disease[1].These include actions of *LIN28B* on insulin sensitivity through the mTOR pathway, $GABA_B$ receptor signaling on inhibition of oxidative stress-related ß-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.

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 datasets if they were poorly imputed or were rare (MAF <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<5×10^{-8}) for association with age at menarche (Figure 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].

Europe PMC Funders Author Manuscripts

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 Mb. 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 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 Figure 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 (r2>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. (2013)[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 Celiac 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 tumor 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 datasets 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 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 datasets.

Pathway analyses

Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) was used to explore pathway-based associations in the full GWAS dataset. 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[22,21] and the GEnetic Factors for OS (GEFOS) consortium[27]. In addition, we

Europe PMC Funders Author Manuscripts

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].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

John RB 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 Coveillo[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 EJ 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 DP 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

Europe PMC Funders Author Manuscripts

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 HR Wolffenbuttel[141], Margaret J Wright[142], Australian Ovarian Cancer Study[42,143], The GENICA Network[61,62,99,144,145,146,147], kConFab[143], The LifeLines Cohort Study, The InterAct Consortium, Early Growth Genetics (EGG) Consortium, Dorret I Boomsma[43], Michael J Econs[44,123], Kay-Tee Khaw[148], Ruth JF Loos[1,149], Mark I McCarthy[3,150,151], Grant W Montgomery[142], John P Rice[40], Elizabeth A Streeten[47,152], Unnur Thorsteinsdottir[5,95], Cornelia M van Duijn[34,39,153], Behrooz Z Alizadeh[28], Sven Bergmann[17,18], Eric Boerwinkle[154], Heather A Boyd[20], Laura Crisponi[29], Paolo Gasparini[38,75], Christian Gieger[15], Tamara B Harris[89], Erik Ingelsson[155], Marjo-Riitta Järvelin[133,156,157,158,159], Peter Kraft[76,160], Debbie Lawlor[26,27], Andres Metspalu[11,36], Craig E Pennell[16], Paul M Ridker[9,10], Harold Snieder[28], Thorkild IA Sørensen[161,162], Tim D Spector[4], David P Strachan[163], André G Uitterlinden[33,34,103], Nicholas J Wareham[1], Elisabeth Widen[19], Marek Zygmunt[164], Anna Murray[2], Douglas F Easton[6], Kari Stefansson[#5,95], Joanne M Murabito[#23,165,*], and Ken K Ong[#1,166]

Affiliations

[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, UK EX1 2LU. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [5]deCODE Genetics, Reykjavik, Iceland. [6]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, UK. [7]Department of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, IN 46202, USA. [8]Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN 46202, USA. [9]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215. [10]Harvard Medical School, Boston, MA 02115. [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, MA 02115, USA. [13]Broad Institute of the Massachusetts Institute of Technology and Harvard University, 140 Cambridge 02142, MA, USA. [14]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [15]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [16]School of Women's and Infants' Health, The University of Western Australia. [17]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [18]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [19]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland. [20]Department of Epidemiology Research, Statens Serum Institut, DK-2300 Copenhagen, Denmark. [21]Department of Epidemiology, University of North Carolina, Chapel Hill, NC.

[22]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. [23]NHLBI's and Boston University's Framingham Heart Study, Framingham, MA. [24]Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden. [25]Boston University School of Public Health, Department of Biostatistics. Boston, MA. [26]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, 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, Groningen, The Netherlands. [29]Institute of Genetics and Biomedical Research, National Research Council, Cagliari, Italy. [30]University of Sassari, Dept. Of Biomedical Sciences, Sassari, Italy. [31]Icelandic Heart Association, Kopavogur, Iceland. [32]University of Iceland, Reykjavik, Iceland. [33]Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. [34]Netherlands Consortium on Health Aging and National Genomics Initiative, Leiden, the Netherlands. [35]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, 17475 Greifswald, Germany. [36]Department of Biotechnology, University of Tartu, Tartu, 51010, Estonia. [37]Hjelt Institute, University of Helsinki, Finland. [38]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo" – Trieste, Italy. [39]Genetic Epidemiology Unit Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. [40]Dept. of Psychiatry, Washington University, St. Louis, MO 63110. [41]The University of Queensland, Queensland Brain Institute, St.Lucia, QLD, Australia. [42]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 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 USA. [45]Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, US. [46]Center for Biological Sequence Analysis, Department of Systems Biology, Technical 142 University of Denmark, Lyngby 2800, Denmark. [47]Program in Personalized and Genomic Medicine, and Department of Medicine, Division of Endocrinology, Diabetes and Nutrition - University of Maryland School of Medicine, USA. Baltimore, MD 21201. [48]Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. [49]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [50]Department of Epidemiology, University of California Irvine, Irvine, California, USA. [51]Sanquin Research, Nijmegen, The Netherlands. [52]Tuscany Regional Health Agency, Florence, Italy, I.O.T. and Department of Medical and Surgical Critical Care, University of Florence, Florence, Italy. [53]Geriatric Unit, Azienda Sanitaria di Firenze, 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, Erlangen, Germany. [55]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [56]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain. [57]Department of Oncology,

