

FREQUENTLY ASKED QUESTIONS

Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour

Mills, M.C., F.C. Tropf, D.M. Brazel, N. van Zydnam, A. Vaez, eQTLGen Consortium, BIOS Consortium, Human Reproductive Behaviour Consortium, Tune H. Pers, H. Snieder, J.R.B. Perry, K.K. Long, M. den Hoed, N. Barban, F.R. Day (2021) Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour, *Nature Human Behaviour*, doi: 10.1038/s41562-021-01135-3.

Upon publication, the paper is available at: <https://www.nature.com/articles/s41562-021-01135-3>

For further information, contact: melinda.mills@nuffield.ox.ac.uk

WHY THIS FAQ?

This a complex interdisciplinary study using multiple advanced methods examining the sensitive and complex traits of age at first sex and first birth. Given the room for misinterpretation, our aim is to provide an accessible document for those interested in the subject matter but not necessarily versed in the terminology and methods. For experts or those seeking more detail and the scientific references underpinning our statements, refer to our main article and the detailed Supplementary material. For a **Glossary of terms** see the end of this document.

WHAT DID WE STUDY?

The Mills et al. (2021) [Nature Human Behaviour article](#) examines the onset of human reproductive behaviour, defined as the age at first sexual intercourse (AFS) and age at when an individual has their first birth (AFB).

Age at first sexual intercourse (AFS) was assessed using questions such as: *What was your age when you first had sexual intercourse?* We included all categories (sexual intercourse includes vaginal, oral or anal intercourse).

Age at first birth (AFB) was asked either asked directly or created from several survey questions (e.g., birthdate of participant and date of birth of first child). The most common question was: *How old were you when you had your first child?* Or *What is the date of birth of your first child?* Individuals were eligible for inclusion if they were asked about their AFB and had given birth to a child.

WHAT WAS THE AIM OF THIS STUDY?

The aim of this study was to:

- identify genetic variants associated with age at first sexual intercourse (AFS) and age at first birth (AFB),
- examine their relationship with other behavioral and medical traits,
- relationship with the socio-environment, and;
- uncover the potential biological function of the variants we discovered.

WHAT ARE THE BIG TAKE-AWAY FINDINGS?

MAIN TAKE-AWAYS: THE SHORT ANSWER

- In the largest genetic discovery to date, we found 371 genetic variants linked to age at first sexual intercourse (AFS) (282) and age at first birth (AFB) (89). This is hundreds more than the previous studies that found 38 (AFS) and 10 (AFB), respectively.
- Combining all results together in a polygenic score, the discovered genetic variants explain 5-6% of the variability in the timing of age at first birth and sex, with whole genome analyses suggesting that genetics may explain up to 15-17%. We also show the effects of genetics can change as the social environment changes.
- The onset of human reproductive behaviour is driven by genetics, but also individual and socio-environmental factors such as level of education, personality, risk taking, substance use, availability of contraception, employment, gender equity and work-life reconciliation.
- The genetic effects we found were sensitive to childhood socioeconomic circumstances, with a higher polygenic score linked to later AFS/AFB and growing up in a household with college-educated parents translating into considerably later ages at first sex and birth.
- There were strong genetic overlaps with AFS/AFB with the genetics of educational attainment, reproductive traits and externalizing behaviour (ADHD, major depressive disorder, age at onset smoking).
- There does not appear to a main third factor (e.g., personality, risk taking) that drives both the genetics of education and age at first birth, although age at initiation of smoking does slightly mediate the relationship.
- Later age at first birth may capture potential metabolic effects protective against later life diseases (type 2 diabetes, coronary artery disease) and is related to parental longevity.
- The biological function of the genetic variants we discovered are related to male and female fertility and infertility markers, suggesting new avenues for reproductive health and fertility research.

MAIN TAKE-AWAYS: THE LONG ANSWER

- We **discovered 371 genetic variants**, of which:
 - 282 were for AFS (272 pooled; 2 women and 8 men only)
 - 89 were for AFB (84 pooled; 1 women and 4 X chromosome only)
- **Heritability (the amount of variability we ascribe to genetics) of AFS and AFB have increased over time**, but this varies by sex. Heritability of AFB increased from 9% for those born in 1940 to ~23% for those born in 1965. It may suggest an increased importance of genetic factors over time (e.g., reproductive choice), access to contraception but additional analyses need to [separate genetic and social factors](#). The effects of genetics can change as the social environment changes, which we [showed previously](#).
- When variants are combined into one polygenic score, which includes information of all contributing genetic factors, **all genetic variants can explain around:**
 - **5% (AFB) and 6% (AFS) of the variability of the timing of age at first birth and sex** respectively.
 - Expressed differently, a one standard deviation change in the polygenic score is associated with a 7.3 and 6.3 month delay in AFS and AFB respectively.

