nature genetics

Physical and neurobehavioral determinants of reproductive onset and success

Felix R Day[1,10], Hannes Helgason[2,3,10], Daniel I Chasman[4,5,10], Lynda M Rose[4], Po-Ru Loh[6,7], Robert A Scott[1], Agnar Helgason[2,8], Augustine Kong[2,3], Gisli Masson[2], Olafur Th Magnusson[2], Daniel Gudbjartsson[2,3], Unnur Thorsteinsdottir[2,9], Julie E Buring[4,5], Paul M Ridker[4,5], Patrick Sulem[2], Kari Stefansson[2,9,10], Ken K Ong[1,10] & John R B Perry[1,10]

The ages of puberty, first sexual intercourse and first birth signify the onset of reproductive ability, behavior and success, respectively. In a genome-wide association study of 125,667 UK Biobank participants, we identify 38 loci associated ($P < 5 \times 10^{-8}$) with age at first sexual intercourse. These findings were taken forward in 241,910 men and women from Iceland and 20,187 women from the Women's Genome Health Study. Several of the identified loci also exhibit associations ($P < 5 \times 10^{-8}$) with other reproductive and behavioral traits, including age at first birth (variants in or near *ESR1* and *RBM6–SEMA3F*), number of children (*CADM2* and *ESR1*), irritable temperament (*MSRA*) and risk-taking propensity (*CADM2*). Mendelian randomization analyses infer causal influences of earlier puberty timing on earlier first sexual intercourse, earlier first birth and lower educational attainment. In turn, likely causal consequences of earlier first sexual intercourse include reproductive, educational, psychiatric and cardiometabolic outcomes.

The age of puberty—the transition from childhood to sexual maturity and reproductive ability—has fallen markedly over the last century in most populations, as illustrated by the decrease in the average age at menarche from 18 years in 1880 to 12.5 years in 1980 (refs. 1,2). This decline was initially observed in industrialized, Western countries and more recently has been seen in countries with more recent economic transitions, often with the decrease in age occurring far more rapidly[3]. These changes likely reflect increases in childhood nutrition and body size, but exposures to endocrine-disrupting chemicals or other specific environmental factors have also been proposed[1]. In contemporary cohorts, earlier puberty timing, in both men and women, is associated with greater propensity for risk-taking behaviors[4,5], lower educational attainment, greater susceptibility to several adverse health outcomes[6] and, in women, increased mortality[7]. Conversely, it has been proposed that earlier puberty timing is a life-history strategy that promotes greater reproductive fitness[8]. Yet, despite some reports that earlier puberty timing is associated with younger age at first sexual intercourse (AFS)[9,10] and younger age at first birth (AFB), there is yet little evidence that this trait is associated with reproductive fitness[11].

In contrast to the small body of evidence on the role of puberty timing in AFS, most research on the predictors or determinants of AFS is contextualized in terms of the social, economic and cultural environment, including the nature of interpersonal relationships. Hence, established correlates of younger AFS include social disadvantage, family instability, low levels of parental monitoring, and lack of religious affiliation and belief[12–14]. In particular, parental and peer norms and behaviors have a strong influence on teenagers' sexual behavior[14–16]. Twins studies have suggested some genetic contribution to AFS[17,18]; however, observations of older AFS among monozygotic twins as compared to dizygotic twins[13] cast doubt on the ability of twins studies to accurately estimate the heritability of this trait.

Recent genome-wide association studies (GWAS) have identified 123 sequence variants independently associated with timing of menarche in females[19], and these signals seem to have concordant effects on puberty timing in males[20]. A valuable application of such GWAS findings is the use of genetic variants as instrumental variables in Mendelian randomization analyses. This allows for the assessment of likely causal relationships, with less risk of confounding when compared to traditionally observed epidemiological associations[21]. Here we use this approach to test the causal relationship of puberty timing[19] with AFS and AFB. We also perform a GWAS to identify sequence variants associated with AFS and AFB and use these findings to test the causal relationships of the ages of onset for reproductive ability, activity and success with other behavioral and health-related outcomes.

[1]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [2]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [3]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [4]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [5]Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [7]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [8]Department of Anthropology, University of Iceland, Reykjavik, Iceland. [9]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [10]These authors contributed equally to this work. Correspondence should be addressed to J.R.B.P. (john.perry@mrc-epid.cam.ac.uk) or K.K.O. (ken.ong@mrc-epid.cam.ac.uk).

Received 17 November 2015; accepted 24 March 2016; published online 18 April 2016; doi:10.1038/ng.3551

RESULTS

Shared genetic architecture for reproductive traits

We used whole-genome LD Score regression[22] to test the genetic correlations between the timings of puberty, first sexual intercourse and first birth; such correlations quantify the extent of shared genetic architecture. Data on genome-wide SNP associations with puberty timing were recently reported[19]. To generate such scores for the other two traits, we performed association tests across a genome-wide panel of ~46 million SNPs for self-reported AFS and AFB (recorded in women only) in 59,357 men and 66,310 women in the UK Biobank study. In this sample, the median (interquartile range, IQR) AFS was 18 years (16–21 years) in men and 18 years (17–21 years) in women, and the median (IQR) AFB was 25 years (22–28 years) (recorded in women only). The SNP-based test REML[23] indicated moderate heritability for AFS, both in men ($h = 0.248$, standard error (s.e.) = 0.010) and women ($h = 0.242$, s.e. = 0.010), and also moderate heritability for AFB ($h = 0.290$, s.e. = 0.015; women only). Using the scores for the three traits, we found moderate positive genetic correlations between puberty timing and AFS, both in women ($r_g = 0.22$, $P = 1.2 \times 10^{-16}$) and men ($r_g = 0.26$, $P = 9.5 \times 10^{-8}$), and between puberty timing and AFB ($r_g = 0.24$, $P = 9.0 \times 10^{-13}$; women only). Furthermore, we found a strong genetic correlation between AFS and AFB ($r_g = 0.86$, $P = 3.1 \times 10^{-136}$).

Genetic correlations have been reported between puberty timing and a range of other health-related traits, including inverse correlations with body mass index (BMI), type 2 diabetes (T2D) and cardiovascular disease (CVD)[20]. In the same way, we applied whole-genome LD Score regression[22] to test genetic correlations between AFS and other health-related traits, using publicly available GWAS data sets or original GWAS findings from the UK Biobank study (Online Methods). We identified genetic correlations between AFS and 22 of the 44 tested outcomes or traits after correction for multiple testing ($P < 1.1 \times 10^{-3}$), including inverse correlations with BMI, T2D and CVD, as well as correlations with a variety of behavioral (for example, smoking or alcohol intake) and neurological (for example, intelligence or risk-taking propensity) traits and psychiatric outcomes (attention deficit hyperactivity disorder (ADHD) or schizophrenia) (**Fig. 1**).

In a sex-combined GWAS in the UK Biobank study, we identified 33 loci with variants associated at $P < 5 \times 10^{-8}$ with AFS (**Fig. 2**, **Table 1** and **Supplementary Figs. 1** and **2**). Nine of these loci showed sex-discordant associations ($P_{heterogeneity} < 0.05$), and subsequent sex-specific models identified an additional five signals, four in men and one in women (**Table 1**). Across these 38 AFS-associated loci, effect sizes ranged from 0.02 to 0.33 s.d. and minor allele frequencies (MAFs) ranged from 0.15% to 49%. Seventeen of these variants were associated with AFS at the more stringent *P*-value threshold of $P < 1 \times 10^{-9}$.

