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[12ϕ], Benjamin M. Neale[1,2,3ϕ], Brendan P. Zietsch[22ϕ]

1. Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

2. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

3. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

4. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

5. Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Meibergdreef 5, 1105 AZ Amsterdam, The Netherlands

6. Department of Biological Psychology, VU University, 1081 BT, Amsterdam, The 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, Colorado 80309-0483, USA

10. Health and Society Program and Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado 80309-0483, USA

11. Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309-0483, USA

12. 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, Florida 33136, USA

16. 23andMe, Inc. Mountain View, CA 94041, USA

17. Centre for Ethics, Law and Mental Health, University of Gothenburg, Sweden

18. Carolina Population Center, University of North Carolina at Chapel Hill

19. Department of Sociology, University of North Carolina at Chapel Hill

20. Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, Evanston, Illinois 60201, USA

21. Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois 60637, USA

22. Centre for Psychology and Evolution, School of Psychology, University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

* Shared first authors, $^{\phi}$ Shared last authors

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 8-25% of variation in same-sex sexual behavior, only partially overlapped between men and women, 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 non-heterosexuals suggests 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.

Across 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 orientation 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 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 ($p<5*10^{-8}$). 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 (e.g., bisexuality)?

In order to identify genetic variants associated with same-sex sexual behavior and explore its genetic architecture and underlying biology we performed a 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, Inc., predominantly located in the U.S.A., 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 United Kingdom residents aged 40 to 70 (see (*14*) and **Tables S1** and **S2**). Our primary phenotype of interest is a binary, self-report measure of whether respondents had ever had sex with someone of the same sex (here termed *non-heterosexuals;* see **Box 1** for a note on terminology) or had not (here termed *heterosexuals*).

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 orientation is taxonic rather than dimensional in structure, with individuals reporting exclusively opposite-sex orientation differing from individuals reporting any same-sex orientation (*15*). However, the binary variable also collapses rich and multifaceted diversity among non-heterosexual individuals (*15*), so we explore finer-scaled measurements and some of the complexities of the phenotype below, though intricacies of the social and cultural

influences on sexuality make it impossible to fully explore this complexity. The 23andMe sample comprised of 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** and **(*14*)**).

We also performed replication analyses in three smaller datasets (*14*): 1) MGSOSO (N=2,308 U.S. adult males), where respondents were asked about their sexual identity, 2) Add Health (N=4,755 U.S. young adults), where respondents were asked whether they ever had same-sex intercourse and whether they were romantically attracted to the same-sex, and 3) CATSS (N=8,093 Swedish adolescents), where participants reported the degree of attraction to the same vs. opposite sex.

We observed in UK Biobank that individuals reporting same-sex sexual behavior had on average fewer offspring than individuals who engaged exclusively in heterosexual behavior, even for individuals reporting only a minority of same-sex partners (**Fig. 1B**). We note that this reduction in number of children is comparable or greater than for other traits that have been linked to lower fertility rates (**Fig. S1** and (*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 modelling 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% CIs: 10.6-54.3) (**Table S4**). This estimate is consistent with previous estimates from smaller twin studies (*7*).

To identify genetic variants (largely single nucleotide polymorphisms, or SNPs) associated with same-sex sexual behavior, we performed a GWAS in the UK Biobank study (N=408,995; (*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 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 (r_g = 0.87, 95% CIs: 0.67-1.06); **Table S5** and (*14*)). Genetic correlations between same-sex sexual behavior and 28 different traits were largely similar in the UK Biobank and 23andMe (see **Fig. S2** and (*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 UK Biobank (r_g = -0.36, 95% CIs: -0.60; -0.09) and 23andMe data (r_g = 0.36, 95% CIs: 0.08-0.65, 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 their genetic correlation to determine the meta-analytic weights, yielding a total sample size of 477,522 (26,827 individuals reporting same-sex sexual behavior).

After standard quality control checks (see (*14*) and **Table S6**), 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 under *'In-silico follow-up of GWAS results'* below. 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: 0.48-0.78) (**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 for 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, effective sample size = 4,887).