- The **genetic effects we isolated are sensitive to childhood socioeconomic status**, with those in the **top 5% of the genetic predisposition (polygenic score) associated with:**
 - For AFS (i.e., later AFS) these people are less likely have a sexual debut before age 19
 - There is also an interaction with socioeconomic status such as there is a 2.08 years delay of sexual debut between the highest and lowest socioeconomic decile
 - AFB (i.e., later AFB): more likely to postpone AFB until approximately age 27
- Examining the genetic overlap (correlations) between 25 other traits by sex, we found the **strongest genetic overlaps were for:**
 - Logically, other **reproductive traits** (number of children)
 - **Educational attainment** –correlation between AFB and educational attainment is a strikingly high (0.74, ± 0.01 in women, also high for men)
 - **Externalizing behaviours** (behaviour that some might view as counter to some social norms) – which can be linked to ADHD, major depressive disorder and age at onset smoking
- Using advanced techniques to unpick whether the strong genetic correlation between years of education and AFB/AFS was the result of another mediating trait such as personality, ADHD, substance use or psychiatric disorders, we found that **the genetic correlation was:**
 - independent of risk tolerance, substance use and psychiatric disorders
 - **partially mediated by age at initiation of smoking** (i.e., a window into risky adolescent behaviour)
 - **a strong bi-directional relationship particularly AFB and years of education** rather than being both downstream of a common identified cause (e.g., risk, psychiatric disorders)
- **Strong bi-directional causality was found with AFS, AFB, years of education, risk taking and age at smoking initiation.** Only age at initiation of smoking was upstream of age at first sex (using bi-directional Mendelian Randomisation).
- **Reproductive timing appears to capture potential metabolic effects protective against later life diseases** (type 2 diabetes, coronary artery disease), serving as a more powerful predictive of later-life disease than the genetics underlying educational attainment.
- Piecing apart whether our genetic signal for AFS/AFB originated from two genetically distinguishable subclusters of reproductive biology versus externalizing behaviour. Here we found that **there was a large overlap in the genetic predisposition to ADHD, age started smoking and other externalizing behaviour measurements) but also with reproductive biology (age at menopause, menarche)**
- The genetics underlying PCOS (polycystic ovarian syndrome) are linked to those related to later AFB, suggesting **a link of the genetic variants isolated in this study with infertility.**
- Those with genetic variants related to a later AFB (one standard deviation increase in PGS), is associated with a 2-4% reduction in parental mortality at any age, suggesting **there is likely a trade-off between the timing of reproduction and longevity.** This is reminiscent of the disposable soma theory of aging, which hypothesizes an evolutionary trade-off between investments in somatic maintenance (e.g., remaining in education) that in turn reduces the resources available for reproduction.
- To understand the biology underlying the variants we discovered, the integrated results found:
 - **Prioritization of 386 unique genes** (314 in 159 loci for AFS; 106 genes in 42 loci for AFB) (prioritization finds promising genes that are likely to be involved)

- 99 of the prioritized genes were expressed at the protein level in cell types of: brain, glands and/or (fe)male reproductive organs
- Prioritization of sex-specific loci resulted in: 11 genes AFB women; 1 gene AFS women; 23 genes AFS men
- 12 of the prioritized genes were expressed at the protein level in relevant tissues
- **Prioritized genes for AFS in men and women combined were those that play a role in:**
 - **Follicle stimulating hormone** (*CGA*),
 - **oocyte development** (*KLF17*),
 - **implantation and placental growth** (*ESR1, SUMO1, ARNT, CAV1, E2F1*)
- **Prioritized genes for AFB were:**
 - [FSHB](#) (Follicle stimulating hormone beta subunit)
 - [ESR1](#) (Estrogen receptor)
- **The study uncovered interesting sex-specific relationships related to infertility**, with genes identified in the pooled meta-analyses expressed at the protein level in:
 - **Men: developing sperm, highlighting role for spermatid differentiation** (*KLF17*) (AFS) and **sperm morphogenesis and binding between acrosome-reacted sperm and zona pellucida** (*ZPBP*) (AFB). *NUP210L* was prioritized in AFS (highly expressed in development and mature sperm), which has been associated with psychological development disorders, intelligence and mathematical ability, illustrating how a testis-specific gene may be linked to brain function in some individuals.
 - **Women: genes related to endometriosis** (*CCR1*), **spontaneous abortion** (*CXCR6*) (AFB)

Several genes prioritized in AFS-associated loci were previously linked to risk-seeking behaviour, sociability and anxiety (*GTF2I, TOP2B, E2F1, NCAM1, NFASC, MEF2C*). Some of the associated genes provide concrete examples of how genetic variants associated with externalizing behavior are associated with the onset of reproductive behavior (*ERBB4* in women; *SLC44A1* and *NR1H3* in men).

