Supplementary Material for

Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior

Andrea Ganna, Karin J. H. Verweij, Michel G. Nivard, Robert Maier, Robbee Wedow, Alexander S. Busch, Abdel Abdellaoui, Shengru Guo, J. Fah Sathirapongsasuti, 23andMe Research Team, Paul Lichtenstein, Sebastian Lundström, Niklas Långström, Adam Auton, Kathleen Mullan Harris, Gary W. Beecham, Eden R. Martin, Alan R. Sanders, John R. B. Perry, Benjamin M. Neale, Brendan P. Zietsch*

*Corresponding author. Email: zietsch@psy.uq.edu.au

Published 30 August 2019, *Science* **365**, eaat7693 (2019)
DOI: 10.1126/science.aat7693

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/365/6456/eaat7693/suppl/DC1)

Table of contents

SUPPLEMENTARY MATERIALS AND METHODS

M1. Information about contributing cohorts

1.1 UK Biobank

The UK Biobank cohort is a population-based cohort of approximately 500,000 participants that were recruited in the United Kingdom between 2006 and 2010 (*49*). Invitations to participate were sent out to approximately 9.2 million individuals aged between 40 and 69 who lived within 25 miles of one of the 22 assessment centers in England, Wales, and Scotland. The participation rate for the baseline assessment was about 5.5%. From these participants, extensive questionnaire data, physical measurements, and biological samples were collected at one of the assessment centers. In this study we used data from 191,737 males and 223,014 females who answered the questions relating to sexual behavior and whose genotype data passed quality control.

Phenotypes

Same-sex sexual behavior (main phenotype) was determined with the question: `Have you ever had sexual intercourse with someone of the same-sex?`, with the note that 'sexual intercourse includes vaginal, oral or anal intercourse.' Cases were participants that answered 'yes', and controls participants that answered 'no'. Participants who reported to have never had a sexual relationship were set at missing. Data were available for 188,825 males and 220,170 females.

Proportion of same-sex to total number of sexual partners among non-heterosexuals was determined based on two open questions: 'About how many sexual partners have you had in your lifetime?' and 'How many sexual partners of the same sex have you had in your lifetime?' The proportion was determined by dividing the number of same-sex partners by the total number of sexual partners. Participants who reported to have never had a sexual relationship, who never had a same-sex partner, those with more than 100 sexual partners, or those with more same-sex than total number of sexual partners were set at missing.

Exclusively same-sex sexual behavior differentiates between participants that had exclusively had sex with same-sex sexual partners versus those that had exclusively had sex with opposite-sex sexual partners and was also determined with the question: 'Have you ever had sexual intercourse with someone of the same-sex?' Participants that responded affirmative to this question and for which their total number of same-sex sexual partners was equal or greater to their total number of sexual partners were considered exclusively homosexual,

whereas those who never had sex with a same-sex partner were considered heterosexual. Participants that reported to have never had a sexual relationship and those with both same-sex and opposite-sex sexual partners were set to missing. Data were available for 1766 homosexual and 180,431 heterosexual males and 693 homosexual and 214,062 heterosexual females.

Tables S1 and **S2** show the prevalence of same-sex sexual behavior in the UK Biobank.

Furthermore, for one sensitivity analysis (**Table S22**) we divided the sample in three groups based on the proportion of their sexual partners being of the same-sex: 1) up to one third same-sex partners, 2) between a third and two third same-sex partners, and 3) more than two third same-sex partners.

Number of children was determined with an open question; for males the question was: 'How many children have you fathered?' and for females the question was: 'How many children have you given birth to? (Please include live births only)'. Among females we considered only participants older than 45 years, among males older than 55 years.

Genotyping and imputation

We used genotype data from the May 2017 release of imputed genetic data from UK Biobank. The quality control and imputation were done by UK Biobank and have been described elsewhere (*49*). Briefly, genotyped variants were filtered based on batch effects, plate effects, departures from HWE, genotype platform, and discordance across control replicates. Participant samples were excluded based on missing rate, inconsistencies in reported versus genetic sex, and heterozygosity based on a set of 605,876 high quality autosomal markers. Imputation was performed using IMPUTE4 with the HRC UK10K and 1000 Genomes Phase 3 dataset used as the reference set. **Table S6** provides an overview of the quality control measures.

For the X-chromosome, HWE was calculated only in females, and the allele frequency was calculated as follows: 1) for variants not in the pseudo-autosomal region of X: (N homozygous in males + N heterozygous in females + 2*N homozygous in females) / (N genotype calls in males + 2*N genotype calls in females)), and 2) for variants in the pseudo-autosomal region of X: (N heterozygous in males + 2*N homozygous in males + N heterozygous in females + 2*N homozygous in females) / (2*N genotype calls in males + 2*N genotype calls in females).

Ethnicity definition in UK Biobank

We used K-means clustering (which aims to partition each individual into one of *k* clusters based on its distance to the cluster mean points) to identify four clusters on the first 4 PCs of the genetic data provided by UK Biobank. The first 4 PCs were chosen because the 5th PC shows substantial spread within the self-identified British population. The four clusters that results from this method were then visually inspected with regard to the self-identified ethnicity of individuals in each cluster. The clusters could loosely be termed: 1. 'African ancestry'; 2. 'Mixed race' (predominantly white-black); 3. 'Asian' (including self-identified Chinese and South Asian individuals); and 4. 'White'. Clusters 1, 2, and 3 were omitted for further analyses. The 4th cluster was the basis for identifying individuals as "White-European' in this analysis. The 'White' cluster completely contained the White British subset previously defined by UK Biobank. Individuals in this cluster who self-reported as different ancestry group were additionally dropped from the analyses set. This resulted in a set of individuals used in the analysis who were genetically determined to be white-European and who also self-identified as white.

1.2 23andMe

23andMe Inc. is a personal genetics company founded in 2006 which, as of June 2018, has now genotyped more than 5 million individuals worldwide. Data for this study were available for approximately 75,000 individuals of European ancestry who provided informed consent and answered surveys online according to a human subjects protocol approved by Ethical & Independent Review Services, a private institutional review board. The number of individuals varied per phenotype (see **Tables S3** and **S5)**. The mean age of the 23andMe sample in our study is 51.3 years (SD = 16.0, median = 51).

Phenotypes

A subset of the 23andMe participants completed a 'Sexual Orientation Survey' that included seven questions about sexuality, sexual behavior, and sexual identity:

1) Sexual Identity: How do you label, identify, or think of yourself?
2) Sexual Attraction: To whom are you sexually attracted?
3) Sexual Experience: With whom have you actually had sex?
4) Sexual Fantasies: Whom do you have sexual fantasies about?
5) Gender and Emotional Connection: Whom do you feel more drawn to or close to emotionally?
6) Gender and Socialization: Which gender do you socialize with?

7) Gender and Time Spent/Comfort: In which community do you like to spend your time? In which do you feel most comfortable?

Questions could be answered on a 7-point Likert scale, ranging from 'Other sex only' to 'Same-sex only' (questions 2-7) and from 'Heterosexual only' to 'Homosexual only' (question 1), or participants could choose the option 'I'd rather not say'. For all phenotypes, only cis-gender individuals were included. That is, participants who answered the question 'what gender do you consider yourself to be': 'Transgender female to male/Other/Transgender male to female' or 'I'd rather not say' were not included in the analyses.

The sexual experience question (item 3) was transformed in a dichotomous variable (to be consistent with the dichotomous variable from UK Biobank) and analyzed as the main phenotype, with participants who answered 0 (other sex only) being considered as heterosexuals and those who answered 1 (other sex mostly) to 6 (same-sex only) as non-heterosexuals. Additionally, we also analyzed this variable as a continuous trait, excluding individuals who reported exclusively having had sex with opposite-sex partners (answer 0); this variable was synonymous to the *proportion of same-sex to total number of sexual partners among non-heterosexuals* in the UK Biobank sample. See **Tables S3 and S5** for more information about the items and their genetic correlation with same-sex sexual behavior in the UK Biobank sample. Note that the 23andMe sample has very higher rates of same-sex sexual behavior (19.5% for males and 17.3% for females) and the distribution of same-sex to total number of sexual partners is different from that in UK Biobank. The high rate of same-sex sexual behavior is probably due to self-selection of participants to answering questions on sexual behavior; individuals who engage in same-sex sexual behavior may be more likely to self-select the 'Sexual Orientation Survey'.

Genotyping and imputation

Genotyping was performed on various genotyping platforms: Illumina HumanHap550+Beadchip (560,000 markers), V3 Illumina OmniExpress+Beadchip (950,000 markers) and V4 custom (570,000 markers). Standard quality control checks were performed prior to imputation: SNPs were excluded if Hardy-Weinberg Equilibrium (HWE) deviation $p < 1x10^{-20}$ and missingness > 5%. Only unrelated individuals of European ancestry were included in the analyses. Participants were excluded if more than 1.5% of their SNPs were missing. Genotype data were imputed with Minimac version 3, using the 1000 Genomes phase 1 release reference set. After imputation, SNPs with low imputation quality or a MAF below 0.01 were excluded. Quality control was performed for each analysis, and the number of individuals that passed quality control for each analysis are listed in **Tables S3** and **S5**. **Table S6** provides an overview of the quality control measures.

1.3 The National Longitudinal Study of Adolescent to Adult Health (Add Health)

Add Health originated as an in-school survey of a nationally representative sample of US adolescents enrolled in grades 7 through 12 during the 1994-1995 school year (*50*). Respondents were born between 1974 and 1983, and a subset of the original Add Health respondents has been followed up with in-home interviews, which allows researchers to assess correlates of outcomes in the transition to early adulthood. In Add Health, the mean birth year of respondents is 1979 (SD = 1.8), and the mean age at the time of assessment (Wave 4) is 29.0 years (SD = 1.8). All phenotypes included in this study come from Wave 4, the latest wave (2007-2009).

Phenotypes

The following two phenotypes were obtained from Add Health: (1) ever had same-sex intercourse and (2) same-sex attraction. The first phenotype is synonymous to the primary same-sex sexual behavior phenotype as defined in UK Biobank and is a binary indicator of whether a respondent has reported ever having had same-sex intercourse. Data were available for 2,529 females and 2,196 males. The second phenotype is a binary indicator of whether respondents reported being romantically attracted to members of the same-sex. Data for this phenotype were available for 2,539 females and 2,216 males.

Genotyping and imputation

Genotyping was performed at Expression Analysis Laboratory in Research Triangle Park, NC, using Illumina's Human Omni1-Quad-BeadChip (*51*). After imputing the genetic data to the Haplotype Reference Consortium (HRC) (*52*) using the Michigan Imputation Server (*53*), only HapMap3 variants were included, which are well imputed and provide good coverage of common variation across the genome. Analyses were limited to individuals of European-ancestry and cryptically related individuals and ancestry outliers were dropped from analyses. Finally, only HapMap3 variants with a call rate above 98% and a minor allele frequency > 1% were used. **Table S6** provides an overview of the quality control measures.

1.4 Molecular Genetic Study of Sexual Orientation (MGSOSO)

The MGSOSO dataset comprised four separate datasets: 1) 372 previously studied multiplex families (i.e. reported to have two or more self-identified homosexual brothers) containing 802 genotyped brothers, including 769 homosexual and 33 heterosexual brothers (*23*); 2) 234 genotyped brothers from 227 additional multiplex families, containing 221 homosexual and 13 heterosexual brothers, 3) 51 homosexual males without homosexual brothers (*23*), largely from community festival venues, and 4) the Molecular Genetics of Schizophrenia (MGS) collaboration controls dataset (the male, European-ancestry portion retained after QC), which included 36 homosexual and 1,185 heterosexual males (*54, 55*). The combined sample of European-ancestry participants included 1,077 self-identified homosexual and self-identified 1,231 heterosexual men. The participants were mostly from the US (99%) and were aged between 18 and 91 years (M=48.6, SD=14.5).

Phenotypes

For the participants from multiplex families, as well as the 51 sporadic participants, sexual orientation was based on their self-reported sexual identity and sexual feelings; for the MGS dataset, only sexual identity was available. Men with a Kinsey score of 5-6 were considered homosexual, those with a Kinsey score of 0-1 heterosexual, and males answering in the more intermediate (bisexual) ranges (Kinsey 2-4) were excluded. The utilized Kinsey question/responses for sexual feelings were: 'Which statement best describes your sexual feelings during the last year: sexual feelings only toward females (0), most sexual feelings toward females but an occasional fantasy about males (1), most feelings toward females but some definite fantasy about males (2), sexual feelings about equally divided between males and females - no strong preference for one or the other (3), most sexual feelings toward males but some definite sexual fantasy about females (4), most sexual feelings toward males but an occasional fantasy about a female (5), sexual feelings toward males only (6)'. The question used to assess identity was: "Do you consider yourself to be heterosexual, bisexual, or gay?'. Participants were categorized as homosexual or heterosexual based on the query regarding their sexual orientation (identity) (and sexual feelings if available), and bisexual males were excluded.

Genotyping and imputation

Samples from the MGSOSO participants were genotyped with the Affymetrix 6.0 SNP array at the Broad Institute, and the remaining samples were genotyped with the Affymetrix 5.0 array at Vanderbilt Microarray Shared Resource. As part of the QC design to help minimize errors due to platform-specific genotype calling differences, 34 participants were genotyped on both platforms, with removal of SNPs discrepant for any of the

34 inter-platform duplicates. Further details about the QC steps can be found elsewhere (*9*). Sample QC included removal of samples with missingness > 5%, failing checks for duplications and relatedness, and ancestry outliers via principal components analysis (PCA). SNP QC included removal of SNPs with minor allele frequency (MAF) < 0.05, missingness ≥ 1%, and HWE deviation of $p < 1x10^{-6}$. We imputed to 1000 Genomes (*56*) using the IMPUTE2 software (*57*) (removing SNPs with an information score <0.6 and a MAF < 0.05). The final QC'd SNP dataset contained 5,642,880 SNPs (200,367 genotyped and 5,442,513 imputed). **Table S6** provides an overview of the quality control measures.

1.5 The Child and Adolescent Twin Study in Sweden (CATSS)

The Child and Adolescent Twin Study in Sweden (CATSS) is an ongoing longitudinal twin study targeting all twins born in Sweden since July 1, 1992. Since 2004, parents of twins are interviewed regarding the children's somatic and mental health and social environment around the twins' 9th or 12th birthdays (CATSS-9/12). By January 2010, 8,610 parental interviews concerning 17,220 twins had been completed, with an overall response rate of 80%. At age 15 (CATSS-15) and 18 (CATSS-18), twins and parents completed questionnaires that, in addition to assessments of somatic and mental health, include measures of personality development and psychosocial adaptation. Individuals at age 15 and age 18 were asked questions about their sexual orientation. Not all twins participated at both ages.

Phenotypes

We considered two questions that were asked to the participants at age 15 and 18: a) What sex do the people have that you have usually felt sexually attracted to? and b) What sex do the people have that you voluntarily had sex with? The possible answers were the following: 1) Only girls/women, 2) Mostly girls/women, seldom boys/men, 3) Both, but girls/women more often than boys/men, 4) Girls/women and boys/men about equally often, 5) Both, but boys/men more often than girls/women, 6) Mostly boys/men, seldom girls/women, 7) Only boys/men. Males engaging in same-sex sexual behavior were defined as those that answered 2, 3, 4, 5, 6, or 7 either at age 15 or age 18. Controls were males that answered 1 at age 15 and age 18 (or answered 1 at age 15 or age 18 and did not participate at the other age). Females engaging in same-sex sexual behavior were defined as those that answered 1, 2, 3, 4, 5, or 6 either at age 15 or age 18. Controls were females that answered 7 at age 15 and age 18 (or answered 7 at age 15 or age 18 and did not participate at the other age). Same-sex sexual attraction was defined in a similar way. We also considered same-sex sexual behavior and attraction only using

the answers collected at age 15. In total, 8109 individuals answered questions about same-sex sexual behavior and attraction: 3503 only at age 15, 1712 only at age 18, and 2894 at both ages.

Genotyping and imputation

Genotyping was performed in 18 batches at SNP&SEQ Technologies in Uppsala, Sweden. The PsychArray includes 265,000 proven tag SNPs found on the Infinium Core-24 BeadChip, 245,000 markers from the Infinium Exome-24 BeadChip, and 50,000 additional markers associated with common psychiatric disorders. We excluded 3,827 markers with a call rate below 98%, 102 markers with over 10% discordant genotypes among 37 cross-batch duplicate samples, 323 markers with more than one discordant genotype among 84 pairs of MZ twins, 2,399 markers that failed HWE test (*p-value* for testing HWE < $1x10^{-6}$), 6 markers with large allele frequency differences from the 1000 Genome European reference samples (absolute difference > 10%) and with mean GenCall scores < 0.5, 35 common variants that were significantly associated with more than 1 genotyping batch (at p < 5e-8), and 1,332 markers on Y-chromosome or mitochondrial markers due to poor variant calling. We also excluded 4 participants with a call rate < 98%, 7 with unusual heterozygosity (autosomal inbreeding coefficient F > |0.2|), 14 with possible sample contamination as indicated in the excessive relatedness with other samples (> 6 standard deviations from the mean of average sample relatedness in a random set of 1000 samples), 22 samples with sex violation (male with X-chromosome F < 0.5 or female with X-chromosome F >= 0.5), and 248 participants identified as non-European ancestral outliers (> 6 standard deviations from the mean values of the first two principal components in 1000 Genome European populations). We also identified within-pair sample swap if both samples within any opposite-sex dizygotic (DZ) pair or parent pair failed the sex check; 6 out of 2255 (0.2%) opposite-sex DZ pairs and 2 out of 82 (2%) parent pairs were corrected. We performed genotype imputation on autosomes using 1000 Genome data (Phase 3 Version 5) as reference panel. More stringent marker QC was applied prior to genotype imputation, which dropped 97,854 monomorphic or singleton sites, 11,921 indels, 32,388 SNPs with strand-ambiguous alleles (A/T, C/G alleles), 7,543 markers with duplicate positions and 208 markers with alleles inconsistent with the ones in the reference panel. Based on the remaining 397,633 autosomal markers, phasing was performed using Shapeit2 on each chromosome, and imputation was performed using Minimac3 on 5Mb chromosomal chunks (with a window of 1Mb on either side). After imputation, ~47M markers were available, and over 7M common variants (MAF > 1%) have high imputation quality (imputation R^2 > 0.8). **Table S6** provides an overview of the quality control measures.

ANALYSES

Our analysis plan was preregistered at the Open Science Framework (https://osf.io/357tn/). The majority of our analyses followed the original plan, although we have now split the paper in two parts: a paper on the genetics of same-sex sexual behavior (this paper) and a paper focusing on the evolutionary basis of same-sex sexual behavior. This second paper also contains the analysis on 'number of lifetime opposite-sex partners among heterosexuals' which was proposed in the preregistered plan.

One analysis that was proposed in the preregistered plan was not included in either papers: we proposed to explore, as secondary outcome, the ratio between number of same-sex partners and overall lifetime sex partners. We decided not to include analyses based on this variable for three reasons. First, fewer individuals had reported their number of sexual partners than that answered the question: 'Have you ever had sex with someone of the same sex?'. Secondly, the ratio variable yielded very low heritability (the observed SNP based heritability was 0.01), which is difficult to interpret given the extreme skew of the variable. Third, the genetic correlation between the ratio variable and the primary dichotomous outcome was very high. In the full sample, the genetic correlation was 0.92 (SE = 0.05), in males it was 0.98 (SE = 0.06), and in females it was 0.79 (SE = 0.09), suggesting that we are detecting largely the same genetic signal with both variables. We did, though, add analyses based on this variable in non-heterosexuals only (i.e. proportion of same-sex to total number of sexual partners among non-heterosexuals); this was in response to reviewer suggestions to explore the complexity of same-sex behavior.

M2. Number of children analysis

We estimated the reproductive deficit of individuals that engaged in same-sex sexual behavior as well as that of several traits that have been linked to lower fertility for comparison. We determined fertility of the following traits obtained from UK Biobank (along with the UK Biobank variable code):

- Same-sex sexual behavior (as defined in **section M1**).
- Anorexia: variable 1470; ICD codes F50.0; self-reported diagnosis of anorexia/bulimia/other eating disorder.
- Autism: ICD codes F84.0 or F84.1 or F84.5 or F84.9. No self-reported information were available.

- Bipolar: variables 1192 and 20122; ICD codes F30 or F31; self-reported diagnosis of mania/bipolar disorder/manic depression/and derived bipolar disorder status from questionnaire.
- Depression: variables 1286, 20124, 20125, 20123; ICD codes F32 or F33; self-reported diagnosis of depression, and derived probable recurrent major depression (moderate or severe), or single episode of probable major depression.
- Polycystic ovary syndrome: variable 1350; ICD codes E28.2; self-reported diagnosis of polycystic ovaries/polycystic ovarian syndrome. This trait was female-specific.
- Schizophrenia: variable 1289; ICD codes F20 or F231 or F232 or F25; self-reported diagnosis of schizophrenia.

Number of children was self-reported in UK Biobank (for males we used variable 2405 and for females variable 2734, see **M1.1**). Among females we considered only participants older than 45 years, among males older than 55 years. We estimated the fertility ratio using a Poisson regression model where the outcome was the number of children and the predictors were the different traits defined above as well as year of birth, to adjust for temporal trends in overall fecundity.

M3. Genetic Association analyses

3.1 Genome-wide association analyses (GWASs)

UK Biobank

For the main association analyses (GWAS of same-sex sexual behavior) we used BOLT-LMM (*58*), adjusting for sex, year of birth, year of birth squared, 10 genetic principal components, genetic relatedness and batch number. Association testing was performed using linear mixed models implemented in BOLT-LMM to account for cryptic population structure and relatedness. Only autosomal genetic variants which were common (MAF > 1%), passed QC in all 106 batches and were present on both genotyping arrays were included in the genetic relationship matrix (GRM). The GWASs were also done separately by sex. We also ran a GWAS for exclusively same-sex sexual behavior, and proportion of same-sex to total number of sexual partners among non-heterosexuals.