The SNPs that reached genome-wide significance had very small effects (odds ratios ~1.1, **Table S7**). For example, in 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% vs. 3.6%). Nevertheless, the contribution of all measured common SNPs in aggregate (i.e. SNP-based heritability) was estimated to be 8-25% (95% CIs: 5-30%) of variation in female and male same-sex sexual behavior, where the range reflects differing estimates using different analysis methods or prevalence assumptions (see **Table S11** and (*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, are 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 (see *(14)* and **Fig. S3**). In contrast to linkage studies that found substantial association of sexual 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 (*24*) (see (*14*)). 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=3.5x10^{-6}) studies (**Tables S12, S13** and **S14**). In CATSS, polygenic scores were also significantly associated with sexual attraction in participants at age 15 (p=6.4x10^{-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. Indeed, 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 unique to the discovery samples or measures. We note that all the SNPs measured, when combined, do not capture the entirety of family-based heritability (8-25% from GWAS vs. 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 3 times larger than SNP-heritability, see (*14*) and **Fig. 3**). There are many possible reasons for this discrepancy,

including, but not limited to, variants not captured by genotyping arrays, non-additive 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 (i.e., having unusually low prevalence of loss of function variants, suggesting stronger evolutionary constraint (*26*), see (*14*)), even after controlling for expression in the brain (see **Table S15**).

At the level of individual loci, we investigated biological pathways by integrating information from eQTL analyses (*27*), PheWAS (*28*) (**Table S16**), and gene-based analysis using MAGMA (*29*) (see (*14*)). 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 (**Fig. S6**, several of which were significantly associated with same-sex sexual behavior in a gene-based test (**Table S18**)). This SNP is correlated (linkage disequilibrium, 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 it is located ~20KB 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 (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 prior to the analyses using summary statistics from other GWASs (see (14), **Fig. 4** and **Table S19**). 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 (e.g. depression, anxiety, substance use) in sexual minority populations, including individuals engaging in same-sex sexual behavior (*35, 36*).

We found several personality traits (loneliness, openness to experience), risky behaviors (smoking, cannabis use) and mental health 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 (e.g. depression: $r_g = 0.44$ (95% CIs: 0.32; 0.55) in females, 0.33 (95% CIs: 0.22; 0.43) in males; schizophrenia: $r_g = 0.17$ (95% CIs: 0.08; 0.35) in females, 0.13 (95% CIs: 0.05; 0.26) in males; all Wald test *p*s<.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, 1.47×10^{-6}, and 3.13×10^{-5} respectively; **Table S19**).

Finally, 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/never having reported 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 both heritable (**Table S20A**) and genetically correlated with each other (r_g = 0.52; 95% CIs: -0.16-1.20 for females and 0.73; 95% CIs: 0.18–1.27 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 (non-heterosexuals) with the binary same-sex sexual behavior variable (r_g = -0.31 (95% CIs: -0.62-0.00) for females and r_g = 0.03 (95% CIs: -0.18-0.23) for males; **Table S20B**). Further, this phenotype showed a

markedly different pattern of genetic correlations with other traits, as compared to corresponding genetic correlations with the binary same-sex sexual behavior variable (**Fig. 5B**; **Table S21**). These findings suggest that the same-sex sexual behavior variable and the proportion of same-sex partners among non-heterosexuals capture aspects of sexuality that are distinct on the genetic level, suggesting that there is no single continuum from opposite-sex to same-sex sexual behavior. We therefore note that 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 non-heterosexuals. To do so, we performed additional GWASs in the UK Biobank data on the following traits: those whose partners were (1) less than a third same-sex; (2) between a third and two-thirds same-sex; (3) more than a third same-sex, and (4) 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 non-heterosexuals.