- Below is the graphical abstract summarising of our findings, which is Figure 6 in our main article.

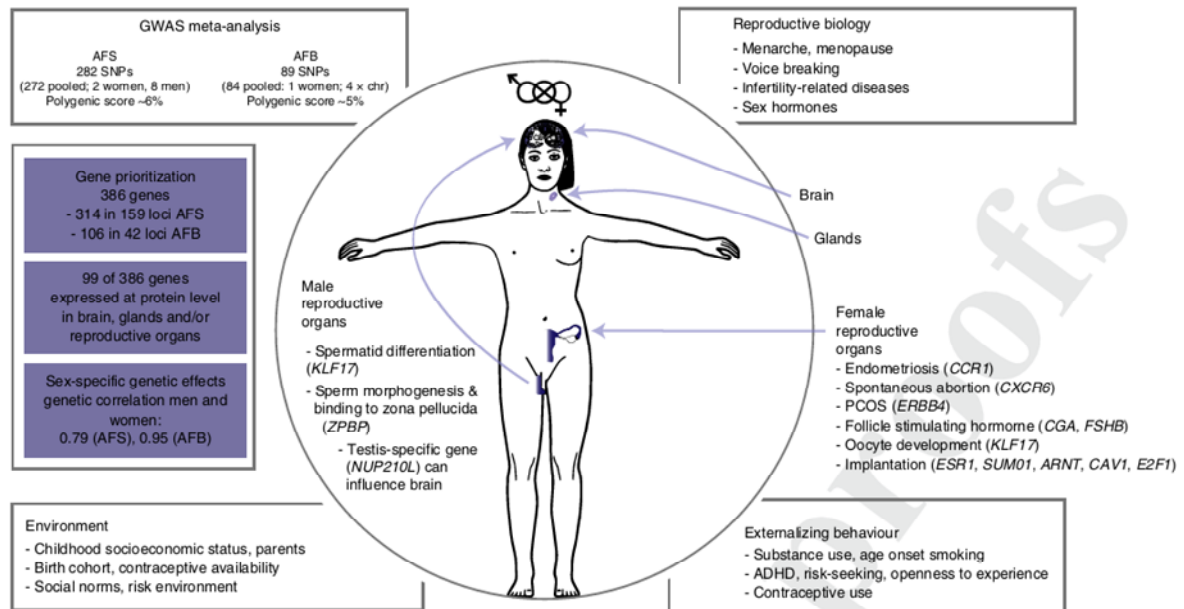


Fig. 6 | Summary GWAS of timing of onset of reproductive behaviour: AFS and AFB. Summary of GWAS meta-analysis results, highlighting that the timing of onset of reproductive behaviour is shaped by the combined effects of reproductive biology and externalizing behaviour but also exogenous environmental factors. Sex-specific genetic correlations are higher for AFB than AFS, with gene prioritization revealing some genes prioritized for men and women as well as sex-specific findings. Of the prioritized genes, 99 were expressed at the protein level in cell types of brain, glands and/or (fe)male reproductive organs. The fact that some prioritised proteins are expressed in some relevant tissues does not provide clear evidence supporting a causal role. Figure created by authors.

WHAT DID WE ALREADY KNOW?

What do we already know about the onset of human reproductive behaviour?

Socio-environmental drivers. Previous research demonstrated that the onset of human reproductive behaviour has a socio-environmental, but also a biological and genetic basis. It is a behavioural and biological trait that is firstly, driven by [socio-environmental factors](#). This includes an individual's societal and historical context, social norms, access to contraception or the ability to reconcile work and family life and have children. Demographers have determined that the key predictors of the timing of reproduction are related to factors such as [socioeconomic background](#), [education and occupation](#), [economic uncertainty and employment](#), [gender equity](#), [family networks](#) and related social and contextual factors.