In the absence of other large GWAS for AFS, we relied on the strong genetic correlation between AFS and AFB in women to collectively confirm our genetic findings in two independent data sets: deCODE (n = 117,626 males and 124,284 females) and the Women's Genome Health Study (WGHS; n = 20,187 women). A weighted SNP genotype score for our 38 new signals for AFS was strongly associated with AFB in both deCODE ($P = 3.3 \times 10^{-21}$) and WGHS ($P = 9.2 \times 10^{-4}$) (**Supplementary Tables 1** and **2**). The subset of 21 less stringently associated AFS variants (those with association *P* values between 5×10^{-8} and 1×10^{-9}) was also collectively associated with AFB (weighted SNP genotype score in deCODE, $P = 3.1 \times 10^{-7}$).

Biological determinants of age at first sexual intercourse

None of the 38 lead AFS-associated variants (or their proxy SNPs with linkage disequilibrium (LD) $r^2 > 0.8$) were nonsynonymous; however, several were located in regions containing promoter or enhancer histone marks, DNase-hypersensitive sites or protein-binding sites (**Supplementary Table 3**). In addition, the majority of these variants either altered regulatory motifs or were associated in *cis* with gene expression (**Supplementary Table 4**).

To identify mechanisms that might regulate AFS, we performed a systematic test of all annotated biological pathways for enrichment of genes located near AFS-associated signals, using MAGENTA (Online Methods and **Supplementary Table 5**). Four pathways were associated with AFS: 'circadian clock system', 'packaging of telomere ends', 'RNA polymerase I promoter opening' and 'NOTCH HLH transcription'.

We then tested puberty timing and body size as specific a priori candidate determinants of AFS, by performing Mendelian randomization analyses in the UK Biobank sample. For each exposure, we created an approximated genetic risk score using summary-level effect estimates for SNPs at reported signals

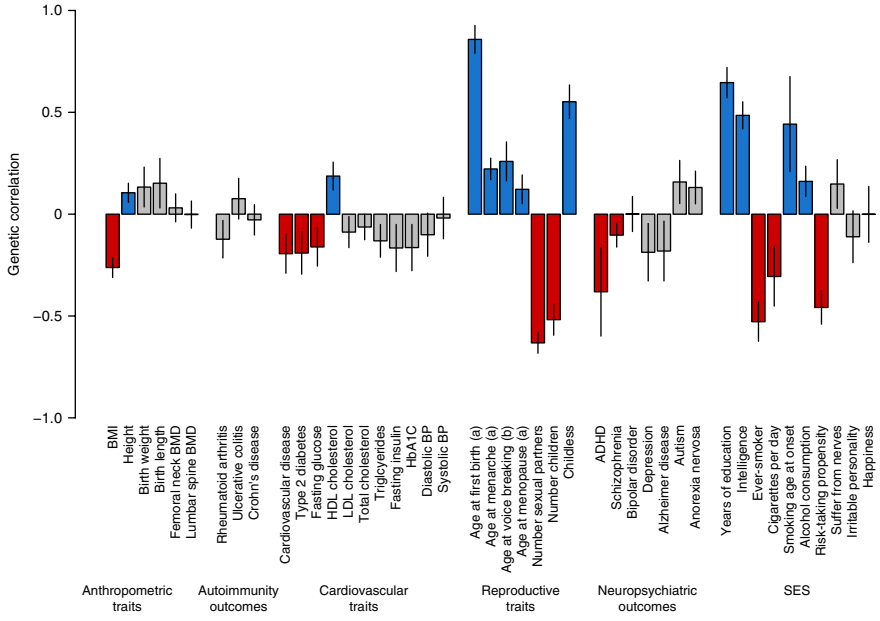

Figure 1 Bar chart of genetic correlations for age at first sexual intercourse. Whole-genome LD Score regression tested genome-wide SNP associations for AFS against similar data for 44 other traits; error bars show 95% confidence intervals on these estimates. Blue (positive correlation) and red (negative correlation) bars represent the 22 traits that showed a significant genetic correlation after correction for multiple testing ($P < 1.1 \times 10^{-3}$). Traits annotated with "a" were analyzed in women only, and those annotated with "b" were analyzed in men only. ADHD, attention deficit hyperactivity disorder; SES, socioeconomic status, behavioral and personality traits; BMD, bone mineral density.

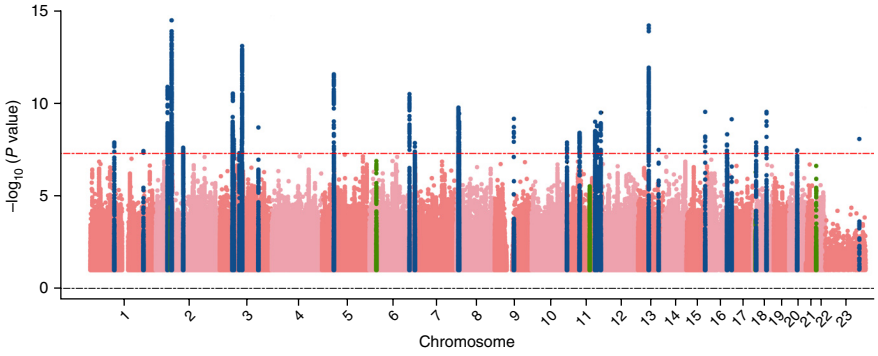

Figure 2 Manhattan plot of the GWAS for age at first sexual intercourse. The plot illustrates the results of the GWAS meta-analysis for AFS in up to 59,357 men and 66,310 women of European descent in the UK Biobank study. Negative $\log_{10}$-transformed *P* values for each SNP (*y* axis) are plotted by chromosomal position (*x* axis). The red dashed line represents the threshold for genome-wide statistical significance ($P = 5 \times 10^{-8}$). Blue dots represent SNPs within a 1-Mb window around each genome-wide significant signal, and green dots represent SNPs with sex-specific effects.

showing robust association with each phenotype (Online Methods and **Supplementary Tables 6–10**). In both men and women, genetically predicted earlier puberty timing ($P_{women} = 2.0 \times 10^{-9}$ and $P_{men} = 4.7 \times 10^{-11}$) and genetically predicted greater BMI ($P_{men} = 5.5 \times 10^{-8}$ and $P_{women} = 2.2 \times 10^{-4}$) seemed to promote earlier AFS (**Supplementary Tables 7** and **8**). Genetically predicted greater height seemed to promote later AFS in men ($P = 1.0 \times 10^{-5}$) and women ($P = 1.1 \times 10^{-3}$) (**Supplementary Table 9**), consistent with reported non-genetic associations between greater height and later AFB in European men and women[24].

