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

**20 authors**, including:

Andrea Ganna
Broad Institute of MIT and Harvard
**142** PUBLICATIONS **4,435** CITATIONS

Karin Verweij
Queensland Institute of Medical Research
**121** PUBLICATIONS **1,863** CITATIONS

Michel Nivard
Vrije Universiteit Amsterdam
**167** PUBLICATIONS **2,794** CITATIONS

Alexander S Busch
Rigshospitalet
**45** PUBLICATIONS **120** CITATIONS

Some of the authors of this publication are also working on these related projects:

The neurobiology of musical expertise View project

Molecular Genetics View project

All content following this page was uploaded by Brendan Zietsch on 03 September 2019.

The user has requested enhancement of the downloaded file.

## HUMAN GENETICS

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

Andrea Ganna[1,2,3,4]*, Karin J. H. Verweij[5]*, Michel G. Nivard[6], Robert Maier[1,2,3], Robbee Wedow[1,3,7,8,9,10,11], Alexander S. Busch[12,13,14], Abdel Abdellaoui[5], Shengru Guo[15], J. Fah Sathirapongsasuti[16], 23andMe Research Team[16], Paul Lichtenstein[4], Sebastian Lundström[17], Niklas Långström[4], Adam Auton[16], Kathleen Mullan Harris[18,19], Gary W. Beecham[15], Eden R. Martin[15], Alan R. Sanders[20,21], John R. B. Perry[12]†, Benjamin M. Neale[1,2,3]†, Brendan P. Zietsch[22]†‡

Twin and family studies have shown that same-sex sexual behavior is partly genetically influenced, but previous searches for specific genes involved have been underpowered. We performed a genome-wide association study (GWAS) on 477,522 individuals, revealing five loci significantly associated with same-sex sexual behavior. In aggregate, all tested genetic variants accounted for 8 to 25% of variation in same-sex sexual behavior, only partially overlapped between males and females, and do not allow meaningful prediction of an individual's sexual behavior. Comparing these GWAS results with those for the proportion of same-sex to total number of sexual partners among nonheterosexuals suggests that there is no single continuum from opposite-sex to same-sex sexual behavior. Overall, our findings provide insights into the genetics underlying same-sex sexual behavior and underscore the complexity of sexuality.

A cross human societies and in both sexes, some 2 to 10% of individuals report engaging in sex with same-sex partners, either exclusively or in addition to sex with opposite-sex partners (1–4). The biological factors that contribute to sexual preference are largely unknown (5), but genetic influences are suggested by the observation that same-sex sexual behavior appears to run in families (6) and is concordant more often in genetically identical (monozygotic) twin pairs than in fraternal twin pairs or siblings (7).

With respect to genetic influences, several questions arise. First, what genes are involved and what biological processes do they affect? Previous reports of genetic variants associated with sexual orientation (8–10) were based on relatively small samples and did not meet current standards of genome-wide significance ($P < 5 \times 10^{-8}$). Identification of robustly associated variants could enable exploration of the biological pathways and processes involved in development of same-sex sexual behavior. One hypothesis suggests that sex hormones are involved (11–13), but little direct genetic or biological evidence is available. Second, to what extent are genetic influences the same or different for females and males; behavior, attraction, and identity; and heterosexuality and different same-sex sexual behaviors (such as bisexuality)?

In order to identify genetic variants associated with same-sex sexual behavior and explore its genetic architecture and underlying biology, we performed a genome-wide association study (GWAS) of same-sex sexual behavior. Analyses were conducted in the UK Biobank from the United Kingdom and a cohort of research participants from 23andMe, predominantly located in the United States, and replications were performed in three other smaller studies. This study is part of a preregistered research plan (Open Science Framework; https://osf.io/357tn), and we explain our deviations from that plan in (14).

### Phenotypic characterization

The UK Biobank study comprises a sample of ~500,000 genotyped UK residents aged 40 to 70 years (tables S1 and S2) (14). Our primary phenotype of interest is a binary, self-reported measure of whether respondents had ever had sex with someone of the same sex (here termed "nonheterosexuals") or had not (here termed "heterosexuals") (Box 1).

In the UK Biobank sample, 4.1% of males and 2.8% of females reported ever having had sex with someone of the same sex (tables S1 and S2), with higher rates among younger participants (Fig. 1A). This binary phenotype follows from previous work proposing that sexual preference is taxonic rather than dimensional in structure, with individuals reporting exclusively opposite-sex preference differing from individuals reporting any same-sex preference (15). However, the binary variable also collapses rich and multi-faceted diversity among nonheterosexual individuals (15), so we explored finer-scaled measurements and some of the complexities of the phenotype, although intricacies of the social and cultural influences on sexuality made it impossible to fully explore this complexity. The 23andMe sample comprised 23andMe customers who consented to participate in research and chose to complete a survey about sexual orientation (from many possible survey topics). Individuals who engage in same-sex sexual behavior may be more likely to self-select the sexual orientation survey, which would explain the unusually high proportion of individuals who had had same-sex sexual partners in this sample (18.9%) (table S3) (14).