Biological and genetic drivers. Yet previous research has also shown that these behaviours are not only driven by socio-environmental factors, but also have a biological and genetic basis. Our [review of twin studies](#) found that the heritability of the timing of having a first child (AFB) was around 25%, and later moving from twins to using whole-genome data we predicted a '[SNP-heritability](#)' of 15% and in the current study, 15-17% for sexual debut (AFS). Both twin and whole genome techniques determined that there was a genetic component of when individuals experienced their sexual debut and had their first child, but were unable to identify which genetic factors and whether they had a biological function. Moving beyond these percentages, our groups published two separate studies in 2016 in *Nature Genetics* finding [10 genetic variants associated with AFB](#) and [38 with AFS](#), using

samples of around 251,151 and 125,667 people respectively. The current study goes substantially beyond what we know about the genetics and biological of reproductive behaviour.

HOW DOES THE CURRENT STUDY EXTEND OUR KNOWLEDGE?

The current study extends our knowledge in several key ways, including:

- it is the **largest genetic discovery on these traits** to date, specifically: AFS (N=397,338 pooled; N=214,547 women; N=182,791) and AFB (N=542,901 pooled; N=418,758 women; N=124,008 men), combining 36 different cohort datasets
- using 1000G imputed genotype data, which in addition to the larger sample, **allows us to detect considerably more signals**
- **including an X-Chromosome analysis**, allowing us to uncover additional novel loci
- an **ability to find more and new biological signals**
- an **extensive analysis of the correlation, causality and aetiology of these traits with other behaviours and diseases** (e.g., infertility related), revealing new information linking reproductive onset to behavioural inhibition (ADHD, risk-taking behaviour, substance use) and infertility (e.g., PCOS, polycystic ovarian syndrome)
- showing that **later AFB has a strong protective effect for later life disease** (type 2 diabetes, cardiovascular) and parental **longevity**
- showing that the **polygenic scores of these behavioural traits are sensitive to socioeconomic circumstances**

WHY STUDY THIS TOPIC NOW?

Reproductive behaviour is a complex trait shaped by behaviour, biology and the social environment. The study of this trait demands a multidisciplinary approach that isolates the common drivers and how they relate to health, reproductive biology, the environment and externalising behaviour.

As shown in our Figure S3, in our Supplementary material (Panel A), also shown below, there has been a shift in the distribution of having a first child (AFB) not only to later to ages, but also a wider spread in the distribution itself. The age at having sexual intercourse has shifted to increasingly earlier over time; with [around one-third of contemporary UK teenagers](#) having sex by the age of 16 years. Conversely, many societies have experienced a shift to a later [average at birth of around 30 years for women](#) and somewhat later for men.

Early sex, teenage birth and externalising behaviour. Figure S3 (Panel B) of age at first sex (AFS) shows that in earlier cohorts, there was a bi-modal distribution, one which had earlier sexual intercourse often tied to [socio-economic circumstances, problem or risky behaviour](#). Early sexual behaviour and teenage pregnancy has also been linked to behavioural disinhibition and externalising behaviour, such as [high substance use \(e.g., smoking, drinking, drugs\) and antisocial behaviours](#). These behaviours are often associated with socioeconomic circumstances in childhood, lower parental monitoring and control. Early sexual behaviour and teenage births have been linked to [health](#), including cervical cancer, depression, sexually transmitted diseases and substance use disorders.

Birth postponement and infertility. Conversely, later reproductive onset has been linked to lower fecundity and subfertility and infertility traits (e.g., endometriosis). Over 20% of women born after 1970 in many countries are now [childless](#). The biological ability to conceive a child starts to steeply decline for some women as of age 25, with almost [50% of women being sterile by the age of 40 unable to conceive naturally](#). A growing number of women start to have their first and subsequent children at the time that their ability to conceive starts to decrease. This has led to an [unprecedented growth in infertility](#) (i.e., involuntary childlessness), which impacts between 10-15% of couples in Western countries. An estimated 48 million couples worldwide are infertile, with a large part of it, particularly in men, remaining unexplained.

[Birth postponement](#) has been [largely attributed to](#) social, economic and cultural environmental factors (i.e., individual and partner characteristics, socioeconomic status), with less attention paid to the genetic or biological underpinnings of this behavior. Later sexual initiation and childbirth is often tied to [higher educational goals and achievement](#), and focussing on longer-term life planning and career goals, with a strong relationship between [higher education and birth postponement](#), particularly for women.

