Physical and neuro-behavioural determinants of reproductive onset and success

Felix R. Day[1]*, Hannes Helgason[2,3]*, Daniel I. Chasman[4,5]*, 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]*, Ken K. Ong[1]*, John R.B. Perry[1]*

Affiliations

1. MRC Epidemiology Unit, University of Cambridge, UK

2. deCODE genetics/Amgen, Reykjavik, Iceland.

3. School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland.

4. Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Ave., East, Boston, MA 02215, USA

5. Harvard Medical School, Boston, MA 02115, 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.

*denotes equal contribution

Correspondence to John R.B. Perry (john.perry@mrc-epid.cam.ac.uk) and Ken Ong (ken.ong@mrc-epid.cam.ac.uk).

ABSTRACT

The ages of puberty, first sexual intercourse and first birth signify the onset of reproductive ability, behaviour and success, respectively. These sequenced events have behavioural, physiological and health significance, and may also influence overall reproductive fitness. In a genome-wide association study of 125,667 white men and women aged 40-69 in the UK Biobank Study, we identify 38 sequence variants with association P-values $<5\times10^{-8}$ with age at first sexual intercourse. Findings were taken forward in up to 241,910 men and women from deCODE Iceland and 20,187 from Women's Genome Health Study. Several of these loci also exhibit strong associations with behavioural traits (rs4856591 in *CADM2* and risk taking propensity: $P=4.3\times10^{-10}$; rs73195303 in *MSRA* and irritable temperament: $P=5.8\times10^{-11}$) and other reproductive traits (rs67229052 in *ESR1* and both age at first birth: $P=1.2\text{x}10^{-13}$ and number of children: $P=4.8\times10^{-12}$; rs2188151 in *SEMA3F* and age at first birth: $P=8.76\times10^{-15}$). In Mendelian randomisation analyses, we demonstrate likely causal influences of earlier puberty timing on earlier first sexual intercourse, earlier first birth and fewer years of education. In turn, likely causal consequences of earlier first sexual intercourse include reproductive, educational, psychiatric and cardiometabolic outcomes. These findings point to the existence of developmental and neuro-behavioural regulators of reproductive activity and success.

MAIN

Introduction

The age of puberty, the transition from childhood to sexual maturity and reproductive ability, has fallen markedly over the last century in most populations, illustrated by the average age at menarche of 18 years in 1880 to 12.5 years in 1980 (ref.[1,2]). This decline was initially observed in industrialized, western countries, and more recently, and often far more rapidly, in countries with more recent economic transitions[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 towards risk-taking behaviours[4,5], lower educational attainment, greater susceptibility to several adverse health outcomes[6] and, in women, to 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 on 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 behaviours have a strong influence on teenagers' sexual behaviour[14-16]. Twin studies have suggested some genetic contribution to AFS[17,18], however observations of older AFS among monozygous compared to dizygous twins[13] casts doubt on the validity 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 appear to have concordant effects on puberty timing also in males[20]. A valuable application of such GWAS findings is the use of genetic variants, with robust association with a specified trait, as instrumental variables in ("Mendelian randomisation") analyses to test the likely causal relationships with that trait with less risk of confounding compared to traditionally observed associations[21]. Here, we use this approach to test the causal relationship of puberty timing[19] to 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 between the timings of the onset of reproductive ability, activity and success to other behavioural and health-related outcomes.

RESULTS

Shared genetic architecture between reproductive onset 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 ~46M 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, median and inter-quartile range (IQR) for AFS was 18 years (16-21) in men and 18 years (17-21) in women, median (IQR) AFB was 25 years (22-28) (recorded in women only), and the SNP-based test REML[23] indicated moderate heritability for AFS both in men (h=0.248, s.e. 0.010) and in 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 three scores, we found moderately strong positive genetic correlations between puberty timing and AFS, both in women (r_g=0.22, P=1.2×10^{-16}) and men (r_g=0.26, P=9.5×10^{-8}), and between puberty timing and AFB (r_g=0.24, P=9.0×10^{-13}; women only). Furthermore, we found a remarkably strong genetic correlation between AFS and AFB (r_g=0.86, P=3.1×10^{-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 used whole-genome LD score regression[22] to test genetic correlations between AFS and other health-related traits using publicly available GWAS datasets or original GWAS findings in UK Biobank (see Methods). We identified genetic correlations between AFS and 22 of the 44 tested outcomes or traits after correction for multiple testing (P<1.1×10^{-3}), including inverse correlations with BMI, T2D and CVD, and also with a variety of behavioural (e.g. smoking; alcohol intake) and neurological traits (e.g. intelligence; risk taking propensity) and psychiatric outcomes (attention deficit hyperactivity disorder (ADHD); schizophrenia) (**Figure 2**).

In a sex-combined GWAS in UK Biobank, we identified 33 loci with variants associated at P<5×10^{-8} with AFS (**Table 1, Figure 1, Figure S1-2**). 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 loci, effect sizes ranged from 0.02-0.33 SDs, minor allele frequencies (MAF) ranged from 0.15% to 49%. Seventeen of these variants were associated with AFS at the more stringent P-value threshold P<1×10^{-9}. In UK Biobank, no loci were found associated at P<5×10^{-8} with AFB in women.

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 datasets: deCODE (N=117,626 males, 124,284 females) and the Women's Genome Health Study (WGHS, N=20,187 women). A weighted SNP genotype score of our 38 novel signals for AFS was strongly associated with AFB in both deCODE ($P=3.3\times10^{-21}$) and WGHS ($P=9.2\times10^{-4}$) (**Table S1, S2 and S6**). The subset of 21 'less stringently associated' AFS variants (those with association P-values: $1\times10^{-9}\geq P<5\times10^{-8}$) was also collectively associated with AFB (weighted SNP genotype score in deCODE: $P=3.1\times10^{-7}$).

