The genetic basis of age at first sex and first birth and the relationship with infertility, psychiatric disorders, risky behaviour, health and longevity

Melinda C. Mills* (melinda.mills@nuffield.ox.ac.uk), David Brazel (david.brazel@nuffield.ox.ac.uk), Nicola Barban (nicola.barban@essex.ox.ac.uk), Felix Tropf (fctropf@gmail.com)

*Corresponding author and presenter

Short Abstract

The onset of age at first sex (AFS) and first birth (AFB) is influenced by social and genetic factors. We exploit recent results of a genome-wide association study (GWAS) of AFS (N=387,338) and AFB (N=542,901), identifying 271 (AFS) and 88 (AFB) genetic loci that together accounting for up to 6% of the total explained variance. We find a strong genetic overlap of AFS and AFB with infertility traits (miscarriage, age at menarche and menopause), and early timing linked with psychiatric disorders (ADHD, major depression) and risky and additive behaviour (risk, smoking and cannabis use). We then move from correlation to causation using bi-directional Mendelian Randomization (IV) and Genomic structural equation modelling to unpack a complex network of mutual causation. Finally, we show that later AFS and AFB is a causal predictor of late-life health and longevity.

Extended Abstract:

The onset of human reproductive behaviour – age at first sexual intercourse (AFS) and age at first birth (AFB) – is tied to social, institutional, cultural and genetic factors (Kohler et al. 2006; Mills & Tropf 2020). Early sexual debut and reproductive onset has been linked to puberty and development, risky behaviour, impulse control, smoking and alcohol consumption, (Karlsson Linnér et al. 2019) with later age at first sex postponement of first child tied to educational attainment (Mills et al. 2011) and infertility related traits (Barban et al. 2016).

Evidence suggests that early sexual debut and teenage pregnancy is linked to risky and externalizing behaviour (Kahn & Anderson 1992; Skinner et al. 2015) and sexual maturation (Cousminer et al. 2014). Demographic research has demonstrated that the postponement of childbirth is driven by higher educational attainment and labour market opportunities of women, coupled with difficulties to conceive at older ages (Balbo et al. 2013). Although the demographic literature has largely focussed on the individual, social, institutional and cultural characteristics predicting age at first birth and age at first sex, several recent studies have uncovered a genetic component.

The current study is the largest genetic discovery (genome-wide association study, GWAS)[1] to date which uncovers hundreds of novel independent SNPs (271 for AFS and 88 for AFB), including some on the X chromosome. Few genetic studies have attempted to verify and understanding the causal genetic component of these underlying relationships and their predictive abilities.

DATA & METHODS

The current study greatly extends two previous studies of age at first sexual intercourse (*n*=125,667) that identified 38 single-nucleotide polymorphisms (SNPs) (Day et al. 2016) and a study of age at first birth (*n*=343,072), which identified 10 loci.(Barban et al. 2016) In this study we draw on results from our most recent GWAS of these traits by accumulating the largest sample size to date examining the genetic underpinning of AFS (*N*=387,338) and AFB (*N*=542,901) gathering 43 different datasets, including AddHealth, HRS, WLS, UKBiobank and other prominent datasets. Here we employed standard GWAS techniques (meta-analysis, COJO analyses, colocalization analyses).

In this EPC paper we explore in more detail three core methods and analyses that examine correlation and disentangle the causal relationships between AFS, AFB and related traits. These methods are: LD-score regression (Finucane et al. 2015), Genomic Structural Equation Modelling (SEM) (Grotzinger et al. 2019) and bi-directional Mendelian Randomization (Mills et al. 2020). We also plan to integrate linear models examining the predictive nature of our AFS and AFB polygenic scores on later-life disease and longevity.

RESULTS

Many genetic loci with predictive score of 4.8-5.8%. We uncover hundreds of novel independent SNPs (271 for AFS and 88 for AFB), including some on the X chromosome. From our GWAS we constructed a polygenic score which can account for up to 4.8% of the variance for age at first birth and 5.8% of the variance for age at first sex (Figure 1), when used for out-of-sample genomic prediction, and demonstrate the robustness of our results. In other words, these genetic variables now reach the explanatory power of even some common non-genetic predictors for AFB and AFS such as educational attainment, socioeconomic status or age at first partnership.