Table 1 Thirty-eight genome-wide significant loci for age at first sexual intercourse in the UK Biobank study

Variant	Location	Nearest gene	Alleles[b]	Allele freq.[c]	Effect (s.e.)	*P* value	$P_{heterogeneity}$[d]
rs115552537	1p22.2	*BARHL2*	A/C	0.22	0.03 (0.005)	1.30×10^{-8}	6.66×10^{-1}
rs10800813	1q32.1	*GPR37L1*	C/T	0.63	0.02 (0.004)	3.80×10^{-8}	5.56×10^{-1}
rs4324362	2p21	*CAMKMT*	G/A	0.38	0.03 (0.004)	1.30×10^{-11}	1.14×10^{-1}
rs1344293	2p16.1	*BCL11A*	G/T	0.48	0.03 (0.004)	3.30×10^{-15}	1.87×10^{-1}
rs1040124	2q12.1	*TMEM182*	A/G	0.57	0.02 (0.004)	2.50×10^{-8}	1.96×10^{-1}
rs1264194	3p21.31	*COL7A1*	C/T	0.71	0.03 (0.004)	5.10×10^{-9}	1.57×10^{-1}
rs2188151	3p21.31	*SEMA3F*	G/T	0.57	0.03 (0.004)	3.00×10^{-11}	7.85×10^{-1}
rs34337122	3p21.1	*CACNA1D*	CG/C	0.85	0.03 (0.005)	8.80×10^{-9}	7.62×10^{-1}
rs6549665	3p12.3	*CNTN3*	G/C	0.18	0.03 (0.005)	4.90×10^{-8}	9.20×10^{-1}
rs12714592	3p12.1	*CADM2*	A/C	0.73	0.03 (0.004)	1.80×10^{-10}	4.34×10^{-1}
rs57401290	3p12.1	*CADM2*	GGTGTGT/G	0.55	0.03 (0.004)	8.00×10^{-14}	1.32×10^{-3}
rs530580221	3q24	*ZIC4*	T/TA	0.38	0.03 (0.004)	2.00×10^{-9}	9.22×10^{-1}
rs12522910	5p12	*HCN1*	C/T	0.17	0.04 (0.005)	2.70×10^{-12}	2.47×10^{-1}
rs726281	6q25.1	*ESR1*	A/G	0.72	0.03 (0.004)	3.20×10^{-11}	6.58×10^{-1}
rs13239969	7p22.3	*MAD1L1*[a]	C/G	0.6	0.02 (0.004)	1.40×10^{-8}	6.39×10^{-1}
rs4840367	8p23.1	*MFHAS1*	A/G	0.41	0.02 (0.004)	1.70×10^{-10}	4.35×10^{-1}
rs658385	8p23.1	*MSRA*	T/C	0.55	0.02 (0.004)	6.70×10^{-9}	2.92×10^{-2}
rs2248699	8p23.1	*BLK*[a]	G/A	0.5	0.02 (0.004)	3.60×10^{-10}	3.28×10^{-1}
rs538498277	9q21.12	*SMC5*	C/G	0.998	0.31 (0.051)	6.90×10^{-10}	4.86×10^{-2}
rs4443996	10q26.3	*LRRC27*	A/C	0.52	0.02 (0.004)	1.30×10^{-8}	8.40×10^{-1}
rs535814333	11p11.2	*ATG13*	TTG/T	0.7	0.02 (0.004)	3.90×10^{-9}	1.39×10^{-2}
rs140976226	11q22.3	*GRIA4*	GTT/G	0.43	0.02 (0.004)	1.00×10^{-9}	2.51×10^{-2}
rs66821824	11q23.2	*NCAM1*	ATTTT/A	0.78	0.03 (0.005)	1.60×10^{-9}	3.39×10^{-1}
rs538200730	11q24.2	*KIRREL3*	T/A	0.29	0.03 (0.004)	3.20×10^{-10}	8.22×10^{-1}
rs341521	13q21.2	*DIAPH3*	G/A	0.3	0.03 (0.004)	6.30×10^{-15}	4.62×10^{-2}
rs9516776	13q32.1	*HS6ST3*	A/T	0.34	0.02 (0.004)	3.30×10^{-8}	4.87×10^{-1}
rs4702	15q26.1	*FURIN*[a]	A/G	0.56	0.02 (0.004)	2.90×10^{-10}	2.79×10^{-2}
rs76513770	16q22.2	*PMFBP1*	C/T	0.13	0.04 (0.006)	4.70×10^{-9}	4.17×10^{-1}
rs369230	16q24.3	*CPNE7*	G/T	0.31	0.02 (0.004)	7.30×10^{-10}	1.02×10^{-4}
rs58749137	18p11.21	*GNAL*	A/G	0.73	0.02 (0.004)	1.30×10^{-8}	1.07×10^{-1}
rs4129322	18q21.2	*DCC*	A/G	0.08	0.04 (0.007)	2.90×10^{-10}	9.35×10^{-1}
rs6058613	20q11.21	*KIF3B*	C/G	0.16	0.03 (0.005)	3.50×10^{-8}	1.67×10^{-2}
rs5932884	Xq26.2	*IGSF1*	G/A	0.47	0.02 (0.005)	8.41×10^{-9}	–
Women only							
rs961522	2p16.1	*VRK2*	C/T	0.61	0.03 (0.005)	2.80×10^{-8}	4.63×10^{-5}
Men only							
rs13194984	6p22.2	*BTN1A1*	G/T	0.86	0.05 (0.009)	3.90×10^{-9}	5.21×10^{-5}
rs201909661	11q14.1	*DLG2*	A/AG	0.02	0.15 (0.02)	7.00×10^{-10}	6.09×10^{-7}
rs138057093	18p11.22	*RAB31*	C/T	0.01	0.20 (0.03)	7.50×10^{-10}	7.25×10^{-6}
rs111837587	22q11.1	*XKR3*	A/G	0.01	0.18 (0.03)	7.20×10^{-9}	8.77×10^{-5}

[a]Gene linked via altered expression (eQTL). [b]Effect allele/other allele. [c]Effect allele frequency. [d]*P* value for heterogeneity in effect estimate between sexes.

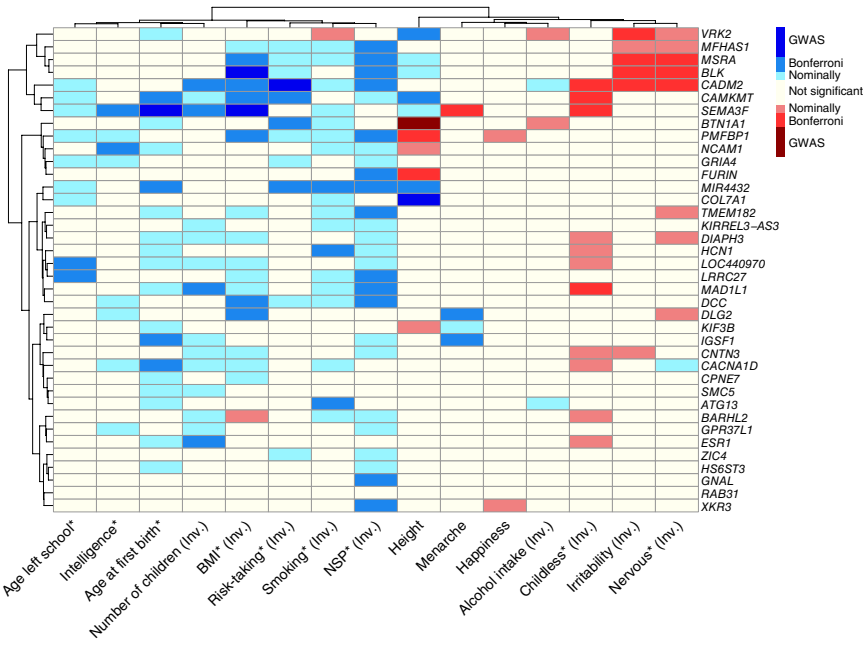

Figure 3 Cluster plot displaying associations between the 38 lead SNPs for age at first sexual intercourse and 15 other behavioral, reproductive and health-related traits in the UK Biobank study. The nine traits that are significantly enriched for AFS signals are annotated with an asterisk. All SNPs are aligned to the AFS-increasing allele. Both SNPs and phenotypes are clustered by patterns of association. To facilitate clustering, some phenotypes have been inverted; these are labeled "(Inv.)" (for example, number of children, BMI, risk-taking propensity). Blue shading indicates a positive association with the indicated variable, and red shading indicates a negative association. Hierarchical clustering of both phenotypes and SNPs was based on the Ward method using Euclidian distances. BMI, body mass index, NSP, number of lifetime sexual partners.

