nature genetics

Genome-wide analysis identifies 12 loci influencing human reproductive behavior

The genetic architecture of human reproductive behavior—age at first birth (AFB) and number of children ever born (NEB)—has a strong relationship with fitness, human development, infertility and risk of neuropsychiatric disorders. However, very few genetic loci have been identified, and the underlying mechanisms of AFB and NEB are poorly understood. We report a large genome-wide association study of both sexes including 251,151 individuals for AFB and 343,072 individuals for NEB. We identified 12 independent loci that are significantly associated with AFB and/or NEB in a SNP-based genome-wide association study and 4 additional loci associated in a gene-based effort. These loci harbor genes that are likely to have a role, either directly or by affecting non-local gene expression, in human reproduction and infertility, thereby increasing understanding of these complex traits.

Human reproductive behavior—AFB and NEB—has been associated with human development[1,2], infertility[3,4] and neuropsychiatric disorders[5]. Reproductive tempo (AFB) and quantum (NEB) are cross-cutting topics in the medical, biological, evolutionary and social sciences and are central in national and international policies[6]. Advanced societies have experienced a rapid postponement of AFB, with the mean AFB of women now being 28–29 years in many countries[7]. This increase in AFB has been linked to lower fertility rates, unprecedented rates of childlessness (~20%) and infertility, which affects 10 to 15% of couples[8]. An estimated 48.5 million couples worldwide are infertile, with a large part of subfertility, particularly in men, remaining unexplained[9]. Although infertility has been related to advanced AFB, ovulation defects, failure of spermatogenesis, and single-gene or polygenic defects, the causal effects for these factors remain unsubstantiated[10].

Recently, genetic and clinical research has focused on proximal infertility phenotypes[3,4,10,11]. AFB and NEB represent accurate measures of complex reproductive outcomes, are frequently recorded and consistently measured, and are key parameters for demographic population forecasting[12]. There is evidence of a genetic component underlying reproduction, with heritability estimates of up to 50% for AFB and NEB (**Supplementary Fig. 1**)[6]. A recent study attributed 15% of the variance in AFB and 10% of the variance in NEB to common genetic variants[13]. There are also sex-specific differences in human reproduction, related to the timing of fertility, fecundability and sex–genotype interactions (**Supplementary Note**). Researchers have given less attention to traits such as NEB because of an erroneous and frequently repeated misinterpretation of Fisher's fundamental theorem of natural selection[14] that the additive genetic variance in fitness should be close to zero. This misreading of the theorem had a naively intuitive appeal: genes that reduce fitness should be passed on less frequently. Fisher, however, actually argues that fitness is moderately heritable in human populations (for a discussion, see the **Supplementary Note**). As no established genes are currently available for clinical testing of infertility[10], isolating genetic loci and their causative effects has the potential to provide new insights into the etiology of reproductive diseases and novel diagnostic and clinical technologies for infertility treatment.

RESULTS

We report a large meta-analysis of genome-wide association studies (GWAS) of 251,151 individuals for AFB and 343,072 individuals for NEB from a total of 62 cohorts of European ancestry. We identify 12 independent loci (10 of which are new and 2 of which were previously identified in a study on age at first sexual intercourse[11]) that are significantly associated with AFB and/or NEB in men, women or both sexes combined (**Table 1**). Follow-up analyses identified a number of genetic variants and genes that likely drive the GWAS associations. We also quantified the genetic overlap with biologically adjacent reproductive, developmental, neuropsychiatric and behavioral phenotypes. A detailed description of all materials and methods is available in the **Supplementary Note**.

Meta-analysis of GWAS

Associations of AFB (mean ± s.d., 26.8 ± 4.78 years) and/or NEB (mean ± s.d., 2.3 ± 1.43 children) with SNPs imputed from NCBI Build 37 HapMap phase 2 data were examined in 62 cohorts using multiple linear regression under an additive model, in men and women separately (**Supplementary Note**). Associations were adjusted for principal components, to reduce confounding by population stratification[15], as well as for the birth year of the respondent and its square and cube to control for nonlinear birth cohort effects (**Supplementary Tables 1** and **2**, and **Supplementary Note**). NEB was assessed only for those who had completed their reproductive period (age ≥45 years for women and ≥55 years for men), while AFB was only assessed for those who were parous. Quality control was conducted in two independent centers using QCGWAS[16] and EasyQC[17] (**Supplementary Note**). Results were subsequently submitted to meta-analysis for the 2.4 million SNPs that passed quality control filters (**Supplementary Note**) and are reported for men and women combined and separately.

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

Received 12 June; accepted 22 September; published online 31 October 2016; doi:10.1038/ng.3698

Table 1 GWAS meta-analysis results for independent loci that are genome-wide significantly ($P < 5.0 \times 10^{-8}$) associated with AFB or NEB in either the combined or sex-specific meta-analysis

SNP	Chr.	Position (bp)	Nearest gene	Annotation	Effect allele/ other allele	EAF	β	*P* value	*n* (pooled)	β (men)	*P* value (men)	β (women)	*P* value (women)
Age at first birth													
rs10908557	1	153,927,052	*CRTC2*	N, R, ctQ, ctM	C/G	0.695	0.091	**5.59×10^{-10}**	249,025	0.185	2.98×10^{-7}	0.076	5.38×10^{-6}
rs1160544	2	100,832,218	*LINC01104*	R, cQ, cM	A/C	0.395	−0.082	**2.90×10^{-9}**	250,330	−0.042	2.12×10^{-1}	−0.092	**5.00×10^{-9}**
rs2777888	3	49,898,000	*CAMKV*	N, R, ctQ, ctM	A/G	0.507	0.106	**4.58×10^{-15}**	250,941	0.155	2.40×10^{-6}	0.095	**6.07×10^{-10}**
rs6885307	5	45,094,503	*MRPS30, HCN1*	R, ctQ, cM	A/C	0.799	−0.107	**2.32×10^{-10}**	248,999	−0.131	2.07×10^{-3}	−0.104	**3.90×10^{-8}**
rs10056247	5	133,898,136	*JADE2*	cQ, cM	T/C	0.289	0.082	**4.37×10^{-8}**	249,429	0.050	1.68×10^{-1}	0.089	1.28×10^{-7}
rs2347867	6	152,229,850	*ESR1*	cM	A/G	0.649	0.091	**1.38×10^{-10}**	248,039	0.098	4.69×10^{-3}	0.097	**1.80×10^{-9}**
rs10953766	7	114,313,218	*FOXP2*	cM	A/G	0.429	0.087	**1.82×10^{-10}**	248,039	0.106	1.31×10^{-3}	0.089	**8.41×10^{-9}**
rs2721195	8	145,677,011	*CYHR1*	R, cQ, ctM	T/C	0.469	−0.073	6.25×10^{-7}	250,493	−0.014	6.85×10^{-1}	−0.099	**6.13×10^{-9}**
rs293566	20	31,097,877	*NOL4L*	cQ, cM	T/C	0.650	0.081	**1.41×10^{-8}**	245,995	0.110	1.47×10^{-3}	0.079	1.31×10^{-6}
rs242997	22	34,503,059	*LARGE1, ISX*		A/G	0.613	−0.084	**3.38×10^{-9}**	238,002	−0.139	8.51×10^{-5}	−0.076	1.82×10^{-6}
Number of children ever born													
rs10908474	1	153,753,725	*SLC27A3, GATAD2B*		A/C	0.384	0.020	**2.08×10^{-8}**	342,340	0.021	8.10×10^{-4}	0.020	7.89×10^{-6}
rs13161115	5	107,050,002	*EFNA5, FBXL17*	cM	C/G	0.234	−0.041	1.34×10^{-2}	341,737	−0.041	**1.37×10^{-8}**	0.005	3.29×10^{-1}
rs2415984	14	46,873,776	*LINC00871*	cM	A/G	0.470	−0.020	**2.34×10^{-8}**	315,167	−0.029	2.41×10^{-6}	−0.016	3.71×10^{-4}

The rows in bold correspond to the independent signals reaching $P < 5 \times 10^{-8}$ in the meta-analysis. Annotation shows for each of the 12 independent lead SNPs (excluding rs10908474 on chromosome 1) whether it is (i) in strong LD ($r^2 > 0.8$) with a nonsynonymous variant (N) or one or more variants prioritized by RegulomeDB (R) with evidence of having functional consequences (defined by a score <4); (ii) associated with an eQTL in *cis* and/or *trans* (ctQ); and (iii) associated with an meQTL in *cis* and/or *trans* (ctM). EAF, effect allele frequency of the pooled meta-analysis; β, effect size in the AFB and NEB analyses. All *P* values are from the sample-size-weighted fixed-effects meta-analysis.

We applied a single genomic control at the cohort level and performed meta-analysis of results using a sample-size-weighted fixed-effect method in METAL (**Supplementary Note**). The PLINK clumping function isolated 'lead SNPs'—those with the lowest *P* value for association that are independently associated—using an r^2 threshold of 0.1 and a distance threshold of 500 kb. For AFB, we identified ten loci associated at genome-wide significance ($P < 5 \times 10^{-8}$ for combined results and $P < 1.67 \times 10^{-8}$ for sex-specific results adjusted for multiple testing), of which 9 were significantly associated in both sexes combined and 1 was associated in women only (n = 154,839) (**Fig. 1a** and **Table 1**). Three loci were significantly associated with NEB: two in both sexes combined and one in men only (n = 103,736) (**Fig. 1b**, **Table 1** and **Supplementary Note**). One locus on chromosome 1 reached significance for association with both AFB and NEB with $r^2 = 0.57$ between the two lead SNPs, suggesting a shared genetic basis for the two traits (**Table 2**). A statistical test of sex-specific effects confirmed that differences are mainly due to variation in sample size and not variation in effect size (**Supplementary Note**).

As for other complex traits[18], the quantile–quantile plots of the meta-analyses exhibited strong inflation of low *P* values (**Fig. 2**), suggesting that, although cohorts controlled for the top principal components and cohort-level genomic control was applied (**Supplementary Note**), residual population stratification may remain. However, the LD Score intercept method[19] as well as a series of individual and within-family regression analyses using polygenic scores as predictors[20,21] (**Supplementary Note**) indicated that the observed inflation was almost entirely attributable to a true polygenic signal, rather than population stratification.

Gene-based GWAS

To increase the power to find statistically significant associations and causal genes, we additionally performed a gene-based GWAS using VEGAS[22,23]. The results confirmed top hits from the single-SNP analyses. For AFB, seven loci from the SNP-based GWAS were also represented in the gene-based analysis (**Supplementary Table 3**), and three additional loci emerged, represented by *SLF2* (chromosome 10), *ENO4* (chromosome 10) and *TRAF3-AMN* (chromosome 14). For NEB, one locus from the SNP-based GWAS was represented in the gene-based analysis—*GATAD2B* (chromosome 1)—and one new locus on chromosome 17 was identified (**Supplementary Table 4**).