We used the FUMA pipeline (*59*) to identify independent loci. In particular, we used pre-calculated LD (linkage disequilibrium) structure based on the European 1000 Genome panel to identify genome-wide

significant SNPs independent from each other at $r^2 < 0.6$. Based on the identified independent significant SNPs, independent lead SNPs are defined if they are independent from each other at $r^2 < 0.1$. Additionally, if LD blocks of independent significant SNPs are closely located to each other (< 250 kb based on the most right and left SNPs from each LD block), they are merged into one genomic locus. Each genomic locus can thus contain multiple independent significant SNPs and lead SNPs.

23andMe

We tested for association between the sexual experiences phenotypes and all SNPs that passed quality control by running linear or logistic regression models in a custom, in-house pipeline. The variable was analyzed in two ways: 1) the main phenotype (same-sex sexual behavior, yes versus no) was analyzed as a dichotomous trait, 2) the proportion of same-sex to total number of sexual partners among non-heterosexuals was analyzed as a continuous trait. Associations were corrected for the effects of sex, age, 5 genetic principal components, and genotype platform. The GWASs were also done separately by sex. We also ran GWASs for the other phenotypes related to sexuality, sexual behavior, and sexual identity (see **M1.2**).

3.2 Multi Trait Analysis of GWAS (MTAG)

To combine the association statistics from UK Biobank and 23andMe, we ran Multi Trait Analysis of GWAS (MTAG) (*17*), a method for the joint analysis of summary statistics from GWASs of different traits. This was done both for the results of the GWASs for same-sex sexual behavior and for proportion of same-sex to total number of sexual partners among heterosexuals. Since the genetic correlation between these two traits was in the order 0.5-0.8, MTAG can provide additional power by relaxing the assumption that the genetic correlation equals 1, as is done in traditional inverse-variance-weighted meta-analyses. MTAG returns two association meta-analyses results, one for same-sex sexual behavior from UK Biobank (enriched by the results from 23andMe) and one for the results from 23andMe (enriched by the results from UK Biobank). The set of results for each trait represents an optimal combination of the information from the single trait summary statistics. Here, we focused on the UK Biobank phenotype (same-sex sexual behavior) because that was our primary sample, as described in our pre-registered study plan. Independent genome-wide significant loci identified for same-sex sexual behavior can be found in **Table S7**. We applied the conventional *p-value* threshold of $5x10^{-8}$ as indication of genome-wide significance.

To mimic what is done in an inverse-variance-weighted meta-analysis we also ran MTAG with a genetic correlation fixed to 1. Findings between the different approaches (i.e. r_g=estimated versus r_g=1) were compared.

This approach resulted in three out of the five loci that were genome-significant with the standard MTAG approach to drop below the conventional *p-value* threshold of $5x10^{-8}$ (**Table S8**).

3.2.1 Transformation of association coefficients to odds ratios

The association results from UK Biobank and 23andMe are based on analyses using BOLT-LMM, which does not account for dichotomous outcomes and does not use a *logit* link function. This complicates the interpretation of the coefficients, which should be interpreted as betas similar to those obtained from linear regression, and not as *log(odds ratios)*. When meta-analyzing the summary statistics from UK Biobank and 23andMe in MTAG, we report the output for the same-sex sexual behavior variable in UK Biobank as enriched by the results from 23andMe. The coefficients obtained from MTAG are on the standardized scale. Therefore, we transformed these coefficients to make them comparable with the observations in the UK Biobank and 23andMe samples; we rescaled the beta with the following formula: $\beta_{SCALED} = k(1-k)\,\beta_{MTAG}$, where *k* is the prevalence of non-heterosexuals (individuals that have had same-sex sexual partners) in UK Biobank and 23andMe.

Next, we transformed β_{SCALED} into odds ratios, which allows us to interpret the magnitude of the associations. To do that we use the approach proposed in Lloyd-Hones et al. (*60*).

$$OR = \frac{[k + \beta_{SCALED}(1-p)][1-k+\beta_{SCALED}\,p]}{[k-\beta_{SCALED}\,p]\,[1-k-\beta_{SCALED}(1-p)]}$$

Where *k* is the prevalence of non-heterosexuals in the UK Biobank and/or 23andMe and *p* is the allele frequency for the risk allele used for calculating β_{SCALED}. This equation approximates the OR expected when running a logistic regression on the same phenotype.

3.3 Test for sex differences in SNP effects

We designed a statistical test to evaluate whether there are sex differences in the association between genome-wide significant SNPs for same-sex sexual behavior. Given the sample size N_1 for females and N_2 for males, the Z-statistics z_1 and z_2 from the GWAS for females and males, and *cti,* the intercept from the LD-score genetic correlation between sexes, we can obtain Z-statistics for the difference between males and females reweighted by the corresponding sample size to allow for difference in scales between the two sexes.

$$Z_{F\ vs\ M} = \frac{\frac{1}{\sqrt{N1}}z_1 - \frac{1}{\sqrt{N2}}z_2}{\sqrt{\frac{1}{N1} + \frac{1}{N2} - 2\sqrt{\frac{1}{N1}\frac{1}{N2}}cti}}$$

We considered SNPs as 'sex-differentiated' when the *p-value* obtained from $Z_{F\ vs\ M}$r for that locus was lower than 0.01 (0.05/5 genome-wide significant loci).

3.4 Replication analyses

We tried to replicate SNPs that were genome-wide significantly ($p < 5x10^{-8}$) associated with same-sex sexual behavior in the MTAG analyses (either in the full sample or in the sex-specific analyses) in three independent replication samples: Add Health, MGSOSO, and CATSS (**Table S10**).

In Add Health, the association analysis was conducted using logistic regression in unrelated individuals with the statistical association software Rvtests (*61*). Covariates used in the association analysis included age, sex, age-by-sex interaction, and the first 10 principle components of the variance-covariance matrix of the genetic data.

In MGSOSO, the association analysis was conducted using logistic regression correcting for family relatedness with the R package Genome-Wide Association analyses with Family (GWAF) (*62*). The first two genetic PCs were included as covariates.

In CATSS, the association analysis was conducted using BOLT-LMM to take into account family relatedness in the dataset, in particular monozygotic and dizygotic twins. Sex, and the 10 first principal components were included as covariates in the model.

We meta-analyzed the results from these three studies using the following approach:

$$Z_{meta} = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i}}$$

where Z_i are the Z-scores of the SNP effect for each study and w_i the square roots of the effective sample size for each study, which is calculated as: $\frac{4}{\frac{1}{ncases}+\frac{1}{ncontrols}}$.

Significance of the SNP effect in the replication samples is based on a Wald test of the parameter as obtained from the (logistic) regression model.

3.5 Gene-based test of association

Gene-based analysis was done using MAGMA (*29*) as implemented in FUMA (*59*) using summary statistics from the MTAG association analysis of same-sex sexual behavior (sexes combined). In gene-based analyses the combined effect of SNPs in protein-coding genes are analyzed, taking into account LD between the SNPs and the size of the gene. MAGMA uses the 1000 Genomes reference-panel (phase 3, 2012) to control for LD. SNPs were mapped to genes if they were located in or within 10 kb from the gene, such that SNPs could be mapped to at least one of 17,715 protein-coding genes in the reference panel. The gene-based analysis is based on a multiple linear principal components regression model, using an F-test to compute the p-value. The significance threshold was set at a Bonferroni corrected $p < 0.05$ ($0.05/17715 = 2.8 \times 10^{-6}$). We used the default MAGMA parameters for this analysis.

M4. Enrichment analysis of evolutionary constrained genes

We tested if there was signal enrichment in 2,929 genes that are under evolutionary constraint. This previously defined set of genes includes genes that are depleted of loss-of-function mutations as compared to what expected under a neutral model and are therefore most likely to contribute to early-onset diseases. The set of genes is defined as having a probability of being loss-of-function intolerant greater than 0.9 (*63*). Individuals with autism, schizophrenia, ADHD, intellectual disability, and lower educational attainment have an increased rate of rare loss-of-function variants in these genes (*64*). In addition, signal enrichment from common SNPs associated with schizophrenia (*65*) and ADHD (*66*) has also been observed in proximity of evolutionarily constrained genes. We evaluated common-variant signal enrichment for same-sex sexual behavior in evolutionarily constrained genes using two approaches: MAGMA (*29*) and LD-score regression (*45, see section M6 for more details on LD-score regression*). For MAGMA we used the standard configuration, i.e. the baseline model adjusting for gene size, log(gene size), gene density, log(gene density), inverse minor allele count (MAC), and log(inverse MAC). In LD-score regression we constructed a new annotation by considering all SNPs within +/- 100kb from the transcription start/end site of the 2,929 genes that are under evolutionary constraint. The baseline model of LD-score regression adjusts for 53 different annotation categories described in Finucane et al. (*25*). In addition to this baseline adjustment, we evaluated if the signal enrichment was independent from brain-tissue specific genes. We therefore conditioned to 12 different annotations derived from the Genotype-Tissue Expression (GTEx) database: Anterior cingulate cortex (BA24), Caudate (basal ganglia),

Cerebellar Hemisphere, Cerebellum, Cortex, Brain Frontal Cortex (BA9), Hippocampus, Hypothalamus, Nucleus accumbens (basal ganglia), Putamen (basal ganglia), Brain Spinal cord (cervical c-1), and Substantia nigra. The method used to derive brain-tissue specific genes is described in (*67*). Briefly, for each gene, we did a t-test to determine whether gene expression differed between brain-tissue versus all non-brain tissue. The top 10% of the genes ranked by t-statistics were considered as brain-tissue specific genes. MAGMA and LD-score regression returned comparable results as shown in **Table S15**.

M5. In-silico follow-up of GWAS results for same-sex sexual behavior

We used a comprehensive approach to try to pinpoint genes that might be tagged by the genome-wide significant SNPs for same-sex sexual behavior. In particular, we combined the PheWAS results with information from eQTL and gene-based analysis using MAGMA (as described in **section M3**).

5.1 Phenome-Wide Association Study (PheWAS)

To examine whether the SNPs we identified for same-sex sexual behavior are also associated with other phenotypes, we conducted a phenome-wide association study (PheWAS; *28*) for the five genome-wide significant SNPs for same-sex sexual behavior (as obtained from the UK Biobank and 23andMe MTAG analysis). We scanned 1352 heritable traits collected in the UK Biobank (http://www.nealelab.is/uk-biobank) and traits from 3675 publications from the GWAS catalog (https://www.ebi.ac.uk/gwas/) for associations with these five SNPs. For the UK Biobank traits we tested both the direction and significance of the SNP associations. The 1352 heritable UK Biobank traits were selected from the full set of 2745 traits as those with a SNP-based heritability (obtained from LD-score regression) *p-value* < 0.05. We reported all associations with *p-value* < 0.05, after correcting for multiple testing by applying Benjamini & Hochberg FDR correction (*68*) to the reported *p-values* across SNPs and all PheWAS phenotypes.

From the GWAS catalog (*69*) we extracted results all the SNPs in LD ($r^2 > 0.6$; (our five SNPs themselves were not represented in the GWAS catalog) with the genome-wide significant SNPs for same-sex sexual behavior. We used the GWAS catalog version e91 from 15-12-2018. We included only results from the GWAS catalog with a *p-value* < 5×10^{-8}. Results can be found in **Table S16**.

5.2 eQTL mapping

eQTL (Expression quantitative trait loci) can help to link non-coding variants (the majority of the genome-wide significant results) with those genes of which their expression is likely to be influenced by these SNPs. eQTL information were obtained by considering all tissues included in GTEx v7 (*47*). Only eQTLs with an FDR-corrected *p-value* < 0.05 were considered. Results of the in-silico follow-up for same-sex sexual behavior (as well as for number of opposite-sex sexual partners in heterosexuals) can be found in **Table S17**.

M6. SNP-based heritability

We used LD-score regression to estimate the SNP-based heritability of the traits and to estimate the genetic correlation between traits.

6.1 SNP-based heritability using LD-score regression

We used LD-score regression (*45*) to estimate the proportion of variance in liability to same-sex sexual behavior that could be explained by the aggregated effect of the SNPs (h^2_{SNPs}). The method is based on the idea that an estimated SNP effect includes effects of all SNPs in LD with that SNP. On average, a SNP that tags many other SNPs will have a higher probability of tagging a causal variant than a SNP that tags few other SNPs. Accordingly, for polygenic traits, SNPs with a higher LD-score have on average stronger effect sizes than SNPs with lower LD-scores. When regressing the effect size obtained from the GWAS against the LD-score for each SNP, the slope of the regression line gives an estimate of the proportion of variance accounted for by all analyzed SNPs. We included 1,217,312 SNPs (those available in the HapMap 3 reference panel). Standard LD-scores were used based on the Hapmap 3 reference panel, restricted to European populations (*45*). Analyses were also done separately for each chromosome. To minimize bias in heritability estimation, we also performed LD-score regression after conditioning on MAF and LD structure.

6.1.1 Accurate estimation of heritability on the liability scale

The UK Biobank is a random, but not necessarily representative, sample of the UK population. It is possible that same-sex sexual behavior in the UK Biobank sample has a different prevalence from that of the UK population as a whole. But even when the exact prevalence of a trait is known, or the sample prevalence is equal the population prevalence, the observed heritability should be transformed to the liability scale. In the liability

threshold model it is assumed that an observed dichotomous phenotype is the product of a latent liability, and that this latent liability has a standard normal distribution. In our example, individuals who exceed a threshold on this liability scale are considered non-heterosexuals while those with a liability score below this threshold are considered heterosexuals. The transformation was derived by Lee et al. (*46*):

$$h^2{}_{liability} = h^2{}_{observed} \frac{K(1-K)}{\varphi(\phi^{-1}[K])^2}$$

Where K is the population prevalence, and the denominator is the squared height of the density of the standard normal distribution (mean = 0, variance = 1) at quantile K. If the sample prevalence P deviates from the population prevalence K the approximation equals:

$$h^2{}_{liability} = h^2{}_{observed} \frac{K(1-K)^2}{P(1-P)\ \ \varphi(\phi^{-1}[K])^2}$$

The sample prevalence can strongly deviate from the population prevalence for several reasons, such as non-random participation, study ascertainment criteria, or non-random dropout. In that case the approximation above is sufficient to correct for a prevalence difference between K and P. However, if the observed prevalence deviates from the population prevalence because participants hide the fact that they have had same-sex intercourse, and falsely answer the question negatively, they in fact change the composition of the control population. Given that same-sex sexual behavior is sometimes stigmatized, it is likely that some participants who have had same-sex partners did not disclose this. Peyrot et al. (*70*) derived an approximation of the heritability on a liability scale in the presence of a subgroup (F) that include false negatives (in this case individuals that disclose they have never had sex with someone from the same sex, when they actually have):

$$h^2{}_{liability} = h^2{}_{observed} \frac{K(1-K)^2}{P(1-P)\ (1-F)^2\ \ \varphi(\phi^{-1}[K])^2}$$

This formula enables us to approximate the heritability of same-sex sexual behavior on the liability scale. To enable transformation to the liability scale we require an educated guess of the population prevalence of same-sex sexual behavior. Here we rely on the national study of sexual attitudes and lifestyles (NATSAL) (*71*). The 2013 wave of the NATSAL data collection estimates the prevalence of same-sex sexual experiences with genital contact at 7.2% between age 45 and 54, 7.3% between age 55 and 64 and at 3.4% between age 65 and 74. The prevalence of non-heterosexual males in the UK Biobank sample equals 4.1%, while the weighted

average percentage in NATSAL at age 45+ for males is 6.3%. For females, the prevalences in NATSAL are 6.6%, 3.5%, and 0.8% for age groups 45-54, 55-64, and 65-74, respectively. The weighted average prevalence in NATSAL above age 45 in females equals 4.1%, while the prevalence in UK Biobank equals 2.8%.

Based on these prevalences, and assuming 50% of the difference between the NATSAL sample and UK Biobank is selection (i.e. those who have had same-sex partners participate less in UK Biobank) and 50% is misreporting (the participant falsely reports as never having had homosexual sex) the heritability of same-sex sexual behavior on the liability scale in males equals 9.6% and in females 11.0%. If we take the prevalence in the UK Biobank sample at face value and assume all responses were truthful we obtain liability-scale heritability estimates of 8-9% (see **Table S11**), so quite similar to the estimates above. Higher values (~17% for males and ~21% for females; **Table S11**) are obtained when the heritability is estimated directly from the genotype data using Bolt-LMM. Bolt-LMM estimates the variance parameter σ_g^2 so that the covariance of the genetic effect on the phenotype is $Cov(g) = \sigma_g^2 K$ where K is the empirical kinship matrix.

23andMe has much higher rates of same-sex sexual behavior (19.5% for males and 17.3% for females). This high rate of having had same-sex partners is probably due to self-selection of participants to answering questions on sexual orientation. Assuming a population prevalence of same-sex sexual behavior in the USA of 3.9% for males and 5.1% for females (*72*), the heritability of same-sex sexual behavior on the liability scale in males equals 14.3% and in females 9.4%. These estimates are somewhat higher than the estimates in the observed scale (~11% in males and 8% in females), but lower than when we take the prevalence in 23andMe at face value (~23% in males and 16% in females) (see **Table S11**).

6.2 Per-chromosome SNP-based heritability

We used BOLT-LMM to estimate the per-chromosome heritability (h^2_{Ol}) and the corresponding standard error: $se(h^2_{Ol})$ for $l = 1 \ldots 23$ chromosomes, where 23 is the X-chromosome. We wanted to test if the observed h^2_l was significantly different from what is expected given the chromosomal length (c_l). First, we regressed $h^2_l = c_l + \varepsilon$. We sampled (with replacement) the data 10,000 times, and each time we fitted the linear regression and predicted the expected heritability h^2_{El} for each chromosome as the average across the 10,000 bootstrapped samples. We also obtained the empirical standard error $se(h^2_{El})$. We then compared the expected with the observed heritability for each chromosome as:

$$Z = \frac{h^2_{Ol} - h^2_{El}}{\sqrt{se(h^2_{Ol})^2 + se(h^2_{El})^2}}.$$

6.3. Heritability partitioning per tissue

To examine whether the genetic variants that play a role in same-sex sexual behavior co-localize in the proximity of genes that are expressed in certain tissues, we used LD-score regression to partition the heritability by functional annotation in 53 different tissues. We used expression data from the GTEx databases, and an approach described by Finucane et al. *(67)*. For each gene we did a t-test to determine whether gene expression differed between that specific tissue versus all other tissues. The analysis were adjusted for 53 different annotation categories described in Finucane et al. *(25)*.

M7. Cross-trait LD-score regression to estimate genetic correlations between traits

We used cross-trait LD-score regression to estimate the genetic covariation between traits using GWAS summary statistics (*16*). The genetic covariance is estimated using the slope from the regression of the product of z-scores from two GWAS studies on the LD-score. The estimate obtained from this method represents the genetic correlation between the two traits based on all polygenic effects captured by SNPs. Standard LD-scores were used as provided by Bulik-Sullivan et al. (*16*) based on the 1000 genomes reference set, restricted to European populations.

Cross-trait LD-score regression was used to estimate the genetic correlation between: 1) males and females, 2) UK Biobank and 23andMe, and 3) same-sex sexual behavior with other relevant traits – i.e. traits that may capture other aspects of sexuality, sexual behavior, and sexual identity, sex-hormonal and reproductive traits, sexually dimorphic traits, and mental health traits. For 3) we used summary statistics from well-powered GWASs for relevant traits, including number of children, age at first birth, various psychiatric disorders, substance use traits, neuroticism, waist-to-hip ratio, risk taking behavior, and sexual/physical developmental traits. As we estimated genetic correlations with 28 traits (separately for males and females), we adopted an alpha level of 8.9×10^{-4}.

Genetic correlations require careful interpretation, especially when dealing with sensitive links such as those between same-sex sexual behavior and psychiatric disorders. Statistically, a positive genetic correlation indicates that variants that increase one trait also tend to increase the second. However, various different (including non-genetic) processes can cause a genetic correlation. For example, it could be that pleiotropic variants have multiple biological effects, which affect both traits (e.g. if the antagonistically linked sex hormone and stress hormone systems are involved in development of same-sex sexual behavior and psychiatric disorders,

respectively). Alternatively, it could be that the first trait causes the second, possibly mediated by environmental influences. In this case, any variants that affect the first trait will also affect the second, causing genetic correlation even though the cause of the correlation may be entirely environmental. In this case, individuals' same-sex sexual behavior could expose them to prejudice and discrimination which in turn increases risk of psychiatric disorder, generating a genetic correlation despite the environmental cause. These and other possibilities are discussed extensively elsewhere (*73*).

7.1 Comparing genetic correlations with other traits between sexes, UK Biobank and 23andMe, and between same-sex sexual behavior and proportion of same-sex to total number of sexual partners among non-heterosexuals

We devised an approach to compare the genetic correlations of same-sex sexual behavior with the other traits between males and females, between UK Biobank and 23andMe, as well as between the two measures of same-sex sexual behavior (i.e. same-sex sexual behavior versus proportion of same-sex to total number of sexual partners among non-heterosexuals). To test whether the genetic correlations were significantly different from each other we adapted the following approach (with the correlations of 23andMe versus UK Biobank used as example): The standard error for the genetic correlation between same-sex sexual behavior in UK Biobank and a trait *i* ($rg_{UKB,i}$) is related to the standard error of the genetic correlation between same-sex sexual behavior in 23andMe and a trait *i* ($rg_{23andme,i}$), because the measurement error contributed by same-sex sexual behavior is shared between the two correlations coefficients. Therefore, to provide unbiased standard errors, we compute the standard error of the difference between $rg_{UKB,i}$ and $rg_{23andme,i}$ by obtaining 200 blocked jackknife estimates of the genetic correlations obtained using the function *--print-delete-vals* in LD-score regression. For example, for trait *i*, we obtained the standard error of the difference in genetic correlations between same-sex sexual behavior and trait *i* in 23andMe *vs* UK Biobank as following:

$$SE_{rg_{23andMe}\ vs\ rg_{UKB},i} = \sqrt{\frac{200-1}{200}\sum_{k=1}^{200}[(rg_{23andMe,i,k} - rg_{UKB,i,k}) - E(rg_{23andMe,i} - rg_{UKB,i})]^2}$$

Using a Wald test we tested whether the genetic correlations were significantly different between sexes, between samples (i.e. UK Biobank versus 23andMe), and the different measures of same-sex sexual behavior, after correcting for multiple testing (number of traits tested). These results are presented in **Fig. S2** (for differences between studies), **Table S19** (for differences between sexes) and **Table S21** (for differences

between same-sex sexual behavior and proportion of same-sex to total sexual partners among non-heterosexuals).