Prompted by our observation of a new AFS-associated locus near *MC1R* (rs369230, $r^2 = 0.12$ with variant rs12931267 that is strongly associated with hair color)[25,26], we tested genetic instrumental variables for skin freckling and hair color, which are traits regulated by this gene[25,26]. Genetically predicted skin freckling seemed to promote later AFS in women ($P = 6.3 \times 10^{-9}$) but not in men ($P = 0.47$), and genetically predicted red hair seemed to promote later AFS in both men ($P = 0.02$) and women ($P = 9.3 \times 10^{-5}$) (**Supplementary Table 10**).

Consequences related to reproductive timing

To test whether puberty timing and AFS might be causally related to other behavioral, reproductive and health outcomes, we performed Mendelian randomization analyses using weighted allele scores calculated from SNP genotypes at signals associated with puberty timing[19] or AFS (described above) as genetic instrumental variables for these traits. To reduce bias, we avoided testing health outcomes in the same data sets that were used to generate the allele weightings (that is, outcomes related to AFS were tested in data sets other than the UK Biobank study). Genetic associations were scaled to indicate the likely causal effect of a 1 s.d. change in AFS and a 1-year change in puberty timing.

Genetically predicted earlier puberty timing decreased the likelihood of remaining in education after 16 years of age (odds ratio (OR) = 0.98 per year, $P = 6.2 \times 10^{-4}$; in the UK Biobank study) (**Supplementary Table 11**). Similarly, genetically predicted earlier AFS decreased the likelihood of attaining university-level education (standardized OR = 0.74, $P = 3.7 \times 10^{-5}$; in publicly available Social Science Genetic Association Consortium data) and increased the likelihood of being an ever-smoker (standardized OR = 1.33, $P = 2.0 \times 10^{-3}$; in publicly available Tobacco and Genetics Consortium data) (**Supplementary Table 9**). For reproductive outcomes, in the deCODE data, genetically predicted earlier AFS promoted earlier AFB (women: standardized effect size (β) = 1.71, $P = 2.2 \times 10^{-17}$; men: standardized β =1.69, $P = 2.6 \times 10^{-13}$; combined $P = 3.3 \times 10^{-21}$), a greater number of children (women: standardized $\beta = 0.035$, $P = 0.006$; men: standardized $\beta = 0.012$, $P = 0.34$; combined $P = 0.0044$) and lower likelihood of being childless (women: standardized OR = 0.67, $P = 0.034$; men: standardized OR = 0.66, $P = 0.009$; combined $P = 0.0022$) (**Supplementary Table 1**). Similarly, earlier puberty timing promoted earlier AFB ($\beta = 0.37$ years per year, $P = 5.8 \times 10^{-8}$; assessed in UK Biobank women only) and earlier age at last birth ($\beta = 0.37$ years per year, $P = 3.7 \times 10^{-7}$; assessed in UK Biobank women) but had little effect on other reproductive outcomes (**Supplementary Tables 6** and **11**).

We noted that several of the 38 new AFS signals were located in or near genes reportedly implicated in brain development (*BARHL2*, *SEMA3F*, *ZIC4–ZIC1*, *DPYSL4* and *DIAPH3*) or neuronal activity and/or susceptibility to schizophrenia or bipolar disorder (*CADM2*, *LRP4*, *GRIA4*, *CACNA1D*, *HCN1*, *GRIA4*, *DRD2*, *FURIN*, *GNAL* and *VRK2*) (**Table 1** and **Supplementary Table 3**), consistent with our observed shared genetic architecture for AFS and ADHD ($r_g = -0.38$, $P = 5.9 \times 10^{-4}$) and for AFS and schizophrenia ($r_g = -0.10$, $P = 7.3 \times 10^{-4}$) (**Fig. 2**). We used a bidirectional Mendelian randomization approach to test likely causal relationships (**Supplementary Tables 12** and **13**): susceptibility to schizophrenia seemed to lower AFS (in the UK Biobank study, with or without exclusion of individuals with self-reported psychiatric illness, $P = 0.005$), but earlier AFS also seemed to increase susceptibility to schizophrenia (in publicly available Psychiatric Genetics Consortium data, $P = 4.1 \times 10^{-11}$), suggesting a pleiotropic relationship between these traits. The substantial shared genetic architecture for AFS and self-reported risk-taking propensity ($r_g = -0.46$, $P = 7.3 \times 10^{-28}$) (**Fig. 1**) gives insights into possible common determinants of AFS and schizophrenia.

To explore potential specific neurobehavioral mechanisms that might contribute to the etiology of AFS, we looked for associations of the 38 individual AFS loci with 15 other behavioral, reproductive and health-related traits in the UK Biobank study and other independent studies (**Fig. 3**, Online Methods and **Supplementary Table 14**).

CADM2 and *MSRA* loci influence multiple behavioral traits

The AFS signal represented by rs57401290 is intronic to *CADM2*, which encodes a neuronal cell adhesion molecule. rs57401290, or highly correlated SNPs in this locus, also showed genome-wide significant associations in the UK Biobank study with self-reported risk-taking propensity (rs57401290: $P = 5.3 \times 10^{-9}$; $r^2 = 0.65$ with the lead *CADM2* SNP for this trait, rs4856591: $P = 4.3 \times 10^{-10}$), number of sexual partners (rs57401290: $P = 6.0 \times 10^{-7}$; $r^2 = 0.60$ with lead SNP rs5850688: $P = 4.1 \times 10^{-8}$) and number of children (rs57401290: $P = 6.2 \times 10^{-7}$; $r^2 = 0.65$ with lead SNP rs4856591: $P = 3.8 \times 10^{-11}$;

Table 2 Association statistics at the *ESR1* and *RBM6–SEMA3F* loci for reproductive outcomes