Causal variants

To identify functional and potentially causal variants, both coding and regulatory, within loci identified in the SNP-based GWAS (**Table 1**), we first performed an *in silico* sequencing annotation analysis using the post-GWAS pipeline reported by Vaez *et al.*[24]. This showed that rs10908557 on chromosome 1 is in high linkage disequilibrium (LD) with nonsynonymous SNPs in *CRTC2* (rs11264680; $r^2 = 0.98$) and *CREB3L4* (rs11264743; $r^2 = 0.94$) (**Supplementary Table 5**). Interestingly, rs11264743 is considered 'deleterious' and 'probably damaging' by SIFT and PolyPhen, respectively (Ensembl release 83). In addition, rs2777888 on chromosome 3 is in high LD with two nonsynonymous SNPs in *MST1R* (rs2230590, $r^2 = 0.95$ and rs1062633, $r^2 = 0.95$) (**Table 1** and **Supplementary Table 5**).

We subsequently performed a comprehensive analysis using results from the Encyclopedia of DNA Elements (ENCODE)[25] and Roadmap Epigenomics[26] projects, as integrated in RegulomeDB[27], to identify variants that likely influence downstream gene expression via regulatory pathways. Among all SNPs that reached $P < 5 \times 10^{-8}$ in the meta-analyses (n = 322), 50 SNPs in five loci showed the most evidence of having functional consequences (**Table 1** and **Supplementary Table 6**).

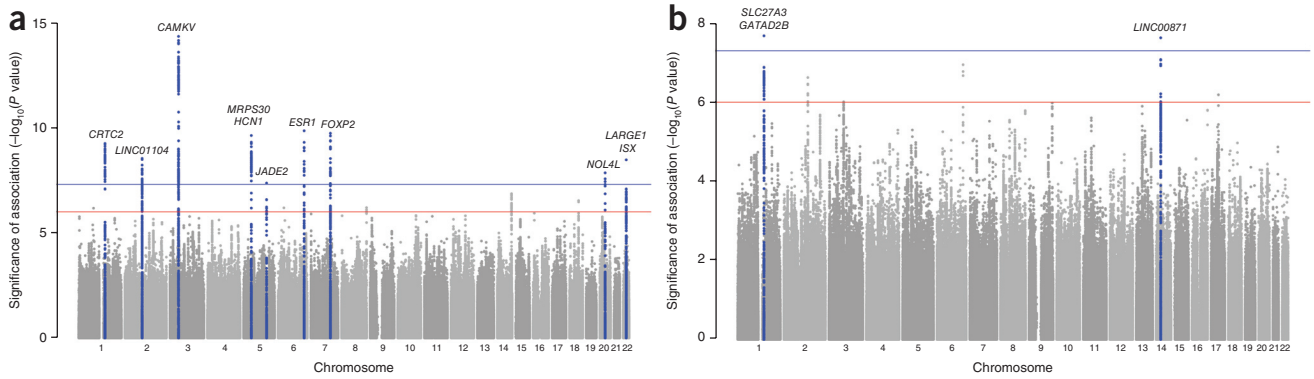

Figure 1 Manhattan plots of SNPs for age at first birth and number of children ever born in single-genomic-control meta-analysis. (**a**,**b**) SNPs are plotted on the *x* axis according to their position on each chromosome against association with AFB (**a**) and NEB (**b**). The solid blue line indicates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$), and the red line represents the threshold for suggestive hits ($P < 5 \times 10^{-6}$). Blue points represent SNPs in a 100- kb region centered on genome-wide significant hits. Loci are annotated with the names of the genes closest to the significant SNPs.

Two sets of SNPs on chromosome 1 (18 SNPs) and chromosome 3 (25 SNPs) stand out in particular. The most promising SNP in the chromosome 1 locus (rs6680140) is located in a site of acetylation of histone H3 at lysine 27 (H3K27ac), often found near active regulatory elements, and lies in a DNase I hypersensitivity cluster where eight proteins are anticipated to bind. One of these proteins is cAMP responsive element binding (CREB)-binding protein, encoded by *CREBBP*. In the chromosome 3 locus, rs2526397 is located in a transcription factor binding site and is an expression quantitative trait locus (eQTL) for *HYAL3* in monocytes, while rs2247510 and rs1800688 are located in H3K27ac sites and DNase I hypersensitivity clusters where a large number of transcription factors are expected to bind (**Supplementary Table 6**). An analysis using Haploplotter showed that rs2247510 and rs7628058 in the chromosome 3 locus are among the 5% of signals showing the most evidence of positive selection in the population. The same applies to the lead SNP of the chromosome 14 locus for NEB (rs2415984).

Causal genes

Information on the function and anticipated relevance of genes in the 12 loci identified in the SNP-based GWAS that are most likely to be causal on the basis of all evidence discussed below is provided in **Table 2**.

Cis- and *trans*-eQTL and meQTL analyses

Identifying alterations in gene methylation status and/or expression levels in relation to GWAS-identified variants may help prioritize causal genes. We examined associations with gene expression and methylation status for the 12 independent lead SNPs in whole-blood BIOS eQTL ($n = 2{,}116$) and methylation quantitative trait locus (meQTL; $n = 3{,}841$) databases in *cis* and *trans*[28,29]. Seven SNPs were associated in *cis* with the expression of 54 unique genes (**Table 1** and **Supplementary Table 7**). Five of these seven SNPs were in high LD ($r^2 > 0.8$) with the strongest eQTL for at least one of the genes within the corresponding locus, indicating that the SNP associated with AFB or NEB and the SNP most significantly associated with expression tag the same functional site: rs10908557 (associated with the expression of *CRTC2* and *SLC39A1*), rs1160544 (*AFF3*), rs2777888 (*RBM6*, *RNF123* and *RBM5*), rs2721195 (*CYHR1*, *GPT*, *RECQL4* and *PPP1R16A*) and rs293566 (*NOL4L*). Three SNPs were associated with the expression of a total of eight genes in *trans* (**Table 1** and **Supplementary Table 8**). Of these SNPs, only rs2777888 was in high LD ($r^2 > 0.8$) with the strongest eQTL for three of its five associated genes: *LRFN1*, *LAMP2* and *FGD3*.

The meQTL analysis showed that 11 of the 12 independent lead SNPs were associated with DNA methylation of a total of 131 unique genes in *cis* (**Table 1** and **Supplementary Table 9**). Seven of the 11 SNPs were in high LD ($r^2 > 0.8$) with the strongest meQTL for one of the corresponding methylation sites: rs10908557 (associated with methylation of *CRTC2*, *SLC39A1*, *CREB3L4*, *DENND4B* and *RAB13*), rs1160544 (*AFF3*), rs2777888 (*CAMKV*), rs6885307 (*C5orf34*), rs10056247 (*JADE2*), rs2721195 (*CYHR1*) and rs13161115 (*EFNA5*). Three of the SNPs were associated with the same genes for both methylation and gene expression in *cis*: rs10908557 (*CRTC2*), rs1160544 (*AFF3*) and rs2721195 (*CYHR1*) (**Supplementary Tables 7** and **9**). Three SNPs were associated with methylation of a total of ten genes in *trans* (**Table 1** and **Supplementary Table 10**). Of these SNPs, only rs2777888 was in high LD ($r^2 > 0.8$) with the strongest meQTL for a corresponding methylation site (*ASAP3*). Of note, rs2777888 was also a *trans*-eQTL.

Gene prioritization

We used four publicly available bioinformatics tools to systematically identify genes that are more likely than neighboring genes to cause the associations identified by our GWAS. Of all genes that reached $P < 0.05$ in analyses using Endeavor[30], MetaRanker[31] and ToppGene[32], eight genes were prioritized for both AFB and NEB: *TPM3*, *GRM7*, *TKT*, *MAGI2*, *PTPRD*, *PTPRM*, *RORA* and *WT1*. DEPICT—a fourth comprehensive and unbiased recently described gene prioritization tool[33]—identified three genes in GWAS significant loci as likely being causal for AFB (*MON1A*, *RBM6* and *U73166.2*) (**Supplementary Tables 11** and **12**).

Gene-based results from RegulomeDB

An analysis using RegulomeDB identified 50 SNPs in five loci that most likely have regulatory consequences (**Supplementary Table 6**). Eighteen and 25 of these SNPs are within the chromosome 1 and chromosome 3 loci, respectively. Among the genes that, at a protein level, bind at the site of one or more of the 18 variants in the locus on chromosome 1 are *CREBBP*, *HNF4A*, *CDX2* and *ERG*. These genes may act upstream in the causal pathway and influence the expression of causal genes at this locus. Of the 25 SNPs on chromosome 3, 10 were eQTLs for *RBM6* in monocytes and 7 were eQTLs for *HYAL3* in monocytes. Among the genes that, at a protein level, bind at rs2247510 and rs1800688 in the chromosome 3 locus are *ARID3A*, *REST* and *TFAP2C*, as well as *HNF4A* and *CDX2*, which also bind at the chromosome 1 locus.

Table 2 Function and potential relevance for genes in GWAS-identified loci that are most likely causal on the basis of all available evidence