We also performed replication analyses in three smaller datasets (14): (i) Molecular Genetic Study of Sexual Orientation (MGSOSO) (n = 2308 U.S. adult males), in which respondents were asked about their sexual identity; (ii) Add Health (n = 4755 U.S. young adults), in which respondents were asked whether they ever had same-sex intercourse and whether they were romantically attracted to the same sex; and (iii) Child and Adolescent Twin Study in Sweden (CATSS) (n = 8093 Swedish adolescents), in which

[1]Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [5]Department of Psychiatry, Amsterdam University Medical Centers (UMC), location AMC, University of Amsterdam, Meibergdreef 5, 1105 AZ Amsterdam, Netherlands. [6]Department of Biological Psychology, Vrije Universiteit Amsterdam, 1081 BT, Amsterdam, Netherlands. [7]Department of Sociology, Harvard University, Cambridge, MA 02138, USA. [8]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. [9]Department of Sociology, University of Colorado, Boulder, CO 80309-0483, USA. [10]Health and Society Program and Population Program, Institute of Behavioral Science, University of Colorado, Boulder, CO 80309-0483, USA. [11]Institute for Behavioral Genetics, University of Colorado, Boulder, CO 80309-0483, USA. [12]Medical Research Council (MRC) Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. [13]Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [14]International Center for Research and Research Training in Endocrine Disruption of Male Reproduction and Child Health (EDMaRC), Rigshospitalet, Copenhagen, Denmark. [15]Department of Human Genetics, University of Miami, Miami, FL 33136, USA. [16]23andMe, Mountain View, CA 94041, USA. [17]Centre for Ethics, Law and Mental Health, Gillberg Neuropsychiatry Centre, University of Gothenburg, Sweden. [18]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, USA. [19]Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [20]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, Evanston, IL 60201, USA. [21]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA. [22]Centre for Psychology and Evolution, School of Psychology, University of Queensland, St. Lucia, Brisbane QLD 4072, Australia.
*These authors contributed equally to this work. †These authors contributed equally to this work.
‡Corresponding author. Email: zietsch@psy.uq.edu.au

**Box 1. Phenotype and sample definition and limitations.**

In this study, we use the term "same-sex sexual behavior," which is defined as having ever had sex with someone of the same sex. Detailed descriptions of the variables used in the different cohorts can be found in the supplementary materials (*14*).

To aid in readability throughout the manuscript, in some places we refer to individuals who have ever had sex with someone of the same sex as "nonheterosexuals," whereas we refer to individuals who have never had sex with someone of the same sex as "heterosexuals."

We acknowledge that the grouping phrase "nonheterosexuals" has the potential to present messages of othering (that is, undesirable marginalization of another person or group on the basis of their sexual expression)—by defining an "outgroup" in reference to an "ingroup" and implying that "nonheterosexual behavior" may have a negative connotation, whereas "heterosexual behavior" may have a positive one. We wish to make clear that our choice of language is not meant to forward messages of othering on the basis of sexual behavior.

Throughout this manuscript, we use the terms "female" and "male" rather than "woman" and "man." This is because our analyses and results relate to biologically defined sex, not to gender.

As is common in genetic analyses, we dropped individuals from our study whose biological sex and self-identified sex/gender did not match. This is an important limitation of our analyses because the analyses do not include transgender persons, intersex persons, and other important persons and groups within the queer community. We hope that this limitation will be addressed in future work.
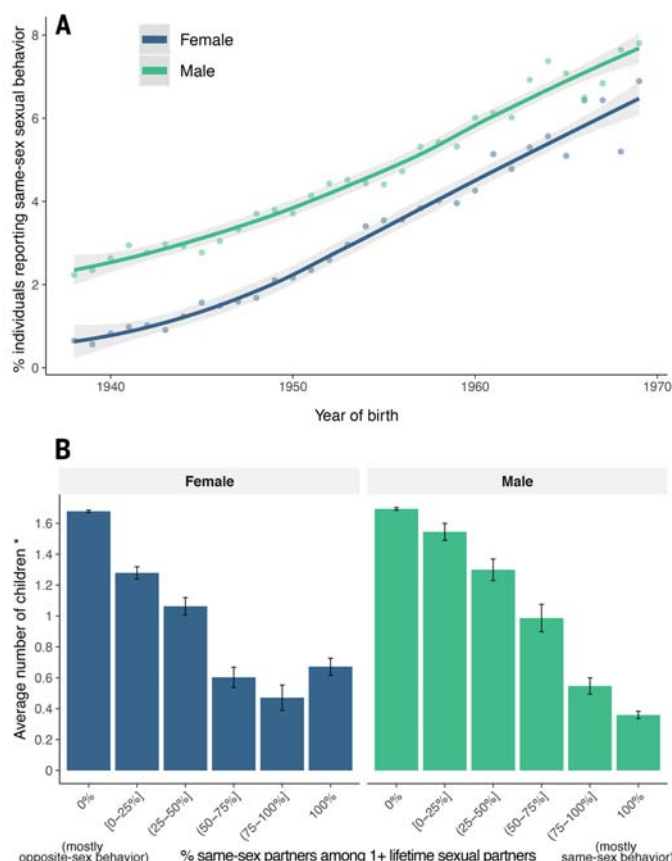
**Fig. 1. Descriptive statistics regarding same-sex sexual behavior in the UK Biobank.** (**A**) The percentage of participants in the UK Biobank who reported having had at least one same-sex sexual partner (*y* axis) increased with participants' year of birth (*x* axis). (**B**) Among participants reporting at least one same-sex partner, those with a greater proportion of same-sex partners (*x* axis) have a larger reproductive disadvantage (lower birth-year adjusted number of children) (*y* axis). Vertical bars represent 95% CIs.

participants reported the degree of attraction to the same versus opposite sex.

We observed in the UK Biobank that individuals who reported same-sex sexual behavior had on average fewer offspring than those of individuals who engaged exclusively in heterosexual behavior, even for individuals reporting only a minority of same-sex partners (Fig. 1B). This reduction in number of children is comparable with or greater than for other traits that have been linked to lower fertility rates (fig. S1) (*14*). This reproductive deficit raises questions about the evolutionary maintenance of the trait, but we do not address these here.

## Genetic architecture of same-sex sexual behavior

We first assessed whether same-sex sexual behavior clustered in families in a manner consistent with genetic influences on the phenotype. Among pairs of individuals in the UK Biobank related at full cousin or closer [as identified by genomic similarity (*14*); *n* pairs = 106,979], more closely related individuals were more likely to be concordant in terms of same-sex sexual behavior. By modeling the correspondence of relatedness among individuals and the similarity of their sexual behavior, we estimated broad-sense heritability—the percentage of variation in a trait attributable to genetic variation—at 32.4% [95% confidence intervals (CIs), 10.6 to 54.3] (table S4). This estimate is consistent with previous estimates from smaller twin studies (*7*).