Finally, using additional measures from 23andMe, we showed strong genetic correlations (all r_gs $\geq$ 0.83; **Fig. 5C** and **Fig. S7**) of same-sex sexual behavior with items assessing same-sex attraction, identity, and fantasies (for full list of items, see **Table S5**), suggesting that these different aspects of sexual orientation are influenced by largely the same genetic variants. For the full set of results of phenotypic and genetic correlations for females, males, and the whole sample, see **Fig. S7** and **Table S5**.

Discussion

Here 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. Importantly, we determined that the genetic effects that differentiate heterosexual from same-sex sexual behavior are not the same as those that differ among non-heterosexuals 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 orientation. The existence of such a dimension, whereby 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 non-sexual preferences); however, we found that these sexual measures are mostly 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.

Though 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, the fact that another replicated SNP (rs34730029-11q12.1) is strongly linked to several genes involved in olfaction raises intriguing questions. Whilst 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/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 and heritability 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 (e.g. importance of testosterone vs. 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 (see **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-analyses 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 based on the meta-analysed 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-analysed using MTAG) on the *proportion of same-sex to total number of sexual partners among non-heterosexuals*.

Heritability estimation

We estimated family-based heritability of same-sex sexual behavior based on known familial relationships in the UK Biobank study. The relatedness between pairs of participants was estimated using KING (*44*). Additive genetic effects as well as common and unique environmental variance components were estimated based on the covariance between different pairs of relatives. Secondly, heritability explained by all measured common SNPs (SNP-based heritability) was estimated using LD-score regression (*45*) and transformed to the liability scale (*46*), assuming different prevalence for same-sex sexual behavior. Using a similar approach, we also estimated the SNP-based heritability per chromosome and evaluated heritability enrichment across various tissues based on GTEx gene-expression results (*47*).

In-silico follow-up

The GWAS results for same-sex sexual behavior were followed-up by gene-based tests of association in MAGMA (*29*), and an enrichment analysis of evolutionarily constrained genes using partitioned LD-score regression (*45*) and MAGMA. We also performed a phenome-wide association study (*28*) to examine whether the SNPs we identified for same-sex sexual behavior have also been associated with other phenotypes and eQTL mapping (Expression quantitative trait loci; (*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 sexual orientation, we looked at the genetic correlations between sexes, between cohorts, and between different measures of sexual orientation.

Science communication strategy

To communicate the results of the study to the broader audience, we engaged with different LGBTQ+ and science communication organizations and created multimedia materials for a lay audience.

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

Supplementary Materials:

Materials and Methods

Figures S1 to S7

Tables S1 to S23

References (49-110)

Main text references (1-48)

1. ACSF investigators, AIDS and sexual behaviour in France. *Nature* **360**, 407-409 (1992).
2. M. Melbye, R. J. Biggar, Interactions between persons at risk for AIDS and the general population in Denmark. *American journal of epidemiology* **135**, 593-602 (1992).
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).
4. J. M. Bailey *et al.*, Sexual Orientation, Controversy, and Science. *Psychological science in the public interest : a journal of the American Psychological Society* **17**, 45-101 (2016).
5. D. Scasta, P. Bialer, American Psychiatric Association, Position statement on issues related to homosexuality. *Arlington County: American Psychiatric Association. Available online: https://www. psychiatry. org/psychiatrists/search-directories-databases/policy-finder (accessed on 26 February 2019)*, (2013).
6. R. C. Pillard, J. M. Bailey, Human sexual orientation has a heritable component. *Human Biology* **70**, 347-365 (1998).
7. N. Langstrom, Q. Rahman, E. Carlstrom, P. Lichtenstein, Genetic and environmental effects on same-sex sexual behavior: a population study of twins in Sweden. *Archives of sexual behavior* **39**, 75-80 (2010).
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 (New York, N.Y.)* **261**, 321-327 (1993).
9. G. Rice, C. Anderson, N. Risch, G. Ebers, Male homosexuality: absence of linkage to microsatellite markers at Xq28. *Science (New York, N.Y.)* **284**, 665-667 (1999).

10. A. R. Sanders *et al.*, Genome-Wide Association Study of Male Sexual Orientation. *Scientific reports* **7**, 16950 (2017).