Lead SNP	Gene	Chr.	Evidence	Gene function and potential role in reproduction and (in)fertility	Ref.
rs10908557	*CRTC2*	1	G, V, ctQ, ctM, Q lymph. (R)	Functions as a Ca^{2+}- and cAMP-sensitive coincidence sensor; promotes CREB target gene expression; signal mediator in FSH and TGF-β1 steroidogenesis in ovarian granulosa cells	42
rs10908557	*CREB3L4*	1	N, V, cQ, cM	Has a role in protein maturation; involved in spermatid differentiation and male germ cell development; expressed in prostate, oocytes, fallopian tube and mammary gland	44,45
rs10908557	*GATAD2B*	1	V, Q monoc. (R)	Transcriptional repressor and a component of nucleosome remodeling complex Mi2/NuRD; increased expression in endometriosis; linked to a common gynecological disorder that causes pelvic pain and infertility	58,59
rs10908557	*SLC39A1*	1	V, cQ, cM	Zinc uptake transporter; major zinc regulator in prostate cells; involved in the regulation of zinc homeostasis by cumulus cells in the oocyte	60,61
rs10908557	*DENND4B*	1	cM	Paralog of *DENN1A*, which has been implicated in polycystic ovary syndrome; expressed at the protein level in the cervix	46,62
rs1160544	*AFF3*	2	cQ, cM	Lymphoid nuclear transcriptional activator implicated in tumorigenesis; same family as *AFF3* and *AFF4* (*FMR2* family member 4); transcriptional regulator in testicular somatic cells; essential for male germ cell differentiation and survival in mice	63,64
rs1160544	*LINC01104*	2	G, V	Unknown	
rs2777888	*HYAL3*	3	cM, Q monoc. (R)	Hyaluronidases, including HYAL3, are involved in degradation of hyaluronan, a major glycosaminoglycan of the extracellular matrix; acquired during sperm maturation in the epididymis and involved in sperm function and the acrosome reaction; required for *in vitro* cumulus penetration in mice, although its absence is not associated with infertility (perhaps compensated for by other hyaluronidases)	65
rs2777888	*RBM6*	3	V, cQ, cM, DEPICT, Q monoc. (R)	Involved in RNA splicing	66
rs2777888	*RNF123*	3	V, cQ, cM, Q liver (R)	Has a role in cellular transitioning from quiescence to a proliferative state through its E3 ubiquitin ligase activity toward cyclin-dependent kinase inhibitor 1B, which controls cell cycle progression in G1 phase	66–68
rs2777888	*RBM5*	3	V, cQ	Involved in cell cycle regulation; regulator of pre-mRNA splicing; involved in spermatogenesis and fertility in mice	47
rs2777888	*MON1A*	3	V, cM, DEPICT	Involved in the movement and trafficking of proteins (for example, ferroportin) through the secretory apparatus	69
rs2777888	*U73166.2*	3	DEPICT	Unknown	
rs2777888	*MST1R*	3	N, V, cM, MetaRanker, ToppGene and Endeavor	Cell surface receptor for MSP with tyrosine kinase activity, expressed on ciliated epithelia of the mucociliary transport apparatus of the lung; involved in host defense, expressed in sperm; may act in a regulatory system of ciliary motility, together with MSP, which sweeps eggs along the oviduct; expressed in mucous membrane and mammary gland	70
rs10056247	*JADE2*	5	G, V, cM	Involved in histone acetylation	
rs13161115	*EFNA5*	5	cM	Involved in development and differentiation of the nervous system and folliculogenesis regulation; required for normal fertility in female mice; expressed in embryonic stem cells and embryoid bodies	50
rs6885307	*HCN1*	5	G, cM	Hyperpolarization-activated cation channel that contributes to the native pacemaker current in, for example, neurons; HCN1 channels are present in kisspeptin (Kiss1) neurons in the rostral periventricular area of the third ventricle (RP3V), which provide an excitatory drive to gonadotropin-releasing hormone (GnRH)-expressing neurons that control fertility	71
rs2347867	*ESR1*	6	G, cM, binds at rs4851269 on chr. 2 (R)	Transcription factor involved in estrogen-responsive gene expression; essential for sexual development and reproductive function in women; genetic variants in *ESR1* may influence susceptibility to endometriosis or female fertility in patients with endometriosis; involved in male fertility by transferring estrogen effect; expressed in myometrium, endometrium, oocytes, uterus and fallopian tube	51,52, 72–74
rs10953766	*FOXP2*	7	G, cM, binds at rs6997 on chr. 3 (R)	Transcription factor expressed in fetal and adult brain that is involved in speech and language development; involved in regulation of neuronal development in the embryonic forebrain; expressed in mucous membrane and myometrium	75
rs2721195	*CYHR1*	8	cQ, cM	Histidine-cysteine-rich protein involved in spermatogenesis	53
rs2721195	*GPT*	8	V, cQ, cM, Q monoc. (R)	Involved in intermediary metabolism of glucose and amino acids; may have a role in spermatogenesis via testicular glucose metabolism, which is pivotal for the normal occurrence of spermatogenesis; levels in the normal range are positively associated with metabolic and endocrine abnormalities in women of reproductive age and negatively associated with FSH levels, independent of obesity	76,77
rs2721195	*RECQL4*	8	V, cQ, cM	Processing of aberrant DNA structures that arise during DNA replication and repair; predominantly expressed in testis	78
rs2721195	*PPP1R16A*	8	V, cQ, cM, Q monoc. (R)	Regulator of protein phosphatase PP1β; present in the sperm tail where it interacts with proteins that are important in sperm structure and motility; expressed in mammary gland and fallopian tube	79
rs293566	*NOL4L*	20	cQ, cM	Component of cytoplasm and nucleoplasm; expressed in umbilical vein	

Evidence categories include the nearest gene (G), nonsynonymous variants (N), gene-based GWAS performed in VEGAS (V), eQTLs in *cis* and/or *trans* (ctQ), meQTLs in *cis* and/or *trans* (ctM), eQTLs (Q) in lymphoblasts (lymph), monocytes (monoc) or liver based on RegulomeDB (R), gene prioritization using either DEPICT or MetaRanker, ToppGene and Endeavor, and protein binding at SNPs based on RegulomeDB (R). Chr., human chromosome on which the gene is located; FSH, follicle-stimulating hormone; CREB, cAMP response element–binding protein; TGF-β1, transforming growth factor β1; MSP, macrophage-stimulating protein. SNIPPER was used for the literature search, with the search terms "fertility," "sperm," "ovum" and "reproduction."
Gene Network was used to find the tissue or organ with high expression for a given gene (AUC > 0.8).

Five genes encode proteins that bind at the site of both SNPs on chromosome 2 that reach $P < 5 \times 10^{-8}$ in the meta-analysis of GWAS. One of these is *REST*; another one, *ESR1*, is the most likely causal gene in the chromosome 6 locus.

Functional network and enrichment analyses

Functional network analysis using five prioritized candidate gene sets as input (**Supplementary Note**) showed no significantly enriched biological function. No biological function was significantly enriched after correction for multiple testing using the Benjamini–Hochberg procedure. Similarly, no reconstituted gene sets and cell or tissue types were significantly enriched in the GWAS meta-analysis results based on results from the DEPICT analysis (**Supplementary Tables 13–20**). The lack of significantly enriched genes, tissue sets and biological functions reflects the need for a larger sample size but also the distal nature of the phenotypes, which are influenced by a mixture of biological, psychological and socioenvironmental factors.

Polygenic prediction

To assess the predictive power of our results, we produced polygenic scores for AFB and NEB with sets of SNPs whose nominal *P* values ranged from $P < 5 \times 10^{-8}$ (using only genome-wide significant SNPs) to 1 (using all SNPs that passed quality control) using PRSice[34] (**Supplementary Note**). We then performed a series of four different out-of-sample predictions in four independent cohorts: HRS, LifeLines, STR and TwinsUK. Across the four cohorts, the mean predictive power of a polygenic score constructed from all measured SNPs is 0.9% for AFB and 0.2% for NEB (**Supplementary Fig. 2**). Despite the low predictive power of the polygenic scores, the results showed that an increase of 1 s.d. in the NEB polygenic score is associated with a 9% (95% confidence interval (CI) = 5–13%) decrease in the probability of women remaining childless, with no significant association in men (**Supplementary Table 21**). When we controlled for right-censored data using a survival model for AFB, we found that an increase of 1 s.d. in the AFB polygenic score was associated with an 8% (95% CI = 7–10%) reduction in the hazard ratio of reproduction in women and a 3% (95% CI = 1–5%) reduction in men (**Supplementary Table 22**). As an additional test, we examined whether the aforementioned polygenic scores for AFB and NEB could predict related fertility traits such as age at menopause and age at menarche (**Supplementary Table 23**). Our estimates indicated that an increase of 1 s.d. in the AFB polygenic score was associated with a 3% decrease in the probability of natural menopause at any age (95% CI = 1–5%) and a 20-d increase in age at menarche (95% CI = 0.4–40 d).

Genetic association with related traits and diseases

Several loci for which the associations with AFB and NEB reached genome-wide significance are associated with behavioral and reproductive phenotypes. The lead SNPs in the chromosome 2 and chromosome 3 loci have been associated with educational attainment[35] and the locus on chromosome 5 has been associated with age at menarche[2], while the locus on chromosome 6 has recently been associated with age at first sexual intercourse[11] (**Supplementary Table 24**). Some of the 38 loci for age at first sexual intercourse that were recently identified[11] in 125,667 UK Biobank participants were also associated with AFB (in or near *RBM6–SEMA3F* and *ESR1*) and NEB (in or near *CADM2* and *ESR1*). The lead SNPs for *RBM6–SEMA3F* (rs2188151) and *ESR1* (rs67229052) are in LD with our lead SNPs for AFB on chromosome 3 ($r^2 = 0.44$) and chromosome 6 ($r^2 = 0.94$), respectively. An *in silico* pleiotropy analysis showed that our lead SNP in the chromosome 3 locus (rs2777888) is in LD ($r^2 = 0.59$) with rs6762477, which has been

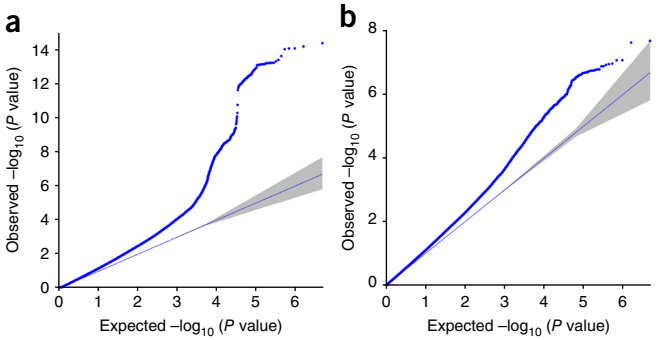

Figure 2 Quantile–quantile plots. (**a**,**b**) SNPs for AFB (**a**) and NEB (**b**) in single-genomic-control meta-analysis. The gray-shaded areas in the quantile–quantile plots represent the 95% confidence bands around the *P* values under the null hypothesis.

associated with age at menarche[2], while other SNPs in the same locus have been associated with HDL cholesterol[36] (rs2013208, $r^2 = 0.81$) and body mass index (BMI)[37] (rs7613875, $r^2 = 0.81$) (**Supplementary Table 5**). We next performed an exploratory analysis using the proxy phenotype method[38] to examine whether the SNPs strongly associated with AFB in women are empirically plausible candidate SNPs for related traits such as age at menarche and age at menopause (**Supplementary Note**). After controlling for multiple testing, we identified three AFB-associated SNPs near rs2777888 on chromosome 3 (rs9589, rs6803222 and rs9858889) that are also associated with age at menarche ($P < 4.10 \times 10^{-4}$). None of the AFB- or NEB-associated SNPs are associated with age at menopause.

We performed a bivariate LD score regression analysis[39] to estimate the pairwise genetic correlation with 27 publicly available GWAS results for traits associated with human reproductive behavior (**Supplementary Note**). AFB showed significant and positive genetic correlation with the human (reproductive) developmental traits of age at menarche, voice breaking, age at menopause, birth weight and age at first sexual intercourse, as well as with years of education. Conversely, having more AFB-increasing alleles was associated with a lower genetic risk of smoking (ever, number of cigarettes and later onset) and with lower insulin-resistance-related phenotypes, that is, BMI, waist–hip ratio adjusted for BMI, fasting insulin, triglyceride levels and risk of diabetes (**Fig. 3** and **Supplementary Table 25**). All genetic correlations remained significant after Bonferroni correction for multiple testing ($P < 2.6 \times 10^{-3}$). Years of education ($P = 6.6 \times 10^{-14}$) and age at first sexual intercourse ($P = 1.14 \times 10^{-15}$) are the only traits that showed significant and negative genetic correlation with NEB. We also observed significant genetic correlations of $r_g = 0.86$ (standard error (SE) = 0.052) for AFB and $r_g = 0.97$ (SE = 0.095) for NEB between men and women, implying that most genetic effects on reproductive behavior resulting from common SNPs are shared by both sexes.