		UK Biobank (up to 59,357 men, 66,310 women)			deCODE (up to 117,626 men, 124,284 women) and WGHS (up to 20,187 women)			Combined		
SNP	Trait	Effect	*P*	*n*	Effect	*P*	*n*	Effect	*P*	Maximum sample
rs67229052[a] *ESR1* TA/T[b] 0.36[c]	Age at menarche	0.01 (0.009)	1.20×10^{-1}	73,397	–	–	–	0.01 (0.009)	1.20×10^{-1}	73,397
	AFS, males	–0.02 (0.006)	3.70×10^{-4}	59,357	–	–	–	–0.02 (0.006)	3.70×10^{-4}	59,357
	AFS, females	–0.03 (0.005)	1.40×10^{-8}	66,310	–	–	–	–0.03 (0.005)	1.40×10^{-8}	66,310
	AFS, combined	–0.03 (0.004)	1.60×10^{-10}	125,667	–	–	–	–0.03 (0.004)	1.60×10^{-10}	125,667
	AFB, males	–	–	–	–0.19 (0.03)	6.73×10^{-9}	117,626	–0.19 (0.03)	6.73×10^{-9}	117,626
	AFB, females	–0.15 (0.029)	2.40×10^{-7}	50,954	–0.08 (0.02)	6.93×10^{-4}	144,471	–0.11 (0.02)	3.58×10^{-9}	195,425
	AFB, combined	–0.15 (0.029)	2.40×10^{-7}	50,954	–0.12 (0.02)	6.96×10^{-8}	262,097	–0.13 (0.02)	1.22×10^{-13}	313,051
	n children, males	0.01 (0.003)	8.20×10^{-2}	66,498	0.009 (0.002)	1.94×10^{-6}	117,626	0.008 (0.002)	7.25×10^{-7}	184,124
	n children, females	0.01 (0.003)	3.70×10^{-8}	75,540	0.005 (0.002)	1.27×10^{-2}	147,498	0.008 (0.002)	2.15×10^{-7}	223,038
	n children, combined	0.01 (0.002)	3.20×10^{-7}	142,038	0.007 (0.001)	1.11×10^{-6}	265,124	0.008 (0.001)	4.82×10^{-12}	407,162
	Childless, males	0.98 (0.01)	1.30×10^{-1}	66,498	0.97 (0.02)	2.12×10^{-1}	97,200	0.98 (0.01)	5.56×10^{-2}	163,698
	Childless, females	0.95 (0.01)	2.50×10^{-6}	75,540	0.93 (0.02)	4.00×10^{-4}	117,972	0.94 (0.01)	4.94×10^{-9}	193,512
	Childless, combined	0.96 (0.008)	1.10×10^{-5}	142,038	0.95 (0.02)	9.04×10^{-4}	215,526	0.96 (0.007)	5.24×10^{-8}	357,564
rs2188151 *RBM6–SEMA3F* T/G 0.43	Age at menarche	0.03 (0.008)	2.20×10^{-5}	73,397	–	–	–	0.03 (0.008)	2.20×10^{-5}	73,397
	AFS, males	–0.03 (0.006)	1.60×10^{-5}	59,357	–	–	–	–0.03 (0.006)	1.60×10^{-5}	59,357
	AFS, females	–0.02 (0.005)	3.60×10^{-7}	66,310	–	–	–	–0.02 (0.005)	3.60×10^{-7}	66,310
	AFS, combined	–0.03 (0.004)	3.00×10^{-11}	125,667	–	–	–	–0.03 (0.004)	3.00×10^{-11}	125,667
	AFB, males	–	–	–	–0.14 (0.032)	1.00×10^{-5}	117,626	–0.14 (0.032)	1.00×10^{-5}	117,626
	AFB, females	–0.16 (0.028)	7.20×10^{-9}	50,954	–0.105 (0.024)	7.88×10^{-6}	144,471	–0.129 (0.018)	9.52×10^{-13}	195,425
	AFB, combined	–0.16 (0.028)	7.20×10^{-9}	50,954	–0.115 (0.022)	9.18×10^{-8}	262,097	–0.132 (0.017)	8.76×10^{-15}	313,051
	n children, males	0.01 (0.003)	2.70×10^{-2}	66,498	0.003 (0.002)	7.47×10^{-2}	117,626	0.004 (0.002)	7.99×10^{-3}	184,124
	n children, females	0.01 (0.003)	1.30×10^{-3}	75,540	0.003 (0.002)	7.45×10^{-2}	147,498	0.004 (0.001)	1.35×10^{-3}	223,038
	n children, combined	0.01 (0.002)	1.60×10^{-4}	142,038	0.003 (0.002)	4.06×10^{-2}	265,124	0.005 (0.001)	9.05×10^{-5}	407,162
	Childless, males	0.98 (0.01)	3.40×10^{-2}	66,498	–0.017 (0.022)	4.27×10^{-1}	97,200	–0.023 (0.01)	2.49×10^{-2}	163,698
	Childless, females	0.97 (0.01)	2.50×10^{-3}	75,540	–0.028 (0.019)	1.52×10^{-1}	117,972	–0.032 (0.01)	8.53×10^{-4}	193,512
	Childless, combined	0.97 (0.008)	2.70×10^{-4}	142,038	–0.023 (0.015)	1.42×10^{-1}	215,526	–0.028 (0.007)	9.38×10^{-5}	357,564

[a]In WGHS, rs67229052 was not imputed, so rs4305732 ($r^2 = 0.98$) was used as a proxy. [b]Effect allele/other allele. [c]Effect allele frequency.

replication in deCODE, $P = 0.006$) (**Supplementary Table 14**). In each case, the AFS-decreasing allele conferred higher values of these outcomes. rs57401290 was modestly correlated with the reported signal at this locus for BMI[27] ($r^2 = 0.11$ with rs13078960; the AFS-decreasing allele also increased BMI) and was strongly correlated with the reported signal in *CADM2* for cognitive processing speed[28] (rs17518584, $r^2 = 0.80$; the AFS-decreasing allele also decreased processing speed). *CADM2* shows highest expression in the prefrontal cortex and is involved in a range of neuronal processes, including glutamate signaling, γ-aminobutyric acid transport and neuron cell–cell adhesion[28].

The AFS-decreasing allele at rs658385 (~25 kb downstream of *MSRA*) was also associated in the UK Biobank study with lower likelihood of self-reported irritable temperament ($P = 3.8 \times 10^{-4}$) and was modestly correlated ($r^2 = 0.14$) with the lead *MSRA* SNP for this trait (rs73195303, $P = 5.8 \times 10^{-11}$). Conditional analyses excluded the presence of independent secondary signals for AFS or irritable temperament at this locus. The enzyme encoded by *MSRA* reduces methionine sulfoxide to methionine and hence repairs proteins that have been inactivated by oxidative stress; blocking this repair represents a candidate mechanism in cognitive impairment, schizophrenia and bipolar disorder[29]. Overexpression of *MsrA* in the fruit fly *Drosophila melanogaster* is reported to markedly delay reproductive capacity and extend life span[30].

ESR1 and *RBM6–SEMA3F* loci influence reproductive traits

The AFS-decreasing allele at rs726281 (intronic to *ESR1*, which encodes the estrogen receptor) was also associated in the UK Biobank study with earlier AFB in women ($P = 6.9 \times 10^{-3}$) and higher number of children in women ($P = 7.0 \times 10^{-5}$) (**Fig. 3** and **Supplementary Fig. 3**). This locus contains a moderately correlated ($r^2 = 0.25$) intronic variant in *ESR1*, rs67229052, that is also associated with AFS ($P = 1.6 \times 10^{-10}$), AFB ($P = 2.4 \times 10^{-7}$) and number of children in women ($P = 3.7 \times 10^{-8}$) in the UK Biobank study (**Table 2**). In deCODE and WGHS, these associations with rs67229052 were robustly confirmed in women and, in deCODE, were extended to include men (rs67229052: AFB in men, $P = 6.7 \times 10^{-9}$; number of children in men, $P = 1.9 \times 10^{-6}$; **Table 2**). Conditional analyses excluded the presence of independent secondary signals at this *ESR1* locus for either AFS or AFB, and, apart from modest correlation between rs726281 and the reported variant for adult height ($r^2 = 0.16$ with rs3020418), rs726281 and rs67229052 were unrelated to the reported GWAS signals in this gene for puberty timing, breast cancer, breast size and bone mineral density (all $r^2 < 0.05$).