University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [58]Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark. [59]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark. [60]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), Milan, Italy. [61]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart. [62]University of Tübingen, Germany. [63]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [64]German Cancer Consortium (DKTK), Heidelberg, Germany. [65]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [66]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [67]Departments of Anatomy and Neurological Surgery, Indiana University school of Medicine, Indianapolis, IN 46202, USA. [68]Stark Neuroscience Research Center, Indiana University school of Medicine, Indianapolis, IN 46202, USA. [69]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia. [70]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. [71]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [72]Department of Biostatistics, University of North Carolina, Chapel Hill, NC. [73]Boston University School of Medicine, Department of Medicine, Sections of Preventive Medicine and Endocrinology, Boston, MA. [74]Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, UK. [75]Department of Clinical Medical Sciences, Surgical and Health, University of Trieste, Italy. [76]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA. [77]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. [78]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minn., 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, UK. [81]National Institute for Health and Welfare, Finland. [82]Department of General Practice and Primary health Care, University of Helsinki, Finland. [83]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland. [84]Folkhalsan Research Centre, Helsinki, Finland. [85]Longitudinal Studies Section, Clinical Research Branch, Gerontology Research Center, National Institute on Aging, Baltimore, Maryland, United States of America. [86]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. [87]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark. [88]Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands. [89]National Insitute on Aging, National Institutes of Health, Baltimore, MD 20892, USA. [90]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK. [91]Breakthrough Breast Cancer

Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London, 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, Australia. [94]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia. [95]Faculty of Medicine, University of Iceland, 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, Villejuif, France. [97]University Paris-Sud, UMRS 1018, Villejuif, France. [98]Department of Obstetrics and Gynecology, Southern Medical University, Guangzhou, China. [99]Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany. [100]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany. [101]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [102]Washington University, Department of Psychiatry, St.Louis, Missouri, USA. [103]Department of Epidemiology, Erasmus MC, Rotterdan, the Netherlands. [104]Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands. [105]Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA. [106]Hebrew SeniorLife Institute for Aging Research, Boston, MA. [107]Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115. [108]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. [109]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [110]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. [111]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. [112]Vesalius Research Center (VRC), VIB, Leuven, Belgium. [113]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium. [114]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [115]Department of Medicine, Stanford School of Medicine, Stanford, USA. [116]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [117]KULeuven (University of Leuven), Department of Oncology, Multidisciplinary Breast Center, University Hospitals Leuven, Belgium. [118]Research Unit of Obstetrics & Gynecology, Institute of Clinical Research, University of Southern denmark, DK. [119]Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [120]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [121]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [122]Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [123]Department of Medicine, Indiana University School of Medicine, Indianapolis,

Indiana USA. [124]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy. [125]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, UK. [126]University Groningen, University Medical Center Groningen, Department Pulmonary Medicine and Tuberculosis, GRIAC Research Institute, Groningen, The Netherlands. [127]Department of Obstetrics and Gynecology, Oulu University Hospital, Finland. [128]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu University Hospital/NordLab Oulu, 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), Milan, Italy. [130]National Institute on Aging, Intramural Research Program, Baltimore, MD, USA. [131]Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands. [132]Department of Pathology, The University of Melbourne, Melbourne, Australia. [133]Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, UK. [134]Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge, United Kingdom. [135]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [136]Department of Obstetrics and Gynaecology, Campus Grosshadern, Ludwig-Maximilians- University, Munich, Germany. [137]Department of Internal Medicine, Lausanne University Hospital, Lausanne, Switzerland. [138]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany. [139]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, 17475 Greifswald, Germany. [140]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [141]Department of Endocrinology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands. [142]Queensland Insitute of Medical Research, Brisbane, Queensland, Australia. [143]Peter MacCallum Cancer Centre, Melbourne, Australia. [144]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany. [145]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. [146]Institute of Pathology, Medical Faculty of the University of Bonn, Bonn, Germany. [147]Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [148]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, CB2 0QQ, UK. [149]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, NY 10029, USA. [150]NIHR Oxford Biomedical Research Centre, Churchill Hospital, OX3 7LE Oxford, UK. [151]Oxford Centre for Diabetes, Endocrinology, & Metabolism, University of Oxford, Churchill Hospital, OX37LJ Oxford, UK. [152]Geriatric Research and