Genetic correlation between traits. Although novel loci have been uncovered, current research lacks a detailed examination of the genetic correlation underlying these traits. We first engaged in LD-score regression (Bulik-Sullivan et al. 2015), finding that these traits have a substantial genetic overlap with other reproductive, behavioural, psychiatric, addiction,

[1] GWAS is a hypothesis-free approach to discover genetic loci that are associated with a particular phenotype. They often combine data from multiple studies and engage in what is called a meta-analysis.

personality and anthropometric traits. The strongest associations were with reproductive traits, psychiatric disorders and addiction (Figure 2).

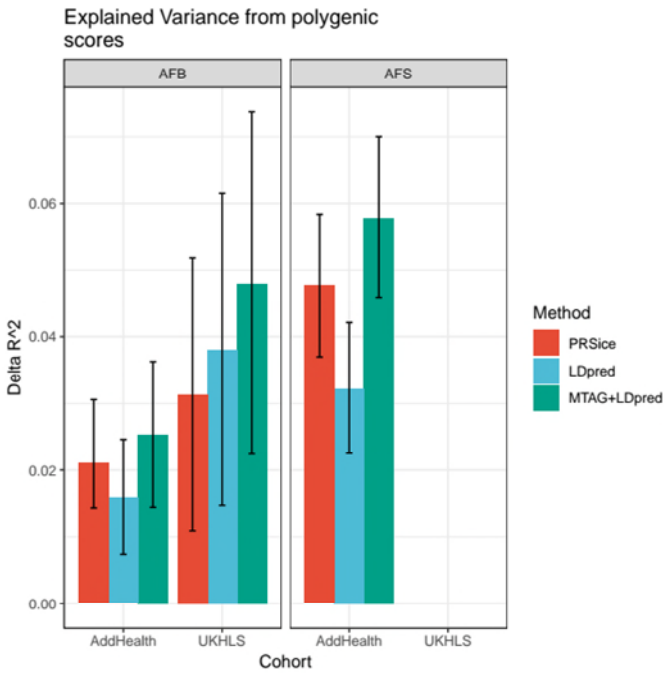

Figure 1. Out-of-sample predictive performance for the AFB and AFS polygenic scores, shown as the incremental increase in R^2 when the scores are added to a linear model which includes common covariates.

For reproductive traits, there was (logically) a negative genetic correlation with number of children ever born (NEB), but stronger for males (AFB males –0.87 ±0.09; AFB females, –0.67 ±0.02) and also with miscarriage or stillbirth (AFB females, –0.51 ±0.06; AFS females, –0.67 ±0.06). In other words later AFB and AFS are genetically correlated with a lower NEB and ever experiencing a miscarriage or stillbirth. There was also a striking correlation between the age at starting oral contraceptives (AFB females, 0.76 ±0.03; AFS females, 0.88 ±0.06), suggesting later sexual and reproductive onset was also linked with later age at using contraceptives. There was likewise a relationship with both age at menarche, menopause and voice breaking (around 0.12 to 0.27). Number of sexual partners was stronger for the related sexual behaviour trait of AFS (AFS males –0.57 ±0.02; AFS females, –0.59 ±0.02), than AFB (AFB males –0.25 ±0.06; AFB females, –0.25 ±0.02). The correlations also diverged with years of education in women, with a considerably stronger overlap of AFB (0.74 ±0.01) compared to AFS (0.53 ±0.01).

The genetic underpinnings of reproductive behaviour is also strongly associated with psychiatric disorders, including impulsive, risky behaviour. One of the strongest associations was ADHD (AFB females, –0.63 ±0.03; AFB males, – 0.68 ±0.09; AFS females, –0.58 ±0.03; AFS males, –0.61 ±0.03), but also Major Depressive Disorder (MDD) (AFB females, –0.42 ±0.03; AFB males, –0.33 ±0.08; AFS females, –0.37 ±0.03; AFS males, –0.32 ±0.03).