11. R. A. Lippa, Sex differences and sexual orientation differences in personality: findings from the BBC Internet survey. *Archives of sexual behavior* **37**, 173-187 (2008).

12. Y. Wang, M. Kosinski, Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of personality and social psychology* **114**, 246-257 (2018).

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. *Journal of personality and social psychology* **78**, 524-536 (2000).

14. See materials and methods and other supplementary materials.

15. A. L. Norris, D. K. Marcus, B. A. Green, Homosexuality as a Discrete Class. *Psychol Sci* **26**, 1843-1853 (2015).

16. B. K. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236-1241 (2015).

17. P. Turley, R. K. Walters, Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics* **50**, 229-237 (2018).

18. K. Rawlik, O. Canela-Xandri, A. Tenesa, Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome biology* **17**, 166 (2016).

19. A. Okbay *et al.*, Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542 (2016).

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).

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. *Nature genetics* **51**, 245-257 (2019).

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. *Proceedings. Biological sciences / The Royal Society* **283**, (2016).

23. A. R. Sanders *et al.*, Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychol Med* **45**, 1379-1388 (2015).

24. B. J. Vilhjalmsson *et al.*, Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American journal of human genetics* **97**, 576-592 (2015).

25. H. K. Finucane *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. **47**, 1228-1235 (2015).

26. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).

27. A. Gusev, A. Ko, Integrative approaches for large-scale transcriptome-wide association studies. **48**, 245-252 (2016).

28. J. C. Denny *et al.*, PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics (Oxford, England)* **26**, 1205-1210 (2010).

29. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, e1004219 (2015).

30. S. R. Jaeger *et al.*, A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Current biology : CB* **23**, 1601-1605 (2013).

31. H. Kische *et al.*, Sex Hormones and Hair Loss in Men From the General Population of Northeastern Germany. *JAMA dermatology* **153**, 935-937 (2017).

32. S. Cui *et al.*, Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development* **131**, 4095-4105 (2004).

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).

34. P. F. Sullivan, M. J. Daly, M. O'Donovan, Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature reviews. Genetics* **13**, 537-551 (2012).

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). *Archives of general psychiatry* **58**, 85-91 (2001).

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).

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).

38. C. W. Burke, D. C. Anderson, Sex-hormone-binding globulin is an oestrogen amplifier. *Nature* **240**, 38-40 (1972).

39. A. C. Kinsey, W. B. Pomeroy, C. E. Martin, *Sexual behavior in the human male*. (W.B. Saunders, Philadelphia, 1948).

40. F. Klein, *The bisexual option, 2nd ed.*, (Routledge, Haworth Press, New York, NY, USA, 1993).

41. R. L. Sell, Defining and measuring sexual orientation: a review. *Archives of sexual behavior* **26**, 643-658 (1997).

42. H. Valdes-Socin *et al.*, Reproduction, smell, and neurodevelopmental disorders: Genetic defects in different hypogonadotropic hypogonadal syndromes. *Frontiers in Endocrinology* **5**, (2014).

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. *Sexuality research & social policy : journal of NSRC : SR & SP* **7**, 176-200 (2010).

44. A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)* **26**, 2867-2873 (2010).

45. B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295 (2015).

46. S. H. Lee, N. R. Wray, M. E. Goddard, P. M. Visscher, Estimating Missing Heritability for Disease from Genome-wide Association Studies. *American journal of human genetics* **88**, 294-305 (2011).

47. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585 (2013).

48. Code is available through DOI: https://doi.org/10.5281/zenodo.3232892

Supplemental references (49-110)

49. C. Bycroft *et al.*, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

50. K. M. Harris, C. T. Halpern, B. C. Haberstick, A. Smolen, The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **16**, 391-398 (2013).

51. M. B. McQueen *et al.*, The National Longitudinal Study of Adolescent to Adult Health (Add Health) sibling pairs genome-wide data. *Behav Genet* **45**, 12-23 (2015).