M8. Comparison of family versus SNP based heritability estimates

8.1 Family-based heritability analysis

We estimated the heritability of same-sex sexual behavior based on known familial relationships in the UK Biobank. The relatedness between pairs of participants is estimated using KING (*44*) and made available by UK Biobank. Based on the fraction of markers for which pairs share zero alleles (Identity by state [IBS]=0) and kinship coefficients, participants were separated into parent-offspring pairs (6273 pairs), sibling pairs (22,666 pairs) 2nd degree related relationships (11,112 pairs; e.g. half-sibs, uncle-cousin) and 3rd degree relationships (66,928 pairs of full-cousins). We only included participants of European ancestry.

We modelled the dichotomous same-sex sexual behavior variable as a function of a normally distributed latent liability. The value of person i on the latent liability equals:

$$y_i = a * A_i + c * C_i + e * E_i$$

Where a, c, and e are regression parameters and A, C, and E are latent variables representing additive genetic (A), shared environmental or family environmental (C), and unshared environmental (E) effects. The fixed effects of sex and year of birth on the liability score were also modelled, but omitted here for clarity.

Assuming unit variance of, and an absence of correlation between, A, C, and E, the total variance of the latent liability (V) is equal to:

$$V(y) = a^2 + c^2 + e^2$$

We can estimate the regression parameters (e.g. a) or variance components (e.g. a^2) directly by specifying the covariance between pairs or relatives (*74*).

$$cov_{parent-offspring} = .5 * a^2 + c^2$$
$$cov_{sibling} = .5 * a^2 + c^2$$

$$cov_{2nd\ degree} = .25 * a^2 + c^2$$
$$cov_{3th\ degree} = .125 * a^2 + c^2$$

We note the following limitation: some participants have more than one relative in the data, which influences inference; to correctly infer the variance of the parameters, we used bootstrapping (by resampling pairs).

Running this model on a number of different traits in the UK Biobank showed that in some cases the model converged at unrealistic estimates, likely due to misspecification. Given that heritability modelling can be mis-specified in idiosyncratic ways for individual traits, we consider a variety of traits to give a sense of the overall pattern of family- and SNP-based heritabilities.

8.1.1 Definition of control phenotypes

To put the family-based and SNP-based heritability estimates of our phenotypes in context, we also estimated both for a diverse set of other phenotypes in the UK Biobank dataset. The phenotypes were selected because they have been analyzed by large GWASs and because they covered different phenotypic categories (e.g. common diseases, behavior phenotypes, reproductive traits).

The following phenotypes were selected (variable numbers and more information about the variables can be found in the UK Biobank data showcase (http://biobank.ctsu.ox.ac.uk/crystal/):

- Age at menarche: variable 2714 (Age when periods started - menarche), only including females and responses lower than 8 and higher than 18 years old were set at missing.
- Age at menopause: variable 3581 (Age at menopause - last menstrual period), only including females, and only females who responded to variable 2724 (Had menopause) that they have had the menopause (response=1).
- Number of children: for females we used variable 2734 (Number of live births), for males we used 2405 (Number of children fathered). We excluded values lower than 0 for both variables and we included only females aged > 45 years old and males aged > 55 at recruitment.
- Standing height: variable 50 (Standing height).
- Body mass index (BMI): variable 21001 (Body mass index - BMI).
- Risk taking: variable 2040 (risk taking) only if values were > 0.
- Neuroticism score: variable 20127 (Neuroticism score).
- Alcohol intake frequency: variable 1558 (Alcohol intake frequency) if values were > 0.

- Fluid intelligence score: variable 20016 (Fluid intelligence score).
- Birth weight: variable 20022 (Birth weight).
- Handedness: variable 1707 (Handedness - chirality/laterality), we only considered left vs right and excluded individuals reporting other options.
- Overall health rating: variable 2178 (Overall health rating) if values were > 0.
- Ever smoked: variable 20160 (Ever smoked).
- Diabetes: we used the definition described by Eastwood et al. (*75*).
- Myocardial infarction: variable 42000 (Date of first myocardial infarction), individuals reporting the date of myocardial infarction were considered as cases, all others as controls.
- Educational attainment: variable 6138 (Qualifications); we defined as cases those individuals that went to college or have a university degree and controls all the others. We considered only values > 0.

Results can be found in **Tables S4** and **S23** and **Fig. 3**.

M9. Polygenic prediction in Add Health, MGSOSO and CATSS

We performed polygenic score analyses to test whether we could use the SNP effects from our GWASs to predict same-sex sexual behavior in three independent replication samples. Furthermore, we used the polygenic scores to predict number of opposite-sex partners in heterosexuals in the Add Health sample. Based on the results from our GWASs we generated polygenic scores for same-sex sexual behavior in the Add Health, MGSOSO, and CATSS samples. Polygenic scores in the target sample were generated by calculating the mean causal effect size of each marker using the SNP effect sizes from our GWAS meta-analysis. Specifically, an individual's polygenic score is a weighted sum of their genotypes at *J* loci, where $\hat{g}_i$ denotes the polygenic score of individual *i*, $\beta_j$is the estimated additive effect size of the effect-coded allele at variant *j*, and g_{ij} is the genotype of individual *i* at variant *j* (coded as having 0, 1, or 2 instances of the effect-coded allele):

$$\hat{g}_i = \sum_{j=1}^{J} \hat{\beta}_j g_{ij}$$

The polygenic scores were constructed in LDpred (*24*), a method shown to have greater prediction accuracy than the conventional risk prediction approach involving LD pruning followed by *p-value* thresholding. LDpred takes into account the genetic architecture by accounting for LD among the SNPs in creating the polygenic scores. We used a Wald test to evaluate the significance of the polygenic scores on the outcomes.

For the Add Health sample, we used the genotyped data from the Add Health prediction cohort to create the LD reference file. After imputing the genetic data to the Haplotype Reference Consortium (HRC) using the Michigan Imputation Server (*53*), we used only HapMap3 variants with a call rate > 98% and a minor allele frequency > 1% to construct the polygenic scores. We limited the analyses to European-ancestry individuals. Polygenic scores were calculated with an expected fractions of causal genetic markers set at 100%. In total, we used 1,177,001 HapMap3 variants to construct the polygenic scores in Add Health. We then used Plink (*76*) to multiply the genotype probability of each variant by the corresponding LDpred posterior mean over all variants. In total, we created three polygenic risk scores, using the summary statistics of: 1) same-sex sexual behavior (males and females combined), 2) same-sex sexual behavior (females), and 3) same-sex sexual behavior (males). We then determined the association of the polygenic score for same-sex sexual behavior with ever had a same-sex partner and same-sex attraction. Prediction accuracy was based on an ordinary least squares regression of the outcome phenotype on the polygenic score and a set of standard controls, which include birth year, sex, an interaction between birth year and sex, and the first 10 genetic principle components. Variance explained by the polygenic risk scores was calculated in regression analyses as the R^2 change (or Nagelkerke's pseudo-R^2 change for the dichotomous variables), i.e. the R^2 of the model including polygenic risk scores and covariates minus the R^2 of the model including only covariates. 95% confidence intervals around all R^2 values are bootstrapped with 1000 repetitions each.

In the MGSOSO sample, polygenic scores for same-sex sexual behavior and number of sexual partners in heterosexuals were created in LDpred in the same way as for the Add Health sample, using summary statistics from 2,882,852 SNPs that overlapped between the UK Biobank and MGSOSO datasets. The association between derived polygenic scores and homosexuality in the MGSOSO dataset was tested using generalized estimating equations (GEE) with a block diagonal working correlation matrix (with kinship coefficients) to adjust for family relationships; analyses were performed in PROC GENMOD in SAS 9.3 (SAS Institute, Cary, NC). The top 10 genetic PCs, age, and age-squared were included as covariates.

In CATSS we used the same approach as described for Add Health. In total, 1,009,809 SNPs were used to construct the polygenic scores in CATSS. The association between derived polygenic scores and same-sex sexual behavior and attraction in the CATSS dataset, both at age 15 and 18, was tested using logistic regression

with robust standard errors to account for the familial relationship in the data (monozygotic and dizygotic twins). The top 10 genetic PCs, age and sex were included as covariates. Results can be found in **Tables S12-S14**.

Supplementary Figures S1-S7

Figure S1 Deficit of number of children (as a ratio to that of general population) among UK Biobank participants who have engaged in same-sex sexual behavior compared with those who have not, in comparison to the corresponding deficit among UK Biobank participants with disorders that have been linked with reduced fertility. Estimates are obtained from a Poisson model adjusted for year of birth. Notice that UK Biobank participants with severe disorders are less likely to participate to the study, so this represents an underestimation of the fertility deficit among individuals with psychiatric disorders. See (*77*) for nation-wide representative results. Bars represent 95% confidence intervals. PCOS = polycystic ovary syndrome.

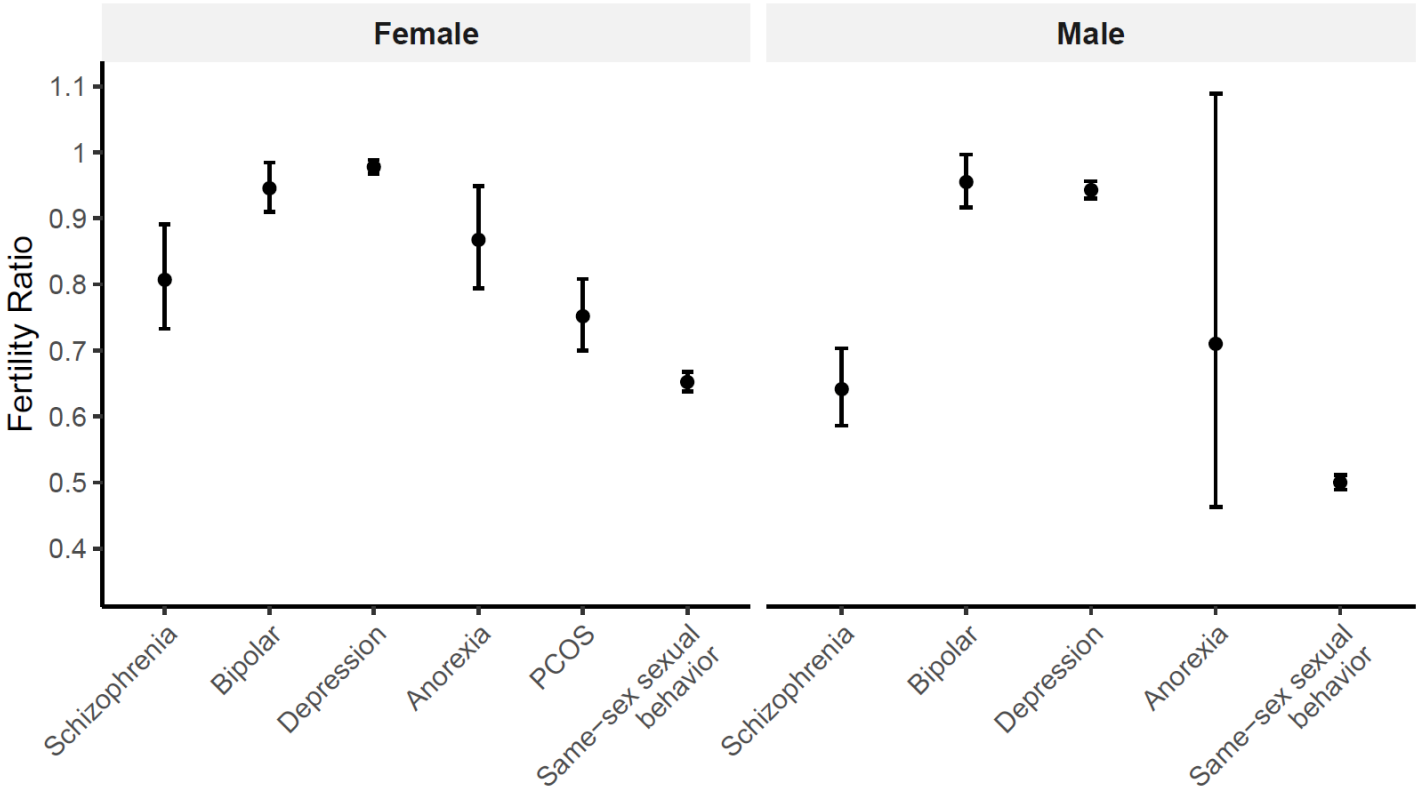


Figure S2. Genetic correlation between same-sex sexual behavior and several traits of interest, separately for UK Biobank and 23andMe. For each trait, we tested if the correlation coefficients (reported above each dot) were significantly different in UK Biobank versus 23andMe; the *p-value* for this test is noted between brackets above the UK Biobank point estimate. Bars represent 95% confidence intervals. The red asterisks indicate traits for which the genetic correlations are significantly different between 23andMe and UK Biobank after adjusting for the number of traits tested.

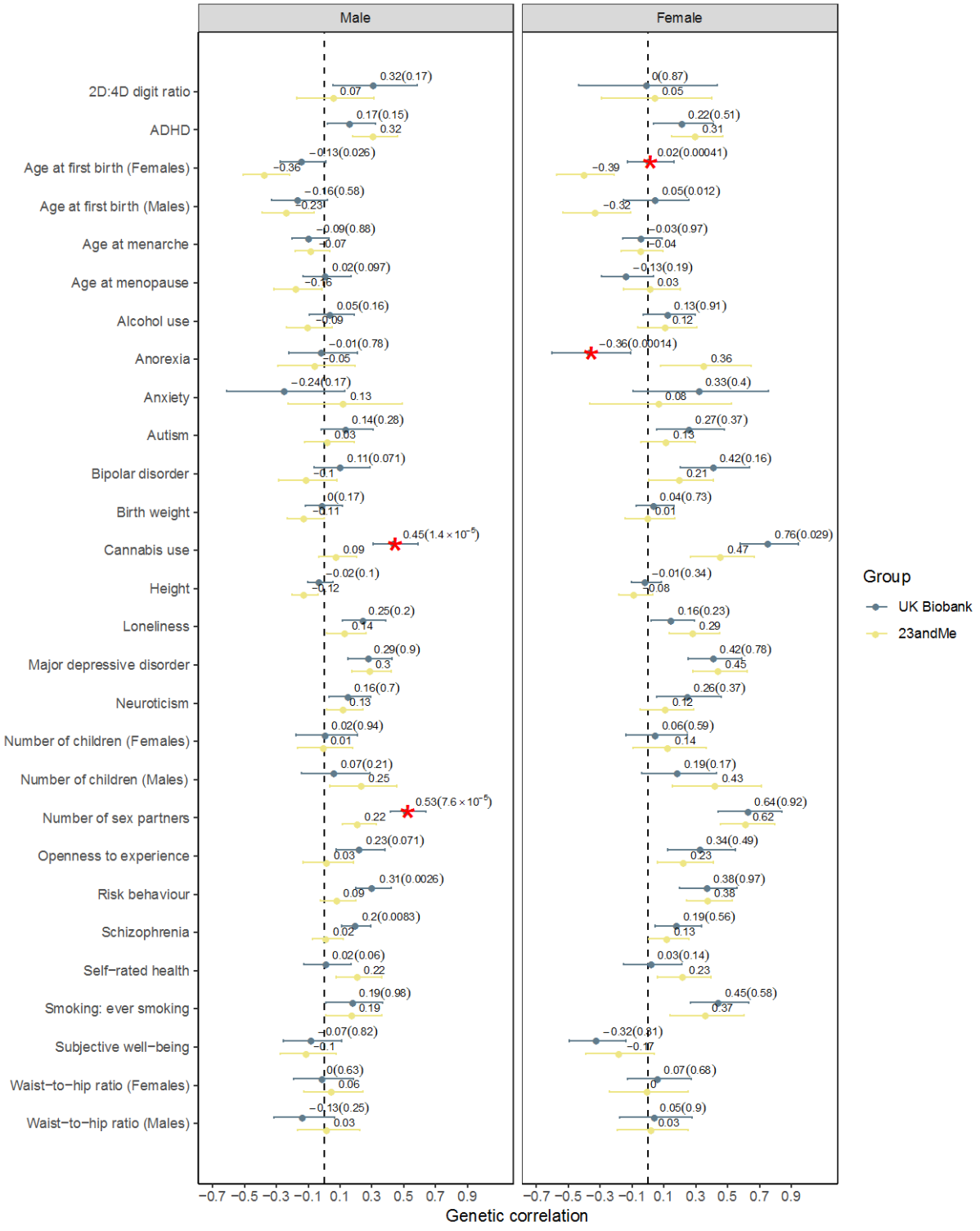


Figure S3. SNP-based heritability estimates (*p-value*) on the observed scale for each chromosome, ordered by the chromosome length as percentage of total genome. The dotted line represent the linear regression line fitted using the plotted values. See Supplementary Material and Methods M6.2 for an explanation on how the per-chromosome heritability was obtained and how we tested whether the heritability was significantly different from what is expected based on chromosome length.

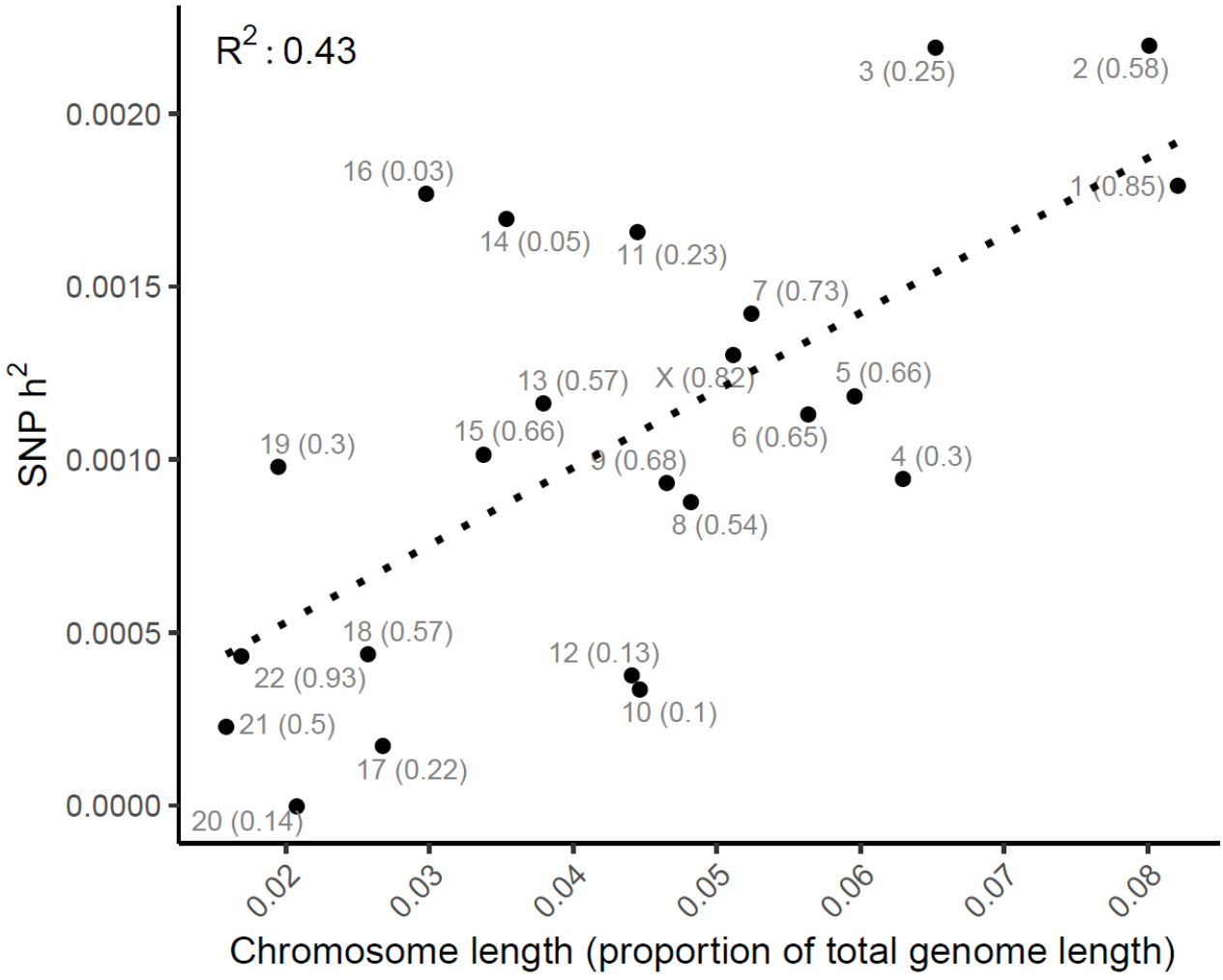


Figure S4. Q-Q plots for the genetic associations results for same-sex sexual behavior for the X chromosome. Results were reported separately for the male-specific (A), female-specific (B), and sexes-combined analyses (C). No genome-wide significant SNPs were identified.