DISCUSSION

This GWAS is a large-scale genetic epidemiological discovery effort for human reproduction, with implications for population fitness and physiological mechanisms linking hypothesized genes and observed phenotypes. Related studies previously focused on reproductive life span[40,41], age at first sexual intercourse[11] and more proximal infertility phenotypes[2–4], largely overlooking AFB and NEB. The rapid postponement of AFB and increased infertility and involuntary childlessness in many societies[7] make it important to uncover the genetic and biological architecture of reproduction. We identify ten

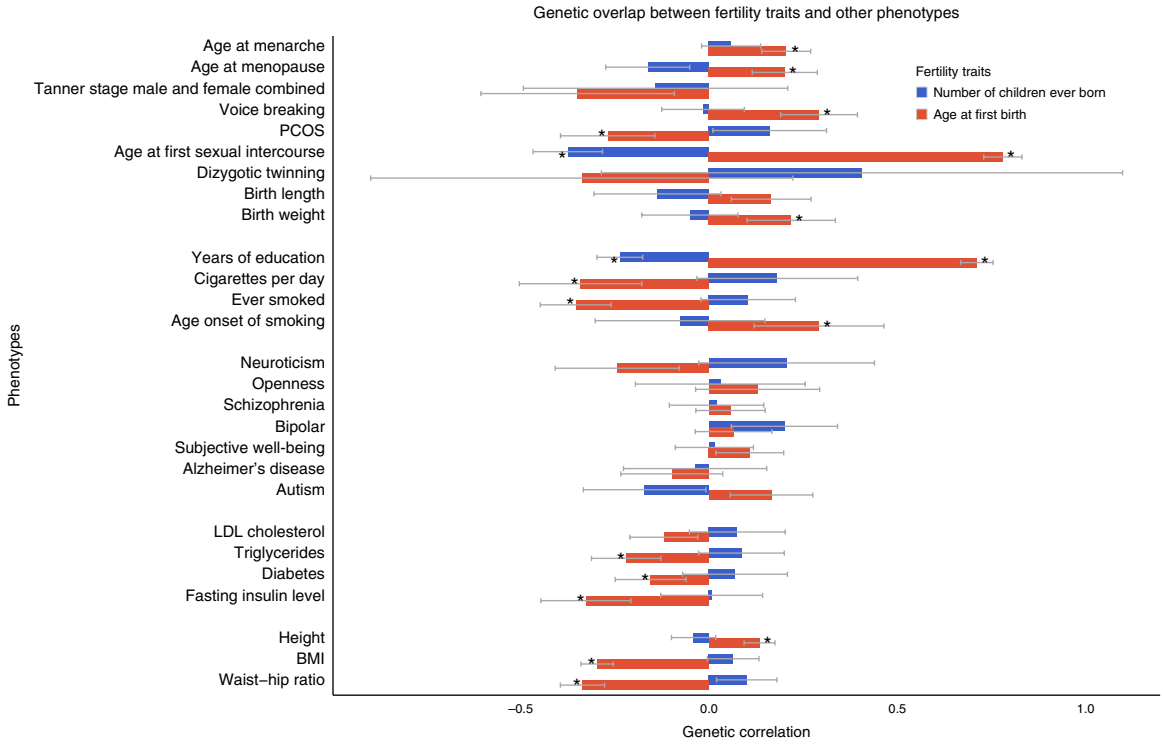

Figure 3 Genetic overlap between AFB or NEB and other related traits. Results from LD Score regressions show estimates of genetic correlation with developmental, reproductive, behavioral, neuropsychiatric and anthropometric phenotypes for which GWAS summary statistics were available in the public domain. The lengths of the bars correspond to estimates of genetic correlation. Gray error bars represent 95% confidence intervals. An asterisk indicates that the estimate of genetic correlation is statistically significant after controlling for multiple testing ($P < 0.05/27 = 1.85 \times 10^{-3}$).

new and confirm two recently identified genetic loci that are robustly associated with AFB and NEB, as well as variants and genes within these loci that potentially drive these associations. Four additional loci were identified in a gene-based GWAS.

Two loci that show interesting results in follow-up analyses are located on chromosomes 1 and 3. The lead SNPs of the chromosome 1 locus for AFB and NEB are in LD with likely functional nonsynonymous SNPs in genes encoding (i) CREB-regulated transcription co-activator 2 (*CRTC2*), which at the protein level acts as a critical signal mediator in follicle-stimulating hormone (FSH) and transforming growth factor (TGF)-β1–stimulated steroidogenesis in ovarian granulosa cells[42]; and (ii) CREB protein 3 like 4 (*CREB3L4*)[43], which in humans is highly expressed in the prostate, ovaries, uterus, placenta and testis and has a role in spermatid differentiation[44] and male germ cell development[45]. The lead SNP and/or functional variants in LD with it are also associated with the methylation status of these two genes and expression of *CRTC2* in whole blood and lymphocytes. Three promising functional variants in the chromosome 1 locus reside in binding sites of the transcriptional co-activator CREBBP. In addition to a direct effect of the above-mentioned nonsynonymous SNPs on protein function, the associations of AFB and NEB with variants in the locus on chromosome 1 may thus be mediated by alterations in cAMP responsive element binding in men and women. The locus on chromosome 1 also harbors *DENND4B*, a paralog of *DENND1A*, implicated in polycistic ovary syndrome (PCOS)[46]. Whereas *DENND1A* is expressed at the protein level in the ovary and testis, DENND4B is expressed in the cervix and its function and role are less well understood.

The lead SNP of the locus on chromosome 3 (rs2777888) is associated with methylation and expression of several genes, in *cis* and *trans*, that are known to have a role in cell cycle progression and/or sperm function. First, rs2777888 is associated with the expression of *RNF123* (or *KPC1*) in *cis*, which has a role in cellular transition from quiescence to a proliferative state. Second, rs2777888 or functional variants in LD with it may influence the cell cycle by altering the expression of *RBM5* and *RBM6* (RNA-binding motif proteins 5 and 6). The former has a role in cell cycle arrest and apoptosis induction and regulates haploid male germ cell pre-mRNA splicing and fertility in mice. *Rmb5*-mutant mice exhibit spermatid differentiation arrest, germ cell sloughing and apoptosis, leading to lack of sperm in ejaculation[47]. Third, rs2777888 affects expression of *LAMP2* in *trans*, which is located on the X chromosome and encodes a lysosomal membrane protein involved in the acrosome reaction, that is, the enzymatic drill allowing sperm to penetrate and fertilize ova[48]. *LAMP2* is expressed at the protein level in male and female reproductive organs with an effect size of rs2777888 for *LAMP2* mRNA expression that is almost twice as large in women than it is in men (**Supplementary Fig. 3**). This suggests an important role for the lysosome in both sperm and ova. Finally, functional variants in the chromosome 3 locus are associated with the mRNA expression of *HYAL3* (hyaluronoglucosaminidase 3) in monocytes. The encoded protein degrades hyaluronan, which also has an important role in sperm function and the acrosome reaction[47,49].

Functional follow-up experiments could disentangle the potential interplay between many candidate genes in the loci on chromosomes 1 and 3 in reproductive behavior and, given our *in silico* results, infertility. This can be extended to candidate genes in the remaining loci identified in the present study, some of which are relevant for fertility: mice lacking *Efna5* (chromosome 5 NEB locus) are subfertile[50], *ESR1* on chromosome 6 encodes an estrogen receptor[51,52] and *CYHR1* on chromosome 8 is involved in spermatogenesis[53]. Such experiments would help in understanding whether binding of estrogen receptor 1,

encoded by *ESR1* in the locus on chromosome 6, at the site of functional variants in the locus on chromosome 2 drives or mediates the association with AFB in the chromosome 2 locus, as well as to identify and characterize causal genes. Recent developments in high-throughput, multiplex mutagenesis using CRISPR/Cas9 allow such experiments to be performed using *in vivo* model systems[54].

AFB and NEB are not only driven by biological processes, but are also subject to individual choice and personal characteristics such as personality traits, as well as by the historical, cultural, economic and social environment (for example, effective contraception and childcare availability). Demographic research has shown a strong positive association between AFB and educational attainment[12]. We show that the associations between fecundity, reproductive behavior and educational attainment are partly driven by genetic factors and identified loci that are associated with AFB as well as with, for example, age at first sexual intercourse[11] and educational attainment[35].

Our findings could lead to insights into how postponing reproduction may be more detrimental for some, on the basis of their genetic make-up, than others, fuel experiments to determine 'how late can you wait' (ref. 55) and stimulate reproductive awareness. Causal genes in the loci we identified could potentially serve as novel drug targets, to prevent or delay age-related declines in fertility and sperm quality or to increase assisted reproductive technology efficiency, but further characterization is needed. Our study examines the genetics of reproductive behavior in both men and women, and, to our knowledge, it is the first that is adequately powered to identify loci in both women and men. We also provide support for Fisher's theorem that fitness is moderately heritable in human populations. Although the effect sizes of the identified common variants are small, there are examples of GWAS-identified loci of small effect that end up leading to important biological insights[56,57]. Many of our findings suggest a role for sperm quality, which is one lead for researchers to pursue. Because current sperm tests remain rudimentary, our findings could serve as a basis for 'good quality' sperm markers. We identified both coding and regulatory variants that are potentially causal, as well as a set of genes that could underlie the associations we identified. Follow-up experiments in animal models are required to confirm and characterize the causal genes in these loci.

URLs. Analysis plan predeposited at the Open Science Framework website, https://osf.io/53tea/; Gene Network, http://129.125.135.180:8080/GeneNetwork/; ReproGen, http://www.reprogen.org/data_download.html; Sociogenome, http://www.sociogenome.com/; Social Science Genetic Association Consortium, http://thessgac.org/.

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

For full acknowledgments, see the **Supplementary Note**. Funding to lead this consortium was provided by grants awarded to M.C.M.: ERC Consolidator Grant SOCIOGENOME (615603), a Dutch Science Foundation (NWO) grant (VIDI grant 452-10-012), a UK ESRC/NCRM SOCGEN grant (ES/N011856/1), the European Union's FP7 Families And Societies project (320116), and the Wellcome Trust ISSF and John Fell Fund. M.d.H. was supported by grants from the Swedish Research Council (2015-03657) and the Swedish Heart-Lung Foundation (20140543). Research was carried out in collaboration with the Social Science Genetic Association Consortium (SSGAC), with funding from the US National Science Foundation (EAGER: 'Workshop for the Formation of a Social Science Genetic Association Consortium'), a Supplementary grant from the National Institutes of Health Office of Behavioral and Social Science Research, the Ragnar Söderberg Foundation (E9/11), the Swedish Research Council (421-2013-1061), the Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810 and T32-AG000186-23 to NBER and R01-AG042568-02 to the University of Southern California. X.S. was supported by a grant from the Swedish Research Council (537-2014-371). We thank X. Ding for research assistance, N. Pirastu, K. Coward and L. Layman for valuable comments, and the University of Oxford Advanced Research Computing (ARC) facility (http://dx.doi.org/10.5281/zenodo.22558). This research has been conducted using the UK Biobank Resource.