Education Clinical Center (GRECC) - Veterans Administration Medical Center, USA. Baltimore, MD 21201. [153]Centre of Medical Systems Biology, Leiden, the Netherlands. [154]Human Genetics Center and Div. of Epidemiology, University of Houston, TX. [155]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [156]Institute of Health Sciences, P.O.Box 5000, FI-90014 University of Oulu, Finland. [157]Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Finland. [158]Department of Children and Young People and Families, National Institute for Health and Welfare, Aapistie 1, Box 310, FI-90101 Oulu, Finland. [159]Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland. [160]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA. [161]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark. [162]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, The Capital Region, Copenhagen, Denmark. [163]Division of Population Health Sciences and Education, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK. [164]Department of Obstetrics and Gynecology, University Medicine Greifswald, 17475 Greifswald, Germany. [165]Boston University School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, MA. [166]Department of Paediatrics,University of Cambridge,Cambridge,UK.

Acknowledgements

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

REFERENCES

1. Prentice P, Viner RM. Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. Int. J. Obes. (Lond). 2013; 37:1036–43. [PubMed: 23164700]
2. Silveira LFG, Latronico AC. Approach to the patient with hypogonadotropic hypogonadism. J. Clin. Endocrinol. Metab. 2013; 98:1781–8. [PubMed: 23650335]
3. Abreu AP, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. N. Engl. J. Med. 2013; 368:2467–75. [PubMed: 23738509]
4. Elks CE, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat. Genet. 2010; 42:1077–85. [PubMed: 21102462]
5. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 2012; 44:369–75. S1–3. [PubMed: 22426310]
6. Kong A, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009; 462:868–74. [PubMed: 20016592]
7. Hindorff, LA., et al. [Accessed 1st Nov 2013] A catalog of published genome-wide association studies. Available at: www.genome.gov/gwastudies
8. Temple IK, Shrubb V, Lever M, Bullman H, Mackay DJG. Isolated imprinting mutation of the DLK1/GTL2 locus associated with a clinical presentation of maternal uniparental disomy of chromosome 14. J. Med. Genet. 2007; 44:637–40. [PubMed: 17601927]

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. 2013; 93:876–90. [PubMed: 24183450]
10. Westra H-J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 2013; 45:1238–43. [PubMed: 24013639]
11. Schaaf CP, et al. Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. Nat. Genet. 2013; 45:1405–8. [PubMed: 24076603]
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. 2007; 16:2591–9. [PubMed: 17704508]
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. Nat. Genet. 2014 advance online publication.
14. Lomniczi A, et al. Epigenetic control of female puberty. Nat. Neurosci. 2013; 16:281–9. [PubMed: 23354331]
15. Partsch C-J, et al. Central precocious puberty in girls with Williams syndrome. J. Pediatr. 2002; 141:441–4. [PubMed: 12219071]
16. Grinspon RP, et al. Early onset of primary hypogonadism revealed by serum anti-Müllerian hormone determination during infancy and childhood in trisomy 21. Int. J. Androl. 2011; 34:e487–98. [PubMed: 21831236]
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-transretinoic acid response element. Brain Res. Mol. Brain Res. 2001; 87:214–22. [PubMed: 11245924]
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. 2009; 297:C1146–56. [PubMed: 19726747]
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. (Oxf). 2004; 60:682–7. [PubMed: 15163330]
20. Constantin S, et al. GnRH neuron firing and response to GABA in vitro depend on acute brain slice thickness and orientation. Endocrinology. 2012; 153:3758–69. [PubMed: 22719049]
21. Speliotes EK, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42:937–948. [PubMed: 20935630]
22. Lango Allen H, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. [PubMed: 20881960]
23. Horikoshi M, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat. Genet. 2013; 45:76–82. [PubMed: 23202124]
24. D'Aloisio AA, DeRoo LA, Baird DD, Weinberg CR, Sandler DP. Prenatal and infant exposures and age at menarche. Epidemiology. 2013; 24:277–84. [PubMed: 23348069]
25. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. 2009; 41:703–7. [PubMed: 19430480]
26. Zheng W, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat. Genet. 2009; 41:324–8. [PubMed: 19219042]
27. Estrada K, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat. Genet. 2012; 44:491–501. [PubMed: 22504420]
28. Parker SE, et al. Menarche, menopause, years of menstruation, and the incidence of osteoporosis: the influence of prenatal exposure to diethylstilbestrol. J. Clin. Endocrinol. Metab. 2014; 99:594–601. [PubMed: 24248183]
29. Huang J, et al. Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. Blood. 2012; 120:4873–81. [PubMed: 22990020]
30. Migliano AB, Vinicius L, Lahr MM. Life history trade-offs explain the evolution of human pygmies. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:20216–9. [PubMed: 18077366]