The AFS signal at rs2188151 was highly correlated with a missense variant in *SEMA3F* ($r^2 = 0.7$ with rs1046956; p.Leu503Met in semaphorin-3F isoform X2), which encodes a semaphorin protein, and is a *cis* expression quantitative trait locus (*cis*-eQTL) for *RBM6* ($P = 5 \times 10^{-143}$), which encodes an RNA-binding protein. rs2188151 was correlated with the reported GWAS signals for HDL ($r^2 = 0.45$ with rs2013208) and puberty timing ($r^2 = 0.18$ with rs2188151); in publicly available ReproGen Consortium data, the AFS-decreasing allele conferred later puberty timing. In both men and women (**Table 2**), the AFS-decreasing allele at rs2188151 was also associated with earlier AFB (sex combined, $P = 8.76 \times 10^{-15}$), greater BMI ($P = 3.6 \times 10^{-15}$; lookup in publicly available GIANT Consortium data, $P = 3.9 \times 10^{-5}$), a greater number of children ($P = 9.05 \times 10^{-5}$) and lower likelihood of being childless ($P = 9.38 \times 10^{-5}$).

DISCUSSION

Here we show that a substantial proportion of variation in AFS is due to genetic factors, which likely act through a variety of biological mechanisms, many of which influence either physical traits, such as puberty timing, or personality characteristics, such as risk-taking propensity. Previous studies have invariably focused on only the sociocultural determinants of AFS and the relevance of early AFS to poor educational achievement and other adverse outcomes[12,13]. We recognize the importance of diverse sociocultural factors, which is reflected by the discordant changes in AFS and AFB seen by year of birth in the UK Biobank study (**Supplementary Fig. 4**). However, despite such marked changes in secular factors, the genetic contribution to AFS has remained stable over time (estimated heritability in men and women born before 1950: $h = 0.262$, s.e. = 0.017; heritability in those born in 1950 and onward: $h = 0.283$, s.e. = 0.015).

The neurobehavioral traits associated with AFS can be broadly categorized as traits corresponding to stimulus-seeking behavior (risk-taking) and moderating traits such as intelligence and neuroticism (irritability). Risk-taking is itself related to an exuberant temperament and is moderated by executive functions[31], which are neurocognitive traits implicated in both AFS[17,18] and schizophrenia[32–34]. Furthermore, our extended findings with the AFS signal at *CADM2* indicate that neurobehavioral traits, such as cognitive processing speed[28] and risk-taking propensity, may also have major relevance to measures of reproductive success, such as number of children. We suggest that future population-based study designs to examine the premorbid personality and cognitive traits associated with schizophrenia and bipolar disorder may inform understanding of the psychological and biological processes that contribute to reproductive behavior and fecundity.

A notable finding was the AFS-associated locus intronic to *ESR1*. Effects of estrogen signaling on reproductive ability in women have long been predicted from models of response to fertility-inducing hormones[35,36], consistent with effects of estrogens in promoting ovarian follicle maturation and uterine receptivity to implantation[37,38]. Estrogen receptors are highly expressed in male pituitary, prostate, testis, breast and liver (**Supplementary Fig. 3**), and disrupted signaling leads to low sperm concentrations and infertility, both in humans[39,40] and a rodent model[41]. However, the variants at this locus that we found associated with reproductive behavior (AFS) and reproductive success (AFB), in both sexes, were largely unrelated to the *ESR1* variants reportedly associated with other traits (puberty timing, breast cancer, breast size and bone mineral density). The possibility of a central, tissue-specific effect of the *ESR1* variant rs67229052 is supported by its demonstration as an eQTL for *ESR1* in only one of ~50 Genotype-Tissue Expression (GTEx) tissues (brain_caudate_basal_ganglia; using the proxy SNP, rs4305732, with $r^2 = 0.98$); the allele associated with higher *ESR1* expression ($P = 0.0004$) is also associated with later AFS, later AFB and fewer children. Central estrogen receptor signaling was recently described as a biological regulator of socioreproductive behaviors in male mice[42]. Our findings support a neurobehavioral role for *ESR1* in both men and women. Furthermore, our findings of robust associations of AFS-associated *ESR1* variants with number of children and likelihood of being childless in mid to late adult life suggest that central processes, such as hypothalamic and pituitary sex hormone signaling, and neurocognitive traits may contribute to reproductive success.

Our genetic findings suggest that both physical maturation and neurobehavioral traits contribute to the timing of reproductive activity and success, with consequences for educational and behavioral outcomes. Consideration of individual variation in pubertal timing and also personality characteristics, such as high risk-taking propensity and low neuroticism, may contribute to targeted and more effective approaches to health education and promotion of safer health-related behaviors.

URLs. Genotype imputation and genetic association studies from UK Biobank (accessed 1 August 2015), http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf; Psychiatric Genomics Consortium data sets used for genetic correlation analyses (accessed 1 August 2015), http://www.med.unc.edu/pgc/downloads; MAGENTA (accessed 1 August 2015), https://www.broadinstitute.org/mpg/magenta/; aggregated GWAS summary statistics from the UK Biobank Data Showcase, http://www.ukbiobank.ac.uk/scientists-3/genetic-data/.

METHODS

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

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

ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource. This work was supported by the Medical Research Council (Unit Programme numbers MC_UU_12015/1 and MC_UU_12015/2).

AUTHOR CONTRIBUTIONS

All authors had full access to all of the data and take responsibility for the integrity of the data and the accuracy of the data analysis. F.R.D., P.M.R., K.S., K.K.O. and J.R.B.P. designed the studies. R.A.S., A.H., A.K., G.M., O.T.M., D.G., U.T. and J.E.B. were responsible for collection and generation of data. F.R.D., H.H., D.I.C., L.M.R., P.-R.L., P.S. and J.R.B.P. performed the statistical analysis; all authors contributed to the interpretation of the findings. F.R.D., K.K.O. and J.R.B.P. drafted the manuscript; all authors contributed to the final version.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Parent, A.S. *et al.* The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24**, 668–693 (2003).
2. Lehmann, A., Scheffler, C. & Hermanussen, M. The variation in age at menarche: an indicator of historic developmental tempo. *Anthropol. Anz.* **68**, 85–99 (2010).
3. Sohn, K. A world record in the improvement in biological standards of living in Korea: evidence from age at menarche. in *Discussion Paper Series* 1–34 (Centre for Economic History, Australian National University, 2015).
4. Waylen, A. & Wolke, D. Sex 'n' drugs 'n' rock 'n' roll: the meaning and social consequences of pubertal timing. *Eur. J. Endocrinol.* **151** (suppl. 3), U151–U159 (2004).
5. Gaudineau, A. *et al.* Factors associated with early menarche: results from the French Health Behaviour in School-aged Children (HBSC) study. *BMC Public Health* **10**, 175 (2010).
6. Day, F.R., Elks, C.E., Murray, A., Ong, K.K. & Perry, J.R. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci. Rep.* **5**, 11208 (2015).
7. Charalampopoulos, D., McLoughlin, A., Elks, C.E. & Ong, K.K. Age at menarche and risks of all-cause and cardiovascular death: a systematic review and meta-analysis. *Am. J. Epidemiol.* **180**, 29–40 (2014).
8. Gluckman, P.D. & Hanson, M.A. Evolution, development and timing of puberty. *Trends Endocrinol. Metab.* **17**, 7–12 (2006).
9. Lam, T.H., Shi, H.J., Ho, L.M., Stewart, S.M. & Fan, S. Timing of pubertal maturation and heterosexual behavior among Hong Kong Chinese adolescents. *Arch. Sex. Behav.* **31**, 359–366 (2002).
10. Baams, L., Dubas, J.S., Overbeek, G. & van Aken, M.A. Transitions in body and behavior: a meta-analytic study on the relationship between pubertal development and adolescent sexual behavior. *J. Adolesc. Health* **56**, 586–598 (2015).
11. Hochberg, Z., Gawlik, A. & Walker, R.S. Evolutionary fitness as a function of pubertal age in 22 subsistence-based traditional societies. *Int. J. Pediatr. Endocrinol.* **2011**, 2 (2011).