AUTHOR CONTRIBUTIONS

Senior investigators who led writing, analysis and study design: M.C.M., H. Snieder and M.d.H. Senior investigators who participated in study design: P.D.K., D.J.B. and D.C. Junior investigator who contributed to the study design and management: N. Barban. Population stratification: N. Barban and F.C.T. Genetic correlations and polygenic score prediction: N. Barban. Meta-analysis and quality control: N. Barban, R.d.V., J.J.M. and I.M.N. Biological annotation: R.J., M.d.H. and A.V. Sex-specific genetic effects: N. Barban and F.C.T. Bivariate and conditional analysis of the two fertility traits: X.S., J.F.W. and D.I.C. Gene-based analysis V.T. and S.W.v.d.L. Authors not listed contributed to recruitment, genotyping or data processing for the meta-analysis (**Supplementary Table 43**).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Elks, C.E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
2. Perry, J.R.B. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
3. Rahmioglu, N. *et al.* Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. *Hum. Reprod. Update* **20**, 702–716 (2014).
4. Day, F.R. *et al.* Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat. Commun.* **6**, 8464 (2015).
5. Mehta, D. *et al.* Evidence for genetic overlap between schizophrenia and age at first birth in women. *JAMA Psychiatry* **73**, 497–505 (2016).
6. Mills, M.C. & Tropf, F.C. The biodemography of fertility: a review and future research frontiers. *Kolner Z. Soz. Sozpsychol.* **67** (Suppl. 1), 397–424 (2015).
7. Mills, M., Rindfuss, R.R., McDonald, P. & te Velde, E. Why do people postpone parenthood? Reasons and social policy incentives. *Hum. Reprod. Update* **17**, 848–860 (2011).
8. Boivin, J., Bunting, L., Collins, J.A. & Nygren, K.G. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum. Reprod.* **22**, 1506–1512 (2007).
9. Mascarenhas, M.N., Flaxman, S.R., Boerma, T., Vanderpoel, S. & Stevens, G.A. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med.* **9**, e1001356 (2012).
10. Venkatesh, T., Suresh, P.S. & Tsutsumi, R. New insights into the genetic basis of infertility. *Appl. Clin. Genet.* **7**, 235–243 (2014).
11. Day, F.R. *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* **48**, 617–623 (2016).
12. Balbo, N., Billari, F.C. & Mills, M.C. Fertility in advanced societies: a review of research. *Eur. J. Popul.* **29**, 1–38 (2012).
13. Tropf, F.C. *et al.* Human fertility, molecular genetics, and natural selection in modern societies. *PLoS One* **10**, e0126821 (2015).
14. Fisher, R.A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
15. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
16. van der Most, P.J. *et al.* QCGWAS: a flexible R package for automated quality control of genome-wide association results. *Bioinformatics* **30**, 1185–1186 (2014).
17. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
18. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
19. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
20. 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).
21. Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
22. Liu, J.Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).

23. Mishra, A. & Macgregor, S. VEGAS2: software for more flexible gene-based testing. *Twin Res. Hum. Genet.* **18**, 86–91 (2015).
24. Vaez, A. *et al. In silico* post genome-wide association studies analysis of C-reactive protein loci suggests an important role for interferons. *Circ Cardiovasc Genet* **8**, 487–497 (2015).
25. ENCODE Project Consortium. ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
26. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
27. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
28. Zhernakova, D. *et al.* Hypothesis-free identification of modulators of genetic risk factors. Preprint at *bioRxiv* http://dx.doi.org/10.1101/033217 (2015).
29. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. Preprint at *bioRxiv* http://dx.doi.org/10.1101/033084 (2015).
30. Tranchevent, L.C. *et al.* ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res.* **36**, W377–W384 (2008).
31. Pers, T.H., Dworzyński, P., Thomas, C.E., Lage, K. & Brunak, S. MetaRanker 2.0: a web server for prioritization of genetic variation data. *Nucleic Acids Res.* **41**, W104–W108 (2013).
32. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
33. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
34. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
35. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
36. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
37. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
38. Rietveld, C.A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. USA* **111**, 13790–13794 (2014).
39. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
40. Day, F.R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and *BRCA1*-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
41. Perry, J.R. *et al.* A genome-wide association study of early menopause and the combined impact of identified variants. *Hum. Mol. Genet.* **22**, 1465–1472 (2013).
42. Fang, W.-L. *et al.* CREB coactivator CRTC2/TORC2 and its regulator calcineurin crucially mediate follicle-stimulating hormone and transforming growth factor β1 upregulation of steroidogenesis. *J. Cell. Physiol.* **227**, 2430–2440 (2012).
43. Cao, G. *et al.* Molecular cloning and characterization of a novel human cAMP response element–binding (CREB) gene (*CREB4*). *J. Hum. Genet.* **47**, 373–376 (2002).
44. El-Alfy, M. *et al.* Stage-specific expression of the Atce1/Tisp40α isoform of CREB3L4 in mouse spermatids. *J. Androl.* **27**, 686–694 (2006).
45. Adham, I.M. *et al.* Reduction of spermatogenesis but not fertility in *Creb3l4*-deficient mice. *Mol. Cell. Biol.* **25**, 7657–7664 (2005).
46. McAllister, J.M. *et al.* Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. *Proc. Natl. Acad. Sci. USA* **111**, E1519–E1527 (2014).
47. O'Bryan, M.K. *et al.* RBM5 is a male germ cell splicing factor and is required for spermatid differentiation and male fertility. *PLoS Genet.* **9**, e1003628 (2013).
48. Tsukamoto, S. *et al.* Functional analysis of lysosomes during mouse preimplantation embryo development. *J. Reprod. Dev.* **59**, 33–39 (2013).
49. Szucs, M., Osvath, P., Laczko, I. & Jakab, A. Adequacy of hyaluronan binding assay and a new fertility index derived from it for measuring of male fertility potential and the efficacy of supplement therapy. *Andrologia* **47**, 519–524 (2015).
50. Buensuceso, A.V. *et al.* Ephrin-A5 is required for optimal fertility and a complete ovulatory response to gonadotropins in the female mouse. *Endocrinology* **157**, 942–955 (2016).
51. Jisa, E. & Jungbauer, A. Kinetic analysis of estrogen receptor homo- and heterodimerization *in vitro*. *J. Steroid Biochem. Mol. Biol.* **84**, 141–148 (2003).
52. O'Donnell, L., Robertson, K.M., Jones, M.E. & Simpson, E.R. Estrogen and spermatogenesis. *Endocr. Rev.* **22**, 289–318 (2001).
53. Ly-Huynh, J.D. *et al.* Importin α2–interacting proteins with nuclear roles during mammalian spermatogenesis. *Biol. Reprod.* **85**, 1191–1202 (2011).
54. Varshney, G.K. *et al.* CRISPRz: a database of zebrafish validated sgRNAs. *Nucleic Acids Res.* **44**, D1, D822–D826 (2016).
55. Menken, J. Age and fertility: how late can you wait? *Demography* **22**, 469–483 (1985).
56. Manolio, T.A., Brooks, L.D. & Collins, F.S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118**, 1590–1605 (2008).
57. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
58. Okkelman, I.A., Sukaeva, A.Z., Kirukhina, E.V., Korneenko, T.V. & Pestov, N.B. Nuclear translocation of lysyl oxidase is promoted by interaction with transcription repressor p66β. *Cell Tissue Res.* **358**, 481–489 (2014).
59. Joshi, N.R. *et al.* Altered expression of microRNA-451 in eutopic endometrium of baboons (*Papio anubis*) with endometriosis. *Hum. Reprod.* **30**, 2881–2891 (2015).
60. Franklin, R.B. *et al.* Human ZIP1 is a major zinc uptake transporter for the accumulation of zinc in prostate cells. *J. Inorg. Biochem.* **96**, 435–442 (2003).
61. Lisle, R.S., Anthony, K., Randall, M.A. & Diaz, F.J. Oocyte–cumulus cell interactions regulate free intracellular zinc in mouse oocytes. *Reproduction* **145**, 381–390 (2013).
62. Shan, B. *et al.* Association of *DENND1A* gene polymorphisms with polycystic ovary syndrome: a meta-analysis. *J. Clin. Res. Pediatr. Endocrinol.* **8**, 135–143 (2016).
63. Impera, L. *et al.* A novel fusion 5′*AFF3*/3′*BCL2* originated from a t(2;18)(q11.2;q21.33) translocation in follicular lymphoma. *Oncogene* **27**, 6187–6190 (2008).
64. Urano, A. *et al.* Infertility with defective spermiogenesis in mice lacking AF5q31, the target of chromosomal translocation in human infant leukemia. *Mol. Cell. Biol.* **25**, 6834–6845 (2005).
65. Reese, K.L. *et al.* Acidic hyaluronidase activity is present in mouse sperm and is reduced in the absence of SPAM1: evidence for a role for hyaluronidase 3 in mouse and human sperm. *Mol. Reprod. Dev.* **77**, 759–772 (2010).
66. Heath, E., Sablitzky, F. & Morgan, G.T. Subnuclear targeting of the RNA-binding motif protein RBM6 to splicing speckles and nascent transcripts. *Chromosome Res.* **18**, 851–872 (2010).
67. Kamura, T. *et al.* Cytoplasmic ubiquitin ligase KPC regulates proteolysis of $p27^{Kip1}$ at G1 phase. *Nat. Cell Biol.* **6**, 1229–1235 (2004).
68. Kato, J.Y., Matsuoka, M., Polyak, K., Massagué, J. & Sherr, C.J. Cyclic AMP–induced G1 phase arrest mediated by an inhibitor ($p27^{Kip1}$) of cyclin-dependent kinase 4 activation. *Cell* **79**, 487–496 (1994).
69. Bagley, D.C., Paradkar, P.N., Kaplan, J. & Ward, D.M. Mon1a protein acts in trafficking through the secretory apparatus. *J. Biol. Chem.* **287**, 25577–25588 (2012).
70. Sakamoto, O. *et al.* Role of macrophage-stimulating protein and its receptor, RON tyrosine kinase, in ciliary motility. *J. Clin. Invest.* **99**, 701–709 (1997).
71. Zhang, C. *et al.* Molecular mechanisms that drive estradiol-dependent burst firing of Kiss1 neurons in the rostral periventricular preoptic area. *Am. J. Physiol. Endocrinol. Metab.* **305**, E1384–E1397 (2013).
72. Ponglikitmongkol, M., Green, S. & Chambon, P. Genomic organization of the human oestrogen receptor gene. *EMBO J.* **7**, 3385–3388 (1988).
73. 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).
74. Lamp, M. *et al.* Polymorphisms in *ESR1*, *ESR2* and *HSD17B1* genes are associated with fertility status in endometriosis. *Gynecol. Endocrinol.* **27**, 425–433 (2011).
75. Chiu, Y.-C. *et al.* Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Dev. Neurobiol.* **74**, 723–738 (2014).
76. Alves, M.G. *et al.* Metabolic fingerprints in testicular biopsies from type 1 diabetic patients. *Cell Tissue Res.* **362**, 431–440 (2015).
77. Mojiminiyi, O.A., Safar, F.H., Al Rumaih, H. & Diejomaoh, M. Variations in alanine aminotransferase levels within the normal range predict metabolic and androgenic phenotypes in women of reproductive age. *Scand. J. Clin. Lab. Invest.* **70**, 554–560 (2010).
78. Van Maldergem, L. *et al.* Revisiting the craniosynostosis-radial ray hypoplasia association: Baller–Gerold syndrome caused by mutations in the *RECQL4* gene. *J. Med. Genet.* **43**, 148–152 (2016).
79. Ruan, Y., Cheng, M., Ou, Y., Oko, R. & van der Hoorn, F.A. Ornithine decarboxylase antizyme Oaz3 modulates protein phosphatase activity. *J. Biol. Chem.* **286**, 29417–29427 (2011).