Europe PMC Funders Author Manuscripts

Additional references cited in the Methods

31. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 2013; 45:353–61. 361e1–2. [PubMed: 23535729]
32. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
33. Yang J, et al. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 2011; 19:807–12. [PubMed: 21407268]
34. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 2011; 88:76–82. [PubMed: 21167468]
35. Langenberg C, et al. 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. 2011; 54:2272–82. [PubMed: 21717116]
36. Göring HHH, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat. Genet. 2007; 39:1208–16. [PubMed: 17873875]
37. Idaghdour Y, et al. Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat. Genet. 2010; 42:62–7. [PubMed: 19966804]
38. Heap GA, et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC Med. Genomics. 2009; 2:1. [PubMed: 19128478]
39. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–8. [PubMed: 18344981]
40. Fehrmann RSN, 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. 2011; 7:e1002197. [PubMed: 21829388]
41. Mehta D, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. Eur. J. Hum. Genet. 2013; 21:48–54. [PubMed: 22692066]
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. 2011; 14:21–6. [PubMed: 20670100]
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. 2013; 8:e54967. [PubMed: 23359819]
44. Dixon AL, et al. A genome-wide association study of global gene expression. Nat. Genet. 2007; 39:1202–7. [PubMed: 17873877]
45. Liang L, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. 2013; 23:716–26. [PubMed: 23345460]
46. Stranger BE, et al. Population genomics of human gene expression. Nat. Genet. 2007; 39:1217–24. [PubMed: 17873874]
47. Kwan T, et al. Genome-wide analysis of transcript isoform variation in humans. Nat. Genet. 2008; 40:225–31. [PubMed: 18193047]
48. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–50. [PubMed: 19644074]
49. Cusanovich DA, 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. 2012; 21:2111–23. [PubMed: 22286170]
50. Grundberg E, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 2012; 44:1084–9. [PubMed: 22941192]
51. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. 2012; 44:502–10. [PubMed: 22446964]
52. Murphy A, et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Hum. Mol. Genet. 2010; 19:4745–57. [PubMed: 20833654]
53. Heinzen EL, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol. 2008; 6:e1. [PubMed: 19222302]

54. Zeller T, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One. 2010; 5:e10693. [PubMed: 20502693]
55. Barreiro LB, et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:1204–9. [PubMed: 22233810]
56. Huang RS, et al. Population differences in microRNA expression and biological implications. RNA Biol. 8:692–701. [PubMed: 21691150]
57. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–4. [PubMed: 22307276]
58. Greenawalt DM, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. 2011; 21:1008–16. [PubMed: 21602305]
59. Kompass KS, Witte JS. Co-regulatory expression quantitative trait loci mapping: method and application to endometrial cancer. BMC Med. Genomics. 2011; 4:6. [PubMed: 21226949]
60. Li Q, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell. 2013; 152:633–41. [PubMed: 23374354]
61. Webster JA, et al. Genetic control of human brain transcript expression in Alzheimer disease. Am. J. Hum. Genet. 2009; 84:445–58. [PubMed: 19361613]
62. Zou F, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet. 2012; 8:e1002707. [PubMed: 22685416]
63. Colantuoni C, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature. 2011; 478:519–23. [PubMed: 22031444]
64. Liu C, et al. Whole-genome association mapping of gene expression in the human prefrontal cortex. Mol. Psychiatry. 2010; 15:779–84. [PubMed: 20351726]
65. Gibbs JR, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 2010; 6:e1000952. [PubMed: 20485568]
66. Zhang B, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell. 2013; 153:707–20. [PubMed: 23622250]
67. Schadt EE, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008; 6:e107. [PubMed: 18462017]
68. Innocenti F, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011; 7:e1002078. [PubMed: 21637794]
69. Sulzbacher S, Schroeder IS, Truong TT, Wobus AM. Activin A-induced differentiation of embryonic stem cells into endoderm and pancreatic progenitors-the influence of differentiation factors and culture conditions. Stem Cell Rev. 2009; 5:159–73. [PubMed: 19263252]
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. 2013; 13:12–20. [PubMed: 22006096]
71. Grundberg E, et al. Population genomics in a disease targeted primary cell model. Genome Res. 2009; 19:1942–52. [PubMed: 19654370]
72. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology. 2013; 144:1488–96. 1496.e1–3. [PubMed: 23474282]
73. Hao K, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet. 2012; 8:e1003029. [PubMed: 23209423]
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. 2010; 87:779–789. [PubMed: 21129726]
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. 2011; 6:e27338. [PubMed: 22102887]

76. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. 2010; 6

77. Cousminer DL, 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. 2014 Epub ahead of print.

Europe PMC Funders Author Manuscripts

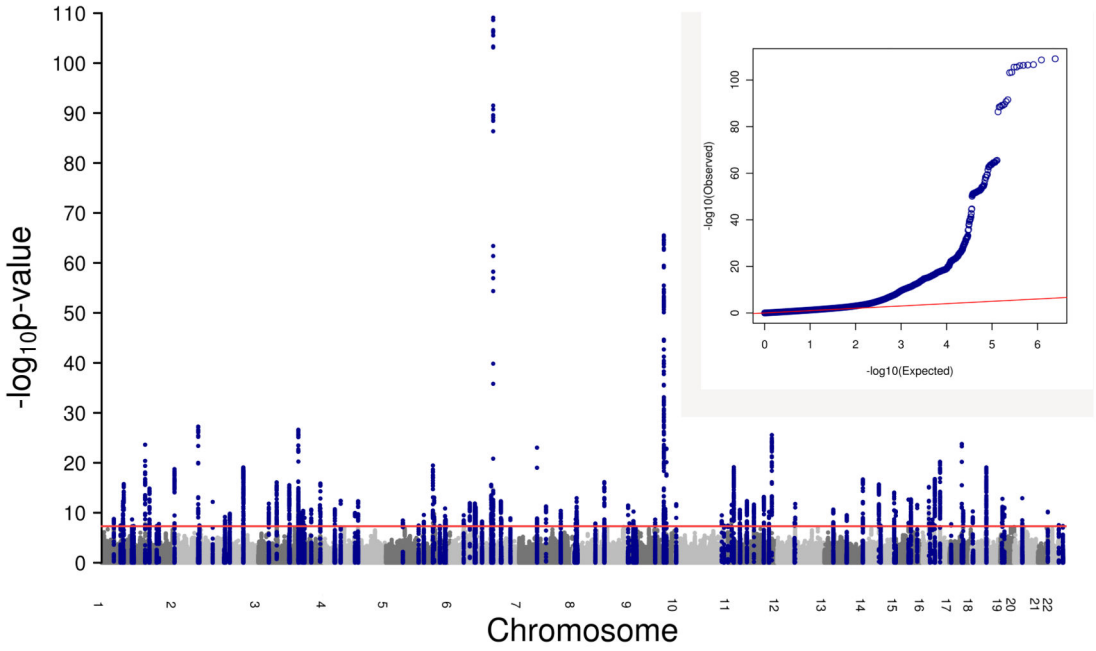

Figure 1. Manhattan and QQ plot of the GWAS for age at menarche

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 ($P=5\times10^{-8}$). 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).