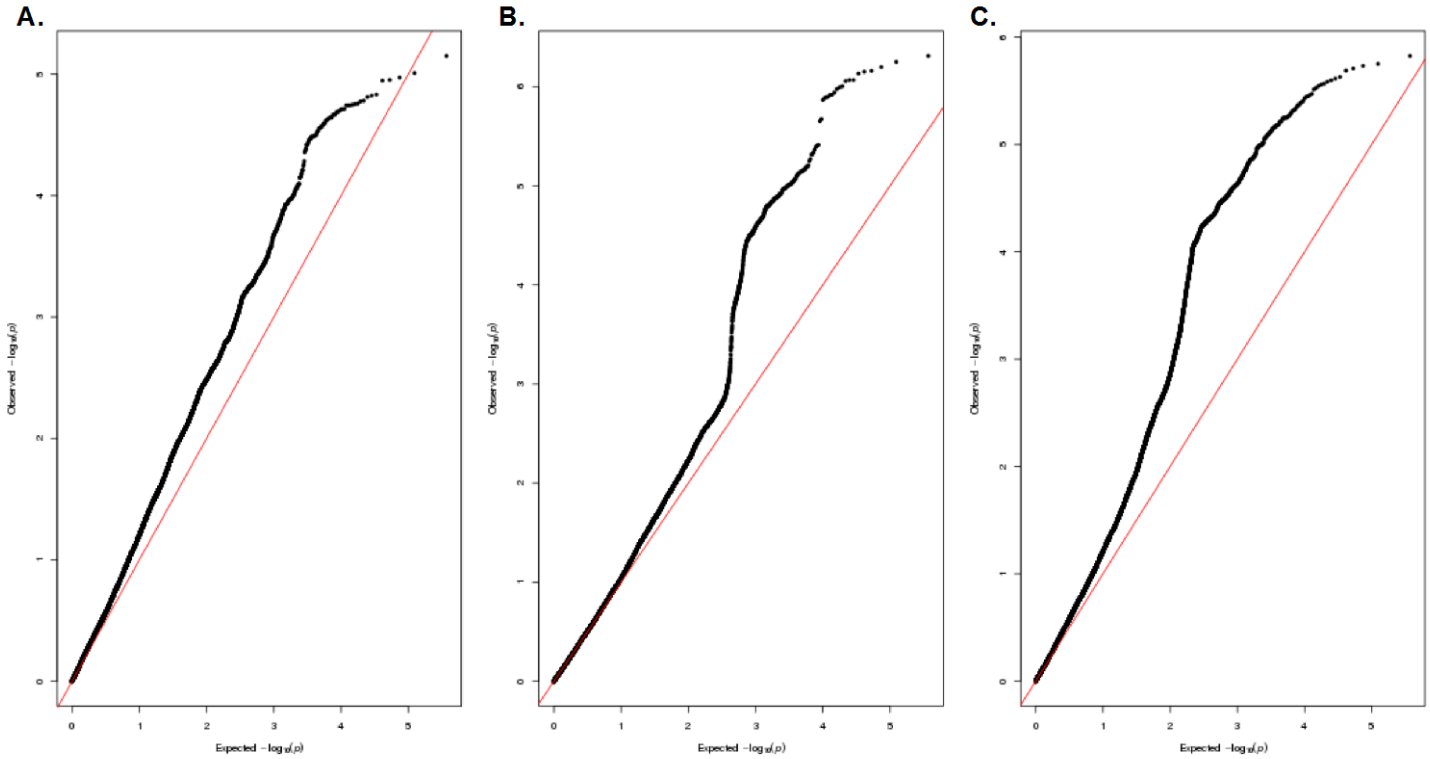


Figure S5. Results of the LD-score based tissue enrichment analysis. The x-axis represents the -log10(*p-value*) of the heritability enrichment for each tissue (y-axis), using the GWAS results for same-sex sexual behavior.

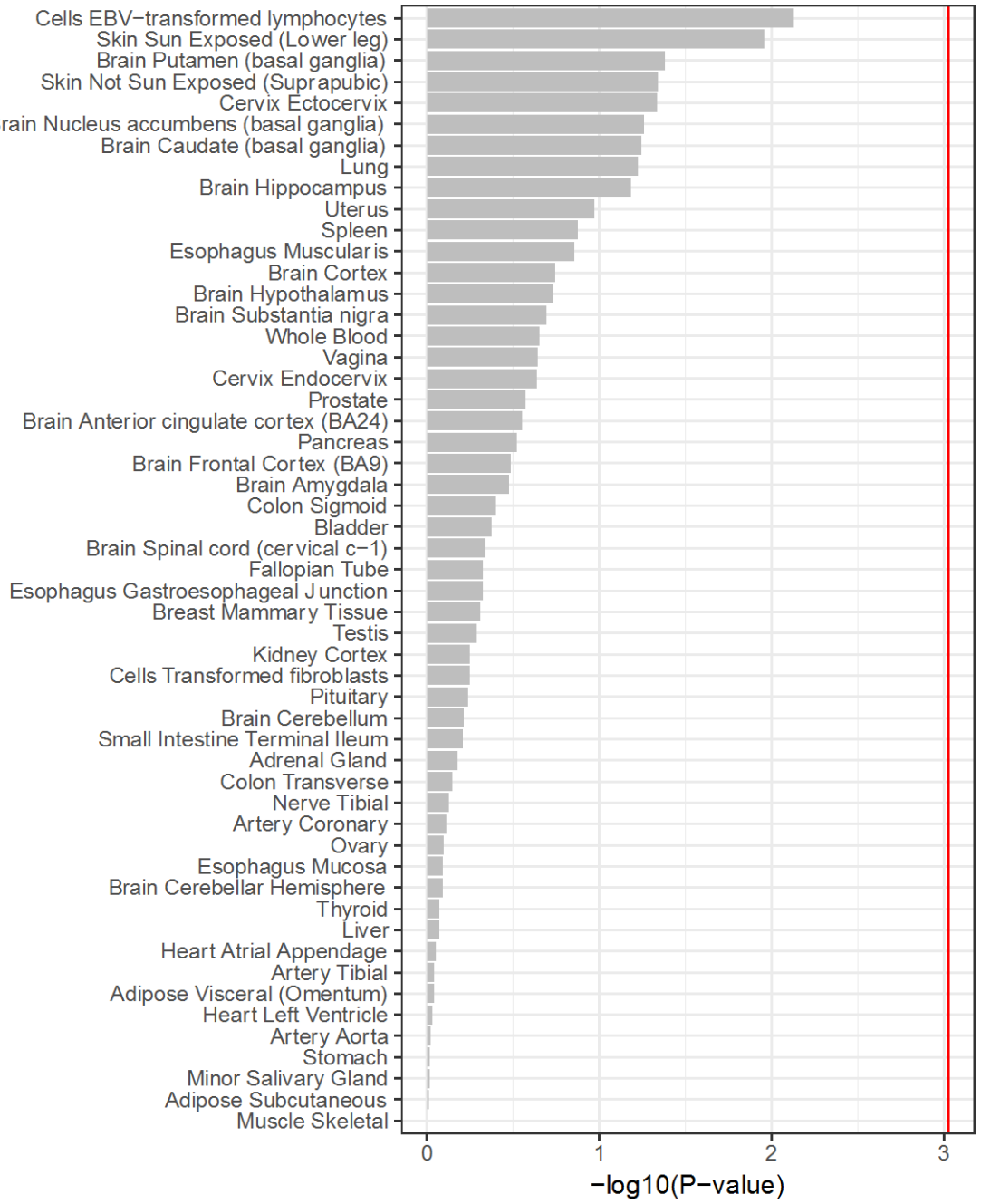


Figure S6: Locus-zoom plots for genome-wide significant loci for same-sex sexual behavior. Each dot is a SNP and the dot's color represents the degree of linkage disequilibrium (LD) with the top lead SNP (in blue). Genes spanning the locus are reported below the x-axis. Panel A: locus 12:81989337-82068452, panel B: locus 7:114940147-115314917, panel C: locus 4:36963942-37032454, panel D: locus 11:59040414-59233752, panel E: locus 15:56999901-57583301.

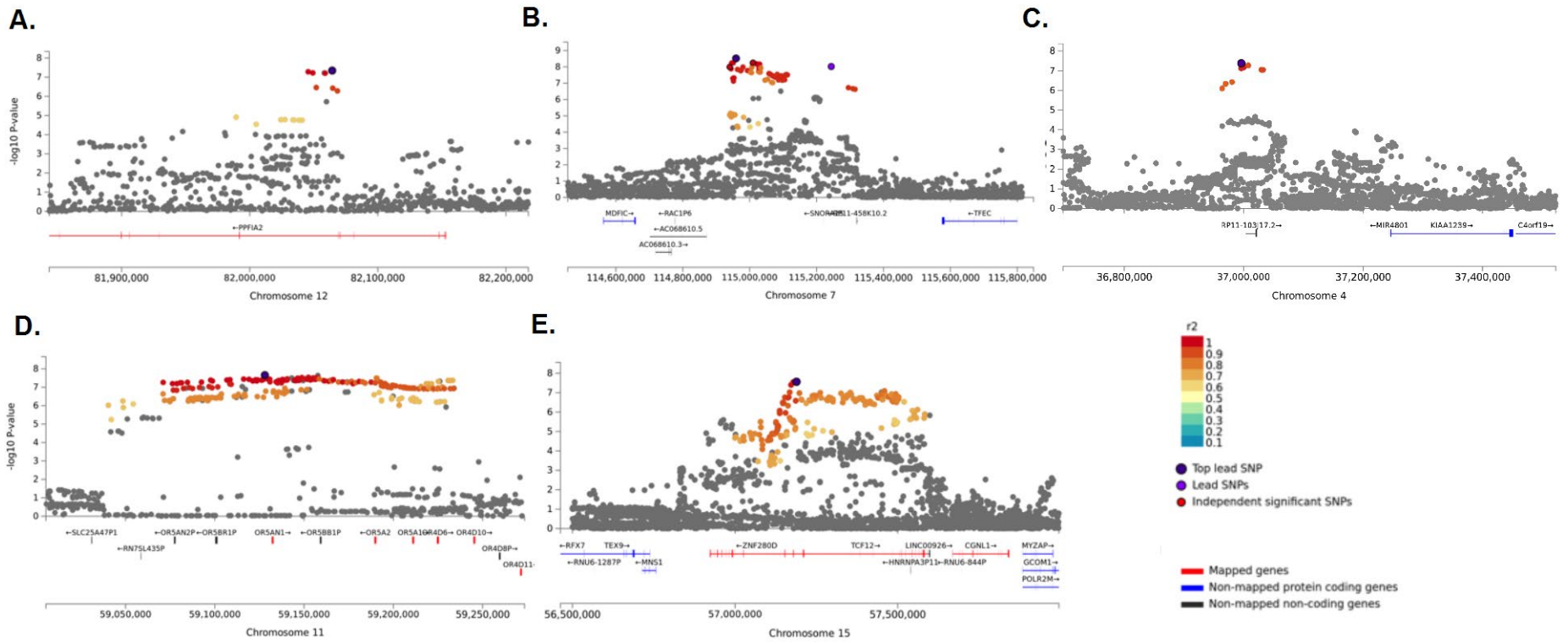
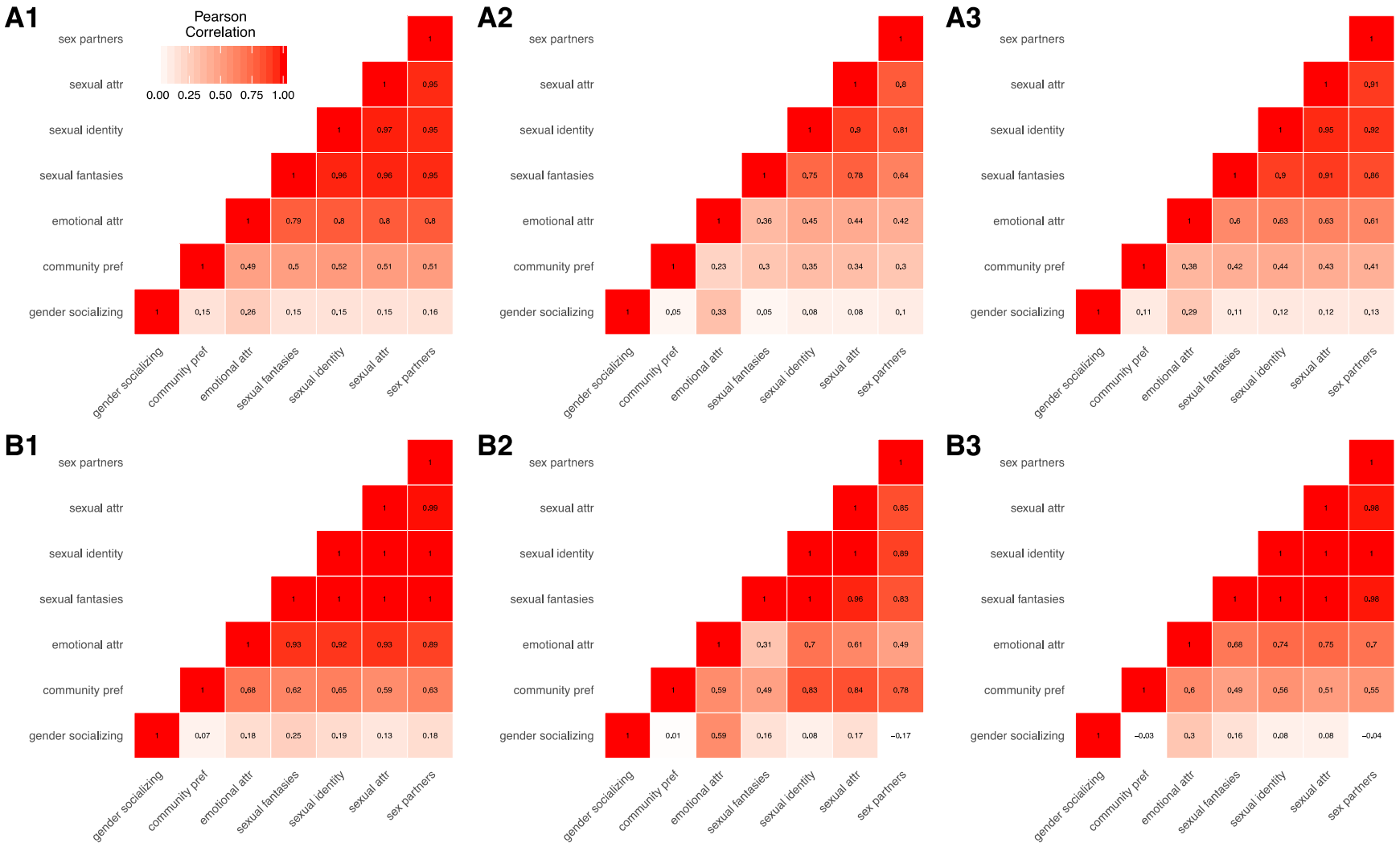


Figure S7: Phenotypic (A) and genetic (B) correlations between different sexuality, sexual behavior, and sexual identity questions in 23andMe. Results are presented separately for the male-specific (1), female-specific (2), and sexes combined analyses (3).

Variables are measured on a Likert scale and include: 1) Sexual partners: With whom have you actually had sex?; 2) Sexual Attr: To whom are you sexually attracted?; 3) Sexual identity: How do you label, identify, or think of yourself?; 4) Sexual fantasies: Whom do you have sexual fantasies about?; 5) Emotional attr: Whom do you feel more drawn to or close to emotionally?; 7) Community pref: In which community do you like to spend your time? In which do you feel most comfortable?; 6) Gender socializing: Which gender do you socialize with?



Supplementary Tables S1-S23

Table S1. Number of individuals included in the UK Biobank GWASs for same-sex sexual behavior.

Phenotype	N. Males		N. Females		
	Non-heterosexuals	Heterosexuals	Non-heterosexuals	Heterosexuals	Total
Same-sex sexual behavior	7831	180994	6063	214107	408995

Table S2. Descriptive statistics for same-sex sexual behavior in UK Biobank (for individuals with genetic data and of European ancestry).

	Males	Females	Total
Individuals with genetic data that passed quality control	207140	245444	452584
Answered sex-related questions	191737	223014	414751
Had sex at least once	189909	220975	410884
Never had same-sex partners*	180994	214107	395101
Number of opposite-sex partners (mean (SD); median)	8.42 (38.19); 4	4.50 (7.97); 3	6.28 (26.45); 3
Had same-sex partners**	7831	6063	13894
Number of opposite-sex partners (mean (SD); median)	12.91 (132.58); 2	12.13 (46.15); 5	12.53 (100.58); 4
Number of same-sex partners (mean (SD); median)	65.42 (433.74); 4	3.35 (10.43); 2	34.66 (309.69); 2
Removing those with over 100 sexual partners in total			
Never had same-sex partners	180431	214062	394493
Number of opposite-sex partners (mean (SD); median)	7.41 (10.97); 4	4.44 (6.06); 3	5.78 (8.76); 3
Had same-sex partners	7359	6029	13388
Number of opposite-sex partners (mean (SD); median)	7.91 (14.44); 2	10.62 (14.80); 5	9.28 (14.68); 4
Number of same-sex partners (mean (SD); median)	11.76 (19.90); 3	3.12 (4.50); 2	7.31 (14.87); 2
Data on those with exclusively same-sex partners			
Exclusively had same-sex partners	1766	693	2459
Number of same-sex partners (mean (SD); median) ***	21.09 (26.11); 10	5.51 (6.23); 4	16.70 (23.45); 6

** Of the 395,101 who reported not to have had a same-sex partner, 36,067 did not provide information on number of sexual partners.*
*** Of the 13,894 who reported to have had same-sex partnership, 2,779 and 2,150 did not report information on total number of sexual partners or number of same-sex partners, respectively (or they reported more same-sex partners than total number of sexual partners and were omitted).*
**** Excluding individuals with over 100 sexual partners in total.*

Table S3. Number of individuals for the phenotypes included in the GWASs from 23andMe.

With whom have you actually had sex?	N. Males	N. Females	N. Total *
Other sex only	29897	27493	56017
Other sex mostly	2095	3582	5550
Other sex slightly	210	760	959
Equal	174	348	511
Same sex slightly	297	228	519
Same sex mostly	1799	608	2376
Same sex only	2768	414	3136

Derived phenotype **	N. Males		N. Females		N. TOTAL *	
	Non-heterosexuals	Heterosexuals	Non-heterosexuals	Heterosexuals	Non-heterosexuals	Heterosexuals
Same-sex sexual behavior - dichotomized as in UK Biobank ***	7241	29821	5736	27343	12933	55594

** Notice that the Total N is not equal to the sum of males and females because the removal of related individuals has been done within each sex for the sex-specific analysis.*

*** The values for this analysis do not match the values in first Table because removal of related individuals is analysis-specific.*

**** Defined as: Other sex only versus Other sex mostly, Other sex somewhat more, Both sexes equally, Same sex somewhat more, Same sex mostly, Same sex only.*

Table S4. Heritability estimates based on the family-based heritability analysis using in UK Biobank data for same-sex sexual behavior.

Variance component	**Males***		**Females***		**Combined**	
	Proportion of Variance (%) *	**Standard error ***	**Proportion of Variance (%)**	**Standard error**	**Proportion of Variance (%)**	**Standard error**
V(a) - additive genetic	37.1	14.7	46.6	18.3	32.4	10.9
V(c) - shared/family environmental	0	1.3	2.9	5.3	2.8	2.9
V(e) - unshared environmental	62.9	14.3	50.5	14.6	64.8	8.9

Table S5. Genetic correlations (r_g) between questions related to sexuality, sexual behaviour, and sexual identity in 23andMe and same-sex sexual behavior in UK Biobank.

Variable in 23andMe	**Response categories**	**Sex**	**N ***	**Genetic correlation with same-sex sexual behavior in UK Biobank *****		
				r_g	**95% CIs**	***p-value***
Whom do you feel more drawn to or close to emotionally?	· Other sex only · Other sex mostly · Other sex somewhat more · Both sexes equally · Same sex somewhat more · Same sex mostly · Same sex only	Females Males Combined	33996 38178 70486	0.36 0.45 0.53	0.08-0.63 0.17-0.74 0.32-0.74	1.0E-02 1.6E-03 5.3E-07
Which gender do you socialize with?	· Other sex only · Other sex mostly · Other sex somewhat more · Both sexes equally · Same sex somewhat more · Same sex mostly · Same sex only	Females Males Combined	34122 38264 70692	0.05 0.04 0.06	-0.23-0.33 -0.36-0.44 -0.17-0.29	7.3E-01 8.4E-01 6.0E-01
In which community do you like to spend your time? In which do you feel most comfortable?	· Heterosexual only · Heterosexual mostly · Heterosexual somewhat more · Both orientations equally · Homosexual somewhat more · Homosexual mostly · Homosexual only	Females Males Combined	34096 38212 70616	0.37 0.35 0.59	-0.03-0.76 0.02-0.68 0.31-0.87	7.2E-02 4.0E-02 4.4E-05
Whom do you have sexual fantasies about?	· Other sex only · Other sex mostly · Other sex somewhat more · Both sexes equally · Same sex somewhat more · Same sex mostly · Same sex only	Females Males Combined	33840 38195 70357	0.98 0.63 0.83	0.37-1.59 0.36-0.89 0.60-1.05	1.6E-03 3.1E-06 3.5E-13
How do you label, identify, or think of yourself?	· Heterosexual only · Heterosexual mostly · Heterosexual somewhat more · Bisexual · Homosexual somewhat more · Homosexual mostly · Homosexual only	Females Males Combined	34136 38272 70716	1.33 0.64 0.79	0.08-2.59 0.38-0.89 0.57-1.01	3.9E-02 1.1E-06 1.2E-12
To whom are you sexually attracted?	· Other sex only · Other sex mostly · Other sex somewhat more · Both sexes equally · Same sex somewhat more · Same sex mostly · Same sex only	Females Males Combined	34129 38261 70699	1.02 0.59 0.75	0.31-1.73 0.33-0.84 0.54-0.96	4.8E-03 4.9E-06 2.7E-12
With whom have you actually had sex?	· Other sex only · Other sex mostly · Other sex somewhat more · Both sexes equally	Females Males	33433 37240	0.97 0.58	0.28-1.67 0.32-0.83	5.8E-03 9.0E-06

	· Same sex somewhat more · Same sex mostly · Same sex only	Combined	69068	0.70	0.49-0.91	3.1E-11
With whom have you actually had sex? Dichotomized as in UK Biobank **	Other sex mostly, other sex somewhat more, both sexes equally, same sex somewhat more, same sex mostly, same sex only versus Other sex only	Females	33079	0.79	0.35-1.23	4.2E-04
		Males	37062	0.70	0.44-0.95	1.4E-07
		Combined	68527	0.87	0.67-1.06	2.3E-18

** Note that the N in the combined analysis is not equal to the sum of males and females because the removal of related individuals has been done within each sex for the sex-specific analysis.*
*** This is the variable used for the main GWAS analysis.*
**** For females, genetic correlations were estimated using the female-specific GWAS results from UK Biobank; for males the male-specific results.*

Table S6. Sample quality control for genome-wide association study cohorts.