Nicola Barban[1,173], Rick Jansen[2], Ronald de Vlaming[3–5], Ahmad Vaez[6,7], Jornt J Mandemakers[8], Felix C Tropf[1], Xia Shen[9–11], James F Wilson[10,11], Daniel I Chasman[12], Ilja M Nolte[6], Vinicius Tragante[13], Sander W van der Laan[14], John R B Perry[15], Augustine Kong[16,17], BIOS Consortium[18], Tarunveer S Ahluwalia[19–21], Eva Albrecht[22], Laura Yerges-Armstrong[23], Gil Atzmon[24–26], Kirsi Auro[27,28], Kristin Ayers[29], Andrew Bakshi[30], Danny Ben-Avraham[25], Klaus Berger[31], Aviv Bergman[32], Lars Bertram[33,34], Lawrence F Bielak[35], Gyda Bjornsdottir[17], Marc Jan Bonder[36], Linda Broer[37], Minh Bui[38], Caterina Barbieri[39], Alana Cavadino[40,41], Jorge E Chavarro[42–44], Constance Turman[44], Maria Pina Concas[45], Heather J Cordell[29], Gail Davies[46,47], Peter Eibich[48], Nicholas Eriksson[49], Tõnu Esko[50,51],

Joel Eriksson[52], Fahimeh Falahi[6], Janine F Felix[4,53,54], Mark Alan Fontana[55], Lude Franke[36], Ilaria Gandin[56], Audrey J Gaskins[42], Christian Gieger[57,58], Erica P Gunderson[59], Xiuqing Guo[60], Caroline Hayward[10], Chunyan He[61], Edith Hofer[62,63], Hongyan Huang[44], Peter K Joshi[11], Stavroula Kanoni[64], Robert Karlsson[9], Stefan Kiechl[65], Annette Kifley[66], Alexander Kluttig[67], Peter Kraft[44,68], Vasiliki Lagou[69–71], Cecile Lecoeur[72], Jari Lahti[73–75], Ruifang Li-Gao[76], Penelope A Lind[77], Tian Liu[78], Enes Makalic[38], Crysovalanto Mamasoula[29], Lindsay Matteson[79], Hamdi Mbarek[80,81], Patrick F McArdle[23], George McMahon[82], S Fleur W Meddens[5,83], Evelin Mihailov[50], Mike Miller[84], Stacey A Missmer[44,85], Claire Monnereau[4,53,54], Peter J van der Most[6], Ronny Myhre[86], Mike A Nalls[87], Teresa Nutile[88], Ioanna Panagiota Kalafati[89], Eleonora Porcu[90,91], Inga Prokopenko[92–94], Kumar B Rajan[95], Janet Rich-Edwards[43,44,96], Cornelius A Rietveld[3–5], Antonietta Robino[97], Lynda M Rose[12], Rico Rueedi[98,99], Kathleen A Ryan[23], Yasaman Saba[100], Daniel Schmidt[38], Jennifer A Smith[35], Lisette Stolk[37], Elizabeth Streeten[23], Anke Tönjes[101], Gudmar Thorleifsson[17], Sheila Ulivi[97], Juho Wedenoja[102], Juergen Wellmann[31], Peter Willeit[65,103,104], Jie Yao[60], Loic Yengo[72,105], Jing Hua Zhao[15], Wei Zhao[35], Daria V Zhernakova[36], Najaf Amin[4], Howard Andrews[106], Beverley Balkau[72], Nir Barzilai[24], Sven Bergmann[98,99], Ginevra Biino[107], Hans Bisgaard[21], Klaus Bønnelykke[21], Dorret I Boomsma[80,81], Julie E Buring[12], Harry Campbell[11], Stefania Cappellani[97], Marina Ciullo[88,108], Simon R Cox[46,47], Francesco Cucca[90,91], Daniela Toniolo[39], George Davey-Smith[109], Ian J Deary[46,47], George Dedoussis[89], Panos Deloukas[64,110], Cornelia M van Duijn[4], Eco J C de Geus[80,81], Johan G Eriksson[75,111–114], Denis A Evans[95], Jessica D Faul[115], Cinzia Felicita Sala[39], Philippe Froguel[72], Paolo Gasparini[56,116], Giorgia Girotto[56,116], Hans-Jörgen Grabe[117], Karin Halina Greiser[118], Patrick J F Groenen[5,119], Hugoline G de Haan[76], Johannes Haerting[67], Tamara B Harris[120], Andrew C Heath[121], Kauko Heikkilä[28], Albert Hofman[4,5,44], Georg Homuth[122], Elizabeth G Holliday[123,124], John Hopper[38], Elina Hyppönen[40,125,126], Bo Jacobsson[86,127], Vincent W V Jaddoe[4,53,54], Magnus Johannesson[128], Astanand Jugessur[86], Mika Kähönen[129], Eero Kajantie[130–132], Sharon L R Kardia[35], Bernard Keavney[29,133], Ivana Kolcic[134], Päivikki Koponen[135], Peter Kovacs[136], Florian Kronenberg[137], Zoltan Kutalik[99,138], Martina La Bianca[97], Genevieve Lachance[139], William G Iacono[84], Sandra Lai[90], Terho Lehtimäki[140], David C Liewald[46], LifeLines Cohort Study[18], Cecilia M Lindgren[93,94,141,142], Yongmei Liu[143], Robert Luben[144], Michael Lucht[111], Riitta Luoto[145], Per Magnus[86], Patrik K E Magnusson[9], Nicholas G Martin[77], Matt McGue[84,146], Ruth McQuillan[11], Sarah E Medland[77], Christa Meisinger[58,147], Dan Mellström[52], Andres Metspalu[50,148], Michela Traglia[39], Lili Milani[50], Paul Mitchell[66], Grant W Montgomery[121,149], Dennis Mook-Kanamori[76,150,151], Renée de Mutsert[76], Ellen A Nohr[152], Claes Ohlsson[52], Jørn Olsen[153], Ken K Ong[15], Lavinia Paternoster[109], Alison Pattie[47], Brenda W J H Penninx[2,81], Markus Perola[27,28,50], Patricia A Peyser[35], Mario Pirastu[45], Ozren Polasek[134], Chris Power[40], Jaakko Kaprio[27,28,102], Leslie J Raffel[154], Katri Räikkönen[73], Olli Raitakari[155], Paul M Ridker[12], Susan M Ring[109], Kathryn Roll[60], Igor Rudan[11], Daniela Ruggiero[88], Dan Rujescu[156], Veikko Salomaa[27], David Schlessinger[157], Helena Schmidt[100], Reinhold Schmidt[62], Nicole Schupf[158], Johannes Smit[2,81], Rossella Sorice[88,108], Tim D Spector[139], John M Starr[46,159], Doris Stöckl[158], Konstantin Strauch[22,160], Michael Stumvoll[101,136], Morris A Swertz[36], Unnur Thorsteinsdottir[17,161], A Roy Thurik[3,5,162], Nicholas J Timpson[109], Joyce Y Tung[49], André G Uitterlinden[3,5,37], Simona Vaccargiu[45], Jorma Viikari[163], Veronique Vitart[10], Henry Völzke[164], Peter Vollenweider[165], Dragana Vuckovic[56,116], Johannes Waage[21], Gert G Wagner[166], Jie Jin Wang[66], Nicholas J Wareham[15], David R Weir[115], Gonneke Willemsen[80,81], Johann Willeit[65], Alan F Wright[10], Krina T Zondervan[167,168], Kari Stefansson[17,161], Robert F Krueger[84], James J Lee[84], Daniel J Benjamin[55,169], David Cesarini[170,171], Philipp D Koellinger[3,5,83], Marcel den Hoed[172,173], Harold Snieder[6,173] & Melinda C Mills[1,173]

[1]Department of Sociology and Nuffield College, University of Oxford, Oxford, UK. [2]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. [3]Department of Applied Economics, Erasmus School of Economics, Rotterdam, the Netherlands. [4]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [5]Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, the Netherlands. [6]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [7]Research Institute for Primordial Prevention of Non-Communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran. [8]Sociology of Consumption and Households, Wageningen University Research, Wageningen, the Netherlands. [9]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [10]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [11]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [12]Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [13]Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. [14]Laboratory of Experimental Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. [15]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [16]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [17]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [18]A list of members and affiliations appears in the **Supplementary Note**. [19]Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [20]Steno Diabetes Center, Gentofte, Denmark. [21]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