Europe PMC Funders Author Manuscripts

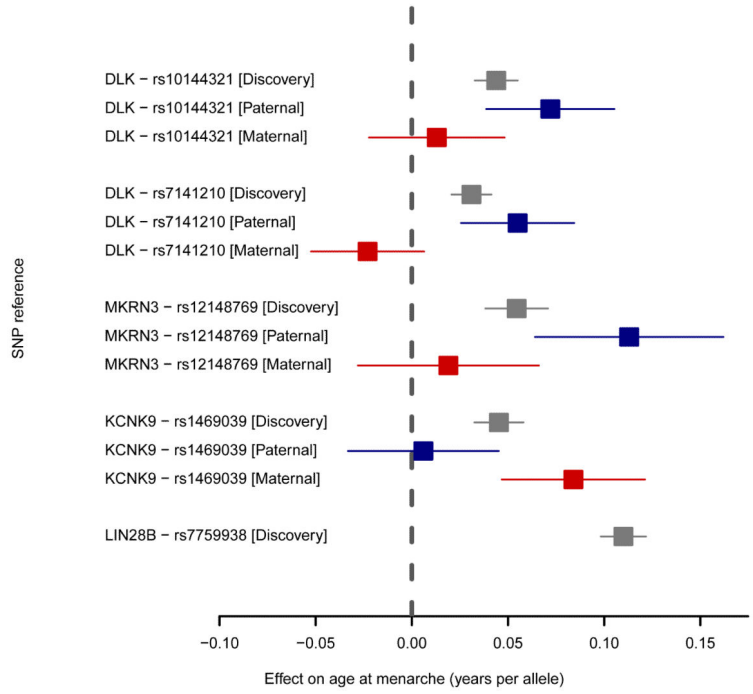

Figure 2. Forest plot of parent-of-origin specific allelic associations at three imprinted menarche loci

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 (Black), 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.

Europe PMC Funders Author Manuscripts

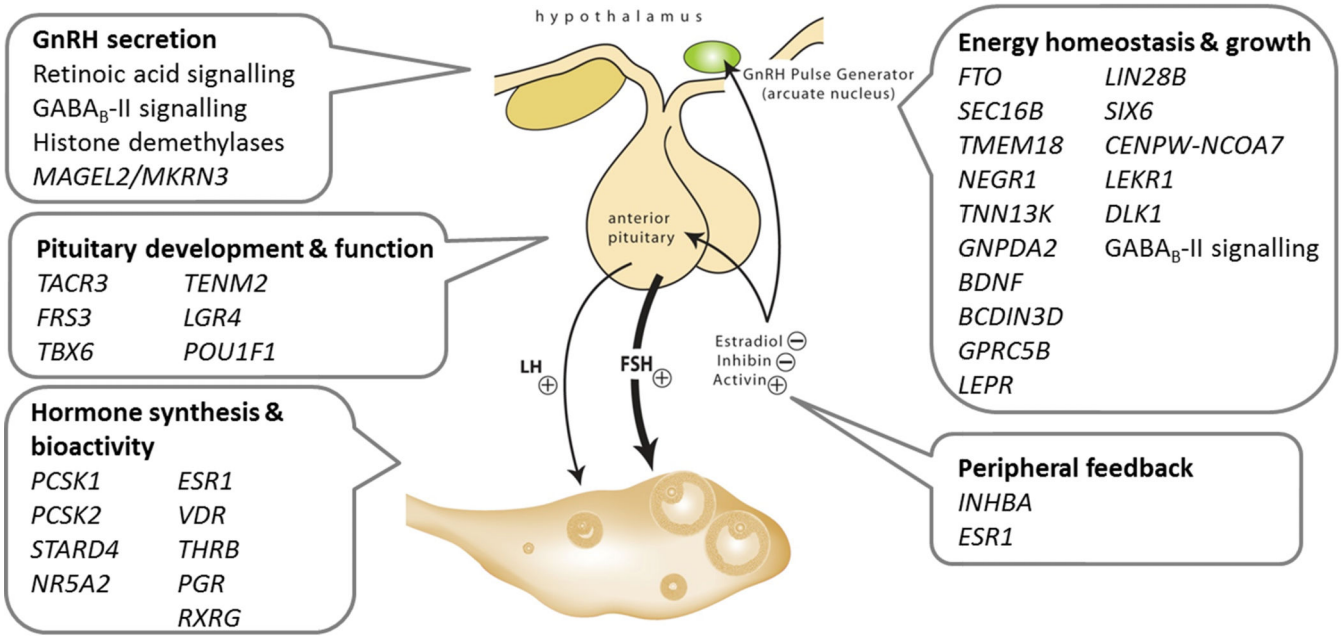

Figure 3. Schematic diagram indicating possible roles in the hypothalamic-pituitary-ovarian axis of several of the implicated genes and biological mechanisms for menarche timing

Europe PMC Funders Author Manuscripts