Study name	Genotyping array	Ancestry	Pre-imputation QC exclusion criteria					Imputation software	Imputation reference data	Post imputation filtering	Association analysis methods	Covariates
			MAF	Call rate	Hardy Weinberg equilibrium	Individual missingness	Other exclusion criteria					
UK Biobank	UK Biobank Axiom UK BiLEVE Axiom (Affymetrix)	European	< 0.0001	<95%	$p < 1x10^{-12}$	<95%	Inconsistencies reported vs genetic sex Controlled for batch and plate effects Discordance across controls Allele frequency differences with ExAC	IMPUTE4	Haplotype Reference Consortium, UK10K & 1000 Genomes Phase 3	INFO score < 0.8 MAF < 0.01 VEP annotated Missense and PTV MAF > $1x10^{-6}$ HWE p-value < $1x10^{-10}$	linear mixed models in BOLT-LMM	Sex year of birth year of birth^2 10 genetic principal components Batch number
23andMe	Illumina HumanHap550+Beadchip V3 Illumina OmniExpress+Beadchip V4 custom	European	NA	<95%	$p < 1x10^{-20}$	<98.5%	Significantly different MAF from 1000 Genome ($p < 1x10^{-15}$)	Minimac version 3	1000 Genomes phase 1	Average imputation r^2 < 0.5 Minimum imputed r^2 < 0.3 batch effect p < $1x10^{-50}$ MAF < 0.01	linear or logistic regression in a custom, in-house pipeline	Sex Age 5 genetic principal components Genotype platform
Add Health	Human Omni1-Quad-BeadChip	European	< 0.01	<98%	$p < 1x10^{-6}$	<95%	Related individuals Allele frequency discrepancy between the meta-analysis and prediction cohort > 15% Limited to HapMap3 SNPs	Michigan Imputation Server (Minimac)	Haplotype Reference Consortium	MAF < 0.05	Logistic regression in Rvtests	Sex Age Age-by-sex 10 genetic prinicpal components
MGSOSO	Affymetrix 6.0	European	< 0.05	<99%	p < $1x10^{-6}$	<95%	SNPs that were discrepant between arrays Individuals failing checks for duplications and relatedness	IMPUTE2	1000 Genomes phase 1	Information quality score < 0.6 MAF < 0.05	Logistic regression with the R package Genome-Wide Association	Correcting for family relatedness 3 genetic principal components

											analyses with Family	
CATSS	PsychArray	European	NA	<98%	$p < 1x10^{-6}$	<98%	Samples with >10% discordant genotypes among cross-batch duplicate samples SNPs with more than one discordant genotypes among MZ twins Allele frequency differences from the 1000 Genome European reference > 10% & GenCall scores < 0.5 Batch effects Autosomal inbreeding coefficient F > |0.2| Excessive relatedness Sex violation	Shapeit 2 for phasing & Minimac3 for imputation	1000 Genomes phase 3	Strand-ambiguous alleles SNPs with alleles inconsistent with reference panel	linear mixed models in BOLT-LMM	Correcting for family relatedness Sex 10 genetic principal components

Table S7. Independent genome-wide significant association signals for same-sex sexual behavior, based on the meta-analyzed UK Biobank and 23andMe samples (using MTAG).

									Males				Females				Males and females combined				
Sex	**Locus**	**rsID**	**Chr**	**Position**	**Effect allele**	**Reference allele**	**Effect allele frequency**	**Independent SNPs**	**Beta**	**Derived OR ***	**SE (OR)**	***p-value***	**Beta**	**Derived OR ***	**SE (OR)**	***p-value***	**Beta**	**Derived OR ***	**SE (OR)**	***p-value***	***p-value* difference between males and females**
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	0.73	rs34730029	0.0155	1.08	0.0028	2.2E-08	0.0036	1.02	0.0028	1.9E-01	0.0098	1.06	0.0019	3.7E-07	2.2E-03**
	15:56999901-57583301	rs28371400	15	57188761	T	A	0.83	rs28371400	0.0180	1.10	0.0032	2.7E-08	-0.0006	1.00	0.0032	8.5E-01	0.0092	1.05	0.0023	4.2E-05	4.5E-05**
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	0.33	rs13135637	-0.0023	0.99	0.0026	3.8E-01	-0.0143	0.91	0.0026	4.2E-08	-0.0078	0.96	0.0018	1.9E-05	1.3E-03**
Males and females combined	7:114940147-115314917	rs10261857	7	114959247	G	A	0.37	rs10261857 rs6968125 rs4377898 rs10275935	0.0115	1.06	0.0026	7.1E-06	0.0103	1.06	0.0025	5.3E-05	0.0105	1.06	0.0018	3.0E-09	7.2E-01
	12:81989337-82068452	rs11114975	12	82064386	C	T	0.90	rs11114975	0.0150	1.08	0.0040	2.0E-04	0.0160	1.11	0.0040	6.4E-05	0.0153	1.09	0.0028	4.5E-08	8.7E-01

** More information can be found in the Supplementary Materials and Methods M3.2.1. Briefly, we transformed the beta from MTAG on the original scale and used the formula from (60) to obtain the corresponding OR. Since the transformation depends on the prevalence of the trait, same betas can result in different ORs depending on whether the analysis was done in males, females, or the combined sample.*

*** A locus is defined as sex-differentiated if the p-value for the difference test between males and females is smaller than 0.01.*

Table S8. Independent genome-wide association signals for same-sex sexual behavior, based on the meta-analyzed UK Biobank and 23andMe samples. MTAG results are shown based on an analysis with the genetic correlation between samples estimated from the data versus fixed to 1.

Sex	**Locus**	**rsID**	**UK Biobank**			**23andMe**			**MTAG (estimated r_g)**			**MTAG (fixed $r_g=1$)**		
			Beta	**SE (Beta)**	***p-value***	**Beta**	**SE (Beta)**	***p-value***	**Beta**	**SE (Beta)**	***p-value***	**Beta**	**SE (Beta)**	***p-value***
Males	11:59040414-59233752	rs34730029	0.0037	0.0007	2.9E-07	0.0566	0.0214	8.0E-03	0.0155	0.0028	2.2E-08	0.0135	0.0025	5.7E-08
	15:56999901-57583301	rs28371400	0.0049	0.0009	8.1E-09	0.0447	0.0242	6.4E-02	0.0180	0.0032	2.7E-08	0.0153	0.0029	1.4E-07
Females	4:36963942-37032454	rs13135637	-0.0021	0.0005	5.7E-05	-0.0876	0.0227	1.1E-04	-0.0143	0.0026	4.2E-08	-0.0135	0.0024	3.1E-08
Males and females combined	7:114940147-115314917	rs10261857	0.0027	0.0004	1.2E-10	0.0229	0.0144	1.1E-01	0.0105	0.0018	3.0E-09	0.0100	0.0017	5.6E-09
	12:81989337-82068452	rs11114975	0.0035	0.0007	1.2E-07	0.0530	0.0238	2.5E-02	0.0153	0.0028	4.5E-08	0.0146	0.0027	6.2E-08

Table S9. Cross-sex genetic correlations (r_g) for same-sex sexual behavior and proportion of same-sex to total partners among non-heterosexuals as obtained from LD-score regression.

		Genetic correlation between males and females			
Phenotype	**Study sample**	**r_g**	**SE (r_g)**	***p-value*** **(test if r_g is significantly different from 0)**	***p-value*** **(test if r_g is significantly different from 1)**
Same-sex sexual behavior	UK Biobank	0.653	0.135	1.28E-06	9.96E-03
Same-sex sexual behavior	23andMe	0.335	0.127	8.50E-03	1.68E-07
Same-sex sexual behavior	UK Biobank + 23andMe (MTAG)	0.625	0.077	3.72E-16	1.03E-06
Proportion of same-sex to total partners among non-heterosexuals	UK Biobank	0.120	0.256	6.39E-01	5.87E-04
Proportion of same-sex to total partners among non-heterosexuals	23andMe	-0.242	0.319	7.46E-02	9.88E-05
Proportion of same-sex to total partners among non-heterosexuals	UK Biobank + 23andMe (MTAG)	0.410	0.187	2.85E-02	1.60E-03

Table S10. Replication analysis of genome-wide significant SNPs for same-sex sexual behavior in MGSOSO, Add Health, and CATSS. The betas are on different scales and not directly interpretable, therefore we meta-analyzed the Z statistics.

Sex	rsID	Chromosome	Position	Effect allele	Reference allele	Effect allele frequency	Beta	Z-score	*p-value*	Effective sample size
						MGSOSO (sexual identity, males only)				
Males	rs34730029	11	59128042	G	T	0.76	0.185	2.39	0.02	2298
Males	rs28371400	15	57188761	T	A	0.81	0.170	2.00	0.05	2298
Females	rs13135637	4	36996064	C	A	*NA*	*NA*	*NA*	*NA*	*NA*
Males and Females Combined	rs10261857	7	114959247	G	A	0.39	0.044	0.66	0.51	2298
Males and Females Combined	rs11114975	12	82064386	C	T	0.90	0.275	2.52	0.01	2298
						Add Health (ever had a same-sex partner)				
Males	rs34730029	11	59128042	G	T	0.73	0.004	0.13	0.90	493
Males	rs28371400	15	57188761	T	A	0.82	0.093	2.37	0.02	493
Females	rs2588543*	4	37000406	C	T	0.32	0.015	0.49	0.62	1314
Males and Females Combined	rs10261857	7	114959247	G	A	0.38	0.007	0.34	0.73	1848
Males and Females Combined	rs11114975	12	82064386	C	T	0.90	0.020	0.60	0.55	1848
						CATSS (ever had a same-sex partner, either at age 15 or 18)				
Males	rs34730029	11	59128042	G	T	0.72	0.002	0.31	0.76	265
Males	rs28371400	15	57188761	T	A	0.83	0.008	0.91	0.36	265
Females	rs13135637	4	36996064	C	A	0.32	-0.002	-0.24	0.81	475
Males and Females Combined	rs10261857	7	114959247	G	A	0.40	0.003	0.59	0.55	741
Males and Females Combined	rs11114975	12	82064386	C	T	0.89	0.013	1.63	0.1	741
						Meta-analysis MGSOSO + Add Health + CATSS				
Males	rs34730029	11	59128042	G	T	-	-	2.22	0.027	-
Males	rs28371400	15	57188761	T	A	-	-	2.96	0.003	-
Females	rs13135637	4	36996064	C	A	-	-	0.30	0.765	-
Males and Females Combined	rs10261857	7	114959247	G	A	-	-	0.89	0.371	-
Males and Females Combined	rs11114975	12	82064386	C	T	-	-	2.73	0.006	-

** rs2588543 is a proxy for rs13135637 (r^2= 0.995), which was not available in this study.*

Table S11. SNP-based heritability estimates (h^2) for same-sex sexual behavior. Shown are estimates based on the general univariate model, the allele frequency/LD stratified analysis from LD-score regression, and estimates obtained from BOLT-LMM. Results are reported both on the observed scale and on the liability scale. The K (prevalence) parameter is inferred from the observed prevalence in the UK Biobank and 23andMe. See Suppl. Material and Methods M6.1.1 for additional information about prevalence assumptions.

		Observed scale					
		Males		**Females**		**Males and females combined ***	
Phenotype	**Study sample**	**h^2 [95% CIs]**	***p-value***	**h^2 [95% CIs]**	***p-value***	**h^2 [95% CIs]**	***p-value***
		LD-score regression - Univariate					
Same-sex sexual behavior - mixed model	UK Biobank	0.0158 [0.0103-0.0213]	1.67E-08	0.0127 [0.0074-0.0180]	2.55E-06	0.0118 [0.0091-0.0145]	3.50E-17
Same-sex sexual behavior - logistic regression	UK Biobank	0.0163 [0.0100-0.0226]	3.51E-07	0.0142 [0.0071-0.0213]	8.00E-05	0.0118 [0.0085-0.0151]	3.89E-12
Same-sex sexual behavior - logistic regression	23andMe	0.1137 [0.0829-0.1445]	4.42E-13	0.0763 [0.0412-0.1114]	2.02E-05	0.0634 [0.0462-0.0806]	5.82E-13
Same-sex sexual behavior - MTAG	UK Biobank + 23andMe	0.0278 [0.0221-0.0335]	9.14E-22	0.0193 [0.0146-0.0240]	8.86E-16	0.0190 [0.0161-0.0219]	9.05E-37
		LD-score regression - MAF and LD-stratified					
Same-sex sexual behavior - mixed model	UK Biobank	0.0177 [0.0095-0.0259]	2.51E-05	0.0171 [0.0083-0.0259]	1.45E-04	0.0124 [0.0085-0.0163]	5.65E-10
Same-sex sexual behavior - logistic regression	UK Biobank	0.0196 [0.0098-0.0294]	8.85E-05	0.0228 [0.0106-0.0350]	2.36E-04	0.0131 [0.0082-0.0180]	1.61E-07
Same-sex sexual behavior - logistic regression	23andMe	0.1190 [0.0671-0.1709]	7.10E-06	0.0835 [0.0231-0.1439]	6.71E-03	0.0613 [0.0335-0.0891]	1.58E-05
Same-sex sexual behavior- MTAG	UK Biobank + 23andMe	0.0302 [0.0220-0.0384]	6.46E-13	0.0196 [0.0127-0.0265]	2.14E-08	0.0195 [0.0150-0.0240]	2.29E-17
		BOLT-LMM					
Same-sex sexual behavior - mixed models	UK Biobank	0.0325 [0.0236-0.0415]	9.77E-13	0.0306 [0.0228-0.0383]	1.16E-14	0.0229 [0.0188-0.0269]	7.80E-29

		Liability Scale (h^2_liability=h^2_observed*K*(1-K)/z^2) **					
		Males		**Females**		**Males and females combined ***	
Phenotype	**Study sample**	**h^2 [95% CIs]**	***p-value***	**h^2 [95% CIs]**	***p-value***	**h^2 [95% CIs]**	***p-value***
		LD-score regression - Univariate					
Same-sex sexual behavior- mixed model	UK Biobank	0.0804 [0.0525-0.1083]	1.67E-08	0.0859 [0.0501-0.1217]	2.55E-06	0.0681 [0.0522-0.0839]	3.50E-17

Same-sex sexual behavior - logistic regression	UK Biobank	0.0829 [0.0510-0.1148]	3.51E-07	0.0960 [0.0483-0.1437]	8.00E-05	0.0681 [0.0489-0.0873]	3.89E-12
Same-sex sexual behavior - logistic regression	23andMe	0.2336 [0.1704-0.2968]	4.42E-13	0.1646 [0.0889-0.2403]	2.02E-05	0.1328 [0.0967-0.1690]	5.82E-13
Same-sex sexual behavior - MTAG	UK Biobank + 23andMe	0.1414 [0.1125-0.1703]	9.14E-22	0.1305 [0.0987-0.1623]	8.86E-16	0.1096 [0.0927-0.1266]	9.05E-37
		LD-score regression - MAF and LD-stratified					
Same-sex sexual behavior - mixed model	UK Biobank	0.0900 [0.0482-0.1319]	2.51E-05	0.1156 [0.0560-0.1753]	1.45E-04	0.0715 [0.0489-0.0942]	5.65E-10
Same-sex sexual behavior - logistic regression	UK Biobank	0.0997 [0.0499-0.1496]	8.85E-05	0.1542 [0.0720-0.2363]	2.36E-04	0.0756 [0.0473-0.1038]	1.61E-07
Same-sex sexual behavior - logistic regression	23andMe	0.2455 [0.1384-0.3527]	7.10E-06	0.1822 [0.0505-0.3140]	6.71E-03	0.1285 [0.0702-0.1869]	1.58E-05
Same-sex sexual behavior - MTAG	UK Biobank + 23andMe	0.1536 [0.1117-0.1955]	6.46E-13	0.1325 [0.0861-0.1789]	2.14E-08	0.1125 [0.0865-0.1385]	2.29E-17
		BOLT-LMM					
Same-sex sexual behavior - mixed model	UK Biobank	0.1655 [0.1201-0.2110]	9.77E-13	0.2066 [0.1542-0.2591]	1.16E-14	0.1318 [0.1087-0.1550]	7.80E-29

** Estimates combining males and females are biased downwards because of the relatively low genetic correlation between the sexes*
*** K for males in UK Biobank = 0.041*
K for females in UK Biobank = 0.028
K for the combined sample (males + females) in UK Biobank = 0.034
K for males in 23andMe = 0.195
K for females in 23andMe = 0.173
K for the combined sample (males + females) in 23andMe = 0.189

Table S12. Association between polygenic scores for same-sex sexual behavior built from the MTAG summary statistics from UK Biobank + 23andMe and same-sex sexual identity in MGSOSO.

	Males
Score (log-odds)	0.132**
Standard Error	0.041
p-value	1.20E-03
Incremental pseudo ΔR^2	0.42%
95% CI - low	0.10%
95% CI - high	0.86%
N	1,077 non-heterosexuals and 1,231 heterosexuals

Analysis includes the control variables age, age^2, and the first 10 genetic PCs. The 95% CIs for the pseudo-R^2 were bootstrapped with 1,000 repetitions each.

Table S13. Associations of polygenic scores for same-sex sexual behavior built from the MTAG summary statistics from UK Biobank + 23andMe with same-sex sexual experience and same-sex attraction in the Add Health sample. Sex-specific analyses are done using the corresponding sex-specific results from UK Biobank + 23andMe.

Males and Females		
	Ever same-sex sex	**Same-sex attraction**
Score (log-odds)	0.08	0.15*
Standard Error	0.057	0.063
p-value	0.206	0.017
Mean/Prevalence	10.98%	7.13%
Standard Deviation	—	—
Incremental ΔR^2	0.2%	0.4%
95% CI - low	0.0%	0.0%
95% CI - high	0.5%	1.0%
N	4725	4755
Females		
Score (log-odds)	0.10*	0.13*
Standard Error	0.051	0.059
p-value	0.049	0.027
Mean/Prevalence	15.34%	10.12%
Standard Deviation	—	—
Incremental ΔR^2	0.3%	0.3%
95% CI - low	0.0%	0.0%
95% CI - high	0.7%	0.8%
N	2529	2539
Males		
Score (log-odds)	0.13	0.17
Standard Error	0.086	0.131
p-value	0.131	0.196
Mean/Prevalence	5.97%	3.70%
Standard Deviation	—	—
Incremental ΔR^2	0.4%	0.6%
95% CI - low	0.0%	0.0%
95% CI - high	1.3%	1.4%
N	2196	2216

*Notes: Analyses include the control variables sex, age, sex*age, and the first 10 PCs of the genetic relatedness matrix. Incremental ΔR^2 indicate incremental pseudo-R^2 values. The 95% CIs for these pseudo-R^2 were bootstrapped with 1000 repetitions each.*

** Estimates were significant at $p<0.05$.*

Table S14. Associations of polygenic scores for same-sex sexual behavior built from the MTAG summary statistics from UK Biobank + 23andMe with different questions in CATSS. Sex-specific analyses were done using the corresponding sex-specific results from UK Biobank + 23andMe. Participants were interviewed at 15 and 18 years old, we considered their responses either at 18 and 15 or only at 15 years old.

	Males and Females			
	Ever had same-sex sex (15 or 18 years old)	**Same-sex attraction (15 or 18 years old)**	**Ever had same-sex sex (15 years old)**	**Same-sex attraction (15 years old)**
Score	0.20*	0.15***	0.00	0.17***
Robust Standard Error	0.079	0.031	0.146	0.043
p-value	0.013	3.5E-06	0.98	6.4E-05
Prevalence	5%	16%	4%	11%
Incremental ΔR^2	0.6%	0.5%	0.0%	0.6%
95% CI - low	0.0%	0.2%	0.0%	0.1%
95% CI - high	1.9%	1.0%	1.4%	1.3%
N	3892	8093	1614	6381
	Females			
Score	0.15	0.14***	-0.10	0.12*
Robust Standard Error	0.096	0.037	0.152	0.050
p-value	0.126	1.7E-04	0.49	0.015
Prevalence	6%	23%	5%	16%
Incremental ΔR^2	0.3%	0.5%	0.0%	0.3%
95% CI - low	0.0%	0.1%	0.0%	0.0%
95% CI - high	1.7%	1.1%	2.4%	1.1%
N	2153	4196	810	3185
	Males			
Score	0.13	0.15**	-0.03	0.25***
Robust Standard Error	0.127	0.057	0.242	0.075
p-value	0.310	0.007	0.90	9.3E-04
Prevalence	0%	9%	2%	6%
Incremental ΔR^2	0.2%	0.4%	0.2%	0.9%
95% CI - low	0.0%	0.0%	0.0%	0.2%
95% CI - high	2.0%	1.2%	4.4%	2.4%
N	1739	3897	801	3196

*Analyses were controlled for age effects and the first 10 genetic PCs. Incremental R^2 values are displayed for continuous phenotypes, incremental pseudo-R^2 values for binary phenotypes. The 95% CIs for the incremental R^2 values were bootstrapped with 1000 repetitions each. Score: standardized beta coefficients are listed for continuous variables, log-odds for binary variables. Significance levels for score effects are indicated as follows: *** p<0.001, ** p<0.01, * p<0.05.*

Table S15. Signal enrichment analysis for same-sex sexual behavior (in 2,929 highly constrained genes), using MAGMA and LD-score regression.

	CONDITIONED ON BASELINE*					
	MAGMA			**LD-score regression**		
Phenotype	**Beta**	**SE**	***p-value***	**Tau coefficient**	**SE**	***p-value***
Same-sex sexual behavior (males and females combined)	0.030	0.017	3.71E-02	1.08E-09	5.32E-10	2.07E-02
	CONDITIONED ON BASELINE + BRAIN-SPECIFIC GENES					
	MAGMA			**LD-score regression**		
	Beta	**SE**	***p-value***	**Tau coefficient**	**SE**	***p-value***
Same-sex sexual behavior (males and females combined)	0.023	0.018	9.47E-02	1.14E-09	5.08E-10	1.27E-02

** MAGMA conditions at baseline on gene size, log(gene size), gene density, log(gene density), inverse MAC, log(inverse MAC). LD-score regression conditions at baseline on 53 annotation categories, described in (25).*

Table S16. Results of the pheWAS (from the Neale Lab database, http://www.nealelab.is/uk-biobank*, top panel) and GWAS catalog (bottom panel) for the SNPs that are genome-wide significantly associated with same-sex sexual behavior (MTAG analysis of UK Biobank + 23andMe). We only present traits with a heritability *p-value* < 0.05.