Biological determinants of age at first sexual intercourse

None of the 38 lead AFS-associated variants (or their $r^2>0.8$ proxies) were non-synonymous SNPs, however several were located in regions containing promoter/enhancer histone marks, DNAse hypersensitive sites or protein binding sites (**Table S3**). In addition, the majority of these variants either altered regulatory motifs or were associated in *cis* with gene expression (**Table S4**)

To identify mechanisms that might regulate AFS, we performed a systematic test of all annotated biological pathways for enrichment of genes located near to AFS-associated signals, using MAGENTA (see Methods, **Table S5**). 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 a genetic instrumental variable by calculating a weighted allele score from the SNP genotypes at reported signals with robust associations with each phenotype (**see methods**, **Tables S6-S10**). In both men and women, genetically-predicted earlier puberty timing ($P_{women}=2.0\times10^{-9}$, $P_{men}=4.7\times10^{-11}$) and genetically-predicted greater BMI ($P_{men}=5.5\times10^{-8}$, $P_{women}=2.2\times10^{-4}$) appeared to promote earlier AFS (**Tables S7-S8**). Genetically-predicted greater height appeared to promote later AFS in men ($P=1.0\times10^{-5}$) and in women ($P=1.1\times10^{-3}$) (**Table S9**), which is consistent with reported non-genetic associations between greater height and later AFB in European men and women[24].

Prompted by our observation of a novel AFS locus near *MC1R* (rs369230 $r^2=0.12$ with the variant rs12931267 that is strongly associated with hair colour)[25,26], we tested genetic instrumental variables for skin freckling and hair colour, which are traits regulated by this gene[25,26]. Genetically-predicted skin freckling appeared to promote later AFS in women ($P=6.3\times10^{-9}$) but not in men (P=0.47), and genetically-predicted red hair appeared to promote later AFS in both men (P=0.02) and women ($P=9.3\times10^{-5}$; **Table S10**).

Relationships to other behavioural, reproductive and health outcomes

To test whether puberty timing and AFS might be causally related to other behavioural, 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 datasets that were used to generate the allele weightings (i.e. outcomes related to AFS were tested in datasets other than UK Biobank). Genetic associations were scaled to indicate the likely causal effect of a one SD change in normalised puberty timing or AFS.

Genetically-predicted earlier puberty timing decreased the age at leaving education (standardised beta: 0.061, $P=4.0\times10^{-7}$; in UK Biobank) (**Table S11**). Similarly, genetically-predicted earlier AFS decreased the likelihood of attaining university-level education (standardised OR=0.74, $P=3.7\times10^{-5}$; in publicly-available Social Science Genetic Association Consortium data) and increased the likelihood of ever-smoking (standardised OR=1.33, $P=2.0\times10^{-3}$; in publicly-available Tobacco and Genetics Consortium data) (**Table S9**). For reproductive outcomes, in the deCODE data, each one SD genetically-predicted earlier AFS promoted earlier AFB (women: standardised beta=1.71, $P=2.2\times10^{-17}$; men: 1.69, $P=2.6\times10^{-13}$; combined $P=3.3\times10^{-21}$), a greater number of children (women: standardised beta= 0.035, P=0.006; men: 0.012, P=0.34; combined P=0.0044), and lower likelihood of being childless (women: standardised OR=0.67, P=0.034; men: OR=0.66, P=0.009; combined P=0.0022) (**Table S1)**. Similarly, each one SD genetically-predicted earlier puberty timing promoted earlier AFB (standardised beta: 0.37, $P=5.8\times10^{-8}$; assessed in UK Biobank women only) and earlier age at last birth (0.37, $P=3.7\times10^{-7}$; in UK Biobank women), but had little effect on other reproductive outcomes (**Tables S6 and S11**).

We noted that several of the 38 novel AFS signals were located in or near genes reportedly implicated in brain development (*BARHL2*, *SEMA3F*, *ZIC4/ZIC1*, *DPYSL4*, *DIAPH3*), or neuronal activity and/or susceptibility to schizophrenia/bipolar disorder (*CADM2*, *LRP4*, *GRIA4*, *CACMA1D*, *HCN1*, *GRIA4*, *DRD2*, *FURIN*, *GNAL* and *VRK2*) (**Table 1 and S3**), consistent with our observed shared genetic architecture between AFS and ADHD (r_g= -0.38, $P=5.9\times10^{-4}$), and between AFS and schizophrenia (r_g= -0.10, $P=7.3\times10^{-4}$) (**Figure 2**). We used a bi-directional Mendelian randomisation approach to test the likely causal relationships: susceptibility to schizophrenia appeared to lower AFS (in UK Biobank with or without exclusion of individuals with self-reported psychiatric illness: P=0.005; **Table S13**), but also, earlier AFS appeared to increase susceptibility to schizophrenia (in publicly available Psychiatry Genetics Consortium data, $P=4.1\times10^{-11}$, **Table S12**), suggesting a pleiotropic relationship between these traits. The substantial shared genetic architecture between AFS and self-reported risk taking propensity (r_g= -0.46, $P=7.3\times10^{-28}$) (**Figure 2**) gives insights into possible common determinants of AFS and schizophrenia.

To explore potential specific neuro-behavioural mechanisms that might contribute to the aetiology of AFS, we performed a look-up of the 38 individual AFS loci for associations with fifteen other behavioural, reproductive and health-related traits in UK Biobank and other independent studies (**see methods; Figure 3; Table S14**).