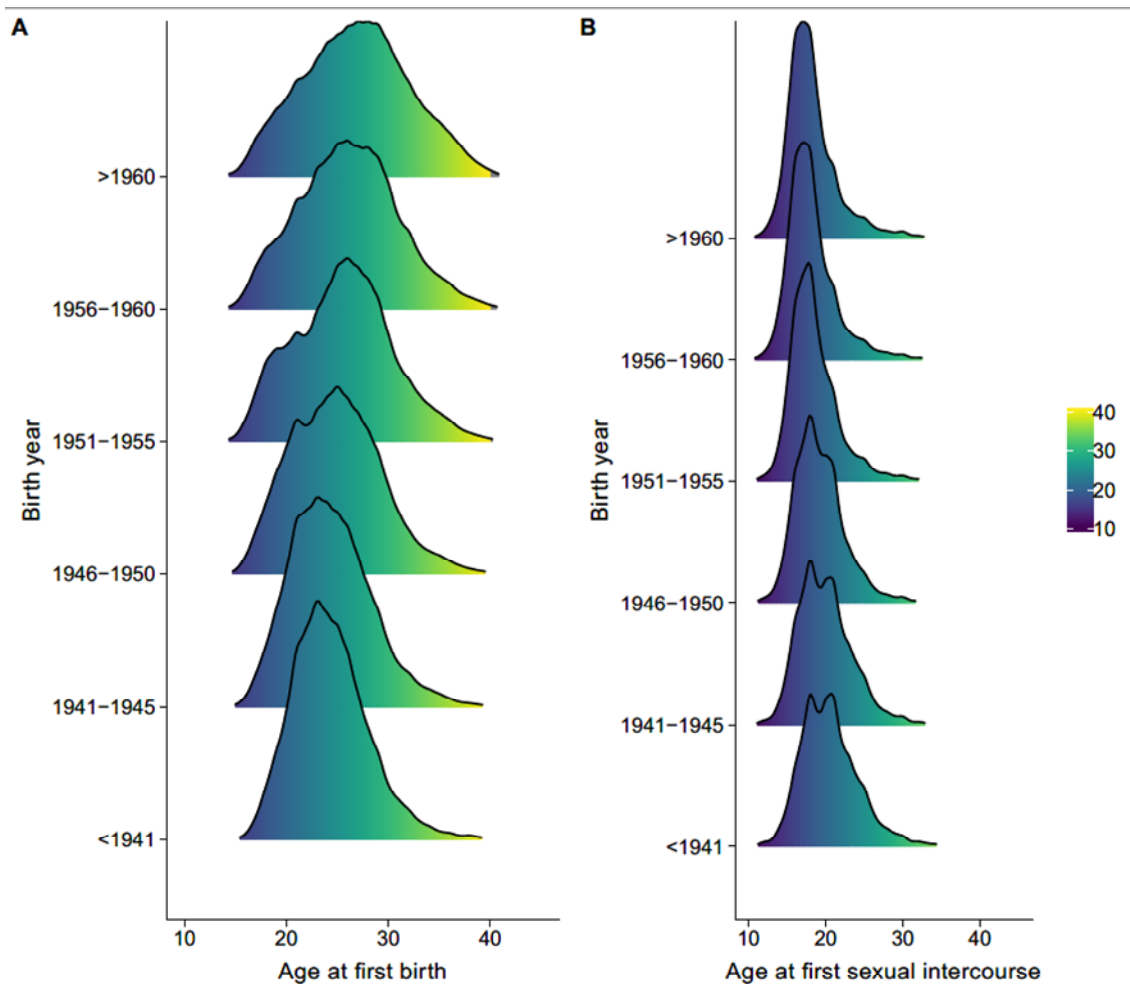


Figure S3. Age at first birth (AFB) panel A and Age at first sex (AFS) panel B by birth cohort, UK Biobank, Source: Supplementary Note, Mills et al. (2021) *Nature Human Behaviour*

HOW DID WE STUDY IT?

The primary analysis that we conducted is called a [Genome-Wide Association Study or GWAS](#) (*pronounced gee-was*), which is a search across the entire human genome, examining each genetic locus (or region) one by one to see if there is a relationship (or what we call an association) between our outcomes and a particular genetic variant. [Variants](#) refer to a specific region of the genome, which differs between two genomes. Different versions of the same variants are termed [alleles](#) and a SNP (single-nucleotide polymorphism) can have two alternative bases or alleles (C and T).

In other words, we study DNA variants that distinguish us from each other. Humans are 99.9% identical to each other, and it is the 0.1% by which we differ that makes us all genetically unique. A small subset of the 0.1% by which we differ genetically is anticipated to influence reproductive behavior.