Sex	Locus	SNP ID	Chr	Position	Effect allele	Reference alelle	Phenotype number	Description phenotype	Beta	SE	*p-value*	Benjamini & Hochberg - adjusted *p-value*	Direction of effect compared to GWAS results
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	5134	6mm strong meridian (left)	2.025E-02	5.578E-03	2.82E-04	3.05E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2110	Able to confide	-2.142E-02	4.637E-03	3.84E-06	1.07E-03	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	20117_2	Alcohol drinker status: Current	-2.137E-03	6.056E-04	4.18E-04	4.00E-02	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	6146_2	Attendance/disability/mobility allowance: Disability living allowance	1.799E-03	5.013E-04	3.33E-04	3.32E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2335	Chest pain or discomfort	3.617E-03	8.939E-04	5.21E-05	8.29E-03	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1239	Current tobacco smoking	7.026E-03	1.335E-03	1.40E-07	7.23E-05	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	20489	Felt loved as a child	-1.450E-02	4.144E-03	4.65E-04	4.32E-02	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1031	Frequency of friend/family visits	9.574E-03	2.703E-03	3.98E-04	3.86E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	4570	Friendships satisfaction	1.220E-02	3.132E-03	9.83E-05	1.43E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1170	Getting up in morning	6.857E-03	1.841E-03	1.96E-04	2.42E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2844	Had other major operations	6.569E-03	1.559E-03	2.52E-05	4.82E-03	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1707_3	Handedness (chirality/laterality): Use both right and left hands equally	1.104E-03	3.150E-04	4.56E-04	4.30E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	6141_1	How are people in household related to participant: Husband, wife or partner	-6.325E-03	1.070E-03	3.41E-09	3.80E-06	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2149	Lifetime number of sexual partners	9.936E-03	2.085E-03	1.87E-06	6.60E-04	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2188	Long-standing illness, disability or infirmity	5.426E-03	1.148E-03	2.27E-06	7.58E-04	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1787	Maternal smoking around birth	5.406E-03	1.209E-03	7.77E-06	1.92E-03	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1418_1	Milk type used: Full cream	2.451E-03	5.946E-04	3.75E-05	6.61E-03	same

Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	709	Number in household	-9.107E-03	1.623E-03	2.02E-08	1.50E-05	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2405	Number of children fathered	-1.665E-02	2.920E-03	1.19E-08	9.96E-06	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	399	Number of incorrect matches in round	-1.319E-02	2.409E-03	4.35E-08	2.91E-05	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	728	Number of vehicles in household	-8.522E-03	2.071E-03	3.87E-05	6.64E-03	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2237	Plays computer games	4.387E-03	1.218E-03	3.17E-04	3.26E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	6138_1	Qualifications: College or University degree	6.547E-03	1.135E-03	8.02E-09	7.66E-06	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	2090	Seen doctor (GP) for nerves, anxiety, tension or depression	4.951E-03	1.150E-03	1.66E-05	3.58E-03	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	20116_2	Smoking status: Current	3.748E-03	7.414E-04	4.29E-07	1.91E-04	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	20116_0	Smoking status: Never	-4.468E-03	1.207E-03	2.14E-04	2.47E-02	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	5084	Spherical power (right)	-1.870E-02	5.160E-03	2.91E-04	3.09E-02	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	1080	Time spent using computer	7.913E-03	2.140E-03	2.18E-04	2.47E-02	same
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	4288	Time to answer	-1.729E-02	4.131E-03	2.84E-05	5.13E-03	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	400	Time to complete round	-1.148E-02	2.332E-03	8.53E-07	3.57E-04	opposite
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	189	Townsend deprivation index at recruitment	8.398E-03	2.429E-03	5.44E-04	4.99E-02	same
Combined	12:81989337-82068452	rs11114975	12	82064386	C	T	6160_3	Leisure/social activities: Religious group	-5.287E-03	1.348E-03	8.73E-05	1.30E-02	opposite
Combined	12:81989337-82068452	rs11114975	12	82064386	C	T	20084_485	Vitamin and/or mineral supplement use: Calcium	8.594E-03	2.303E-03	1.91E-04	2.40E-02	same
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	1558	Alcohol intake frequency.	1.366E-02	3.672E-03	2.00E-04	2.43E-02	opposite
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	4290	Duration screen displayed	2.189E-02	4.130E-03	1.15E-07	7.02E-05	opposite
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	20453	Ever taken cannabis	-1.458E-02	4.058E-03	3.27E-04	3.31E-02	same
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	20016	Fluid intelligence score	-1.715E-02	4.358E-03	8.30E-05	1.26E-02	same

Females	4:36963942-37032454	rs13135637	4	36996064	C	A	4291	Number of attempts	6.937E-03	1.726E-03	5.86E-05	9.12E-03	opposite
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	20018	Prospective memory result	8.188E-03	2.208E-03	2.09E-04	2.47E-02	opposite
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	6138_1	Qualifications: College or University degree	-4.532E-03	1.168E-03	1.04E-04	1.48E-02	same
Females	4:36963942-37032454	rs13135637	4	36996064	C	A	1080	Time spent using computer	-1.162E-02	2.206E-03	1.38E-07	7.23E-05	same
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	20160	Ever smoked	6.195E-03	1.290E-03	1.56E-06	6.15E-04	same
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	1249	Past tobacco smoking	-1.497E-02	3.419E-03	1.20E-05	2.86E-03	opposite
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	20116_0	Smoking status: Never	-6.141E-03	1.309E-03	2.72E-06	8.67E-04	opposite
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	20116_1	Smoking status: Previous	4.556E-03	1.253E-03	2.77E-04	3.04E-02	same
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	1488	Tea intake	-1.148E-02	2.690E-03	1.97E-05	3.98E-03	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23124	Arm fat mass (left)	-1.158E-02	3.072E-03	1.64E-04	2.15E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23120	Arm fat mass (right)	-1.249E-02	3.071E-03	4.74E-05	7.73E-03	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23123	Arm fat percentage (left)	-1.073E-02	2.361E-03	5.50E-06	1.47E-03	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23119	Arm fat percentage (right)	-1.108E-02	2.370E-03	2.95E-06	8.98E-04	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23099	Body fat percentage	-8.942E-03	2.362E-03	1.53E-04	2.05E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	21001	Body mass index (BMI)	-1.162E-02	3.091E-03	1.71E-04	2.20E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23104	Body mass index (BMI)	-1.154E-02	3.115E-03	2.13E-04	2.47E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	1408	Cheese intake	-1.365E-02	3.348E-03	4.59E-05	7.67E-03	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	1687	Comparative body size at age 10	-8.114E-03	2.131E-03	1.40E-04	1.91E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	2277	Frequency of solarium/sunlamp use	5.172E-03	1.115E-03	3.51E-06	1.02E-03	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	1747_4	Hair colour (natural, before greying): Dark brown	-6.586E-03	1.509E-03	1.28E-05	2.86E-03	opposite

Males	15:56999901-57583301	rs28371400	15	57188761	T	A	2395_1	Hair/balding pattern: Pattern 1	-1.713E-02	2.139E-03	1.14E-15	2.54E-12	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	2395_4	Hair/balding pattern: Pattern 4	9.197E-03	1.772E-03	2.11E-07	1.01E-04	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	46	Hand grip strength (left)	9.137E-03	2.156E-03	2.25E-05	4.43E-03	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	47	Hand grip strength (right)	7.973E-03	2.160E-03	2.24E-04	2.49E-02	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	30100	Mean platelet (thrombocyte) volume	-3.232E-02	3.167E-03	1.87E-24	1.25E-20	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	30080	Platelet count	2.846E-02	3.066E-03	1.64E-20	5.50E-17	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	30090	Platelet crit	1.441E-02	3.014E-03	1.75E-06	6.52E-04	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	30110	Platelet distribution width	-2.087E-02	3.135E-03	2.83E-11	4.73E-08	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23128	Trunk fat mass	-1.132E-02	3.132E-03	3.02E-04	3.15E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23127	Trunk fat percentage	-1.080E-02	2.835E-03	1.40E-04	1.91E-02	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	2267	Use of sun/UV protection	1.180E-02	2.700E-03	1.24E-05	2.86E-03	same
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	48	Waist circumference	-1.228E-02	2.738E-03	7.28E-06	1.87E-03	opposite
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	23100	Whole body fat mass	-1.079E-02	3.044E-03	3.94E-04	3.86E-02	opposite

							GWAS Catalog	
Sex	**Locus**	**SNP ID**	**Chr**	**Position**	**Effect allele**	**Reference allele**	**Phenotype description**	**Reference**
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	Tonsillectomy	(*78*)
Combined	7:114940147-115314917	rs10261857	7	114959247	G	A	Tonsillectomy	(*79*)
Males	11:59040414-59233752	rs34730029	11	59128042	G	T	Odorant perception (& beta;-ionone)	(80)
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	Platelet count	(*81*)

Males	15:56999901-57583301	rs28371400	15	57188761	T	A	Platelet distribution width	(*81*)
Males	15:56999901-57583301	rs28371400	15	57188761	T	A	Mean platelet volume	(*81*)

** All summary statistics from the Neale Lab database are also available at GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics)*

Table S17. In-silico biological follow-up. Identification of target genes for each genome-wide significant SNP associated with same-sex sexual behavior. Different types of information were used (eQTL, MAGMA) to identify potential candidate target genes.

Sex	Locus	Plausible gene(s)	eQTL	Literature	Gene-based test (MAGMA)
Males and females combined	7:114940147-115314917	FOXP2	-	The FOXP2 locus is in juxtaposition. The FOXP2 locus is associated with mathematical ability (*82*), self-reported educational attainment (*82*) and BMI (*83*). FOXP2 encodes a member of the forkhead/winged-helix (FOX) family of transcription factors. This gene is required for proper development of speech and language regions of the brain during embryogenesis, and may be involved in a variety of biological pathways and cascades that may ultimately influence language development (RefSeq). Mutations of FOXP2 are associated with developmental verbal dyspraxia [OMIM: 602081].	-
	12:81989337-82068452	PPFIA2	-	rs11114975 is an intron variant of PPFIA2. The PPFIA2 locus is associated with Schizophrenia (*84*). The protein encoded by this gene is a member of the LAR protein-tyrosine phosphatase-interacting protein (liprin) family. Liprins interact with members of LAR family of transmembrane protein tyrosine phosphatases, which are known to be important for axon guidance and mammary gland development. It has been proposed that liprins are multivalent proteins that form complex structures and act as scaffolds for the recruitment and anchoring of LAR family of tyrosine phosphatases. This protein has been shown to bind the calcium/calmodulin-dependent serine protein kinase (MAGUK family) protein (also known as CASK) and proposed to regulate higher-order brain functions in mammals (RefSeq). Deletions of PPFIA2 were reported to be associated with Intellectual disability (*85*).	-
Males	11:59040414-59233752	OR5A1	-	SNP in LD (rs6591536: r^2=0.70) with a SNP that has been associated with sensitivity with β-ionone (*30*). β-ionone is also an aroma produced by carotenoid degradation in cigarettes (*86*). Studies in mice and fruitflies have shown a relationship between sex hormone signaling and olfactory circuits (*87*). rs6591536 genotype also influences emotional associations and explains differences in food and product choices (*30*).	OR5A1, OR4D6 and OR5AN1
	15:56999901-57583301	TCF12	-	The TCF12 and the ZNF280D locus are in proximity. TCF12 is a member of the basic helix-loop-helix (bHLH) E-protein family that recognizes the consensus binding site (E-box) CANNTG. This encoded protein is expressed in many tissues, among them skeletal muscle, thymus, B- and T-cells, and may participate in regulating lineage-specific gene expression through the formation of heterodimers with other bHLH E-proteins (RefSeq). TCF12 is the primary heterodimerization partner for TCF21, binding the same primary sequence as TCF21. TCF21 transcription factor alone and in combination with TCF12, have been shown in vivo to be a downstream targets of the SRY gene (responsible for the initiation of male sex determination in humans) during male sex determination (*33*). This result has been independently confirmed by Xu et al. (*88*).	-
Females	4:36963942-37032454	-	-	An adjacent locus (LOC105374400 / LOC100508631) is associated with self-reported educational attainment (*82*).	-

eQTL; Expression quantitative trait loci.

Table S18. Results of the gene-based test of association for same-sex sexual behavior (performed in MAGMA).
See external file S18.xlsx

Table S19. Genetic correlations (r_g) of same-sex sexual behavior (MTAG results of UK Biobank + 23andMe) with a range of traits as estimated using LD-score regression.

Phenotype	Reference	N	Genetic correlation with same-sex sexual behavior (males)			Genetic correlation with same-sex sexual behavior (females)			*p-value* for difference in r_g between males and females**
			r_g	Standard Error	*p-value*	r_g	Standard Error	*p-value*	
ADHD	(*66*)	20,183 cases, 35,191 controls	0.272	0.060	6.29E-06	0.253	0.059	1.90E-05	7.98E-01
Age at first birth (females)	(*89*)	189,656	-0.235	0.059	5.93E-05	-0.146	0.063	2.01E-02	2.87E-01
Age at first birth (males)	(*89*)	48,408	-0.193	0.062	1.68E-03	-0.090	0.080	2.60E-01	3.26E-01
Age at Menarche	(*90*)	87,802	-0.086	0.044	5.32E-02	-0.049	0.050	3.31E-01	5.64E-01
Age at Menopause	(*91*)	69,360	-0.079	0.061	1.93E-01	-0.069	0.064	2.75E-01	9.09E-01
Alcohol use	(*92*)	112,117	-0.025	0.057	6.66E-01	0.149	0.065	2.22E-02	2.95E-02
Anorexia	(*93*)	17,767	-0.020	0.095	8.33E-01	-0.085	0.100	3.96E-01	5.85E-01
Anxiety	(*94*)	18,186	-0.050	0.142	7.27E-01	0.251	0.179	1.61E-01	1.63E-01
Autism	(*95*)	18,381 cases, 27,969 controls	0.098	0.069	1.59E-01	0.210	0.068	2.06E-03	1.96E-01
Bipolar	(*96*)	11,974 cases, 51,792 controls	0.019	0.076	8.08E-01	0.340	0.077	1.09E-05	1.05E-03
Birth Weight	(*97*)	143,677	-0.046	0.048	3.31E-01	0.035	0.050	4.88E-01	2.05E-01
Cannabis use	(*98*)	~69,878 cases, 92,204 controls	0.308	0.058	1.09E-07	0.671	0.062	1.20E-27	1.47E-06
Height	(*99*)	253,280	-0.075	0.033	2.27E-02	-0.043	0.039	2.73E-01	5.05E-01
Loneliness	http://www.nealelab.is/uk-biobank	355,583	0.218	0.053	3.83E-05	0.220	0.053	3.61E-05	9.81E-01
Major depressive disorder (MDD)	(*100*)	59,851 cases, 113,154 controls	0.328	0.055	1.83E-09	0.438	0.059	1.04E-13	1.51E-01
Neuroticism	(*101*)	170,911	0.159	0.052	2.20E-03	0.220	0.067	1.08E-03	4.52E-01
Number of Children (females)	(*89*)	225,230	0.013	0.076	8.63E-01	0.086	0.085	3.12E-01	4.90E-01

Number of Children (males)	(*89*)	103,909	0.178	0.091	4.96E-02	0.296	0.107	5.87E-03	3.74E-01
Number of sex partners	(*21*)	370,711	0.407	0.040	3.33E-24	0.681	0.053	1.00E-37	3.13E-05
Openness to experience	(*102, 103*)	76,551	0.135	0.060	2.37E-02	0.312	0.066	2.72E-06	4.06E-02
Risk behavior	UK-Biobank (ran GWAS ourselves)	~440,000	0.224	0.047	1.77E-06	0.402	0.052	9.54E-15	6.60E-03
Schizophrenia	(*104*)	36,989 cases, 113,075 controls	0.129	0.038	7.83E-04	0.173	0.046	1.65E-04	4.25E-01
Self-rated Health*	(*105*)	111,749	0.126	0.059	3.30E-02	0.108	0.062	7.93E-02	8.38E-01
Smoking: ever smoking	(*106*)	41,969 cases, 32,066 controls	0.205	0.073	4.93E-03	0.446	0.082	4.58E-08	2.02E-02
Subjective well-being	(*101*)	298,420	-0.091	0.075	2.22E-01	-0.286	0.079	2.69E-04	4.16E-02
Waist-to-hip ratio (females)	(*107*)	42,734	0.029	0.074	6.98E-01	0.048	0.085	5.74E-01	8.60E-01
Waist-to-hip ratio (males)	(*107*)	34,635	-0.049	0.080	5.42E-01	0.019	0.094	8.40E-01	5.50E-01
2D:4D finger ratio	(*108*)	12,579	0.237	0.099	1.71E-02	0.002	0.139	9.87E-01	1.45E-01

** Note that self-rated health in this GWAS has been coded such that a higher score indicates worse health.*

*** Traits for which the genetic correlations with same-sex sexual behavior differ significantly between sexes (after correcting for multiple testing) are highlighted in grey.*

Table S20. Genetic analyses of the proportion of same-sex to total partners among non-heterosexuals: SNP-based heritability (A), genetic correlation with same-sex sexual behavior within (B) and across samples (C).

A) SNP-based heritability of proportion of same-sex to total partners among non-heterosexuals.

Study sample	Sex	h^2	Standard error	*p-value*
UK Biobank	Males	0.218	0.080	6.36E-03
	Females	0.267	0.087	2.06E-03
	Combined	0.138	0.042	1.03E-03
23andMe	Males	0.216	0.070	2.16E-03
	Females	0.102	0.075	1.73E-01
	Combined	0.086	0.034	1.14E-02
UK Biobank + 23andMe (MTAG)	Males	0.404	0.092	1.11E-05
	Females	0.324	0.086	1.60E-04
	Combined	0.273	0.047	6.08E-09

B) Genetic correlations (r_g) of proportion of same-sex to total partners among non-heterosexuals with same-sex sexual behavior within cohorts.

		Genetic correlation with same-sex sexual behavior		
Study sample	Sex	r_g	Standard error	*p-value*
UK Biobank	Males	0.036	0.174	8.36E-01
	Females	-0.336	0.184	6.78E-02
	Combined	-0.285	0.154	6.51E-02
23andMe	Males	0.015	0.150	9.18E-01
	Females	-0.441	0.335	1.89E-01
	Combined	-0.150	0.186	4.20E-01
UK Biobank + 23andMe (MTAG)	Males	0.028	0.104	7.88E-01
	Females	-0.309	0.156	4.76E-02
	Combined	-0.185	0.083	2.56E-02

C) Genetic correlations (r_g) of proportion of same-sex to total partners among non-heterosexuals with same-sex sexual behavior across cohorts.

		Genetic correlation with same-sex sexual behavior		
	Sex	r_g	Standard error	*p-value*
UK Biobank vs 23andMe	Males	0.726	0.279	9.40E-03
	Females	0.518	0.346	1.34E-01
	Combined	1.260	0.382	1.00E-03

Table S21. Genetic correlations (r_g) of same-sex sexual behavior (MTAG results of UK Biobank + 23andMe) and proportion of same-sex to total number of sexual partners among non-heterosexuals (MTAG results of UK Biobank + 23andMe) with a range of traits as estimated using LD-score regression.