CADM2 and MSRA loci influence multiple behavioural traits

The AFS signal represented by rs57401290 is intronic in *CADM2*, which encodes a neuronal cell-adhesion molecule. rs57401290, or highly correlated SNPs in this locus, also showed genome-wide significant associations in UK Biobank with: self-reported risk taking propensity (rs57401290: $P=5.3\times10^{-9}$; $r^2=0.65$ with the lead *CADM2* SNP for this trait rs4856591: $P=4.3\times10^{-10}$), number of sexual partners (rs57401290: $P=6.0\times10^{-7}$; $r^2=0.60$ with lead SNP rs5850688: $P=4.1\times10^{-8}$) and number of children (rs57401290: $P=6.2\times10^{-7}$, $r^2=0.65$ with lead SNP rs4856591: $P=3.8\times10^{-11}$, replication in deCODE P=0.006) (**Table S14**). In each case, the AFS-decreasing allele conferred higher values of those outcomes. rs57401290 is modestly correlated with the reported signal at this locus for BMI[27] ($r^2=0.11$ with rs13078960, the AFS-decreasing allele also increases BMI) and is strongly correlated with the reported signal in *CADM2* for cognitive processing speed[28] (rs17518584 $r^2=0.80$, the AFS-decreasing allele also decreases processing speed). *CADM2* shows highest expression in the prefrontal cortex and is involved in a range of neuronal processes, including glutamate signaling, gamma-aminobutyric acid transport and neuron cell-cell adhesion[28].

The AFS-decreasing allele at rs658385 (~25 kb downstream of *MSRA*) was also associated in UK Biobank with lower likelihood of self-reported irritable temperament ($P=3.8\times10^{-4}$) and was modestly correlated ($r^2=0.14$) with the lead *MSRA* SNP for this trait (rs73195303, $P=5.8\times10^{-11}$). 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, which is a candidate mechanism in cognitive impairment and schizophrenia/bipolar disorder[29]. Overexpression of *msra* in the fruit fly Drosophila 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 in *ESR1*, which encodes the estrogen receptor) was also associated in UK Biobank with earlier AFB in women ($P=6.9\times10^{-3}$), higher number of children in women ($P=7.0\times10^{-5}$) (**Figure 3**). This locus contains a moderately correlated intronic variant rs67229052 in *ESR1* ($r^2=0.25$) that is also associated with AFS ($P=1.6\times10^{-10}$), AFB ($P=2.4\times10^{-7}$), and number of children in women ($P=3.7\times10^{-8}$) in UK Biobank (**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 $P=6.7\times10^{-9}$; number of children in men $P=1.9\times10^{-6}$; **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 adult height variant ($r^2=0.16$ 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 is highly correlated with a missense variant in *SEMA3F* (r^2=0.7 with rs1046956; Leu503Met in semaphorin-3F isoform X2), which encodes a semaphorin protein, and is a *cis* eQTL for *RBM6* (P=5×10^{-143}), which encodes an RNA binding protein. rs2188151 is 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 confers 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=8.76×10^{-15}), greater BMI (P=3.6×10^{-15}; look-up in publicly available GIANT consortium data: P=3.9×10^{-5}), a greater number of children (P=9.05×10^{-5}) and lower likelihood of being childless (P=9.38×10^{-5}).

DISCUSSION

Here, we show that a substantial proportion of the 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 socio-cultural determinants of AFS and the relevance of early AFS to poor educational achievement and other adverse outcomes[12,13]. We recognise the importance of diverse socio-cultural factors, which are reflected by the discordant changes in AFS and AFB seen by year of birth in the UK Biobank study (**Figure S4**). However, despite such marked secular changes, the genetic contribution to AFS has remained stable over time (estimated heritability in men and women born pre-1950: h=0.262, s.e. 0.017; in those born in 1950 onwards: h=0.283, s.e. 0.015).

The neuro-behavioural traits associated with AFS can be broadly categorised as stimulus-seeking (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 function[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 neuro-behavioural traits, such as cognitive processing speed[28] and risk taking propensity, may also have important relevance to measures of reproductive success, such as number of children. We suggest that future population-based study designs to study the pre-morbid personality and cognitive traits associated with schizophrenia and bipolar disorder may inform the psychological and biological processes that contribute to reproductive behaviour and fecundity.

A notable finding was the AFS locus intronic in *ESR1.* Effects of estrogen signalling on reproductive ability in women have been long predicted from models of response to fertility-inducing hormones[35,36], consistent with effects of estrogens on promoting ovarian follicle maturation and uterine receptivity to implantation[37,38]. Estrogen receptors are highly expressed in male pituitary, prostate, testis, breast and liver (**Figure S3)**, and disrupted signalling leads to low sperm concentrations and infertility, both in humans[39,40] and in a rodent model[41]. However, the variants at this locus that we found associated with

reproductive behaviour (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 this *ESR1* variant rs67229052 is supported by its demonstration as an eQTL for *ESR1* in only one of ~50 GTEX tissues (with "brain_caudate_basal_ganglia", using the r2=0.98 proxy rs4305732); the allele associated with higher *ESR1* expression (P=0.0004) is also associated with later AFS, later AFB and fewer children. Central estrogen receptor signalling was recently described as a biological regulator of socio-reproductive behaviours in male mice[42]. Our findings support a neuro-behavioural role for *ESR1* in both men and women. Furthermore, our findings of robust associations between AFS-associated *ESR1* variants and number of children and likelihood of being childless in mid-late adult life suggest that central processes, such as hypothalamic-pituitary sex hormone signalling and neuro-cognitive traits, may contribute to reproductive success.