A comprehensive interdisciplinary study such as this demanded multiple analytical approaches, namely:

- **Phenotypic and genotypic historical trend analysis:** phenotypic (trait) changes in age at sex, birth across birth cohort and genotypic changes, estimating heritability for sex and birth cohort
- **Polygenic score (PGS) construction and prediction:** using GWAS results we produced a variety of scores, tested out of sample prediction, looking for population stratification (using a method called LD score regression), survival models (to examine those who not experience event by time of survey), sensitivity of PGS to sex and childhood socioeconomic status predictions
- **Correlation, etiology, causality and prediction analysis:** Here we used five different types of methods:
 1. We were able to identify the shared genetic background between a number of traits (using a method called [LD score regression](#)) to examine genetic correlation between traits,
 2. We tried to explore the links between these phenotypes more thoroughly (using a method called [Genomic SEM](#))
 3. [Bi-directional Mendelian Randomization](#) allowed us to understand causal pathways of our phenotypes. We were particularly interested by the links with educational attainment, age at initiation of smoking and risk taking and whether the genetics variants had independent effects on later life diseases (type 2 diabetes, coronary artery disease). We also considered a model accounting for the effect of educational attainment and BMI.
 4. [Exploratory Factor Analysis](#) to breakdown the underlying components of reproductive behaviour into those related to externalizing or disinhibition versus reproductive biology components.
 5. Models to understand if there was a link to parental longevity (done using [survival models](#))
- **Biological annotation.** We also carried out a variety of biological analyses, namely:
 1. [DEPICT](#) for candidate gene identification and tissue enrichment.
 2. [CELLECT using RNAseq data from mouse brain](#) and [Tabula muris](#) to identify enriched cell types.
 3. [Phenolyzer](#) to prioritize candidates using prior knowledge of these phenotypes using machine learning on seed genes and predicted gene rankings.

4. [In silico sequencing](#) to identify non-synonymous variants and summary-based Mendelian Randomization (SMR) using [eQTL](#) data from [brain](#) and [whole blood](#) to examine if a transcript and phenotype are likely associated because of a shared causal variant.
5. [LD score bivariate regression](#) to identify evidence of sex-specific effects, followed by identifying sex-specific loci based on heterogeneity in effect by sex.
6. Integration of results across all gene prioritisation approaches and overlaying results with tissues in which encoded proteins are expressed allowed us to: a) identify key genes related to reproduction and externalizing behaviour as well as; b) prioritize genes in sex-specific loci.

WHO ARE WE AND WHO FUNDED THIS STUDY?

This transdisciplinary study united researchers from the social science disciplines of demography with researchers within genetic epidemiology, molecular genetics, bioinformatics and medical sciences. A study of a complex phenotype such as this is only possible with integration of experts across the social, biological, medical and statistical sciences.

This study combined data from 36 studies, including authors who collected and analysed the data, who are listed as authors as part of the Human Reproductive Behaviour Consortium. Many participated in our original [2016 study in Nature Genetics](#) and we are grateful for their continued contribution to this team science study.

In 2010, demographer Melinda Mills submitted (and was later awarded) a VIDI grant from the Dutch Science Foundation, proposing to conduct a GWAS on reproductive behaviour (timing of reproductive onset, number of children). Funding was extended by an ERC (European Research Council) Consolidator Grant to Mills in 2013 for the SOCIOGENOME (www.sociogenome.com) project at the University of Oxford and Nuffield College, extended in 2015 with a UK ESRC National Center for Research Methods Grant, Wellcome Trust ISSF Grant. The first GWAS results were published in [2016 in Nature Genetics](#). Work progressed further on different but related topics within Mills' ERC Advanced Grant CHRONO in 2019 and further support from the Leverhulme Trust to realize a large interdisciplinary demographic science centre at Oxford. Without this continued generous funding, this research would not have been possible. Further funding is acknowledged in the Supplementary Material.

Several years ago, the Oxford team put their [200 year rivalry of the boat race](#) competition aside and united with the Cambridge team of Felix Day, John Perry and Ken Ong, with expertise in reproductive phenotypes, teenage development and previous publication of a GWAS on [age at first sexual initiation](#), exploring similar topics of neurobehavioural determinants of reproductive onset and success.

ADDITIONAL GENERAL QUESTIONS

Are the genetic effects small or large?

Reproductive and sexual behavior is complex phenomenon that is not only genetically based, but is a complex interplay with other individual and socio-environmental contextual traits. Genetics is only one piece of this larger puzzle and in this study we only examine one type of genetic variant (SNPs) and consider only one of the many possible biological and genetic ways in which individuals may vary.

For many of these lifestyle or behavioural traits, genetics is just one piece of the puzzle that we continue to put together. This does not impact the biological importance of the findings, as our results have the potential to substantially improve our understanding of human biology. In the context of human disease for example, variants identified by GWAS for diabetes and cardiovascular diseases ‘tag’ genes that encode well known drug targets for the treatment of such diseases. This implies that a further understanding of the genes underlying the associations we identified for reproductive behavior may someday result in new strategies for infertility and assisted reproductive technology (ART) treatment.