To identify genetic variants [largely single-nucleotide polymorphisms (SNPs)] associated with same-sex sexual behavior, we performed a GWAS in the UK Biobank study (*n* = 408,995 individuals) (*14*). To increase power and generalizability of our results, we also performed a GWAS in the cohort from 23andMe using an equivalent variable (individuals who reported having had sex with "Other sex only" versus the other options on a seven-point scale regarding participants' sexual partners) (*n* = 68,527 individuals, of which 12,933 reported same-sex sexual behavior) (table S3) (*14*). We estimated the genetic correlation (*16*) between different heritable traits to determine the degree of consistency of genetic influences on same-sex sexual behavior in the two studies, which was high [genetic correlation coefficient ($r_g$) = 0.87; 95% CIs, 0.67 to 1.06] (table S5) (*14*). Genetic correlations between same-sex sexual behavior and 28 different traits were largely similar in the UK Biobank and 23andMe (fig. S2) (*14*), although a few differences were observed; for example, in females, the genetic correlations between same-sex sexual behavior and anorexia were in opposite directions in the UK Biobank ($r_g$ = −0.36; 95% CIs, −0.60 and −0.09) and 23andMe data ($r_g$ = 0.36; 95% CIs, 0.08 to 0.65; Wald test *P* value for differences = 0.0001). Overall, these results indicate that the genetic influences on same-sex sexual behavior in the two samples is similar, although there is some suggestion of phenotypic heterogeneity. We meta-analyzed the two sample sets using MTAG (*17*), which models
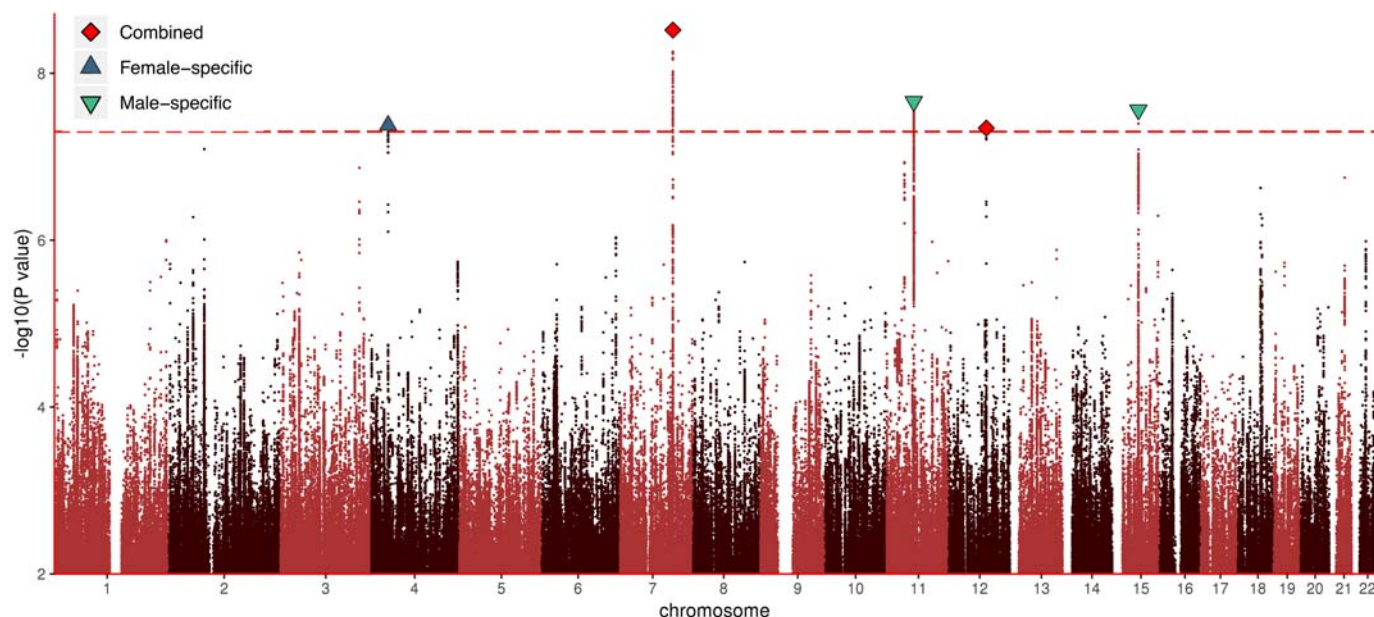
**Fig. 2. Manhattan plot for a GWAS of same-sex sexual behavior.** Diamonds (red) represent genome-wide significant signals from analysis of males and females combined, and triangles represent genome-wide significant signals that are female (pointing up, blue) or male (pointing down, green) specific.

their genetic correlation to determine the meta-analytic weights, yielding a total sample size of 477,522 individuals (26,827 individuals reporting same-sex sexual behavior).

After standard quality control checks (table S6) (14), we identified two genome-wide significant signals for same-sex sexual behavior (rs11114975-12q21.31 and rs10261857-7q31.2) (Fig. 2 and tables S7 and S8). We discuss these SNPs further in the section "In-silico follow-up of GWAS results." To assess differences in effects between females and males, we also performed sex-specific analyses. These results suggested only a partially shared genetic architecture across the sexes; the across-sex genetic correlation was 0.63 (95% CIs, 0.48 to 0.78) (table S9). This is noteworthy given that most other studied traits show much higher across-sex genetic correlations, often close to 1 (18–21). Through the sex-specific analyses, we identified two additional signals in males (rs28371400-15q21.3 and rs34730029-11q12.1), which showed no significant association in females, and one in females (rs13135637-4p14), which showed no significant association in males. Overall, three of the SNPs replicated at a nominal P value in the meta-analyzed replication datasets (Wald test P = 0.027 for rs34730029, P = 0.003 for rs28371400, and P = 0.006 for rs11114975) (table S10), despite the much smaller sample size (MGSOSO, Add Health, and CATSS; total sample size = 15,156 individuals, effective sample size = 4887 individuals).