52. S. McCarthy, S. Das, A reference panel of 64,976 haplotypes for genotype imputation. **48**, 1279-1283 (2016).

53. S. Das *et al.*, Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284-1287 (2016).

54. A. R. Sanders *et al.*, The Internet-based MGS2 control sample: self report of mental illness. *The American journal of psychiatry* **167**, 854-865 (2010).

55. J. Shi *et al.*, Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753-757 (2009).

56. G. R. Abecasis *et al.*, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).

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).

58. P. R. Loh *et al.*, Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics* **47**, 284-290 (2015).

59. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).

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. *Genetics* **208**, 1397-1408 (2018).

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 (Oxford, England)* **32**, 1423-1426 (2016).

62. M. H. Chen, Q. Yang, GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics (Oxford, England)* **26**, 580-581 (2010).

63. K. E. Samocha *et al.*, A framework for the interpretation of de novo mutation in human disease. *Nature genetics* **46**, 944-950 (2014).

64. A. Ganna *et al.*, Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *American Journal of Human Genetics* **102**, 1204-1211 (2018).

65. A. F. Pardinas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. **50**, 381-389 (2018).

66. D. Demontis *et al.*, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75 (2019).

67. H. K. Finucane, Y. A. Reshef, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. **50**, 621-629 (2018).

68. K. I. Kim, M. A. van de Wiel, Effects of dependence in high-dimensional multiple testing problems. *BMC bioinformatics* **9**, 114 (2008).

69. J. MacArthur *et al.*, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). **45**, D896-d901 (2017).

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. *American journal of human genetics* **98**, 382-391 (2016).

71. C. H. Mercer *et al.*, 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 (London, England)* **382**, 1781-1794 (2013).

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

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

74. M. Lynch, B. Walsh, *Genetics and analysis of quantitative traits*. (Sinauer Sunderland, MA, 1998), vol. 1.

75. S. V. Eastwood *et al.*, Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. **11**, e0162388 (2016).

76. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575 (2007).

77. R. A. Power *et al.*, Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry* **70**, 22-30 (2013).

78. J. K. Pickrell, T. Berisa, J. Z. Liu, Detection and interpretation of shared genetic influences on 42 human traits. **48**, 709-717 (2016).

79. C. Tian *et al.*, Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature communications* **8**, 599 (2017).

80. J. F. McRae *et al.*, Identification of regions associated with variation in sensitivity to food-related odors in the human genome. *Current biology : CB* **23**, 1596-1600 (2013).

81. W. J. Astle *et al.*, The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e1419 (2016).

82. J. J. Lee, *et al.*, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121 (2018).

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).

84. D. F. Levinson *et al.*, Genome-wide association study of multiplex schizophrenia pedigrees. *The American journal of psychiatry* **169**, 963-973 (2012).

85. D. T. Uehara *et al.*, SNP array screening of cryptic genomic imbalances in 450 Japanese subjects with intellectual disability and multiple congenital anomalies previously negative for large rearrangements. **61**, 335-343 (2016).

86. G. M. Lowe *et al.*, The degradation of (all-E)-beta-carotene by cigarette smoke. *Free radical research* **43**, 280-286 (2009).

87. D. S. Manoli, P. Fan, E. J. Fraser, N. M. Shah, Neural control of sexually dimorphic behaviors. *Current opinion in neurobiology* **23**, 330-338 (2013).

88. Z. Xu *et al.*, Synergistic effect of SRY and its direct target, WDR5, on Sox9 expression. *PloS one* **7**, e34327 (2012).

89. D. Demontis, R. K. Walters, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. **51**, 63-75 (2019).

90. N. Barban, R. Jansen, Genome-wide analysis identifies 12 loci influencing human reproductive behavior. **48**, 1462-1472 (2016).

91. J. R. Perry *et al.*, Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97 (2014).

92. F. R. Day *et al.*, Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature genetics* **47**, 1294-1303 (2015).

93. T. K. Clarke *et al.*, Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Molecular psychiatry* **22**, 1376-1384 (2017).