Could genetic results alone be used at the individual level to predict when someone will first have sex or their first child?

Genetic scores alone are not useful to predict complex individual disease and behavioural outcomes. Since each individual SNP or genetic variant has such a small effect, prediction of using genetic results alone is not possible. Even if we combine the genetic variants together into our index of a ‘polygenic score’ using all approximately 9 million SNPs in our data, we predict 5-6% of the variance across individuals. When we conduct analyses on whole-genome data, we see that the ceiling of prediction is likely more 15-17%.

Even the ‘gold standard’ social science predictors when entered alone as a single variable in a regression equation also have low predictive power, generally under 10%. It is therefore unhelpfully reductive to only enter one single variable as a predictor without considering additional factors (what we term in multivariate regression).

Put another way, if we asked you to only use one social or lifestyle variable such as someone’s educational level or whether they smoked to predict when someone would have their first sexual encounter or age at first birth, you would find it dubious. (Or, at least you should). The same holds for genetic predictors. We discuss this in [detail elsewhere](#).

In reality, complex outcomes are a culmination of multiple factors such as genetics, parental background, lifestyle, level of education and national institutional configurations that constrain or enable behavior (e.g., in the case of reproduction it is childcare, work-life reconciliation). We have even shown that the explanation of genetics can [vary across country and when you were born](#).

Are there genetic differences between men and women in your study?

We the GWAS in men and women combined and separately, and uncovered 11 sex-specific loci (2 in women, 8 in men only (AFS); and 1 in women only (AFB)). Surprisingly less is known about men’s reproduction in comparison with the wide array of studies on women, so this study is rare in having such a large sample of men. Genes can have a different function depending on the cell type and tissues. Such genes may thus influence fertility in men and women via currently ill-understood and possibly different mechanisms. We also examined whether there were significant genetic correlations for AFS and AFB between men and women and found that most genetic effects on reproductive behavior resulting from common SNPs are shared by both sexes. However, the male-female correlations for AFB were 0.95, but lower at 0.79 for AFS, which we hope will be explored further.

How could fertility be genetically possible? Wouldn't those with fertility problems simply die out over time?

This is an age-old question and is known as Fisher's Fundamental Theorem of Natural Selection. This states that the genetic variance in fitness should be zero, how does our study fit?

A common question is how it could even be possible for those with infertility problems to pass it on to the next generation. Our traits have received less attention, likely due to a frequent erroneous interpretation of Fisher's Fundamental Theorem of Natural Selection. It has often been misinterpreted to mean that the additive genetic variance in fitness itself should always be close to zero. A close reading of the text shows that Fisher actually argued that fitness is moderately heritable in human populations. The misinterpretation of Fisher's theorem is likely repeated so often due to its intuitive appeal. It may seem that genes that reduce fitness should have been less frequently passed on, leading to the elimination of genetic variability in traits such as fertility. Nevertheless, we find that fitness traits such as AFB (or number of children ever born, which we also study in [2016 here](#) and [extend here](#)), have significant narrow-sense heritabilities – yet these traits are still not as heritable as morphological traits such as height.

Several reasons have been put forward to explain the persistent genetic variance in fertility. One argument is that new mutations suffice to restore any genetic variance lost to selection. Additional aspects to consider are sexual antagonistic genetic effects, non-additive genetic effects, environment and gene-environment interaction, which [we studied here](#). Sexual antagonism, which is the existence of opposite genotypic effects among sexes, has been often theorized as one of the possible explanations for genetic differences in fertility. In other words, particular genes might influence men and women differently and will therefore still be transmitted to the next generation. Genes that contribute to the fecundability of men may therefore be inherited via women's lineage and those for women via the men's lineage. Certain damaging genetic factors may also only become relevant with age (e.g, related to endometriosis), which arises now with very late fertility.

We have also shown that there have been recent shifts in the patterns of reproductive timing, and thus the environmental factors are likely different to what they were in the past. Given this changing environment the identified variants may have had different effects in the past, meaning that the selective pressures might have been different.

Is it nature or nurture?

The timing of sexual and reproductive behaviour is not nature versus nurture, but rather a combination of nature and nurture. Just as complex diseases such as obesity or diabetes are neither purely genetically or socially determined, the timing of reproductive onset are complex outcomes related not only to biological fecundity, but also have a highly behavioral component in that they are driven by personality, partners, and simultaneously shaped by the social, cultural, economic and historical environment. Genetic factors influence the first two factors of biological fecundity and choice, with the social and historical environment filtering the types of behavior that are possible (e.g., via contraceptive legislation and availability, social norms).