The SNPs that reached genome-wide significance had very small effects (odds ratios ~1.1) (table S7). For example, in the UK Biobank, males with a GT genotype at the rs34730029 locus had 0.4% higher prevalence of same-sex sexual behavior than those with a TT genotype (4.0 versus 3.6%). Nevertheless, the contribution of all
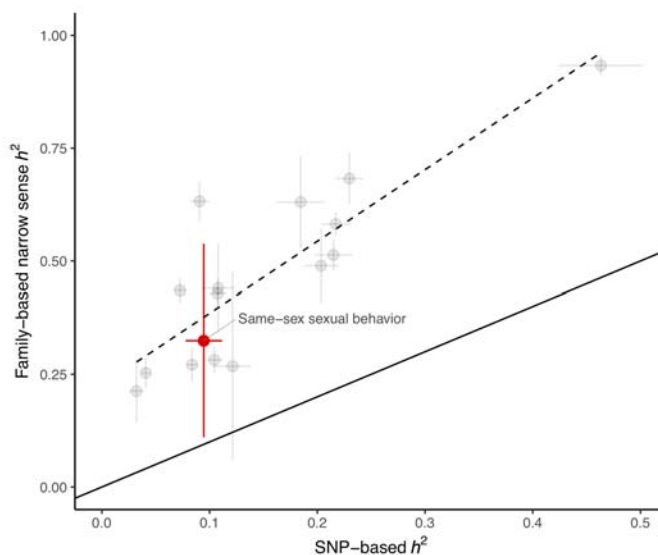


**Fig. 3. SNP-based versus family-based heritability estimates for same-sex sexual behavior compared with a variety of other traits.** Heritability, $h^2$; same-sex sexual behavior, red dot; other traits, gray dots. The estimates for all traits are provided in table S23. Horizontal bars represent 95% CIs for the SNP-based estimate, and vertical bars represent 95% CIs for the family-based estimate. Dashed and solid lines represent the observed (obtained by linear regression) and expected relationship between family-based and SNP-based heritability, respectively.

measured common SNPs in aggregate (SNP-based heritability) was estimated to be 8 to 25% (95% CIs, 5 to 30%) of variation in female and male same-sex sexual behavior, in which the range reflects differing estimates by using different analysis methods or prevalence assumptions (table S11) (14). The discrepancy between the variance captured by the significant SNPs and all common SNPs suggests that same-sex

sexual behavior, like most complex human traits, is influenced by the small, additive effects of very many genetic variants, most of which cannot be detected at the current sample size (22). Consistent with this interpretation, we show that the contribution of each chromosome to heritability is broadly proportional to its size (fig. S3) (14). In contrast to linkage studies that found substantial association of sexual
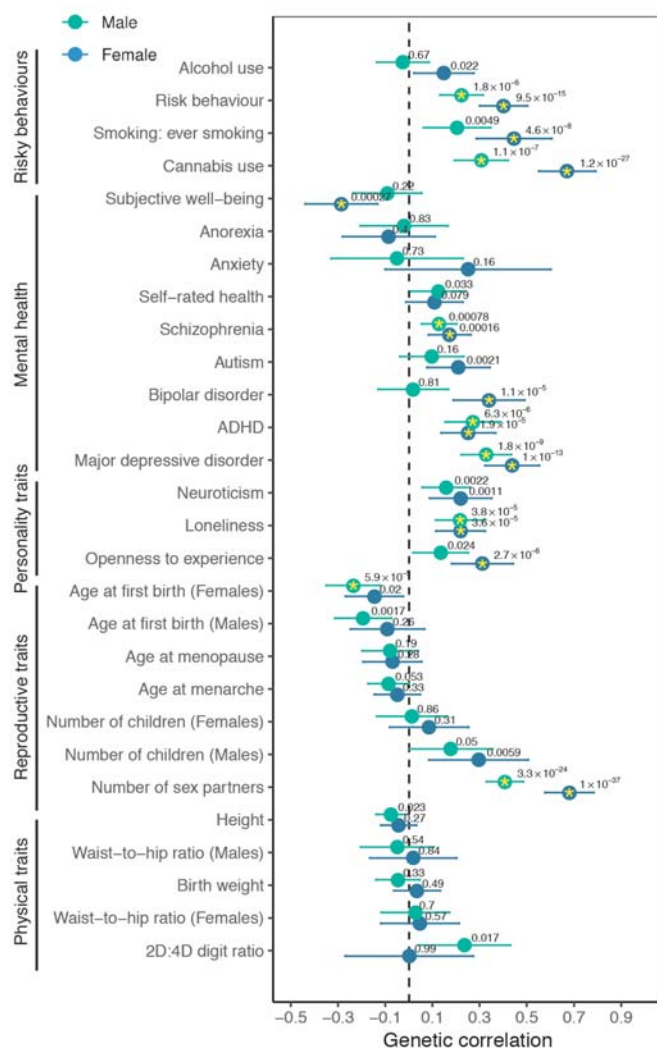
**Fig. 4. Genetic correlations of same-sex sexual behavior with various preselected traits and disorders, separately for males and females.** Males, green; females, blue. Yellow asterisks denote the genetic correlations that were experiment-wise significant ($P < 8.9 \times 10^{-4}$; references, definitions, and full results can be found in table S19). Wald test $P$ values for the genetic correlations are reported above each dot. Horizontal bars represent 95% CIs.

orientation with variants on the X-chromosome ([8], [23]), we found no excess of signal (and no individual genome-wide significant loci) on the X-chromosome (fig. S4).