Our genetic findings indicate that both physical maturation and neuro-behavioural traits contribute to the timing of reproductive activity and success, with consequences for educational and behavioural 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 behaviours.

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 centres were invited to participate in 2006-10. Overall, about 9.2 million invitations were mailed in order to recruit 503,325 participants (i.e. 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 analysed here are shown in **Table S15**. All participants provided informed written 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 UK Biobank

We analysed data from the May 2015 release of imputed genetic data from UK Biobank, containing ~73M 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 by 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 studies. Removal of SNPs with missing data, multi-allelic SNPs, SNPs with a minor allele frequency (MAF) <1%, and 1,037 sample outliers, resulted in a dataset 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 Phase 3 reference panel.

In addition to the quality control metrics performed centrally by UK Biobank, we defined a subset of “white European” ancestry samples using a K-means clustering approach applied to the first four principle components calculated from genome-wide SNP genotypes. All individuals defined in this group also self-identified by questionnaire as being of white ancestry. Autosomal SNPs were analysed 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 analysed using SNPTEST[45]. Genotyping chip was included as a binary covariate in all models. Any SNPs with an imputation quality < 0.4 or MAF < 0.1% were excluded post-analysis. After application of QC 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 those individuals with any reported psychiatric illness. Genomic loci were defined on the basis of physical proximity using a 1 Mb window. 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 analysed N=99,241) using Restricted Estimate Maximum Likelihood (REML) models in BOLT-LMM[46]. Genetic variance was calculated on all QC'd genotyped autosomal SNPs, adjusting for chip status and the top 5 genetically determined principal components.

Replication studies

deCODE Genetics: Whole genomes of 8,453 Icelanders were sequenced using Illumina technology to a mean depth of at least 10X (median 32X) and SNPs and indels 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 and long range phased. Around 30M sequenced variants were then imputed into 150,656 Icelanders who had been genotyped using the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips[48]. SNPs were excluded if they had (i) yield <95%, (ii) MAF <1% in the population or (iii) significant deviation from Hardy–Weinberg equilibrium (HWE) (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 (from just a single chip if that resolved all differences, but from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. Using genealogic information, the sequence variants were imputed into 294,212 un-typed relatives of chip-typed individuals to further increase the sample size for association analysis and increased 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): WGHS derives from the 23,294 Women's Health Study (WHS) European ancestry participants who provided baseline blood samples. They 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]. Genotyping was performed using the HumanHap300 Duo ''+'' platform (Illumina, San Diego, CA) with the Infinium II protocol. For quality control, all samples were required to have successful genotyping using the BeadStudio v.3.3 software (Illumina, San Diego, CA) 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 dataset, a total of 339,596 SNPs were retained with: MAF >1%, call rate >90%, and HWE $P<10^{-6}$. Genotypes for a total of 30,052,423 (autosomes) + 1,264,493 (X) SNPs were imputed from the experimental genotypes and phase information from the 1000G phase I v.3 release (March 2012) ALL

panel using MaCH (v. 1.0.16) and Minimac (release 5/29/2012). 332,927 genotyped SNPs that were selected by HWE $P>10^{-6}$ but unrestricted by MAF could be reconciled with the 1000G 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/diseases in publicly-available datasets using LD Score Regression[22] (see URLs). Genome-wide SNP association were also generated in UK Biobank 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 of these phenotypes are described in **Table S15**. A conservative Bonferroni corrected P-value threshold of $P<1.1\times10^{-3}$ [=0.05/44] was used to define significant associations.

Mendelian Randomization

Mendelian randomization is an analytical method to infer the likely causal unconfounded causal relationship between an exposure trait and an outcome. It is considered to be more accurate than estimate causal , using genetic variants that are associated with the exposure trait and do not influence the outcome by other unrelated biological pathways ('pleiotropy')[21]. We calculated using weighted allele scores from SNP genotypes at signals robustly associated with each modelled exposure as genetic instrumental variables for those traits. SNP genotypes were based on reported GWAS for adult height[50], BMI[27], puberty timing[19], schizophrenia[51], skin freckling and hair colour[25], or from the current GWAS for AFS in UK Biobank. To avoid bias, outcomes were tested in datasets (UK Biobank, deCODE, WGHS or publicly-available datasets) 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 one SD change in the normalised exposure variable (**Table S6**).

Pathway analyses and functional insight of SNPs

Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA - https://www.broadinstitute.org/mpg/magenta/) was used to test the full genome-wide discovery dataset for genetic associations with the 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 a 110 kb upstream, 40 kb downstream window. 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 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, 3216 pathways were tested for enrichment of multiple modest associations with AFS.

Each AFS-associated locus was annotated for possible genomic functions using ENCODE and Epigenome Roadmap project data in HaploReg v4.1[52].

URLs

Genotype imputation and genetic association studies of UK Biobank www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf (accessed 01/08/2015)

Datasets used for genetic correlation analyses http://www.med.unc.edu/pgc/downloads (accessed 01/08/2015).