	MALES						
	Genetic correlation with same-sex sexual behavior			**Genetic correlation with proportion of same-sex to total sexual partners**			***p-value* for difference in r_g*** **
**Phenotype ** **	**r_g**	**Standard Error**	***p-value***	**r_g**	**Standard Error**	***p-value***	
ADHD	0.272	0.060	6.29E-06	-0.089	0.073	2.26E-01	2.20E-04
Age at first birth (females)	-0.235	0.059	5.93E-05	0.114	0.081	1.60E-01	3.75E-04
Age at first birth (males)	-0.193	0.062	1.68E-03	0.201	0.101	4.70E-02	9.75E-04
Age at Menarche	-0.086	0.044	5.32E-02	0.027	0.062	6.68E-01	1.35E-01
Age at Menopause	-0.079	0.061	1.93E-01	0.041	0.086	6.34E-01	2.41E-01
Alcohol use	-0.025	0.057	6.66E-01	-0.097	0.091	2.83E-01	4.79E-01
Anorexia	-0.020	0.095	8.33E-01	-0.006	0.127	9.60E-01	9.32E-01
Anxiety	-0.050	0.142	7.27E-01	-0.057	0.215	7.93E-01	9.79E-01
Autism	0.098	0.069	1.59E-01	-0.245	0.101	1.48E-02	6.11E-03
Bipolar	0.019	0.076	8.08E-01	-0.014	0.102	8.93E-01	8.03E-01
Birth Weight	-0.046	0.048	3.31E-01	-0.033	0.086	6.98E-01	8.94E-01
Cannabis use	0.308	0.058	1.09E-07	-0.433	0.095	5.53E-06	5.10E-11
Height	-0.075	0.033	2.27E-02	-0.111	0.051	2.96E-02	5.65E-01
Loneliness	0.218	0.053	3.83E-05	-0.106	0.084	2.07E-01	7.99E-04
Major depressive disorder (MDD)	0.328	0.055	1.83E-09	-0.222	0.079	4.75E-03	3.34E-09
Neuroticism	0.159	0.052	2.20E-03	0.014	0.073	8.52E-01	9.93E-02

Number of Children (females)	0.013	0.076	8.63E-01	-0.108	0.118	3.59E-01	4.09E-01
Number of Children (males)	0.178	0.091	4.96E-02	-0.644	0.155	3.17E-05	3.85E-06
Number of sex partners	0.407	0.040	3.33E-24	-0.404	0.078	1.92E-07	2.35E-18
Openness to experience	0.135	0.060	2.37E-02	-0.433	0.112	1.18E-04	5.30E-06
Risk Behavior	0.224	0.047	1.77E-06	-0.245	0.078	1.83E-03	2.60E-08
Schizophrenia	0.129	0.038	7.83E-04	-0.031	0.061	6.12E-01	3.24E-02
Self-rated Health*	0.126	0.059	3.30E-02	-0.029	0.080	7.19E-01	1.25E-01
Smoking: ever smoking	0.205	0.073	4.93E-03	-0.209	0.109	5.64E-02	1.32E-03
Subjective well-being	-0.091	0.075	2.22E-01	0.019	0.096	8.45E-01	3.37E-01
Waist-to-hip ratio (females)	0.029	0.074	6.98E-01	0.012	0.113	9.13E-01	9.06E-01
Waist-to-hip ratio (males)	-0.049	0.080	5.42E-01	0.052	0.116	6.57E-01	4.80E-01
2D:4D finger ratio	0.237	0.099	1.71E-02	0.218	0.143	1.28E-01	9.08E-01

	FEMALES						
	Genetic correlation with same-sex sexual behavior			**Genetic correlation with proportion of same-sex to total sexual partners**			***p-value* for difference in r_g*** **
**Phenotype ** **	**r_g**	**Standard Error**	***p-value***	**r_g**	**Standard Error**	***p-value***	
ADHD	0.253	0.059	1.90E-05	-0.003	0.097	9.74E-01	2.52E-02
Age at first birth (females)	-0.146	0.063	2.01E-02	-0.146	0.099	1.42E-01	9.98E-01
Age at first birth (males)	-0.090	0.080	2.60E-01	-0.070	0.117	5.48E-01	8.92E-01
Age at Menarche	-0.049	0.050	3.31E-01	-0.186	0.074	1.19E-02	1.32E-01

Age at Menopause	-0.069	0.064	2.75E-01	-0.211	0.103	4.14E-02	2.50E-01
Alcohol use	0.149	0.065	2.22E-02	-0.032	0.108	7.67E-01	1.46E-01
Anorexia	-0.085	0.100	3.96E-01	0.026	0.149	8.60E-01	5.09E-01
Anxiety	0.251	0.179	1.61E-01	0.279	0.306	3.61E-01	9.34E-01
Autism	0.210	0.068	2.06E-03	-0.159	0.104	1.25E-01	3.16E-03
Bipolar	0.340	0.077	1.09E-05	0.031	0.111	7.77E-01	1.97E-02
Birth Weight	0.035	0.050	4.88E-01	-0.171	0.077	2.61E-02	3.41E-02
Cannabis use	0.671	0.062	1.20E-27	-0.430	0.094	5.11E-06	6.16E-22
Height	-0.043	0.039	2.73E-01	-0.024	0.055	6.57E-01	7.99E-01
Loneliness	0.220	0.053	3.61E-05	-0.075	0.092	4.12E-01	4.42E-03
Major depressive disorder (MDD)	0.438	0.059	1.04E-13	-0.142	0.090	1.15E-01	4.08E-08
Neuroticism	0.220	0.067	1.08E-03	0.026	0.087	7.68E-01	9.88E-02
Number of Children (females)	0.086	0.085	3.12E-01	0.016	0.131	9.03E-01	6.64E-01
Number of Children (males)	0.296	0.107	5.87E-03	-0.047	0.156	7.65E-01	7.62E-02
Number of sex partners	0.681	0.053	1.00E-37	-0.251	0.070	3.00E-04	9.70E-26
Openness to experience	0.312	0.066	2.72E-06	-0.278	0.104	7.31E-03	4.71E-06
Risk Behavior	0.402	0.052	9.54E-15	-0.217	0.079	6.01E-03	3.03E-10
Schizophrenia	0.173	0.046	1.65E-04	0.075	0.065	2.51E-01	2.49E-01
Self-rated Health*	0.108	0.062	7.93E-02	0.184	0.101	6.84E-02	5.20E-01
Smoking: ever smoking	0.446	0.082	4.58E-08	-0.069	0.123	5.73E-01	8.05E-04

Subjective well-being	-0.286	0.079	2.69E-04	-0.014	0.106	8.95E-01	4.94E-02
Waist-to-hip ratio (females)	0.048	0.085	5.74E-01	0.058	0.121	6.32E-01	9.41E-01
Waist-to-hip ratio (males)	0.019	0.094	8.40E-01	0.179	0.140	2.01E-01	3.47E-01
2D:4D finger ratio	0.002	0.139	9.87E-01	-0.010	0.153	9.49E-01	9.51E-01

** Note that self-rated health in this GWAS has been coded such that a higher score indicates worse health.*
** References and sample sizes of the GWASs can be found in Table S19.*
**** Traits for which the genetic correlations with same-sex sexual behavior differ significantly between the two variables (after correcting for multiple testing) are highlighted in grey.*

Table S22. Genetic correlation (r_g) between different degrees of same-sex sexual behavior (defined by the ratio of number of same-sex sexual partners over total number of lifetime sex partners) and exclusively same-sex sexual behavior.

Proportion of same-sex sexual partners / total lifetime sexual partners	**Genetic correlation with exclusively same-sex sexual behavior**	
	r_g [95% CIs]	***p-value***
Less than 1/3 same-sex partners	0.129 [-0.198 - 0.457]	4.39E-01
Between 1/3 and 2/3 same-sex partners	0.795 [0.049 - 1.540]	3.66E-02
More than 2/3 same-sex partners	0.945 [0.797 - 1.094]	8.27E-36

Table S23. Family-based versus SNP-based heritability (h^2) estimates for same-sex sexual behavior as well as a variety of other traits measured in UK Biobank.

Phenotype	Family-based heritability		SNP-based heritability	
	h^2 estimate	SE	h^2 estimate	SE
Same-sex sexual behavior	0.324	0.109	0.095	0.009
Age at menarche	0.513	0.018	0.215	0.009
Age at menopause	0.440	0.050	0.108	0.008
Number of children	0.253	0.017	0.041	0.002
Standing height	0.933	0.011	0.463	0.020
Body mass index	0.581	0.013	0.217	0.007
Risk taking	0.270	0.025	0.089	0.004
Neuroticism	0.282	0.015	0.104	0.004
Alcohol intake frequency	0.435	0.013	0.073	0.003
Fluid intelligence score	0.489	0.042	0.204	0.008
Birth weight	0.632	0.022	0.091	0.005
Handedness	0.213	0.035	0.032	0.004
Overall health rating	0.271	0.019	0.084	0.003
Ever smoked	0.428	0.016	0.107	0.004
Diabetes	0.630	0.052	0.185	0.011
Myocardial infarction	0.268	0.106	0.122	0.009
Educational attainment	0.683	0.029	0.229	0.007

References and Notes

1. ACSF investigators, AIDS and sexual behaviour in France. ACSF investigators. *Nature* **360**, 407–409 (1992). doi:10.1038/360407a0 Medline

2. M. Melbye, R. J. Biggar, Interactions between persons at risk for AIDS and the general population in Denmark. *Am. J. Epidemiol.* **135**, 593–602 (1992). doi:10.1093/oxfordjournals.aje.a116338 Medline

3. S. W. Semenyna, D. P. VanderLaan, L. J. Petterson, P. L. Vasey, Familial patterning and prevalence of male androphilia in Samoa. *J. Sex Res.* **54**, 1077–1084 (2017). doi:10.1080/00224499.2016.1218416 Medline

4. J. M. Bailey, P. L. Vasey, L. M. Diamond, S. M. Breedlove, E. Vilain, M. Epprecht, Sexual orientation, controversy, and science. *Psychol. Sci. Public Interest* **17**, 45–101 (2016). doi:10.1177/1529100616637616 Medline

5. D. Scasta, P. Bialer, American Psychiatric Association, Position statement on issues related to homosexuality (American Psychiatric Association, 2013); www.psychiatry.org/psychiatrists/search-directories-databases/policy-finder (accessed on 26 February 2019).

6. R. C. Pillard, J. M. Bailey, Human sexual orientation has a heritable component. *Hum. Biol.* **70**, 347–365 (1998). Medline

7. N. Långström, Q. Rahman, E. Carlström, P. Lichtenstein, Genetic and environmental effects on same-sex sexual behavior: A population study of twins in Sweden. *Arch. Sex. Behav.* **39**, 75–80 (2010). doi:10.1007/s10508-008-9386-1 Medline

8. D. H. Hamer, S. Hu, V. L. Magnuson, N. Hu, A. M. Pattatucci, A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* **261**, 321–327 (1993). doi:10.1126/science.8332896 Medline

9. G. Rice, C. Anderson, N. Risch, G. Ebers, Male homosexuality: Absence of linkage to microsatellite markers at Xq28. *Science* **284**, 665–667 (1999). doi:10.1126/science.284.5414.665 Medline

10. A. R. Sanders, G. W. Beecham, S. Guo, K. Dawood, G. Rieger, J. A. Badner, E. S. Gershon, R. S. Krishnappa, A. B. Kolundzija, J. Duan, P. V. Gejman, J. M. Bailey, E. R. Martin, MGS Collaboration, Genome-wide association study of male sexual orientation. *Sci. Rep.* **7**, 16950 (2017). doi:10.1038/s41598-017-15736-4 Medline

11. R. A. Lippa, Sex differences and sexual orientation differences in personality: Findings from the BBC Internet survey. *Arch. Sex. Behav.* **37**, 173–187 (2008). doi:10.1007/s10508-007-9267-z Medline

12. Y. Wang, M. Kosinski, Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J. Pers. Soc. Psychol.* **114**, 246–257 (2018). doi:10.1037/pspa0000098 Medline

13. J. M. Bailey, M. P. Dunne, N. G. Martin, Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. *J. Pers. Soc. Psychol.* **78**, 524–536 (2000). doi:10.1037/0022-3514.78.3.524 Medline

14. Materials and methods are available as supplementary materials.

15. A. L. Norris, D. K. Marcus, B. A. Green, Homosexuality as a discrete class. *Psychol. Sci.* **26**, 1843–1853 (2015). doi:10.1177/0956797615598617 Medline

16. B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, L. Duncan, J. R. B. Perry, N. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, B. M. Neale, ReproGen Consortium; Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015). doi:10.1038/ng.3406 Medline

17. P. Turley, R. K. Walters, O. Maghzian, A. Okbay, J. J. Lee, M. A. Fontana, T. A. Nguyen-Viet, R. Wedow, M. Zacher, N. A. Furlotte, P. Magnusson, S. Oskarsson, M. Johannesson, P. M. Visscher, D. Laibson, D. Cesarini, B. M. Neale, D. J. Benjamin, 23andMe Research Team, Social Science Genetic Association Consortium, Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018). doi:10.1038/s41588-017-0009-4 Medline

18. K. Rawlik, O. Canela-Xandri, A. Tenesa, Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* **17**, 166 (2016). doi:10.1186/s13059-016-1025-x Medline

19. A. Okbay *et al.*, Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016). doi:10.1038/nature17671 Medline

20. J. Martin *et al.*, A Genetic investigation of sex bias in the prevalence of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **83**, 1044–1053 (2018). doi:10.1016/j.biopsych.2017.11.026 Medline

21. R. Karlsson Linnér, P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, M. Lebreton, S. P. Tino, A. Abdellaoui, A. R. Hammerschlag, M. G. Nivard, A. Okbay, C. A. Rietveld, P. N. Timshel, M. Trzaskowski, R. Vlaming, C. L. Zünd, Y. Bao, L. Buzdugan, A. H. Caplin, C.-Y. Chen, P. Eibich, P. Fontanillas, J. R. Gonzalez, P. K. Joshi, V. Karhunen, A. Kleinman, R. Z. Levin, C. M. Lill, G. A. Meddens, G. Muntané, S. Sanchez-Roige, F. J. V. Rooij, E. Taskesen, Y. Wu, F. Zhang, A. Auton, J. D. Boardman, D. W. Clark, A. Conlin, C. C. Dolan, U. Fischbacher, P. J. F. Groenen, K. M. Harris, G. Hasler, A. Hofman, M. A. Ikram, S. Jain, R. Karlsson, R. C. Kessler, M. Kooyman, J. MacKillop, M. Männikkö, C. Morcillo-Suarez, M. B. McQueen, K. M. Schmidt, M. C. Smart, M. Sutter, A. R. Thurik, A. G. Uitterlinden, J. White, H. Wit, J. Yang, L. Bertram, D. I. Boomsma, T. Esko, E. Fehr, D. A. Hinds, M. Johannesson, M. Kumari, D. Laibson, P. K. E. Magnusson, M. N. Meyer, A. Navarro, A. A. Palmer, T. H. Pers, D. Posthuma, D. Schunk, M. B. Stein, R. Svento, H. Tiemeier, P. R. H. J. Timmers, P. Turley, R. J. Ursano, G. G. Wagner, J. F. Wilson, J. Gratten, J. J. Lee, D. Cesarini, D. J. Benjamin, P. D. Koellinger, J. P. Beauchamp, 23and Me Research Team, eQTLgen Consortium, International Cannabis Consortium, Social Science Genetic Association Consortium, 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**, 245–257 (2019). doi:10.1038/s41588-018-0309-3 Medline

22. M. E. Goddard, K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, B. J. Hayes, Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. Biol. Sci.* **283**, 20160569 (2016). doi:10.1098/rspb.2016.0569 Medline

23. A. R. Sanders, E. R. Martin, G. W. Beecham, S. Guo, K. Dawood, G. Rieger, J. A. Badner, E. S. Gershon, R. S. Krishnappa, A. B. Kolundzija, J. Duan, P. V. Gejman, J. M. Bailey, Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychol. Med.* **45**, 1379–1388 (2015). doi:10.1017/S0033291714002451 Medline

24. B. J. Vilhjálmsson, J. Yang, H. K. Finucane, A. Gusev, S. Lindström, S. Ripke, G. Genovese, P. R. Loh, G. Bhatia, R. Do, T. Hayeck, H. H. Won, S. Kathiresan, M. Pato, C. Pato, R. Tamimi, E. Stahl, N. Zaitlen, B. Pasaniuc, G. Belbin, E. E. Kenny, M. H. Schierup, P. De Jager, N. A. Patsopoulos, S. McCarroll, M. Daly, S. Purcell, D. Chasman, B. Neale, M. Goddard, P. M. Visscher, P. Kraft, N. Patterson, A. L. Price, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, biology, and risk of inherited variants in breast cancer (DRIVE) study, modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015). doi:10.1016/j.ajhg.2015.09.001 Medline

25. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi:10.1038/ng.3404 Medline

26. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016). doi:10.1038/nature19057 Medline

27. A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for

large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016). doi:10.1038/ng.3506 Medline

28. J. C. Denny, M. D. Ritchie, M. A. Basford, J. M. Pulley, L. Bastarache, K. Brown-Gentry, D. Wang, D. R. Masys, D. M. Roden, D. C. Crawford, PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010). doi:10.1093/bioinformatics/btq126 Medline

29. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015). doi:10.1371/journal.pcbi.1004219 Medline

30. S. R. Jaeger, J. F. McRae, C. M. Bava, M. K. Beresford, D. Hunter, Y. Jia, S. L. Chheang, D. Jin, M. Peng, J. C. Gamble, K. R. Atkinson, L. G. Axten, A. G. Paisley, L. Tooman, B. Pineau, S. A. Rouse, R. D. Newcomb, A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Curr. Biol.* **23**, 1601–1605 (2013). doi:10.1016/j.cub.2013.07.030 Medline

31. H. Kische, A. Arnold, S. Gross, H. Wallaschofski, H. Völzke, M. Nauck, R. Haring, Sex hormones and hair loss in men from the general population of northeastern Germany. *JAMA Dermatol.* **153**, 935–937 (2017). doi:10.1001/jamadermatol.2017.0297 Medline

32. S. Cui, A. Ross, N. Stallings, K. L. Parker, B. Capel, S. E. Quaggin, Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development* **131**, 4095–4105 (2004). doi:10.1242/dev.01266 Medline

33. R. K. Bhandari, I. Sadler-Riggleman, T. M. Clement, M. K. Skinner, Basic helix-loop-helix transcription factor TCF21 is a downstream target of the male sex determining gene SRY. *PLOS ONE* **6**, e19935 (2011). doi:10.1371/journal.pone.0019935 Medline

34. P. F. Sullivan, M. J. Daly, M. O'Donovan, Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nat. Rev. Genet.* **13**, 537–551 (2012). doi:10.1038/nrg3240 Medline

35. T. G. M. Sandfort, R. de Graaf, R. V. Bijl, P. Schnabel, Same-sex sexual behavior and psychiatric disorders: Findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Arch. Gen. Psychiatry* **58**, 85–91 (2001). doi:10.1001/archpsyc.58.1.85 Medline

36. J. Semlyen, M. King, J. Varney, G. Hagger-Johnson, Sexual orientation and symptoms of common mental disorder or low wellbeing: Combined meta-analysis of 12 UK population health surveys. *BMC Psychiatry* **16**, 67 (2016). doi:10.1186/s12888-016-0767-z Medline

37. A. D. Coviello, R. Haring, M. Wellons, D. Vaidya, T. Lehtimäki, S. Keildson, K. L. Lunetta, C. He, M. Fornage, V. Lagou, M. Mangino, N. C. Onland-Moret, B. Chen, J. Eriksson, M. Garcia, Y. M. Liu, A. Koster, K. Lohman, L.-P. Lyytikäinen, A.-K. Petersen, J. Prescott, L. Stolk, L. Vandenput, A. R. Wood, W. V. Zhuang, A. Ruokonen, A.-L. Hartikainen, A. Pouta, S. Bandinelli, R. Biffar, G. Brabant, D. G. Cox, Y. Chen, S. Cummings, L. Ferrucci, M. J. Gunter, S. E. Hankinson, H. Martikainen, A. Hofman, G. Homuth, T. Illig, J.-O. Jansson, A. D. Johnson, D. Karasik, M. Karlsson, J. Kettunen, D. P. Kiel, P. Kraft, J. Liu, Ö. Ljunggren, M. Lorentzon, M. Maggio, M. R. P. Markus, D.

Mellström, I. Miljkovic, D. Mirel, S. Nelson, L. Morin Papunen, P. H. M. Peeters, I. Prokopenko, L. Raffel, M. Reincke, A. P. Reiner, K. Rexrode, F. Rivadeneira, S. M. Schwartz, D. Siscovick, N. Soranzo, D. Stöckl, S. Tworoger, A. G. Uitterlinden, C. H. van Gils, R. S. Vasan, H.-E. Wichmann, G. Zhai, S. Bhasin, M. Bidlingmaier, S. J. Chanock, I. De Vivo, T. B. Harris, D. J. Hunter, M. Kähönen, S. Liu, P. Ouyang, T. D. Spector, Y. T. van der Schouw, J. Viikari, H. Wallaschofski, M. I. McCarthy, T. M. Frayling, A. Murray, S. Franks, M.-R. Järvelin, F. H. de Jong, O. Raitakari, A. Teumer, C. Ohlsson, J. M. Murabito, J. R. B. Perry, A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLOS Genet.* **8**, e1002805 (2012). doi:10.1371/journal.pgen.1002805 Medline

38. C. W. Burke, D. C. Anderson, Sex-hormone-binding globulin is an oestrogen amplifier. *Nature* **240**, 38–40 (1972). doi:10.1038/240038a0 Medline

39. A. C. Kinsey, W. B. Pomeroy, C. E. Martin, *Sexual Behavior in the Human Male* (W.B. Saunders, 1948).

40. F. Klein, *The Bisexual Option* (Routledge, ed. 2, 1993).

41. R. L. Sell, Defining and measuring sexual orientation: A review. *Arch. Sex. Behav.* **26**, 643–658 (1997). doi:10.1023/A:1024528427013 Medline

42. H. Valdes-Socin, M. Rubio Almanza, M. Tomé Fernández-Ladreda, F. G. Debray, V. Bours, A. Beckers, Reproduction, smell, and neurodevelopmental disorders: Genetic defects in different hypogonadotropic hypogonadal syndromes. *Front. Endocrinol.* **5**, 109 (2014). doi:10.3389/fendo.2014.00109 Medline

43. G. M. Herek, A. T. Norton, T. J. Allen, C. L. Sims, Demographic, Psychological, and Social Characteristics of Self-Identified Lesbian, Gay, and Bisexual Adults in a US Probability Sample. *Sex. Res. Social Policy* **7**, 176–200 (2010). doi:10.1007/s13178-010-0017-y Medline

44. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010). doi:10.1093/bioinformatics/btq559 Medline

45. B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, Schizophrenia Working Group of the Psychiatric Genomics Consortium, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015). doi:10.1038/ng.3211 Medline

46. S. H. Lee, N. R. Wray, M. E. Goddard, P. M. Visscher, Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011). doi:10.1016/j.ajhg.2011.02.002 Medline

47. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G.

Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013). doi:10.1038/ng.2653 Medline

48. A. Ganna, andgan/sexual_orientation_GWAS v1.0. Zenodo (2019); doi:10.5281/zenodo.3232892

49. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018). doi:10.1038/s41586-018-0579-z Medline

50. K. M. Harris, C. Tucker Halpern, E. A. Whitsel, J. M. Hussey, L. A. Killeya-Jones, J. Tabor, S. C. Dean, Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). *Int. J. Epidemiol.* 10.1093/ije/dyz115 (2019).

51. M. B. McQueen, J. D. Boardman, B. W. Domingue, A. Smolen, J. Tabor, L. Killeya-Jones, C. T. Halpern, E. A. Whitsel, K. M. Harris, The National Longitudinal Study of Adolescent to Adult Health (Add Health) sibling pairs genome-wide data. *Behav. Genet.* **45**, 12–23 (2015). doi:10.1007/s10519-014-9692-4 Medline

52. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C. Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S. Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. van Duijn, C. E. Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D. Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S. Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A. E. Farmaki, L. Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L. Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L. Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B. Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T. Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van den Berg, W. Van Rheenen, U. Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P. Gasparini, M. G. Sampson, J. F. Wilson, T. Frayling, P. I. de Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler, C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A.