To test whether these aggregate estimates of genetic effects correlate with sexuality in other samples, we constructed polygenic scores for same-sex sexual behavior ([14], [24]). These polygenic scores were significantly associated with sexual identity in MGSOSO (Wald test, $P = 0.001$) and same-sex attraction in the Add Health ($P = 0.017$) and CATSS ($P = 3.5 \times 10^{-6}$) studies (tables S12, S13, and S14). In CATSS, polygenic scores were also significantly associated with sexual attraction in participants at age 15 years ($P = 6.4 \times 10^{-5}$), suggesting that at least some of the genetic influences on same-sex sexual behavior manifest early in sexual development. The purpose of these analyses is to further characterize the genetic influences on same-sex sexual behavior and not

to predict same-sex sexual behavior on the individual level. In all cases, the variance explained by the polygenic scores was extremely low (<1%); these scores could not be used to accurately predict sexual behavior in an individual.

Overall, these findings suggest that genetic influences on same-sex sexual behavior are highly polygenic and are not specific to the discovery samples or measures. All the SNPs measured, when combined, do not capture the entirety of family-based heritability (8 to 25% from GWAS versus 32% from family-based methods). In this, same-sex sexual behavior is similar to many other complex traits; the ratio between family-based heritability and SNP-heritability estimated in the same sample is consistent with empirical findings for the other 16 traits we tested (family heritability approximately three times larger than SNP-heritability) (Fig. 3) ([14]). There are many possible reasons for this dis-

crepancy, including, but not limited to, variants not captured by genotyping arrays, nonadditive genetic effects, and phenotypic heterogeneity.

### In silico follow-up of GWAS results

To explore the biological processes that may influence same-sex sexual behavior, we performed cell- and tissue-type enrichment analyses using the GWAS discovery dataset ([14], [25]). We did not find clear evidence of enrichment for any particular cell or tissue (fig S5). However, we did find that genes near variants associated with same-sex sexual behavior are more likely than chance to be highly constrained [having unusually low prevalence of loss-of-function variants, suggesting stronger evolutionary constraint ([14], [26])], even after controlling for expression in the brain (table S15).

At the level of individual loci, we investigated biological pathways by integrating information from expression quantitative trait loci (eQTL) analyses ([27]), phenome-wide association study (PheWAS) (table S16) ([28]), and gene-based analysis by using MAGMA ([14], [29]). A full report can be found in table S17. Here, we highlight findings relating to the two SNPs associated with male same-sex sexual behavior: rs34730029 and rs28371400. First, the locus encompassing rs34730029-11q12.1 contains a number of olfactory receptor genes (several of which were significantly associated with same-sex sexual behavior in a gene-based test) (fig. S6 and table S18). This SNP is correlated [linkage disequilibrium, coefficient of determination ($R^2$) = 0.70] with a missense variant (rs6591536) in *OR5A1* that has been reported to have a substantial effect on the sensitivity to certain scents ([30]). Second, rs28371400-15q21.3 had several indications of being involved in sex hormone regulation: The allele positively associated with same-sex sexual behavior is associated with higher rate of male pattern balding [in which sex-hormone sensitivity is implicated ([31])] and is located ~20 kb upstream of the *TCF12* gene. *TCF12* is the primary heterodimerization partner for *TCF21*, a transcription factor essential for normal development of the gonads in mice ([32]), and is involved in the downstream actions of the *SRY* gene (which is responsible for the initiation of male sex determination) in humans ([33]).

### Genetic correlations with other traits

Next, we explored the genetic correlations between same-sex sexual behavior and 28 other relevant traits chosen before the analyses, using summary statistics from other GWASs (Fig. 4 and table S19) ([14]). In particular, we included mental health traits because they are substantially heritable ([34]), and previous population surveys have shown elevated risk of adverse mental health outcomes (such as depression, anxiety, or substance use) in sexual minority populations, including individuals engaging in same-sex sexual behavior ([35], [36]).