References

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**, 668-93 (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. The 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**, U151-9 (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., Elks, C.E., Murray, A.M., 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. *Scientific Reports* **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**, 359-66 (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**, 586-98 (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**, 137-52 (2010).
13. Waldron, M. *et al.* Parental separation, parental alcoholism, and timing of first sexual intercourse. *J Adolesc Health* **56**, 550-6 (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**, 607-18 (2008).
15. Ingham, R., Woodcock, A. & Stenner, K. Getting to know you . . . young people's knowledge of their partners at first intercourse. *Journal of Community & Applied Social Psychology* **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**, 1866-80 (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**, 91-7 (1977).
18. Harden, K.P. & Mendle, J. Why don't smart teens have sex? A behavioral genetic approach. *Child Dev* **82**, 1327-44 (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. *Nature Communications* **6**, 8842 (2015).
21. Burgess, S., Butterworth, A., Malarstig, A. & Thompson, S.G. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *Bmj* **345**, e7325 (2012).
22. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
23. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (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 Sci* **282**(2015).
25. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* **39**, 1443-52 (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* (2015).
29. Emiliani, F.E., Sedlak, T.W. & Sawa, A. Oxidative stress and schizophrenia: recent breakthroughs from an old story. *Curr Opin Psychiatry* **27**, 185-90 (2014).
30. Ruan, H. *et al.* High-quality life extension by the enzyme peptide methionine sulfoxide reductase. *Proc Natl Acad Sci U S A* **99**, 2748-53 (2002).
31. Lahat, A. *et al.* Temperamental exuberance and executive function predict propensity for risk taking in childhood. *Dev Psychopathol* **24**, 847-56 (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**, 104-17 (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**, 421-9 (2007).
34. Hoptman, M.J. Impulsivity and aggression in schizophrenia: a neural circuitry perspective with implications for treatment. *CNS Spectr* **20**, 280-6 (2015).
35. Altmae, S. *et al.* Allelic estrogen receptor 1 (ESR1) gene variants predict the outcome of ovarian stimulation in in vitro fertilization. *Mol Hum Reprod* **13**, 521-6 (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**, 939-80 (2013).
39. Hess, R.A. *et al.* A role for oestrogens in the male reproductive system. *Nature* **390**, 509-12 (1997).
40. Carreau, S. & Hess, R.A. Oestrogens and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* **365**, 1517-35 (2010).
41. Couse, J.F. *et al.* Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. *Science* **286**, 2328-31 (1999).
42. Lee, H. *et al.* Scalable control of mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. *Nature* **509**, 627-32 (2014).
43. Allen, N.E., Sudlow, C., Peakman, T., Collins, R. & Biobank, U.K. 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**, 906-13 (2007).
46. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-92 (2015).
47. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
48. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* **47**, 435-44 (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**, 249-55 (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* (2014).
51. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (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**, D930-4 (2012).

Figure Legends

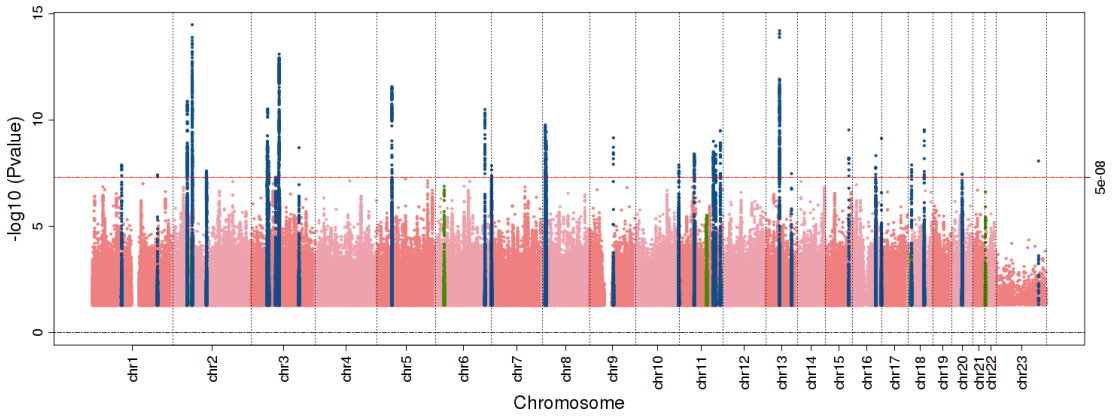

Figure 1 | Manhattan plot of the GWAS for age at first sexual intercourse

Manhattan plot illustrating results of the genome-wide association study (GWAS) meta-analysis for age at first sexual intercourse in up 59,357 men and 66,310 women of European descent in the UK Biobank study. -log10 P-values for each SNP (Y-axis) are plotted by chromosomal position (X-axis). The red line indicates the threshold for genome-wide statistical significance ($P=5\times10^{-8}$). Blue dots represent SNPs within a 1Mb base pair window around the genome-wide significant signals, green dots indicate sex specific effects.