12. Hawes, Z.C., Wellings, K. & Stephenson, J. First heterosexual intercourse in the United Kingdom: a review of the literature. *J. Sex Res.* **47**, 137–152 (2010).
13. Waldron, M. *et al.* Parental separation, parental alcoholism, and timing of first sexual intercourse. *J. Adolesc. Health* **56**, 550–556 (2015).
14. Lenciauskiene, I. & Zaborskis, A. The effects of family structure, parent–child relationship and parental monitoring on early sexual behaviour among adolescents in nine European countries. *Scand. J. Public Health* **36**, 607–618 (2008).
15. Ingham, R., Woodcock, A. & Stenner, K. Getting to know you… young people's knowledge of their partners at first intercourse. *J. Community Appl. Soc. Psychol.* **1**, 117–132 (1991).
16. Allen, J.P., Schad, M.M., Oudekerk, B. & Chango, J. What ever happened to the "cool" kids? Long-term sequelae of early adolescent pseudomature behavior. *Child Dev.* **85**, 1866–1880 (2014).
17. Martin, N.G., Eaves, L.J. & Eysenck, H.J. Genetical, environmental and personality factors influencing the age of first sexual intercourse in twins. *J. Biosoc. Sci.* **9**, 91–97 (1977).
18. Harden, K.P. & Mendle, J. Why don't smart teens have sex? A behavioral genetic approach. *Child Dev.* **82**, 1327–1344 (2011).
19. Perry, J.R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
20. Day, F.R. *et al.* Genetic determinants of puberty timing in men and women: shared genetic aetiology between sexes and with health-related outcomes. *Nat. Commun.* **6**, 8842 (2015).
21. Burgess, S., Butterworth, A., Malarstig, A. & Thompson, S.G. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *Br. Med. J.* **345**, e7325 (2012).
22. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
23. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
24. Stulp, G., Barrett, L., Tropf, F.C. & Mills, M. Does natural selection favour taller stature among the tallest people on earth? *Proc. Biol. Soc.* **282**, 20150211 (2015).
25. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
26. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
27. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
28. Ibrahim-Verbaas, C.A. *et al.* GWAS for executive function and processing speed suggests involvement of the *CADM2* gene. *Mol. Psychiatry* **21**, 189–197 (2016).
29. Emiliani, F.E., Sedlak, T.W. & Sawa, A. Oxidative stress and schizophrenia: recent breakthroughs from an old story. *Curr. Opin. Psychiatry* **27**, 185–190 (2014).
30. Ruan, H. *et al.* High-quality life extension by the enzyme peptide methionine sulfoxide reductase. *Proc. Natl. Acad. Sci. USA* **99**, 2748–2753 (2002).
31. Lahat, A. *et al.* Temperamental exuberance and executive function predict propensity for risk taking in childhood. *Dev. Psychopathol.* **24**, 847–856 (2012).
32. Martin, A.K., Robinson, G., Dzafic, I., Reutens, D. & Mowry, B. Theory of mind and the social brain: implications for understanding the genetic basis of schizophrenia. *Genes Brain Behav.* **13**, 104–117 (2014).
33. Dinzeo, T.J. & Docherty, N.M. Normal personality characteristics in schizophrenia: a review of the literature involving the FFM. *J. Nerv. Ment. Dis.* **195**, 421–429 (2007).
34. Hoptman, M.J. Impulsivity and aggression in schizophrenia: a neural circuitry perspective with implications for treatment. *CNS Spectr.* **20**, 280–286 (2015).
35. Altmäe, S. *et al.* Allelic estrogen receptor 1 (*ESR1*) gene variants predict the outcome of ovarian stimulation in *in vitro* fertilization. *Mol. Hum. Reprod.* **13**, 521–526 (2007).
36. de Mattos, C.S. *et al. ESR1* and *ESR2* gene polymorphisms are associated with human reproduction outcomes in Brazilian women. *J. Ovarian Res.* **7**, 114 (2014).
37. Drummond, A.E. The role of steroids in follicular growth. *Reprod. Biol. Endocrinol.* **4**, 16 (2006).
38. Zhang, S. *et al.* Physiological and molecular determinants of embryo implantation. *Mol. Aspects Med.* **34**, 939–980 (2013).
39. Hess, R.A. *et al.* A role for oestrogens in the male reproductive system. *Nature* **390**, 509–512 (1997).
40. Carreau, S. & Hess, R.A. Oestrogens and spermatogenesis. *Phil. Trans. R. Soc. Lond. B* **365**, 1517–1535 (2010).
41. Couse, J.F. *et al.* Postnatal sex reversal of the ovaries in mice lacking estrogen receptors α and β. *Science* **286**, 2328–2331 (1999).
42. Lee, H. *et al.* Scalable control of mounting and attack by $Esr1^{+}$ neurons in the ventromedial hypothalamus. *Nature* **509**, 627–632 (2014).

ONLINE METHODS

UK Biobank. The UK Biobank study design has been reported[43]. Briefly, all people aged 40–69 years who were registered with the National Health Service and living up to ~25 miles from one of the 22 study assessment centers were invited to participate in 2006–2010. Overall, about 9.2 million invitations were mailed to recruit 503,325 participants (a response rate of 5.47%)[44]. Extensive self-reported baseline data were collected by questionnaire, in addition to anthropometric assessments. Details of the phenotypes analyzed here are shown in **Supplementary Table 15**. All participants provided written informed consent, the study was approved by the National Research Ethics Service Committee North West–Haydock and all study procedures were performed in accordance with the World Medical Association Declaration of Helsinki ethical principles for medical research.

Genetic analysis in the UK Biobank study. We analyzed data from the May 2015 release of imputed genetic data from UK Biobank, comprising ~73 million SNPs, short indels and large structural variants in 152,249 individuals. Full details are published (see URLs). Briefly, the samples were genotyped on two slightly different arrays. Approximately 50,000 were genotyped on a custom UL BiLEVE study array, and the remaining samples (~100,000) were genotyped on the UK Biobank Axiom array from Affymetrix, which was specifically designed to optimize imputation performance in GWAS. Removal of SNPs with missing data, multiallelic SNPs, SNPs with MAF <1% and 1,037 sample outliers resulted in a data set with 641,018 autosomal SNPs in 152,256 samples for phasing and imputation. Imputation was performed using a reference panel created by merging the UK10K haplotype panel with the 1000 Genomes Project Phase 3 reference panel.