[22]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [23]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [24]Department of Medicine, Institute for Aging Research and the Diabetes Research Center, Albert Einstein College of Medicine, Bronx, New York, USA. [25]Department of Genetics, Institute for Aging Research and the Diabetes Research Center, Albert Einstein College of Medicine, Bronx, New York, USA. [26]Department of Natural Science, University of Haifa, Haifa, Israel. [27]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [28]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland. [29]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK. [30]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [31]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [32]Departments of Systems and Computational Biology, of Pathology and of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, USA. [33]Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. [34]School of Public Health, Faculty of Medicine, Imperial College, London, UK. [35]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA. [36]Department of Genetics, Genomics Coordination Center, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [37]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [38]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. [39]Division of Genetics and Cell Biology, San Raffaele Research Institute, Milan, Italy. [40]Population, Policy and Practice, UCL Institute of Child Health, London, UK. [41]Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventative Medicine, Queen Mary University of London, London, UK. [42]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [43]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [44]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [45]Institute of Genetic and Biomedical Research, National Research Council, UOS of Sassari, Sassari, Italy. [46]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [47]Department of Psychology, University of Edinburgh, Edinburgh, UK. [48]Health Economics Research Centre, University of Oxford, Oxford, UK. [49]23andMe, Inc., Mountain View, California, USA. [50]Estonian Genome Center, University of Tartu, Tartu, Estonia. [51]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [52]Centre for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [53]Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [54]Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [55]Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA. [56]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy. [57]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [58]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [59]Cardiovascular and Metabolic Conditions Section, Division of Research, Kaiser Permanente Northern California, Oakland, California, USA. [60]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA. [61]Department of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA. [62]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. [63]Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. [64]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [65]Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria. [66]Centre for Vision Research, Department of Ophthalmology and Westmead Institute for Medical Research, University of Sydney, Westmead, New South Wales, Australia. [67]Institute of Medical Epidemiology, Biostatistics and Informatics, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [68]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [69]Department of Neurosciences, KU Leuven, Leuven, Belgium. [70]Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium. [71]Translational Immunology Laboratory, VIB, Leuven, Belgium. [72]University of Lille, CNRS, Institut Pasteur of Lille, Lille, France. [73]Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland. [74]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland. [75]Folkhälsan Research Centre, Helsinki, Finland. [76]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [77]Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Herston Brisbane, Queensland, Australia. [78]Center for Lifespan Psychology, Max Planck Institute for Human Development and Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [79]Minnesota Center for Twin and Family Research, Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA. [80]Department of Biological Psychology, VU University, Amsterdam, the Netherlands. [81]EMGO+ Institute for Health and Care Research, Amsterdam, the Netherlands. [82]School of Social and Community Medicine, University of Bristol, Bristol, UK. [83]Complex Trait Genetics, VU University, Amsterdam, the Netherlands. [84]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA. [85]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [86]Department of Genetics and Bioinformatics, Area of Health Data and Digitalization, Institute of Public Health, Oslo, Norway. [87]Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [88]Institute of Genetics and Biophysics 'A. Buzzati-Traverso', CNR, Naples, Italy. [89]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [90]Istituto di Ricerca Genetica e Biomedica, CMR, Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy. [91]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy. [92]Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK. [93]Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [94]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [95]Rush University Medical Center, Chicago, Illinois, USA. [96]Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [97]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [98]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [99]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [100]Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria. [101]Department of Medicine, University of Leipzig, Leipzig, Germany. [102]Department of Public Health, University of Helsinki, Helsinki, Finland. [103]King's British Heart Foundation Centre, King's College London, London, UK. [104]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [105]Centre for Neurogenetics and Statistical Genomics, University of Queensland, Brisbane, Queensland, Australia. [106]Data Coordinating Center, New York State Psychiatric Institute, New York, New York, USA. [107]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. [108]IRCCS Neuromed, Pozzilli, Isernia, Italy. [109]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [110]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [111]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [112]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [113]Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. [114]Vasa Central Hospital, Vasa, Finland. [115]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA. [116]Division of Experimental Genetics, Sidra, Doha, Qatar. [117]Department of Psychiatry, University Medicine Greifswald, Greifswald, Germany. [118]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. [119]Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [120]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, USA. [121]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [122]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [123]School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia. [124]Hunter Medical Research Institute, Newcastle, New South Wales, Australia. [125]Centre for Population Health Research, Sansom Institute of Health Research and School of Health Sciences, University of South Australia, Adelaide, South Australia, Australia. [126]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [127]Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden. [128]Department of Economics, Stockholm School of Economics, Stockholm, Sweden. [129]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. [130]Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. [131]Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland. [132]Department of Obstetrics and Gynecology, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. [133]Institute of Cardiovascular Sciences, University of Manchester, Manchester, UK. [134]Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia. [135]Health Monitoring Unit, National Institute for Health and Welfare, Helsinki, Finland. [136]IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany. [137]Division of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. [138]Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [139]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [140]Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland. [141]NIHR Oxford Biomedical Research Centre, Oxford, UK. [142]Li Ka Shing Centre for Health Information and Discovery, Big Data Institute, University of Oxford, Oxford, UK. [143]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [144]Strangeways Research Laboratory, University of Cambridge, Cambridge, UK. [145]UKK Institute for

Health Promotion,Tampere, Finland. [146]Danish Aging Research Center and Danish Twin Registry, Institute of Public Health, University of Southern Denmark, Odense, Denmark. [147]MONICA/KORA Myocardial Infarction Registry, Central Hospital of Augsburg, Augsburg, Germany. [148]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [149]Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [150]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [151]Department of BESC, Epidemiology Section, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. [152]Research Unit for Gynecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense, Denmark. [153]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark. [154]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [155]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [156]Department of Psychiatry, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [157]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. [158]Departments of Epidemiology and Psychiatry, Columbia University Medical Center, New York, New York, USA. [159]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [160]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians Universität, Munich, Germany. [161]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [162]Montpellier Business School, Montpellier, France. [163]Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland. [164]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [165]Department of Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [166]German Socio-Economic Panel Study (SOEP), Max Planck Institute for Human Development and Berlin University of Technology (TUB), Berlin, Germany. [167]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [168]Endometriosis CaRe Centre, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK. [169]National Bureau of Economic Research, Cambridge, Massachusetts, USA. [170]Department of Economics, New York University, New York, New York, USA. [171]Research Institute for Industrial Economics, Stockholm, Sweden. [172]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [173]These authors contributed equally to this work. Correspondence should be addressed to M.C.M. (melinda.mills@nuffield.ox.ac.uk), N.B. (nicola.barban@sociology.ox.ac.uk), H.S. (h.snieder@umcg.nl) or M.d.H. (marcel.den_hoed@medsci.uu.se).

ONLINE METHODS

GWAS of reproductive behavior study design in brief. Genome-wide association analyses of AFB and NEB were performed at the cohort level according to a prespecified analysis plan (**Supplementary Note**). Cohort-uploaded results were imputed using the HapMap 2 CEU (r22.b36) or 1000 Genomes Project reference sample. Cohorts were asked to only include participants of European ancestry, with no missing values for all relevant covariates (sex, birth year and cohort-specific covariates), who were successfully genotyped over the whole genome and passed cohort-specific quality control filters. We followed the quality control protocol of the GIANT Consortium's recent study of human height[20] and employed QCGWAS[16] and EasyQC[17] software, which allowed us to harmonize the files and identify possible sources of error in association results.

Cohort association results (after applying the quality control filters) were combined using sample-size-weighted meta-analysis with genomic control correction within each study, implemented in METAL[80]. SNPs were considered genome-wide significant at P values smaller than 5×10^{-8} ($\alpha = 5\%$, Bonferroni corrected for 1 million tests). The meta-analyses were carried out by two independent analysts. Detailed results for each genome-wide significant locus are shown in in **Supplementary Figures 4–29**.

The total sample size of the meta-analyses is $n = 251{,}151$ for AFB pooled and $n = 343{,}072$ for NEB pooled. The PLINK clumping function[81] was used to identify the most significant SNPs in associated regions (termed lead SNPs). Detailed cohort descriptions, information about cohort-level genotyping and imputation procedures, cohort-level measures and quality control filters are shown in **Supplementary Tables 26** and **27** and discussed in the **Supplementary Note**.

Dominant genetic variation in fertility. We applied a method recently developed by Zhu *et al.*[82] to estimate dominant genetic effects on the basis of the genetic relatedness of unrelated individuals. Our results, based on combined TwinsUK and LifeLines samples, showed no evidence of dominant genetic effects for either NEB (1.0×10^{-7}, SE = 0.07; $P = 0.45$) or AFB (0.02, SE = 0.08; $P = 0.43$). Results are shown in **Supplementary Table 28** and discussed in the **Supplementary Note**.

Bivariate and conditional analyses. As joint analysis of correlated traits may boost power for mapping functional loci, we applied a recently developed multiple-trait analysis method[83] to test the association between each variant and the two correlated traits AFB and NEB simultaneously using multivariate analysis of variance (MANOVA) (**Supplementary Table 29** and **Supplementary Note**). The analysis was performed on the basis of the genome-wide meta-analysis summary statistics for each single trait. Although this analysis did not identify additional genome-wide significant loci ($\lambda = 0.995$), it did account for the correlation between the two phenotypes, thus improving the strength of two signals on chromosomes 1 and 5, indicating possible pleiotropic architecture for AFB and NEB (**Supplementary Fig. 30**). The analysis also provided a conditional association test of the genetic effect of each variant on AFB including NEB as a covariate and the genetic effect on NEB including AFB as a covariate (**Supplementary Fig. 31**).

Population stratification. We used two methods to assess whether our GWAS results exhibited signs of population stratification (**Supplementary Note**). First, we used the LD Score intercept method described in Bulik-Sullivan *et al.*[19] to test whether inflation in χ^2 statistics was due to confounding biases such as cryptic relatedness and population stratification. In all six cases, the intercept estimates were not significantly different from 1, suggesting no appreciable inflation of the test statistics attributable to population stratification. Second, we conducted a series of individual and within-family regressions using polygenic scores as predictors[20,21,38] on a data set of dizygotic twins (STR and TwinsUK). The regression analyses showed that within-family regression coefficients for both AFB and NEB were statistically different from 0 when the P-value threshold was sufficiently high (**Supplementary Figs. 32** and **33**, and **Supplementary Tables 30** and **31**).

Sex-specific effects. In addition to the pooled GWAS for which results are presented in the main text, we also ran sex-specific GWAS meta-analyses for AFB and NEB (**Supplementary Note**). The sample sizes for sex-specific analysis were as follows: AFB in women, $n = 189{,}656$; AFB in men, $n = 48{,}408$; NEB in women, $n = 225{,}230$; NEB in men, $n = 103{,}909$. Our results identified six genome-wide significant ($P < 5 \times 10^{-8}$) independent SNPs for AFB in women and one genome-wide significant independent SNP for NEB in men (**Supplementary Figs. 34** and **35**, and **Supplementary Table 32**). We also used LD Score bivariate regression and GREML bivariate analysis to estimate the genetic correlation between men and women on the basis of the sex-specific summary statistics from the AFB and NEB meta-analyses. Our estimates based on LD bivariate regression indicated genetic correlations between the sexes of $r_g = 0.86$ (SE = 0.052) for AFB and $r_g = 0.97$ (SE = 0.095) for NEB. Results are shown in **Supplementary Tables 33** and **34** and discussed in the **Supplementary Note**.

Polygenic score prediction. We performed out-of-sample prediction and calculated polygenic scores for AFB and NEB, on the basis of genome-wide association meta-analysis results, and used regression models to predict the same phenotypes in four independent cohorts: HRS, LifeLines, STR and TwinsUK (**Supplementary Fig. 2** and **Supplementary Note**). We ran ordinary least-squares (OLS) regression models and report R^2 as a measure of goodness of fit for the model. In addition, we tested how well our polygenic scores for NEB could predict childlessness at the end of the reproductive period (using age 45 for women and 55 for men; **Supplementary Table 21**). Because AFB is observed only in parous women, we adopted an additional statistical model to account for censoring (Cox proportional hazard model; **Supplementary Table 22**) and selection (Heckman selection model; **Supplementary Table 35**). We additionally tested the predictive value of our polygenic scores for AFB on age at menarche (TwinsUK) and age at menopause (LifeLines) (**Supplementary Table 23**). Finally, we examined whether variants associated with menopause are associated with AFB. We calculated a polygenic score for age at menopause based on recent GWAS results from Day *et al.*[40] and applied the predictor to the LifeLines and TwinsUK cohorts (**Supplementary Table 36**).