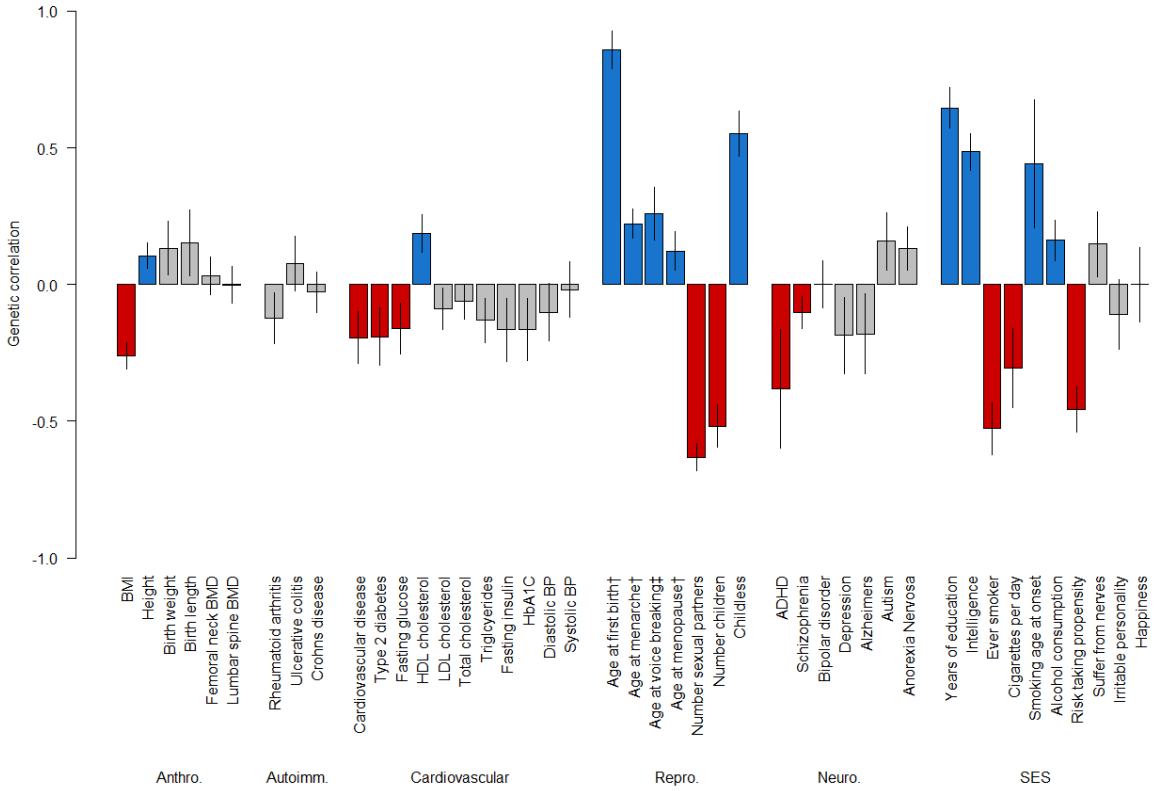

Figure 2 | Bar chart of genetic correlations age at first sexual intercourse

Whole-genome LD score regression tested genome-wide SNP associations for age at first sexual intercourse against similar data for 44 other traits. Blue (positive correlation) and red (negative correlation) bars indicate the 22 traits that showed a significant genetic correlation after correction for multiple testing ($P<1.1\times10^{-3}$). †Women only. ‡Men only. Abbreviations: ADHD – attention deficit hyperactivity disorder; Repro. – reproductive traits; Neuro. – neuro-psychiatric outcomes; SES – socio-economic status, behavioural and personality traits.

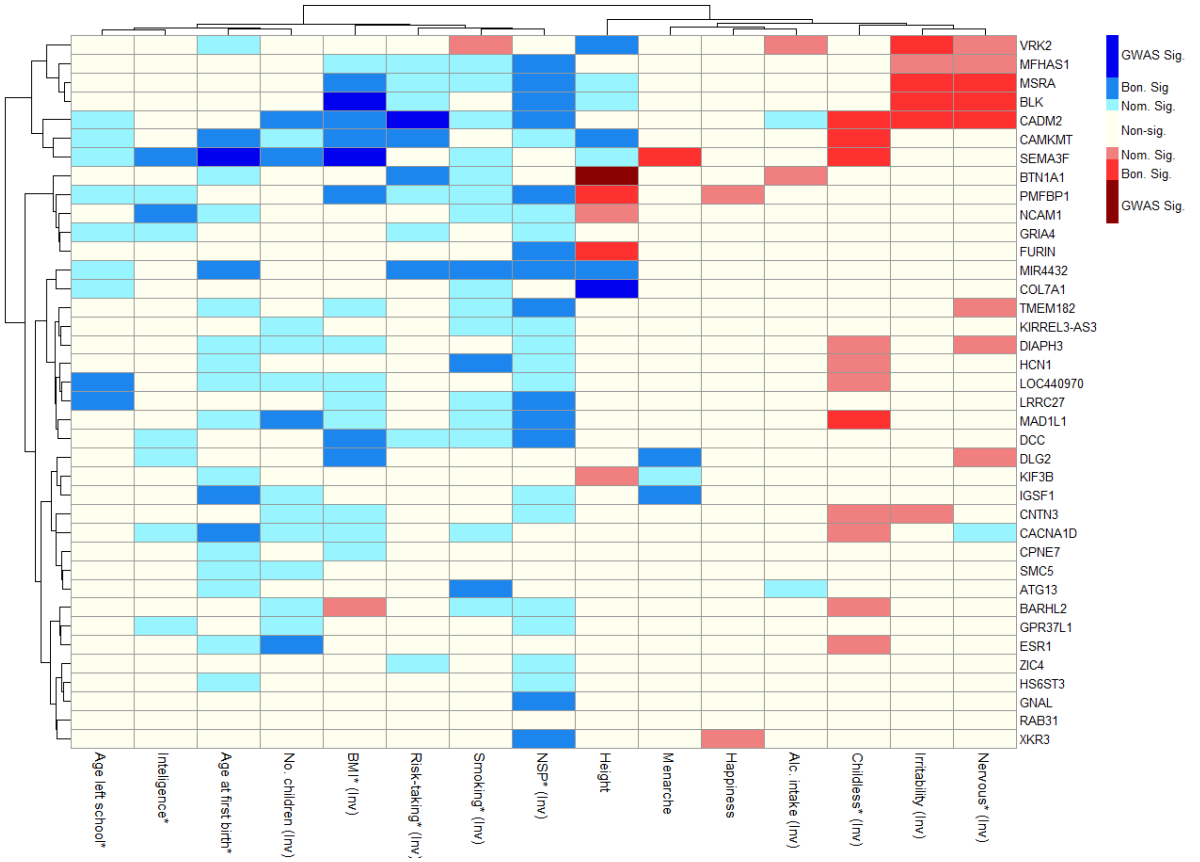

Figure 3 | Cluster plot displaying associations between the 38 'age at first sexual intercourse' lead SNPs and 15 other behavioural, reproductive and health-related traits in UK Biobank.