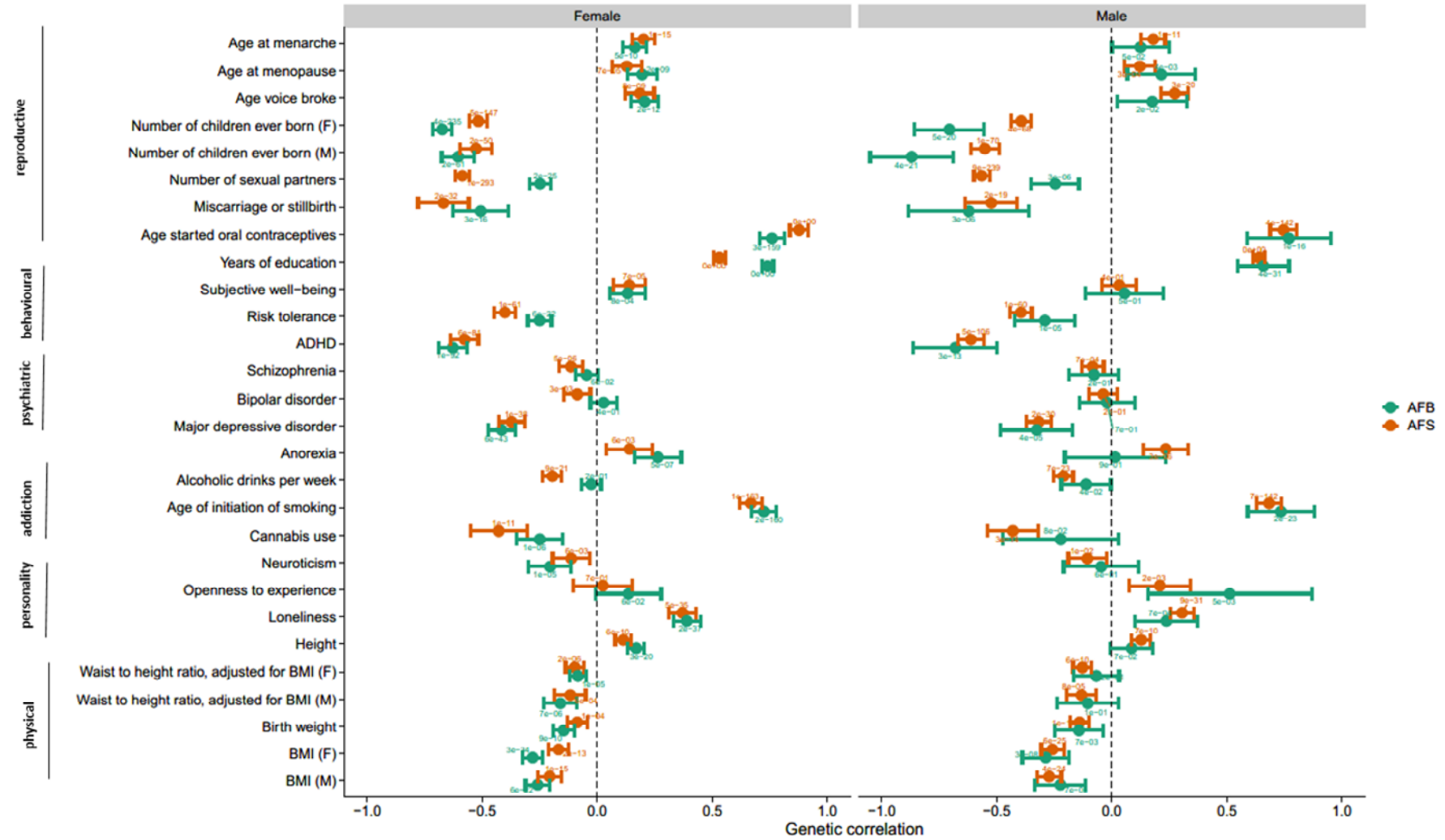

Figure 3. The sex-specific genetic correlations between AFB and AFS and a selection of related traits.

Addictive and substance use also had striking correlations, particularly with age at onset of smoking (AFB females, 0.73 ±0.03; AFB males, 0.74 ±0.07; AFS females, 0.67 ±0.03; AFS males, 0.68 ±0.03), which provides a window into adolescence and timing of early sexual behavior and pregnancies. Related to this is risk tolerance (AFB females, –0.25 ±0.03; AFB males, – 0.29 ±0.07; AFS females, –0.40 ±0.03; AFS males, –0.40 ±0.02) and cannabis use (AFB females, –0.25 ±0.05; AFB males, – 0.22 ±0.13; AFS females, –0.43 ±0.06; AFS males, – 0.43 ±0.06). Personality traits correlated with late AFS and AFB were loneliness (AFB females, 0.40 ±0.03; AFB males, 0.24 ±0.07; AFS females, 0.37 ±0.03; AFS males, 0.31 ±0.03) and to some extent neuroticism in females (AFB females, –0.21 ±0.05), which may be related to the ability to find a partner.