Are there societal or medical implications of this study?

In the longer term, this study offers a better understanding of the genetic architecture of human reproductive behavior and its relation to the environment. It likewise has the potential to enable the

discovery of predictors of infertility, which would in turn greatly improve family planning but also increase the effectiveness of costly and invasive ART treatments as well as allow couples to realize their fertility intentions. Showing that particularly the timing of AFB can be protective for later life disease could have ramifications for the study of many health outcomes, but our findings are especially the etiology of diseases related to the reproductive tract. Furthermore, it is important to understand whether and which proportion of these traits are driven by genetic, behavioral and environmental factors. Relatively little is known about the relationship between indicators of women's reproductive lifespan (menarche, menopause) and reproductive success – or in other words 'How late can you wait?' We anticipate that our study has identified and prioritized several candidates for numerous follow-up experimental studies. The fact that we also found a strong relationship between particularly age at first sex and externalizing factors suggests that continued research in this area is warranted to aid in teenage reproductive health.

What are the limitations of this study?

Although we open up new avenues of research, there are limitations and although not exhaustive or exclusive to this type of GWAS study, the central ones are listed here:

- Sample sizes for men are smaller, calling for need to collect fertility data from men
- Lack of availability of summary statistics for key infertility traits means we were unable to examine relationships with other traits (e.g., larger studies of endometriosis)
- Focus on European-ancestry individuals only, a problem we have [highlighted elsewhere](#)

After conducting a [scientometric review of all GWAS](#) and realizing that 72% of genetic discoveries come from 3 countries, we set up the [GWASDiversityMonitor](#) described in our [Nature Genetics](#) article. We have also considered [problematic biological race and genetic essentialism narratives](#).

Data availability

The summary statistics are available on the GWAS Catalog website:

<https://www.ebi.ac.uk/gwas/downloads/summary-statistics>

The phenotype and genotype data for separate studies used in this GWAS are available upon application to each of the participating cohorts who can be contacted directly to follow their different data access policies. Access to the UK Biobank is available through application with information available at: <http://www.ukbiobank.ac.uk>.

GLOSSARY

From Mills et al. (2020) [An Introduction to Statistical Genetic Data Analysis](#), Cambridge: MIT Press.

Phenotype or trait. The observable characteristic of an individual, ranging from physical traits (hair colour, height) to disease status (diabetic) to behaviour (risk-taker, age at first sexual intercourse, educational attainment).

Genotype. Describes part of an individual's DNA that influences their phenotype.

Genome-wide association study (GWAS). A GWAS is designed to adopt an unbiased hypothesis-free approach to discover genetic variants associated with a trait. They often combine data from multiple studies to gather the largest sample possible. An updated and searchable list of all GWAS discoveries to date can be found at www.gwasdiversity.com, linked to [this article](#), with summary statistics available at the [GWAS Catalog](#).

Single-nucleotide polymorphism (SNP). A variation in a single nucleotide (i.e., A, C, G, or T) that occurs at a specific position in the genome. A SNP exists as two different forms (e.g., A vs. T). These different forms are called alleles. A SNP with two alleles has three different genotypes (e.g., AA, AT, and TT).

Genetic variant. Refers to a specific region of the genome that differs between two genomes.

Heritability. A population measure defining the proportion of variance in a phenotype explained by genetic variance within a population. We can differentiate between broad-sense heritability, including both additive and non-additive genetic effects such as epistasis and dominance, and narrow-sense heritability focusing on additive genetic effects only.

SNP-heritability. The fraction of phenotypic variance of a trait explained by all SNPs in the analysis.

GWAS-heritability. The fraction of phenotypic variance of a trait explained by genome-wide significant genetic variants—sometimes also by polygenic scores based on GWAS findings.

Polygenic score. A single quantitative variable that summarizes genetic association to a phenotype by combining multiple genetic variants and their associated weights, derived from a GWAS.

Population stratification. The presence of multiple subpopulations (e.g., individuals with different ancestral background) in a study. Because allele frequencies can differ between subpopulations, population stratification can lead to false positive associations and/or mask true associations. An example is the chopstick gene, where a SNP, due to population stratification, would be wrongly assumed to be a true association due to differences in allele frequencies of those of Asian and European ancestry who have a different usage of chopsticks for purely cultural rather than biological reasons.

Externalising behaviour. A wide variety of acts that generally violate social norms. They include acts that might be victimless such as substance use (smoking, drug use), targeted at another individual (aggression). They frequently co-occur or are related to mental health or self-control (risk-taking, ADHD) and psychiatric traits (depression).