*Indicates the nine traits that are significantly enriched for age at first sexual intercourse (AFS) signals. 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 and these have the label "(Inv.)" (e.g. number of children, BMI, risk-taking propensity). Blue shading indicates a positive association with the indicated variable; red shading indicates a negative association. Hierarchical clustering of both phenotypes and SNPs was based on the Ward method using Euclidian distances. Abbreviations: BMI - body mass index, NSP – number of lifetime sexual partners.

Table 1 | The 38 genome-wide significant loci for Age at First Sexual Intercourse (AFS) in UK Biobank

Variant	Location	Nearest gene	Alleles[2]	Effect (s.e)	P-value	Sex Het
rs115552537	1p22.2	*BARHL2*	A/C/0.22	0.03 (0.005)	1.30E-08	6.66E-01
rs10800813	1q32.1	*GPR37L1*	C/T/0.63	0.02 (0.004)	3.80E-08	5.56E-01
rs4324362	2p21	*CAMKMT*	G/A/0.38	0.03 (0.004)	1.30E-11	1.14E-01
rs1344293	2p16.1	*BCL11A*	G/T/0.48	0.03 (0.004)	3.30E-15	1.87E-01
rs1040124	2q12.1	*TMEM182*	A/G/0.57	0.02 (0.004)	2.50E-08	1.96E-01
rs1264194	3p21.31	*COL7A1*	C/T/0.71	0.03 (0.004)	5.10E-09	1.57E-01
rs2188151	3p21.31	*SEMA3F*	G/T/0.57	0.03 (0.004)	3.00E-11	7.85E-01
rs34337122	3p21.1	*CACNA1D*	CG/C/0.85	0.03 (0.005)	8.80E-09	7.62E-01
rs6549665	3p12.3	*CNTN3*	G/C/0.18	0.03 (0.005)	4.90E-08	9.20E-01
rs12714592	3p12.1	*CADM2*	A/C/0.73	0.03 (0.004)	1.80E-10	4.34E-01
rs57401290	3p12.1	*CADM2*	GGTGTGT/G/0.55	0.03 (0.004)	8.00E-14	1.32E-03
rs530580221	3q24	*ZIC4*	T/TA/0.38	0.03 (0.004)	2.00E-09	9.22E-01
rs12522910	5p12	*HCN1*	C/T/0.17	0.04 (0.005)	2.70E-12	2.47E-01
rs726281	6q25.1	*ESR1*	A/G/0.72	0.03 (0.004)	3.20E-11	6.58E-01
rs13239969	7p22.3	*MAD1L1[e]*	C/G/0.6	0.02 (0.004)	1.40E-08	6.39E-01
rs4840367	8p23.1	*MFHAS1*	A/G/0.41	0.02 (0.004)	1.70E-10	4.35E-01
rs658385	8p23.1	*MSRA*	T/C/0.55	0.02 (0.004)	6.70E-09	2.92E-02
rs2248699	8p23.1	*BLK[e]*	G/A/0.5	0.02 (0.004)	3.60E-10	3.28E-01
rs538498277	9q21.12	*SMC5*	C/G/0.998	0.31 (0.051)	6.90E-10	4.86E-02
rs4443996	10q26.3	*LRRC27*	A/C/0.52	0.02 (0.004)	1.30E-08	8.40E-01
rs535814333	11p11.2	*ATG13*	TTG/T/0.7	0.02 (0.004)	3.90E-09	1.39E-02
rs140976226	11q22.3	*GRIA4*	GTT/G/0.43	0.02 (0.004)	1.00E-09	2.51E-02
rs66821824	11q23.2	*NCAM1*	ATTTT/A/0.78	0.03 (0.005)	1.60E-09	3.39E-01
rs538200730	11q24.2	*KIRREL3*	T/A/0.29	0.03 (0.004)	3.20E-10	8.22E-01
rs341521	13q21.2	*DIAPH3*	G/A/0.3	0.03 (0.004)	6.30E-15	4.62E-02
rs9516776	13q32.1	*HS6ST3*	A/T/0.34	0.02 (0.004)	3.30E-08	4.87E-01
rs4702	15q26.1	*FURIN[e]*	A/G/0.56	0.02 (0.004)	2.90E-10	2.79E-02
rs76513770	16q22.2	*PMFBP1*	C/T/0.13	0.04 (0.006)	4.70E-09	4.17E-01
rs369230	16q24.3	*CPNE7*	G/T/0.31	0.02 (0.004)	7.30E-10	1.02E-04
rs58749137	18p11.21	*GNAL*	A/G/0.73	0.02 (0.004)	1.30E-08	1.07E-01
rs4129322	18q21.2	*DCC*	A/G/0.08	0.04 (0.007)	2.90E-10	9.35E-01
rs6058613	20q11.21	*KIF3B*	C/G/0.16	0.03 (0.005)	3.50E-08	1.67E-02
rs5932884	Xq26.2	*IGSF1*	G/A/0.47	0.02 (0.005)	8.41E-09	-
Women only						
rs961522	2p16.1	*VRK2*	C/T/0.61	0.03 (0.005)	2.80E-08	4.63E-05
Men Only						
rs13194984	6p22.2	*BTN1A1*	G/T/0.86	0.05 (0.009)	3.90E-09	5.21E-05
rs201909661	11q14.1	*DLG2*	A/AG/0.02	0.15 (0.02)	7.00E-10	6.09E-07
rs138057093	18p11.22	*RAB31*	C/T/0.01	0.20 (0.03)	7.50E-10	7.25E-06
rs111837587	22q11.1	*XKR3*	A/G/0.01	0.18 (0.03)	7.20E-09	8.77E-05