We found several personality traits (loneliness and openness to experience), risky behaviors (smoking and cannabis use) and mental health
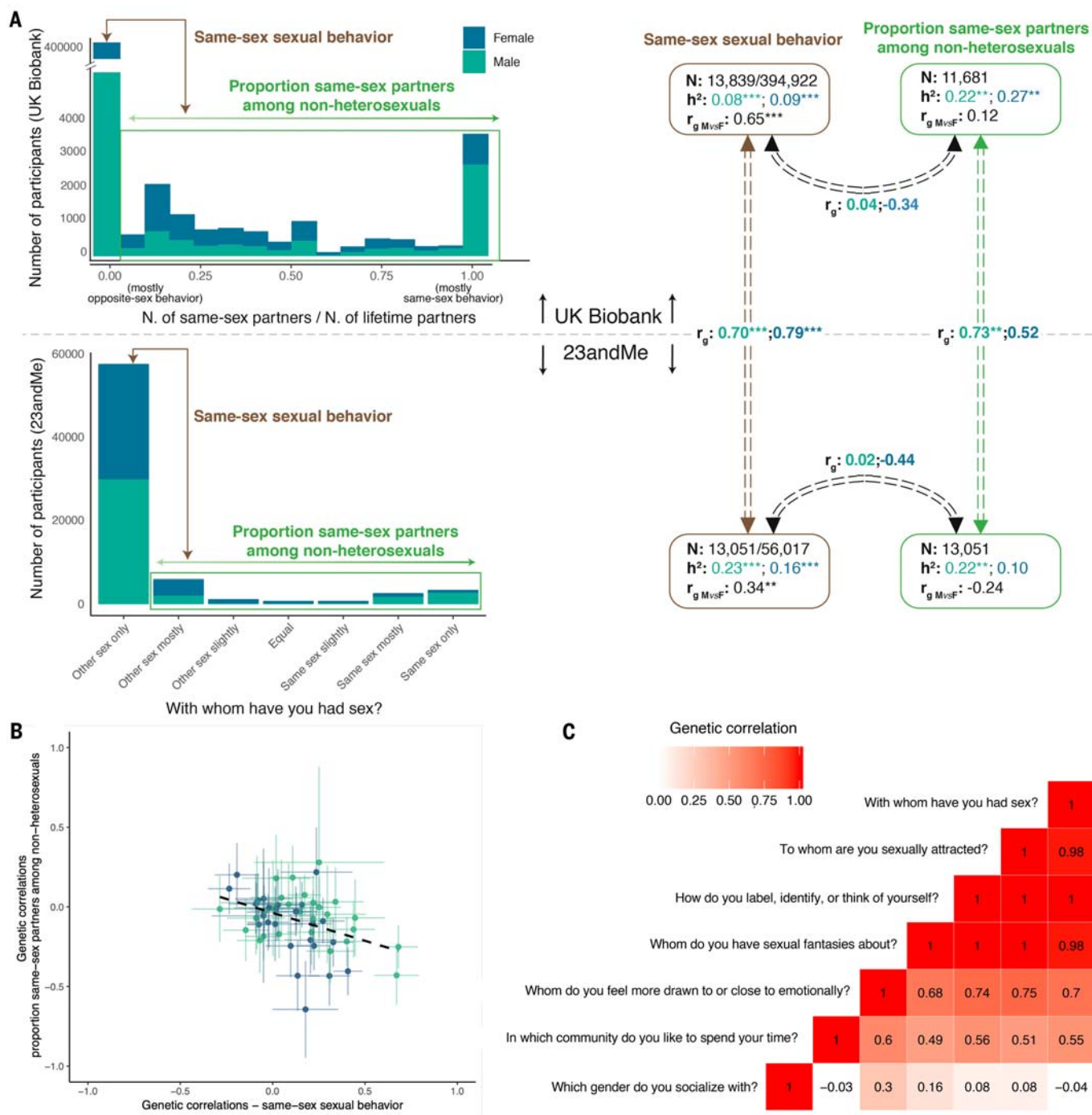
**Fig. 5. Complexity and heterogeneity of genetic influences.** (**A**) Genetic correlations between the main phenotype (same-sex sexual behavior; heterosexuals versus nonheterosexuals) and proportion of same-sex to total sexual partners among nonheterosexuals, in the UK Biobank and 23andMe samples. (**B**) Scatterplot showing genetic correlations of the main phenotype (*x* axis) and the proportion of same-sex to total partners among nonheterosexuals (*y* axis) with various other traits (table S21). (**C**) Genetic correlations among different sexual preference items in the 23andMe sample.

disorders, but not physical traits, to be significantly genetically correlated with same-sex sexual behavior. We found in both sexes that same-sex sexual behavior was positively genetically correlated with several psychiatric or mental health traits [for example, depression, $r_g = 0.44$ in females (95% CIs 0.32 and 0.55), $r_g = 0.33$ in males (95% CIs 0.22 and 0.43); schizophrenia,

$r_g = 0.17$ in females (95% CIs 0.08 and 0.35), $r_g = 0.13$ in males (95% CIs 0.05 and 0.26); all Wald test $P < 0.001$]. We emphasize that the causal processes underlying these genetic correlations are unclear and could be generated by environmental factors relating to prejudice against individuals engaging in same-sex sexual behavior, among other possibilities, which we

discuss in (*14*). Some associations were sex specific. In particular, the genetic correlations with bipolar disorder, cannabis use, and number of sexual partners were significantly higher in females than in males (Wald test $P = 0.001$, $1.47 \times 10^{-6}$, and $3.13 \times 10^{-5}$ respectively) (table S19).

Last, given the potential roles of sex hormones in sexual behaviors, we directly explored whether

there is a genetic correlation with serum sex-hormone–binding globulin (SHBG) levels (37), which are thought to be inversely related to bioactive testosterone and estrogen in females and males, respectively (38). There was a significant correlation in females ($r_g$ = 0.25, Wald test $P$ = 0.03) but not in males ($r_g$ = 0.10, Wald test $P$ = 0.32).

### Complexity and heterogeneity

To maximize our sample size and increase the power to detect SNP associations, we defined our primary phenotype as ever or never having had a same sex partner. Such a measure fails to capture the multifaceted richness and complexity of human sexual orientation. To explore the consequences of this simplification, we pursued genetic analyses across different aspects of sexual orientation and behavior.

First, within participants reporting same-sex sexual behavior, we performed a GWAS on the proportion of same-sex partners to total partners, with a higher value indicating a higher proportion of same-sex partners (14). In the UK Biobank, this is measured directly from participants' reported number of same-sex and all partners, whereas in 23andMe, we used participants' raw responses to the item "With whom have you had sex?", which in individuals reporting same-sex sexual behavior could be "other sex mostly," "other sex slightly," "equal," "same sex slightly," "same sex mostly," or "same sex only." The UK Biobank and 23andMe variables were heritable (table S20A) and genetically correlated with each other ($r_g$ = 0.52 and 95% CIs, –0.16 to 1.20 for females; $r_g$ = 0.73 and 95% CIs, 0.18 to 1.27 for males) (Fig. 5A and table S20C), so we used MTAG to meta-analyze across the two studies for subsequent analyses.

We found little evidence for genetic correlation of the proportion of same-sex to total partners among individuals reporting same-sex sexual behavior (nonheterosexuals) with the binary same-sex sexual behavior variable [$r_g$ = –0.31 (95% CIs, –0.62 to 0.00) for females and $r_g$ = 0.03 (95% CIs, –0.18 to 0.23) for males] (table S20B). Further, this phenotype showed a markedly different pattern of genetic correlations with other traits, as compared with corresponding genetic correlations with the binary same-sex sexual behavior variable (Fig. 5B and

table S21). These findings suggest that the same-sex sexual behavior variable and the proportion of same-sex partners among nonheterosexuals capture aspects of sexuality that are distinct on the genetic level, which in turn suggests that there is no single continuum from opposite-sex to same-sex sexual behavior. Interpretations of any one set of results in our study must consider this complexity.