Swaroop, F. Cucca, C. A. Anderson, R. M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin, Haplotype Reference Consortium, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016). doi:10.1038/ng.3643 Medline

53. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016). doi:10.1038/ng.3656 Medline

54. A. R. Sanders, D. F. Levinson, J. Duan, J. M. Dennis, R. Li, K. S. Kendler, J. P. Rice, J. Shi, B. J. Mowry, F. Amin, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. Freedman, C. R. Cloninger, P. V. Gejman, The Internet-based MGS2 control sample: Self report of mental illness. *Am. J. Psychiatry* **167**, 854–865 (2010). doi:10.1176/appi.ajp.2010.09071050 Medline

55. J. Shi, D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'er, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. R. Crowe, J. R. Oksenberg, D. B. Mirel, K. S. Kendler, R. Freedman, P. V. Gejman, Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009). doi:10.1038/nature08192 Medline

56. G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, G. A. McVean, 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010). doi:10.1038/nature09534 Medline

57. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009). doi:10.1371/journal.pgen.1000529 Medline

58. P. R. Loh, G. Tucker, B. K. Bulik-Sullivan, B. J. Vilhjálmsson, H. K. Finucane, R. M. Salem, D. I. Chasman, P. M. Ridker, B. M. Neale, B. Berger, N. Patterson, A. L. Price, Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015). doi:10.1038/ng.3190 Medline

59. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017). doi:10.1038/s41467-017-01261-5 Medline

60. L. R. Lloyd-Jones, M. R. Robinson, J. Yang, P. M. Visscher, Transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio, transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio. *Genetics* **208**, 1397–1408 (2018). doi:10.1534/genetics.117.300360 Medline

61. X. Zhan, Y. Hu, B. Li, G. R. Abecasis, D. J. Liu, RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016). doi:10.1093/bioinformatics/btw079 Medline

62. M. H. Chen, Q. Yang, GWAF: An R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581 (2010). doi:10.1093/bioinformatics/btp710 Medline

63. K. E. Samocha, E. B. Robinson, S. J. Sanders, C. Stevens, A. Sabo, L. M. McGrath, J. A. Kosmicki, K. Rehnström, S. Mallick, A. Kirby, D. P. Wall, D. G. MacArthur, S. B. Gabriel, M. DePristo, S. M. Purcell, A. Palotie, E. Boerwinkle, J. D. Buxbaum, E. H. Cook Jr., R. A. Gibbs, G. D. Schellenberg, J. S. Sutcliffe, B. Devlin, K. Roeder, B. M. Neale, M. J. Daly, A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014). doi:10.1038/ng.3050 Medline

64. A. Ganna, F. K. Satterstrom, S. M. Zekavat, I. Das, M. I. Kurki, C. Churchhouse, J. Alfoldi, A. R. Martin, A. S. Havulinna, A. Byrnes, W. K. Thompson, P. R. Nielsen, K. J. Karczewski, E. Saarentaus, M. A. Rivas, N. Gupta, O. Pietiläinen, C. A. Emdin, F. Lescai, J. Bybjerg-Grauholm, J. Flannick, J. M. Mercader, M. Udler, M. Laakso, V. Salomaa, C. Hultman, S. Ripatti, E. Hämäläinen, J. S. Moilanen, J. Körkkö, O. Kuismin, M. Nordentoft, D. M. Hougaard, O. Mors, T. Werge, P. B. Mortensen, D. MacArthur, M. J. Daly, P. F. Sullivan, A. E. Locke, A. Palotie, A. D. Børglum, S. Kathiresan, B. M. Neale, GoT2D/T2D-GENES Consortium, SIGMA Consortium Helmsley IBD Exome Sequencing Project, FinMetSeq Consortium, iPSYCH-Broad Consortium, Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *Am. J. Hum. Genet.* **102**, 1204–1211 (2018). doi:10.1016/j.ajhg.2018.05.002 Medline

65. A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, GERAD1 Consortium, CRESTAR Consortium, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi:10.1038/s41588-018-0059-2 Medline

66. D. Demontis, R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, C. Churchhouse, A. Dumont, N. Eriksson, M. Gandal, J. I. Goldstein, K. L. Grasby, J. Grove, O. O. Gudmundsson, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, D. P. Howrigan, H. Huang, J. B. Maller, A. R. Martin, N. G. Martin, J. Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, E. B. Robinson, F. K. Satterstrom, H. Stefansson, C. Stevens, P. Turley, G. B. Walters, H. Won, M. J. Wright, O. A. Andreassen, P. Asherson, C. L. Burton, D. I. Boomsma, B. Cormand, S. Dalsgaard, B. Franke, J. Gelernter, D. Geschwind, H. Hakonarson, J. Haavik, H. R. Kranzler, J. Kuntsi, K. Langley, K.-P. Lesch, C. Middeldorp, A. Reif, L. A.

Rohde, P. Roussos, R. Schachar, P. Sklar, E. J. S. Sonuga-Barke, P. F. Sullivan, A. Thapar, J. Y. Tung, I. D. Waldman, S. E. Medland, K. Stefansson, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, M. J. Daly, S. V. Faraone, A. D. Børglum, B. M. Neale, ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019). doi:10.1038/s41588-018-0269-7 Medline

67. H. K. Finucane, Y. A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, S. Gazal, P.-R. Loh, C. Lareau, N. Shoresh, G. Genovese, A. Saunders, E. Macosko, S. Pollack, J. R. B. Perry, J. D. Buenrostro, B. E. Bernstein, S. Raychaudhuri, S. McCarroll, B. M. Neale, A. L. Price, Brainstorm Consortium, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018). doi:10.1038/s41588-018-0081-4 Medline

68. K. I. Kim, M. A. van de Wiel, Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics* **9**, 114 (2008). doi:10.1186/1471-2105-9-114 Medline

69. J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, H. Parkinson, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45** (D1), D896–D901 (2017). doi:10.1093/nar/gkw1133 Medline

70. W. J. Peyrot, D. I. Boomsma, B. W. Penninx, N. R. Wray, Disease and polygenic architecture: Avoid trio design and appropriately account for unscreened control subjects for common disease. *Am. J. Hum. Genet.* **98**, 382–391 (2016). doi:10.1016/j.ajhg.2015.12.017 Medline

71. C. H. Mercer, C. Tanton, P. Prah, B. Erens, P. Sonnenberg, S. Clifton, W. Macdowall, R. Lewis, N. Field, J. Datta, A. J. Copas, A. Phelps, K. Wellings, A. M. Johnson, Changes in sexual attitudes and lifestyles in Britain through the life course and over time: Findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet* **382**, 1781–1794 (2013). doi:10.1016/S0140-6736(13)62035-8 Medline

72. GALLUP, In U.S., Estimate of LGBT population rises to 4.5% (2018); https://news.gallup.com/poll/234863/estimate-lgbt-population-rises.aspx.

73. B. P. Zietsch, in *Psychiatric Disorders—Worldwide Advances*, T. Uehara, Ed. (InTech, 2011), pp. 277–300.

74. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Sunderland, 1998), vol. 1.

75. S. V. Eastwood, R. Mathur, M. Atkinson, S. Brophy, C. Sudlow, R. Flaig, S. de Lusignan, N. Allen, N. Chaturvedi, Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLOS ONE* **11**, e0162388 (2016). doi:10.1371/journal.pone.0162388 Medline

76. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome

association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795 Medline

77. R. A. Power, S. Kyaga, R. Uher, J. H. MacCabe, N. Långström, M. Landen, P. McGuffin, C. M. Lewis, P. Lichtenstein, A. C. Svensson, Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* **70**, 22–30 (2013). doi:10.1001/jamapsychiatry.2013.268 Medline

78. J. K. Pickrell, T. Berisa, J. Z. Liu, L. Ségurel, J. Y. Tung, D. A. Hinds, Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016). doi:10.1038/ng.3570 Medline

79. C. Tian, B. S. Hromatka, A. K. Kiefer, N. Eriksson, S. M. Noble, J. Y. Tung, D. A. Hinds, Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* **8**, 599 (2017). doi:10.1038/s41467-017-00257-5 Medline

80. J. F. McRae, S. R. Jaeger, C. M. Bava, M. K. Beresford, D. Hunter, Y. Jia, S. L. Chheang, D. Jin, M. Peng, J. C. Gamble, K. R. Atkinson, L. G. Axten, A. G. Paisley, L. Williams, L. Tooman, B. Pineau, S. A. Rouse, R. D. Newcomb, Identification of regions associated with variation in sensitivity to food-related odors in the human genome. *Curr. Biol.* **23**, 1596–1600 (2013). doi:10.1016/j.cub.2013.07.031 Medline

81. W. J. Astle, H. Elding, T. Jiang, D. Allen, D. Ruklisa, A. L. Mann, D. Mead, H. Bouman, F. Riveros-Mckay, M. A. Kostadima, J. J. Lambourne, S. Sivapalaratnam, K. Downes, K. Kundu, L. Bomba, K. Berentsen, J. R. Bradley, L. C. Daugherty, O. Delaneau, K. Freson, S. F. Garner, L. Grassi, J. Guerrero, M. Haimel, E. M. Janssen-Megens, A. Kaan, M. Kamat, B. Kim, A. Mandoli, J. Marchini, J. H. A. Martens, S. Meacham, K. Megy, J. O'Connell, R. Petersen, N. Sharifi, S. M. Sheard, J. R. Staley, S. Tuna, M. van der Ent, K. Walter, S.-Y. Wang, E. Wheeler, S. P. Wilder, V. Iotchkova, C. Moore, J. Sambrook, H. G. Stunnenberg, E. Di Angelantonio, S. Kaptoge, T. W. Kuijpers, E. Carrillo-de-Santa-Pau, D. Juan, D. Rico, A. Valencia, L. Chen, B. Ge, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yang, R. Guigo, S. Beck, D. S. Paul, T. Pastinen, D. Bujold, G. Bourque, M. Frontini, J. Danesh, D. J. Roberts, W. H. Ouwehand, A. S. Butterworth, N. Soranzo, The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016). doi:10.1016/j.cell.2016.10.042 Medline

82. J. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Karlsson Linnér, M. A. Fontana, T. Kundu, C. Lee, H. Li, R. Li, R. Royer, P. N. Timshel, R. K. Walters, E. A. Willoughby, L. Yengo, M. Alver, Y. Bao, D. W. Clark, F. R. Day, N. A. Furlotte, P. K. Joshi, K. E. Kemper, A. Kleinman, C. Langenberg, R. Mägi, J. W. Trampush, S. S. Verma, Y. Wu, M. Lam, J. H. Zhao, Z. Zheng, J. D. Boardman, H. Campbell, J. Freese, K. M. Harris, C. Hayward, P. Herd, M. Kumari, T. Lencz, J. Luan, A. K. Malhotra, A. Metspalu, L. Milani, K. K. Ong, J. R. B. Perry, D. J. Porteous, M. D. Ritchie, M. C. Smart, B. H. Smith, J. Y. Tung, N. J. Wareham, J. F. Wilson, J. P. Beauchamp, D. C. Conley, T. Esko, S. F. Lehrer, P. K. E. Magnusson, S. Oskarsson, T. H. Pers, M. R. Robinson, K. Thom, C. Watson, C. F. Chabris, M. N. Meyer, D. I. Laibson, J. Yang, M. Johannesson, P. D. Koellinger, P.

Turley, P. M. Visscher, D. J. Benjamin, D. Cesarini, 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018). doi:10.1038/s41588-018-0147-3 Medline

83. T. W. Winkler *et al.*, The influence of age and sex on genetic associations with adult body size and shape: A large-scale genome-wide interaction study. *PLOS Genet.* **11**, e1005378 (2015). doi:10.1371/journal.pgen.1005378 Medline

84. D. F. Levinson, J. Shi, K. Wang, S. Oh, B. Riley, A. E. Pulver, D. B. Wildenauer, C. Laurent, B. J. Mowry, P. V. Gejman, M. J. Owen, K. S. Kendler, G. Nestadt, S. G. Schwab, J. Mallet, D. Nertney, A. R. Sanders, N. M. Williams, B. Wormley, V. K. Lasseter, M. Albus, S. Godard-Bauché, M. Alexander, J. Duan, M. C. O'Donovan, D. Walsh, A. O'Neill, G. N. Papadimitriou, D. Dikeos, W. Maier, B. Lerer, D. Campion, D. Cohen, M. Jay, A. Fanous, P. Eichhammer, J. M. Silverman, N. Norton, N. Zhang, H. Hakonarson, C. Gao, A. Citri, M. Hansen, S. Ripke, F. Dudbridge, P. A. Holmans, Schizophrenia Psychiatric GWAS Consortium, Genome-wide association study of multiplex schizophrenia pedigrees. *Am. J. Psychiatry* **169**, 963–973 (2012). doi:10.1176/appi.ajp.2012.11091423 Medline

85. D. T. Uehara, S. Hayashi, N. Okamoto, S. Mizuno, Y. Chinen, R. Kosaki, T. Kosho, K. Kurosawa, H. Matsumoto, H. Mitsubuchi, H. Numabe, S. Saitoh, Y. Makita, A. Hata, I. Imoto, J. Inazawa, SNP array screening of cryptic genomic imbalances in 450 Japanese subjects with intellectual disability and multiple congenital anomalies previously negative for large rearrangements. *J. Hum. Genet.* **61**, 335–343 (2016). doi:10.1038/jhg.2015.154 Medline

86. G. M. Lowe, K. Vlismas, D. L. Graham, M. Carail, C. Caris-Veyrat, A. J. Young, The degradation of (all-E)-beta-carotene by cigarette smoke. *Free Radic. Res.* **43**, 280–286 (2009). doi:10.1080/10715760802691497 Medline

87. D. S. Manoli, P. Fan, E. J. Fraser, N. M. Shah, Neural control of sexually dimorphic behaviors. *Curr. Opin. Neurobiol.* **23**, 330–338 (2013). doi:10.1016/j.conb.2013.04.005 Medline

88. Z. Xu, X. Gao, Y. He, J. Ju, M. Zhang, R. Liu, Y. Wu, C. Ma, C. Ma, Z. Lin, X. Huang, Q. Zhao, Synergistic effect of SRY and its direct target, WDR5, on Sox9 expression. *PLOS ONE* **7**, e34327 (2012). doi:10.1371/journal.pone.0034327 Medline

89. N. Barban *et al.*, Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016). doi:10.1038/ng.3698 Medline

90. J. R. Perry *et al.*, Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014). doi:10.1038/nature13545 Medline

91. F. R. Day *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). doi:10.1038/ng.3412 Medline

92. T. K. Clarke, M. J. Adams, G. Davies, D. M. Howard, L. S. Hall, S. Padmanabhan, A. D. Murray, B. H. Smith, A. Campbell, C. Hayward, D. J. Porteous, I. J. Deary, A. M.

McIntosh, Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol. Psychiatry* **22**, 1376–1384 (2017). doi:10.1038/mp.2017.153 Medline

93. V. Boraska *et al.*, A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085–1094 (2014). doi:10.1038/mp.2013.187 Medline

94. T. Otowa, K. Hek, M. Lee, E. M. Byrne, S. S. Mirza, M. G. Nivard, T. Bigdeli, S. H. Aggen, D. Adkins, A. Wolen, A. Fanous, M. C. Keller, E. Castelao, Z. Kutalik, S. Van der Auwera, G. Homuth, M. Nauck, A. Teumer, Y. Milaneschi, J.-J. Hottenga, N. Direk, A. Hofman, A. Uitterlinden, C. L. Mulder, A. K. Henders, S. E. Medland, S. Gordon, A. C. Heath, P A F. Madden, M. L. Pergadia, P. J. van der Most, I. M. Nolte, F. V. A. van Oort, C. A. Hartman, A. J. Oldehinkel, M. Preisig, H. J. Grabe, C. M. Middeldorp, B. W. J. H. Penninx, D. Boomsma, N. G. Martin, G. Montgomery, B. S. Maher, E. J. van den Oord, N. R. Wray, H. Tiemeier, J. M. Hettema, Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016). doi:10.1038/mp.2015.197 Medline

95. J. Grove, S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, S. Awashti, R. Belliveau, F. Bettella, J. D. Buxbaum, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, J. H. Christensen, C. Churchhouse, K. Dellenvall, D. Demontis, S. De Rubeis, B. Devlin, S. Djurovic, A. L. Dumont, J. I. Goldstein, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, S. Hope, D. P. Howrigan, H. Huang, C. M. Hultman, L. Klei, J. Maller, J. Martin, A. R. Martin, J. L. Moran, M. Nyegaard, T. Nærland, D. S. Palmer, A. Palotie, C. B. Pedersen, M. G. Pedersen, T. dPoterba, J. B. Poulsen, B. S. Pourcain, P. Qvist, K. Rehnström, A. Reichenberg, J. Reichert, E. B. Robinson, K. Roeder, P. Roussos, E. Saemundsen, S. Sandin, F. K. Satterstrom, G. Davey Smith, H. Stefansson, S. Steinberg, C. R. Stevens, P. F. Sullivan, P. Turley, G. B. Walters, X. Xu, K. Stefansson, D. H. Geschwind, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, B. M. Neale, M. J. Daly, A. D. Børglum; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; BUPGEN; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; 23andMe Research Team, Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019). doi:10.1038/s41588-019-0344-8 Medline

96. Psychiatric GWAS Consortium Bipolar Disorder Working Group, Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011). doi:10.1038/ng.943 Medline

97. M. Horikoshi *et al.*, Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248–252 (2016). doi:10.1038/nature19806 Medline

98. J. A. Pasman, K. J. H. Verweij, Z. Gerring, S. Stringer, S. Sanchez-Roige, J. L. Treur, A. Abdellaoui, M. G. Nivard, B. M. L. Baselmans, J.-S. Ong, H. F. Ip, M. D. van der Zee, M. Bartels, F. R. Day, P. Fontanillas, S. L. Elson, H. de Wit, L. K. Davis, J. MacKillop, J. L. Derringer, S. J. T. Branje, C. A. Hartman, A. C. Heath, P. A. C. van Lier, P. A. F. Madden, R. Mägi, W. Meeus, G. W. Montgomery, A. J. Oldehinkel, Z. Pausova, J. A. Ramos-Quiroga, T. Paus, M. Ribases, J. Kaprio, M. P. M. Boks, J. T. Bell, T. D. Spector, J. Gelernter, D. I. Boomsma, N. G. Martin, S. MacGregor, J. R. B. Perry, A. A. Palmer,

D. Posthuma, M. R. Munafò, N. A. Gillespie, E. M. Derks, J. M. Vink; 23andMe Research Team; Substance Use Disorders Working Group of the Psychiatric Genomics Consortium; International Cannabis Consortium, GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. *Nat. Neurosci.* **21**, 1161–1170 (2018). doi:10.1038/s41593-018-0206-1 Medline

99. A. R. Wood *et al.*, Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). doi:10.1038/ng.3097 Medline

100. N. R. Wray *et al.*, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018). doi:10.1038/s41588-018-0090-3 Medline

101. A. Okbay *et al.*, Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016). doi:10.1038/ng.3552 Medline

102. M. H. M. de Moor, P. T. Costa, A. Terracciano, R. F. Krueger, E. J. C. de Geus, T. Toshiko, B. W. J. H. Penninx, T. Esko, P A F. Madden, J. Derringer, N. Amin, G. Willemsen, J.-J. Hottenga, M. A. Distel, M. Uda, S. Sanna, P. Spinhoven, C. A. Hartman, P. Sullivan, A. Realo, J. Allik, A. C. Heath, M. L. Pergadia, A. Agrawal, P. Lin, R. Grucza, T. Nutile, M. Ciullo, D. Rujescu, I. Giegling, B. Konte, E. Widen, D. L. Cousminer, J. G. Eriksson, A. Palotie, L. Peltonen, M. Luciano, A. Tenesa, G. Davies, L. M. Lopez, N. K. Hansell, S. E. Medland, L. Ferrucci, D. Schlessinger, G. W. Montgomery, M. J. Wright, Y. S. Aulchenko, A. C. J. W. Janssens, B. A. Oostra, A. Metspalu, G. R. Abecasis, I. J. Deary, K. Räikkönen, L. J. Bierut, N. G. Martin, C. M. van Duijn, D. I. Boomsma, Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* **17**, 337–349 (2012). doi:10.1038/mp.2010.128 Medline

103. M. T. Lo, D. A. Hinds, J. Y. Tung, C. Franz, C.-C. Fan, Y. Wang, O. B. Smeland, A. Schork, D. Holland, K. Kauppi, N. Sanyal, V. Escott-Price, D. J. Smith, M. O'Donovan, H. Stefansson, G. Bjornsdottir, T. E. Thorgeirsson, K. Stefansson, L. K. McEvoy, A. M. Dale, O. A. Andreassen, C.-H. Chen, Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017). doi:10.1038/ng.3736 Medline

104. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014). doi:10.1038/nature13595 Medline

105. S. E. Harris, S. P. Hagenaars, G. Davies, W. David Hill, D. C. M. Liewald, S. J. Ritchie, R. E. Marioni, C. L. M. Sudlow, J. M. Wardlaw, A. M. McIntosh, C. R. Gale, I. J. Deary; METASTROKE Consortium, International Consortium for Blood Pressure Genome-Wide Association Studies; International Consortium for Blood Pressure Genome-Wide Association Studies; CHARGE Consortium Aging and Longevity Group; CHARGE Consortium Cognitive Group, Molecular genetic contributions to self-rated health. *Int. J. Epidemiol.* **46**, 994–1009 (2017). Medline

106. Tobacco and Genetics Consortium, Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010). doi:10.1038/ng.571 Medline

107. J. C. Randall *et al.*, Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLOS Genet.* **9**, e1003500 (2013). doi:10.1371/journal.pgen.1003500 Medline

108. N. M. Warrington, E. Shevroja, G. Hemani, P. G. Hysi, Y. Jiang, A. Auton, C. G. Boer, M. Mangino, C. A. Wang, J. P. Kemp, G. McMahon, C. Medina-Gomez, M. Hickey, K. Trajanoska, D. Wolke, M. A. Ikram, G. W. Montgomery, J. F. Felix, M. J. Wright, D. A. Mackey, V. W. Jaddoe, N. G. Martin, J. Y. Tung, G. Davey Smith, C. E. Pennell, T. D. Spector, J. van Meurs, F. Rivadeneira, S. E. Medland, D. M. Evans; 23andMe Research Team, Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Hum. Mol. Genet.* **27**, 2025–2038 (2018). doi:10.1093/hmg/ddy121 Medline