1. [e] refers to a gene linked via altered expression (eQTL);2 effect allele / other allele / effect allele frequency

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

		UK Biobank (up to 59,357 men and 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* *ESR1* TA/T/0.36	Age at menarche	0.01 (0.009)	1.20E-01	73,397	-	-	-	0.01 (0.009)	1.20E-01	73,397
	AFS - Males	-0.02 (0.006)	3.70E-04	59,357	-	-	-	-0.02 (0.006)	3.70E-04	59,357
	AFS - Females	-0.03 (0.005)	1.40E-08	66,310	-	-	-	-0.03 (0.005)	1.40E-08	66,310
	AFS - Combined	-0.03 (0.004)	1.60E-10	125,667	-	-	-	-0.03 (0.004)	1.60E-10	125,667
	AFB - Males	-	-	-	-0.19 (0.03)	6.73E-09	117,626	-0.19 (0.03)	6.73E-09	117,626
	AFB - Females	-0.15 (0.029)	2.40E-07	50,954	-0.08 (0.02)	6.93E-04	144,471	-0.11 (0.02)	3.58E-09	195,425
	AFB - Combined	-0.15 (0.029)	2.40E-07	50,954	-0.12 (0.02)	6.96E-08	262,097	-0.13 (0.02)	1.22E-13	313,051
	Num Child - Males	0.01 (0.003)	8.20E-02	66,498	0.009 (0.002)	1.94E-06	117,626	0.008 (0.002)	7.25E-07	184,124
	Num Child - Females	0.01 (0.003)	3.70E-08	75,540	0.005 (0.002)	1.27E-02	147,498	0.008 (0.002)	2.15E-07	223,038
	Num Child - Combined	0.01 (0.002)	3.20E-07	142,038	0.007 (0.001)	1.11E-06	265,124	0.008 (0.001)	4.82E-12	407,162
	Childless - Males	0.98 (0.01)	1.30E-01	66,498	0.97 (0.02)	2.12E-01	97,200	0.98 (0.01)	5.56E-02	163,698
	Childless - Females	0.95 (0.01)	2.50E-06	75,540	0.93 (0.02)	4.00E-04	117,972	0.94 (0.01)	4.94E-09	193,512
	Childless - Combined	0.96 (0.008)	1.10E-05	142,038	0.95 (0.02)	9.04E-04	215,526	0.96 (0.007)	5.24E-08	357,564
rs2188151 *RBM6-SEMA3F* **T/G/0.43**	Age at menarche	0.03 (0.008)	2.20E-05	73,397	-	-	-	0.03 (0.008)	2.20E-05	73,397
	AFS - Males	-0.03 (0.006)	1.60E-05	59,357	-	-	-	-0.03 (0.006)	1.60E-05	59,357
	AFS - Females	-0.02 (0.005)	3.60E-07	66,310	-	-	-	-0.02 (0.005)	3.60E-07	66,310
	AFS - Combined	-0.03 (0.004)	3.00E-11	125,667	-	-	-	-0.03 (0.004)	3.00E-11	125,667
	AFB - Males	-	-	-	-0.14 (0.032)	1.00E-05	117,626	-0.14 (0.032)	1.00E-05	117,626
	AFB - Females	-0.16 (0.028)	7.20E-09	50,954	-0.105 (0.024)	7.88E-06	144,471	-0.129 (0.018)	9.52E-13	195,425
	AFB - Combined	-0.16 (0.028)	7.20E-09	50,954	-0.115 (0.022)	9.18E-08	262,097	-0.132 (0.017)	8.76E-15	313,051
	Num Child - Males	0.01 (0.003)	2.70E-02	66,498	0.003 (0.002)	7.47E-02	117,626	0.004 (0.002)	7.99E-03	184,124
	Num Child - Females	0.01 (0.003)	1.30E-03	75,540	0.003 (0.002)	7.45E-02	147,498	0.004 (0.001)	1.35E-03	223,038
	Num Child - Combined	0.01 (0.002)	1.60E-04	142,038	0.003 (0.002)	4.06E-02	265,124	0.005 (0.001)	9.05E-05	407,162
	Childless - Males	0.98 (0.01)	3.40E-02	66,498	-0.017 (0.022)	4.27E-01	97,200	-0.023 (0.01)	2.49E-02	163,698
	Childless - Females	0.97 (0.01)	2.50E-03	75,540	-0.028 (0.019)	1.52E-01	117,972	-0.032 (0.01)	8.53E-04	193,512
	Childless - Combined	0.97 (0.008)	2.70E-04	142,038	-0.023 (0.015)	1.42E-01	215,526	-0.028 (0.007)	9.38E-05	357,564

* In WGHS rs67229052 was not imputed so rs4305732 (r2=0.98) was used as a proxy. 1 effect allele / other allele / effect allele frequency

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.

FD, KO and JP designed the study. FD, HH, DC, LR, PS and JP performed the statistical analysis and all authors contributed to the interpretation of the findings. FD, KO and JP drafted the paper and all authors contributed to the final version.

Conflict of interests: The authors declare no conflicts of interests.