With this in mind, we examined the possibility of different genetic variants distinguishing heterosexual behavior from differing proportions of same-sex partners within nonheterosexuals. To do so, we performed additional GWASs in the UK Biobank data on the following traits: those whose partners were (i) less than a third same-sex, (ii) between a third and two-thirds same-sex, (iii) more than a third same-sex, and (iv) exclusively same-sex. Genetic correlations of the first three categories with the fourth were 0.13, 0.80, and 0.95 (table S22), indicating partly different genetic variants distinguishing heterosexual behavior from differing proportions of same-sex partners within nonheterosexuals.

Last, using additional measures from 23andMe, we showed strong genetic correlations (all $r_g$ ≥ 0.83) (Fig. 5C and fig. S7) of same-sex sexual behavior with items assessing same-sex attraction, identity, and fantasies (a full list of items is provided in table S5), suggesting that these different aspects of sexual orientation are influenced by largely the same genetic variants. The full set of results of phenotypic and genetic correlations for females, males, and the whole sample is available in fig. S7 and table S5.

### Discussion

We identified genome-wide significant loci associated with same-sex sexual behavior and found evidence of a broader contribution of common genetic variation. We established that the underlying genetic architecture is highly complex; there is certainly no single genetic determinant (sometimes referred to as the "gay gene" in the media). Rather, many loci with individually small effects, spread across the whole genome and partly overlapping in females and males, additively contribute to individual differences in predisposition to same-sex sexual behavior. All measured common variants together explain only part of the genetic heritability at the population level and do not

allow meaningful prediction of an individual's sexual preference.

The knowledge that the variants involved are numerous and spread across the genome enabled us to leverage whole-genome analytic techniques to explore human sexual behavior in ways previously impossible. We determined that the genetic effects that differentiate heterosexual from same-sex sexual behavior are not the same as those that differ among nonheterosexuals with lower versus higher proportions of same-sex partners. This finding suggests that on the genetic level, there is no single dimension from opposite-sex to same-sex preference. The existence of such a dimension, in which the more someone is attracted to the same-sex the less they are attracted to the opposite-sex, is the premise of the Kinsey scale (39), a research tool ubiquitously used to measure sexual orientation. Another measure, the Klein Grid (40), retains the same premise but separately measures sexual attraction, behavior, fantasies, and identification (as well as nonsexual preferences); however, we found that these sexual measures are influenced by similar genetic factors. Overall, our findings suggest that the most popular measures are based on a misconception of the underlying structure of sexual orientation and may need to be rethought. In particular, using separate measures of attraction to the opposite sex and attraction to the same sex, such as in the Sell Assessment of Sexual Orientation (41), would remove the assumption that these variables are perfectly inversely related and would enable more nuanced exploration of the full diversity of sexual orientation, including bisexuality and asexuality.

Although we emphasize the polygenicity of the genetic effects on same-sex sexual behavior, we identified five SNPs whose association with same-sex sexual behavior reached genome-wide significance. Three of these replicated in other independent samples whose measures related to identity and attraction rather than behavior. These SNPs may serve to generate new lines of enquiry. In particular, the finding that one of the replicated SNPs (rs28371400-15q21.3) is linked to male pattern balding and is nearby a gene (TCF12) relevant to sexual differentiation strengthens the idea that sex-hormone regulation may be involved in the development of same-sex sexual behavior. Also, that another replicated SNP (rs34730029-11q12.1) is strongly linked to several genes involved in olfaction raises intriguing questions. Although the underlying mechanism at this locus is unclear, a link between olfaction and reproductive function has previously been established. Individuals with Kallmann syndrome exhibit both delayed or absent pubertal development and an impaired sense of smell because of the close developmental origin of fetal gonadotropin-releasing hormone and olfactory neurons (42).

Our study focused on the genetic basis of same-sex sexual behavior, but several of our results point to the importance of sociocultural context as well. We observed changes in prevalence of

reported same-sex sexual behavior across time, raising questions about how genetic and sociocultural influences on sexual behavior might interact. We also observed partly different genetic influences on same-sex sexual behavior in females and males; this could reflect sex differences in hormonal influences on sexual behavior (for example, importance of testosterone versus estrogen) but could also relate to different sociocultural contexts of female and male same-sex behavior and different demographics of gay, lesbian, and bisexual groups (*43*). With these points in mind, we acknowledge the limitation that we only studied participants of European ancestry and from a few Western countries; research involving larger and more diverse samples will afford greater insight into how these findings fare across different sociocultural contexts.

Our findings provide insights into the biological underpinnings of same-sex sexual behavior but also underscore the importance of resisting simplistic conclusions (Box 2)—because the behavioral phenotypes are complex, because our genetic insights are rudimentary, and because there is a long history of misusing genetic results for social purposes.

## Materials and methods summary
### Study samples

We used data from genotyped individuals from five cohorts (total $n$ = 492,678) who provided self-report information using different questionnaire-based measurement scales. Informed consent was provided from all individuals participating in the studies, which were approved by their local research ethic committee.

### Genetic association analyses

After standard quality control, we performed GWASs for "same-sex sexual behavior" (defined as ever versus never having had sex with a same-sex partner) in the UK Biobank and 23andMe samples, which we meta-analysed using MTAG (*17*). We also conducted GWASs separately by sex. Genome-wide significant SNPs were replicated in three independent samples. Also, using LD-pred (*24*), we derived polygenic score for same-sex sexual behavior according to the meta-analyzed GWAS results and tested the association between this polygenic score and same-sex sexual behavior in three independent samples. To explore diversity among individuals reporting same-sex sexual behavior, we also conducted GWASs in the UK-Biobank and 23andMe samples (meta-analyzed using MTAG) on the "proportion of same-sex to total number of sexual partners among nonheterosexuals."