Genetic correlations. We used information from 27 publicly available GWAS data sets to estimate the number of genetic correlations between AFB or NEB and related traits (**Fig. 3** and **Supplementary Table 25**) via LD Score bivariate regression. Details on these phenotypes are provided in the **Supplementary Note**. A conservative Bonferroni-corrected P-value threshold of $P < 1.85 \times 10^{-3}$ (= 0.05/27) was used to define significant associations. We also tested the correlation between NEB and AFB using bivariate GREML analysis on the Women's General Health Study (WGHS; $n = 40{,}621$).

Lookups and proxy phenotypes. Following up on the results of genetic overlap with other phenotypes, we tested in a quasi-phenotype replication setting whether the SNPs strongly associated with AFB in women were empirically plausible candidate SNPs for age at menarche and age at menopause (**Supplementary Note**). We used a two-stage approach applied in other contexts[38,84]. In the first stage, we conducted a meta-analysis of AFB excluding cohorts that were part of the meta-analysis for the phenotype we intended to replicate. We merged the SNPs from this meta-analysis with the publically available association results for the most recent GWAS on age at menarche[2] and age at menopause[40] from the ReproGen consortium website[1]. SNPs that were not present in both studies considered were dropped from the analysis. We aligned alleles and directions of effect using EasyStrata software[85]. We then selected independent SNPs with $P < 1 \times 10^{-5}$, using the clump procedure in PLINK[81] (window size of 1,000 kb and LD threshold of $r^2 > 0.1$) to identify the most significant SNPs in the associated regions included in both files. We defined 'prioritized SNP associations' as those that passed the Bonferroni correction for the number of SNPs tested ($0.05/122 = 4.10 \times 10^{-4}$, for both age at menarche and age at menopause). Our results identified three SNPs after Bonferroni correction that could be used as good candidates for age at menarche. We did not find any clear 'candidate SNP' for age at menopause (**Supplementary Fig. 36**).

Gene-based GWAS analysis. We performed gene-based testing with the full GWAS set (~2.5 million HapMap-imputed SNPs) for both phenotypes using VEGAS (**Supplementary Tables 3** and **4**, and **Supplementary Note**)[22,23].

doi:10.1038/ng.3698

This software has the advantage of accounting for LD structure and allowing a gene to be defined as a range with boundaries beyond the edges of the gene to include intergenic regions in the analysis. We defined genes including an additional 50-kb window around each gene. We considered every SNP for the gene-based analysis, ran the analyses for each chromosome with up to 10^6 permutations and considered $P < 2.5 \times 10^{-6}$ as the threshold for significance (0.05/~20.000 genes).

eQTL and meQTL analyses. For each of the 12 SNPs identified in the GWAS, local (*cis*; exons or methylation sites <1 Mb from the SNP) and genome-wide (*trans*; exons or methylation sites >5 Mb from the SNP) effects were identified by computing Spearman rank correlations between SNPs and local or global exons and methylation sites (**Supplementary Note**). Bonferroni correction for multiple testing was performed for the 12 SNPs tested ($P < 2.5 \times 10^{-6}$ for *cis*-meQTL analysis, $P < 1 \times 10^{-8}$ for *trans*-meQTL analysis, $P < 1.2 \times 10^{-6}$ for *cis*-eQTL analysis, $P < 1.3 \times 10^{-8}$ for *trans*-eQTL analysis). For each of the significant associations, the corresponding exons or methylation sites were selected, the strongest eQTLs were identified for these elements and the LD between the strongest eQTLs and the corresponding SNP identified in the GWAS was computed. LD was computed using BIOS genotypes (genotypes used for eQTL and meQTL mapping).

Functional variant analysis using RegulomeDB. We used RegulomeDB[27] to identify variants among the 322 SNPs that reached $P < 5 \times 10^{-8}$ for association with AFB and/or NEB in the meta-analysis of GWAS results that likely influence regulation of gene expression (**Supplementary Note**). RegulomeDB integrates results from the Roadmap Epigenomics[26] and ENCODE[86] projects. SNPs showing the most evidence of being functional—defined by having a RegulomeDB score <4—were subsequently examined in more detail in terms of effects on gene expression (eQTLs) and their protein-binding capacity (**Supplementary Table 6**).

Gene prioritization. Potentially causal genes for the associations identified by GWAS were identified using four previously described bioinformatics tools: ToppGene[4], Endeavor[5], MetaRanker[6] and DEPICT[7]. To this end, we first retrieved positional coordinates for all lead SNPs according to GRCh37/hg19 using Ensembl BioMart. These coordinates were used to extract all genes located within 40 kb of lead SNPs from the UCSC table browser. The identified genes then served as input for ToppGene and Endeavor. Genes with established roles in fertility served as training genes in this procedure, that is, *BRCA1*, *EGFR*, *ERBB2*, *ERBB3*, *ERBB4*, *HSD17B1*, *RBM5*, *ESR1*, *ESR2* and *FSHB*. For MetaRanker, we provided SNPs that reached $P < 5 \times 10^{-4}$ and their chromosomal positions as input, together with the above set of training genes. Because ToppGene, Endeavor and MetaRanker are biased toward larger and well-described genes, we also performed a gene prioritization procedure using DEPICT[7]. All SNPs that reached $P < 5 \times 10^{-4}$ in the meta-analysis served as input, and information on prioritized genes, gene set enrichment, and tissue and cell type enrichment was extracted. Genes were subsequently prioritized if they (i) reached $P < 0.05$ in DEPICT or (ii) reached $P < 0.05$ in ToppGene, Endeavor and MetaRanker (**Supplementary Table 37**).

Functional network and enrichment analyses. DEPICT was used to identify gene set, cell type and tissue enrichment, using the GWAS-identified SNPs with $P < 5 \times 10^{-4}$ as input (**Supplementary Note**). Because of the relatively small number of identified loci, DEPICT was only able to perform these analyses for AFB and NEB pooled and for AFB in women. To construct a functional association network, we combined five prioritized candidate gene sets into a single query gene set that was then used as input for functional network analysis[24]. We applied the GeneMANIA algorithm together with its large set of accompanying functional association data[87]. We used the Cytoscape software platform[88], extended by the GeneMANIA plugin (data version 8/12/2014, accessed 24 April 2016)[89]. All the genes in the composite network, from either the query or resulting gene sets, were then used for functional enrichment analysis against Gene Ontology (GO) terms[90] to identify the most relevant terms, using the same plugin[89].

Gene–environment interactions. Previous research based on twin studies shows differential heritability of fertility behavior across birth cohorts[91,92]. We used the Swedish Twin Register (STR) to examine whether the effect of a polygenic score for AFB or NEB varied across birth cohort. We followed the analysis presented in the recent GWAS of education[35] and divided the sample into six groups on the basis of year of birth. Each group spanned five birth years, with the oldest ranging from 1929–1933 and the youngest born from 1954–1958. **Supplementary Table 38** reports the estimated coefficients from these regressions. The results indicate a U-shaped trend in AFB and a linear decline in NEB, but they do not provide any clear evidence of interaction effects between the polygenic scores and birth cohort. We additionally tested the interaction effects for educational level and the polygenic scores for AFB and NEB in three different samples (LifeLines, STR and HRS). **Supplementary Table 39** reports the estimated coefficients from these regressions. The results indicate that years of education are positively associated with AFB in both the LifeLines and STR cohorts and negatively associated with NEB in the HRS cohort. With the exception of NEB in the HRS cohort, we found no evidence of gene–environment effects with education.

Robustness checks. To estimate the robustness of our results for AFB, we conducted two additional analyses. First, we estimated how the coefficients changed if we controlled for educational attainment. Using data from deCODE, we ran an additional association analysis using the ten loci that were genome-wide significant in the meta-analysis ($P < 5 \times 10^{-8}$). The analysis was restricted to individuals born between 1910 and 1975 who also had data available on completed education. The total sample size was 42,187 (17,996 males and 24,191 females). The analysis was adjusted for sex, year of birth (linear, squared and cubed), interaction between sex and year of birth, and the first ten principal components. Education is measured by years of education, ranging between 10 and 20 years. **Supplementary Table 40** reports the association results before and after adjusting for educational attainment. Our analysis shows that effect sizes shrink after including educational attainment as a covariate, with an average reduction of around 15%. We also estimated the effect of a polygenic risk score for AFB calculated from meta-analysis data excluding the deCODE cohort. The polygenic risk score remained highly significant. The effect of 1 s.d. for the AFB score decreased from 0.19 years (69 d) without controlling for education to 0.16 years (59 d) when we controlled for years of education. Second, we estimated how the coefficients changed after controlling for educational attainment and age at first sexual intercourse using the UK Biobank cohort ($n = 50{,}954$). We ran two association models: the first followed the GWAS analysis plan with no additional covariates, and the second added years of education and age at first sexual intercourse as covariates. The results are presented in **Supplementary Figure 37** and **Supplementary Table 41**. Our analysis shows that the effect sizes of our top hits are highly concordant ($R^2 = 0.94$). The inclusion of educational attainment and age at first sexual intercourse as covariates weakened the effect sizes on average by 40% and increased the *P* values of the estimated coefficients. Overall, we interpret this additional analysis as a robustness test that confirms that the top hits from our meta-analysis are robust to the inclusion of the confounding factors of educational attainment and age at first sexual intercourse.

Positive selection. We performed Haploplotter analysis[93] to examine whether lead SNPs and/or functional variants identified using RegulomeDB showed evidence of positive selection. Three variants showed standardized integrated haplotype scores <–2 or >2, indicating that these variants represent the top 5% of signals in the population. These SNPs are (i) rs7628058 on chromosome 3 for AFB, an eQTL for *RBM6* in monocytes; (ii) rs2247510 on chromosome 3 for AFB, an eQTL for *RBM6* and *HYAL3* in monocytes and a binding site for a range of transcription factors; and (iii) rs2415984, the lead SNP in the chromosome 14 locus for NEB. Results are presented in **Supplementary Table 42**.

Data availability. Results can be downloaded from the SOCIOGENOME and SSGAC website. Data come from multiple studies, most of which are subject to a MTA, and are listed in the **Supplementary Note**. Correspondence and requests for materials should be addressed to the corresponding authors or info@sociogenome.com.

80. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
81. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
82. Zhu, Z. *et al*. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).
83. Shen, X. *et al*. Simple multi-trait analysis identifies novel loci associated with growth and obesity measures. Preprint at *bioRxiv* http://dx.doi.org/10.1101/022269 (2015).
84. Okbay, A. *et al*. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
85. Winkler, T.W. *et al*. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
86. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
87. Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* **9** (Suppl. 1), S4 (2008).
88. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nat. Methods* **9**, 1069–1076 (2012).
89. Montojo, J. *et al*. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927–2928 (2010).
90. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
91. Kohler, H.-P., Rodgers, J.L. & Christensen, K. Is fertility behavior in our genes? Findings from a Danish twin study. *Popul. Dev. Rev.* **25**, 253–288 (1999).
92. Tropf, F.C., Barban, N., Mills, M.C., Snieder, H. & Mandemakers, J.J. Genetic influence on age at first birth of female twins born in the UK, 1919–68. *Popul. Stud. (Camb.)* **69**, 129–145 (2015).
93. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).