In addition to the quality control metrics performed centrally by UK Biobank, we defined a subset of samples of 'white European' ancestry using a *k*-means clustering approach applied to the first four principal components calculated from genome-wide SNP genotypes. All individuals included in this group also self-identified by questionnaire as being of white ancestry. Autosomal SNPs were analyzed by linear mixed models implemented in BOLT-LMM[23] to account for cryptic population structure and relatedness within this group in our genetic association tests. X-chromosome SNPs were analyzed using SNPTEST[45]. Genotyping chip was included as a binary covariate in all models. Any SNPs with imputation quality <0.4 or MAF <0.1% were excluded after analysis. After the application of quality control criteria, a maximum of 142,630 individuals were available for analysis with genotype and phenotype data. There was no substantial effect on test statistics after exclusion from the models of individuals with any reported psychiatric illness. Genomic loci were defined on the basis of physical proximity using a 1-Mb window centered on each SNP. Signals were excluded from consideration if they were significantly associated with genotyping chip.

Variance-component analyses were performed in the subset of individuals of only 'white British' genetic ancestry (maximum analyzed $n = 99{,}241$) using REML models in BOLT-LMM[46]. Genetic variance was calculated for all genotyped autosomal SNPs for which quality control was performed, adjusting for chip status and the top five genetically determined principal components.

Replication studies. *deCODE Genetics.* The whole genomes of 8,453 Icelanders were sequenced using Illumina technology to a mean depth of at least 10× (median depth of 32×), and SNPs and indels were identified and their genotypes called using joint calling with the Genome Analysis Toolkit HaplotypeCaller (GATK version 3.3.0)[47]. Genotype calls were improved by using information about haplotype sharing, taking advantage of the fact that all the sequenced individuals had also been chip-typed with genotypes submitted to long-range phasing. Around 30 million sequence variants were then imputed into 150,656 Icelanders who had been genotyped using an Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni1, Omni2.5 or OmniExpress array[48]. SNPs were excluded if they had (i) yield <95%, (ii) MAF <1% in the population or (iii) significant deviation from Hardy–Weinberg equilibrium ($P < 0.001$), (iv) if they produced an excessive inheritance error rate (over 0.001) or (v) if there was substantial difference in allele frequency between chip types (SNPs were removed from just a single chip if that resolved all differences, but were removed from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. Using genealogical information, the sequence variants were imputed into 294,212 untyped relatives of chip-typed individuals to further increase the sample size for association analysis and to increase the power to detect associations. The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. All subjects gave their written informed consent.

Women's Genome Health Study. WGHS comprises 23,294 participants derived from the Women's Health Study (WHS) of European-ancestry individuals who provided baseline blood samples. These individuals represent approximately 72% of the 39,876 initially healthy female healthcare professionals, aged >45 years at baseline, who participated in a randomized, placebo-controlled trial of aspirin and vitamin E in primary prevention over 10 years of incident CVD. The Institutional Review Board of Brigham and Women's Hospital, Boston, approved all analyses[49], and all participants gave informed consent for research. Genotyping was performed using the HumanHap300 Duo + platform (Illumina) with the Infinium II protocol. For quality control, all samples were required to have successful genotyping using BeadStudio v.3.3 software (Illumina) for at least 98% of the SNPs. The subset of 23,294 women had self-reported European ancestry that could be verified by multidimensional scaling analysis of identity by state using 1,443 ancestry-informative markers in PLINK v.1.06. In the final data set, a total of 339,596 SNPs were retained with MAF >1%, call rate >90% and Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$. Genotypes for a total of 30,052,423 (autosomes) and 1,264,493 (X-chromosome) SNPs were imputed from the experimental genotypes and phase information from the 1000 Genomes Project phase I v.3 release (March 2012) ALL panel using MaCH (v. 1.0.16) and Minimac (29 May 2012 release). 332,927 genotyped SNPs that were selected by Hardy–Weinberg equilibrium $P > 1 \times 10^{-6}$ but unrestricted by MAF could be reconciled with the 1000 Genomes Project ALL panel and were used for imputation.

Genetic correlations. Genetic correlations (r_g) were calculated between puberty timing, AFS, AFB and 44 other complex traits or diseases in publicly available data sets using LD Score regression[22] (see URLs). Genome-wide SNP associations were also generated in the UK Biobank study for the following traits: number of children, childlessness, number of sexual partners, smoking status, alcohol intake, years of education, risk-taking propensity, suffering from nerves, irritability, happiness and intelligence. Details on these phenotypes are provided in **Supplementary Table 15**. A conservative Bonferroni-corrected *P*-value threshold of $P < 1.1 \times 10^{-3}$ (= 0.05/44) was used to define significant associations.

Mendelian randomization. Mendelian randomization is an analytical method to infer the likely causal, non-confounded relationship between an exposure trait and an outcome. This model assumes that the included genetic variants are associated with the exposure trait and do not influence the outcome by other unrelated biological pathways (pleiotropy)[21]. We calculated approximated genetic risk scores using summary statistics for SNPs reported at genome-wide significance for the modeled traits. Effect estimates were based on data from reported GWAS for adult height[50], BMI[27], puberty timing[19], schizophrenia[51], and skin freckling and hair color[25] or were from the current GWAS for AFS in the UK Biobank study. To avoid bias, outcomes were tested in data sets (UK Biobank, deCODE, WGHS or publicly available data sets) that were independent of the discovery GWAS for each exposure. The associations with weighted allele scores were scaled to indicate the causal effect of a 1 s.d. change in the normalized exposure variable, unless otherwise stated (**Supplementary Table 6**).

Pathway analyses and functional insight on SNPs. MAGENTA (see URLs) was used to test the full genome-wide discovery data set for genetic associations with biological pathways defined by Gene Ontology, PANTHER, KEGG and Ingenuity. MAGENTA implements a gene set enrichment analysis (GSEA)-based approach, where each gene in the genome is mapped to a single index SNP with the lowest *P* value within the window ranging from 110 kb upstream to 40 kb downstream of the gene. This *P* value, representing a gene score, is then corrected in a regression model for confounding factors such as gene size, SNP density and LD-related properties. Genes within the HLA region were excluded from analysis because of difficulties in accounting

doi:10.1038/ng.3551

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 one calculated from 1,000,000 randomly permuted pathways of identical size. This comparison generates an empirical GSEA *P* value for the pathway. An individual pathway was defined as being significantly enriched when it reached false discovery rate (FDR) <0.05 in either analysis. In total, 3,216 pathways were tested for enrichment of multiple modest associations with AFS.

Each AFS-associated locus was annotated for possible genomic functions using ENCODE and Epigenomics Roadmap data in HaploReg v4.1 (ref. 52).

43. Allen, N.E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).
44. Sudlow, C. *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
45. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
46. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
47. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
48. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
49. Ridker, P.M. *et al.* Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. *Clin. Chem.* **54**, 249–255 (2008).
50. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
51. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
52. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).