### Heritability estimation

We estimated family-based heritability of same-sex sexual behavior on the basis of known familial relationships in the UK Biobank study. The relatedness between pairs of participants was estimated by using KING (*44*). Additive genetic effects as well as shared and unshared environmental variance components were esti-

mated on the basis of the covariance between different pairs of relatives. Second, heritability explained by all measured common SNPs (SNP-based heritability) was estimated by using linkage disequilibrium (LD) score regression (*45*) and transformed to the liability scale (*46*). Using a similar approach, we also estimated the SNP-based heritability per chromosome and evaluated heritability enrichment across various tissues on the basis of Genotype-Tissue Expression (GTEx) gene-expression results (*47*).

### In silico follow-up

The GWAS results for same-sex sexual behavior were followed up with gene-based tests of association in MAGMA (*29*) and an enrichment analysis of evolutionarily constrained genes by using partitioned LD score regression (*45*) and MAGMA. We also performed a PheWAS (*28*) to examine whether the SNPs we identified for same-sex sexual behavior have also been associated with other phenotypes and eQTL mapping (*27*) to link SNPs with gene expression.

### Genetic correlations and phenotypic heterogeneity

Using cross-trait LD score regression (*16*), we estimated the genetic correlations of same-sex sexual behavior and proportion of same-sex to total number of sexual partners among non-heterosexuals with a range of traits, including mental health, personality, and sexually dimorphic traits. To examine heterogeneity of genetic influences, we looked at the genetic correlations between sexes, between cohorts, and between different measures of sexual preference.

### Science communication strategy

To communicate the results of the study to the broader audience, we engaged with different LGBTQIA+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other+) and science communication organizations and created multimedia materials for a lay audience.

Detailed materials and methods can be found in the supplementary materials (*14*).

## REFERENCES AND NOTES

1. ACSF investigators, AIDS and sexual behaviour in France. ACSF investigators. *Nature* **360**, 407–409 (1992). doi: 10.1038/360407a0; pmid: 1448162
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; pmid: 1580235
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; pmid: 27593894
4. J. M. Bailey *et al.*, Sexual orientation, controversy, and science. *Psychol. Sci. Public Interest* **17**, 45–101 (2016). doi: 10.1177/1529100616637616; pmid: 27113562
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). pmid: 9549243
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; pmid: 18536986
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; pmid: 8332896
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; pmid: 10213693
10. A. R. Sanders *et al.*, Genome-wide association study of male sexual orientation. *Sci. Rep.* **7**, 16950 (2017). doi: 10.1038/s41598-017-15736-4; pmid: 29217827
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; pmid: 18074219
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; pmid: 29389215
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; pmid: 10743878
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; pmid: 26499203
16. B. Bulik-Sullivan *et al.*, An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015). doi: 10.1038/ng.3406; pmid: 26414676
17. P. Turley *et al.*, Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018). doi: 10.1038/s41588-017-0009-4; pmid: 29292387
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; pmid: 27473438
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; pmid: 27225129
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; pmid: 29325848
21. R. Karlsson Linnér *et al.*, 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; pmid: 30643258
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; pmid: 27440663
23. A. R. Sanders *et al.*, Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychol. Med.* **45**, 1379–1388 (2015). doi: 10.1017/S0033291714002451; pmid: 25399360
24. B. J. Vilhjálmsson *et al.*, 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; pmid: 26430803
25. H. K. Finucane *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678
26. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016). doi: 10.1038/nature19057; pmid: 27535533
27. A. Gusev *et al.*, Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016). doi: 10.1038/ng.3506; pmid: 26854917
28. J. C. Denny *et al.*, PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010). doi: 10.1093/bioinformatics/btq126; pmid: 20335276
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; pmid: 25885710

30. S. R. Jaeger *et al.*, 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; pmid: 23910657

31. H. Kische *et al.*, 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; pmid: 28403384

32. S. Cui *et al.*, Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development* **131**, 4095–4105 (2004). doi: 10.1242/dev.01266; pmid: 15289436

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; pmid: 21637323

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; pmid: 22777127

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; pmid: 11146762

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; pmid: 27009565

37. A. D. Coviello *et al.*, 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; pmid: 22829776

38. C. W. Burke, D. C. Anderson, Sex-hormone-binding globulin is an oestrogen amplifier. *Nature* **240**, 38–40 (1972). doi: 10.1038/240038a0; pmid: 4120573

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; pmid: 9415799

42. H. Valdes-Socin *et al.*, Reproduction, smell, and neurodevelopmental disorders: Genetic defects in different hypogonadotropic hypogonadal syndromes. *Front. Endocrinol.* **5**, 109 (2014). doi: 10.3389/fendo.2014.00109; pmid: 25071724

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; pmid: 20835383

44. A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010). doi: 10.1093/bioinformatics/btq559; pmid: 20926424

45. B. K. Bulik-Sullivan *et al.*, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015). doi: 10.1038/ng.3211; pmid: 25642630

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; pmid: 21376301

47. J. Lonsdale *et al.*, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013). doi: 10.1038/ng.2653; pmid: 23715323

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

## ACKNOWLEDGMENTS

B. Alipanahi, A. Auton, R. K. Bell, K. Bryc, S. L. Elson, P. Fontanillas, N. A. Furlotte, B. Hicks, K. E. Huber, E. M. Jewett, Y. Jiang, A. Kleinman, K.-H. Lin, N. K. Litterman, J. C. McCreight, M. H. McIntyre, K. F. McManus, J. L. Mountain, E. S. Noblin, C. A. M. Northover, S. J. Pitts, G. D. Poznik, A. J. Shastri, J. F. Shelton, S. Shringarpure, C. Tian, J. Y. Tung, V. Vacic, X. Wang, and C. H. Wilson.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6456/[page]/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S23
References (*49–108*)