The causal relationship between reproductive, behavioural, psychiatric and addiction traits are likely both complex and bi-directional and it is highly likely that there are pleiotropic variants with biological pathways implicated across these traits. For this reason we explore issues of causality and prediction.

From correlation to causality. We first used a bi-directional Mendelian randomization to provide a causal measure of the associations between AFB, AFS, EA (educational attainment), adult risk tolerance, and age at initiation of smoking (AI) (and others highly correlated above). We find that AFB, AFS, and related traits form a complex causal web where each of the assessed traits appears to have an important impact on the others. We find that for some traits, such as ADHD and early AFS, the MR-Egger intercept estimate between ADHD and AFS is significantly different than zero. This statistic indicates that the signal of the causal relationship is likely due to pleiotropic effects. This is when one gene influences two or more phenotypes.

In this paper we will also report separate models we also engage in Genomic SEM which interrogates the underlying causality of correlations when fitting genetic multivariate regression models (Grotzinger et al. 2019). This attempts to untangle the genetic relationship between the aforementioned genetic correlations described previously such as AFB and AFS with ADHD, educational attainment, risky behaviour and personality traits.

Prediction of later-life health and longevity. In a second set of MR analyses, we examined whether AFB and AFS influence later-in-life disease, after controlling for their associations with EA. We find that AFB is in fact more predictive of coronary artery disease and type 2 diabetes than EA, and that AFB is the better index of a “common soil” of early-life health-related behaviours.

For the EPC presentation, using a linear regression model using the UK Biobank we will also show how the polygenic scores of AFS and AFB relate to reproductive, health and (parental) longevity traits (adjusted for controls). For a sensitivity analysis we also control for additional socioeconomic environmental variables such as educational attainment, and lifestyle factors. Preliminary results show that the polygenic scores of AFB and AFS predict

many of the psychiatric disorders, but also reproductive and health traits. Even after controlling for socioeconomic factors we find that the signals generally remain. Finally we find that the polygenic score for late AFB is particularly predictive of parent's longevity, suggesting that AFB is linked with senescence and has a protective element.

References

Balbo N, Billari FC, Mills M. 2013. Fertility in Advanced Societies: A Review of Research. *Eur. J. Popul. / Rev. Eur. Démographie*. 29(1):1–38

Barban N, Jansen R, De Vlaming R, Vaez A, Mandemakers JJ, et al. 2016. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* 48(12):1–7

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, et al. 2015. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*

Cousminer DL, Stergiakouli E, Berry DJ, Ang W, Groen-Blokhuis MM, et al. 2014. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum. Mol. Genet.* 23(16):4452–64

Day FR, Helgason H, Chasman DI, Rose LM, Loh P-R, et al. 2016. Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* doi:10.1038/ng.3551

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47(11):1228–35

Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, et al. 2019. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* 3(5):513–25

Kahn, J. R., & Anderson KE. 1992. Intergenerational patterns of teenage fertility. *Demography*. 29:39–57

Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, et al. 2019. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* 51(2):245–57

Kohler H-P, Rodgers JL, Miller WB, Skytthe A, Christensen K. 2006. Bio-social determinants of fertility. *Int. J. Androl.* 29(1):46–53

Mills MC, Barban N, Tropf FC. 2020. *An Introduction to Applied Statistical Genetic Analysis*. Cambridge, MA: The MIT Press

Mills MC, Rindfuss RR, McDonald P, te Velde E. 2011. Why do people postpone parenthood? Reasons and social policy incentives. *Hum. Reprod. Update*. 17(6):848–60

Mills MC, Tropf FC. 2020. Sociology, Genetics and the Coming of Age of Sociogenomics. *Annu. Rev. Sociol.*

Skinner SR, Robinson M, Smith MA, Robbins SCC, Mattes E, et al. 2015. Childhood Behavior Problems and Age at First Sexual Intercourse: A Prospective Birth Cohort Study. *Pediatrics*. 135(2):255–63