94. V. Boraska *et al.*, A genome-wide association study of anorexia nervosa. *Molecular psychiatry* **19**, 1085 (2014).

95. T. Otowa *et al.*, Meta-analysis of genome-wide association studies of anxiety disorders. **21**, 1391-1399 (2016).

96. J. Grove *et al.*, Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444 (2019).

97. G. C. B. D. W. G. Psychiatric *et al.*, Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics* **43**, 977 (2011).

98. M. Horikoshi *et al.*, Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248 (2016).

99. J. A. Pasman *et al.*, GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature Neuroscience* **21**, 1161-1170 (2018).

100. A. R. Wood *et al.*, Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* **46**, 1173 (2014).

101. N. R. Wray *et al.*, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetic* **50**, 668-681 (2018).

102. A. Okbay *et al.*, Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics* **48**, 624 (2016).

103. R. Karlsson Linner, P. Biroli, Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. **51**, 245-257 (2019).

104. M. H. M. De Moor *et al.*, Meta-analysis of genome-wide association studies for personality. *Molecular psychiatry* **17**, 337-349 (2010).

105. M. T. Lo *et al.*, Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. **49**, 152-156 (2017).

106. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

107. S. E. Harris *et al.*, Molecular genetic contributions to self-rated health. *International journal of epidemiology* **46**, 994-1009 (2017).

108. Tobacco and Genetics Consortium, Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics* **42**, 441-447 (2010).

109. 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).

110. N. M. Warrington *et al.*, Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Human molecular genetics* **27**, 2025-2038 (2018).

Acknowledgements

We thank N.J. Cox, M.C. Keller and E.S. Lander for carefully reading and commenting on the manuscript. We thank P. Turley and R. Walters for constructive discussion about MTAG. This research was conducted using the UK Biobank Resource under applications 25995. We thank all cohort participants for making this study possible.

Funding. A.R.S. received funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development specifically to investigate the genetics of sexual orientation: R01HD041563 (Sanders PI) and R21HD080410 (Sanders, Martin MPI). E.R.M., G.W.B., and S.G. are also supported by R21HD080410.

No other member of the group received funding specifically for this study, but members of our team received salary funding from organizations as well as our own universities.

B.P.Z. received funding from The Australian Research Council (FT160100298). A.G. was supported by the Knut and Alice Wallenberg Foundation (2015.0327) and the Swedish Research Council (2016-00250). A.G., R.M., and B.M.N. were supported by National Institutes of Health (NIH) grant 1R01MH107649-03 (to B.M.N.).

R.W. was supported by the National Science Foundation's Graduate Research Fellowship Program (DGE 1144083). Any opinion, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

M.G.N. is supported by ZonMw grants 849200011 and 531003014 from The Netherlands Organisation for Health Research and Development.

Study-specific acknowledgements: This research uses data from Add Health, a program project directed by K.M.H. (PI) and designed by J.R. Udry, P.S. Bearman, and K.M.H. at the University of North Carolina at Chapel Hill, and funded by grant P01-HD031921 from the Eunice Kennedy Shriver National Institute of Child Health and Human

Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). This research uses Add Health GWAS data funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grants R01 HD073342 to K.M.H. (PI) and R01 HD060726 to K.M.H., J.D. Boardman, and M.B. McQueen (MPIs).

The genetic part of the CATSS study was supported by grant 2014-0834 from the Swedish Council for Working life and Social Research.

We thank the research participants of 23andMe and the other contributing cohorts for making this study possible. Collaborators for the 23andMe Research Team are: M. Agee, 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.

Author contributions. A.G., K.J.H.V., M.G.N., J.R.B.P., B.M.N., and B.P.Z. were responsible for the study concept and the design of the study. A.G. was the main analyst; he performed the majority of analyses, and created most figures and tables. J.F.S., 23andMe Research Team, and A.A. contributed to data acquisition and analysis of the 23andMe dataset. R.W. and K.M.R. contributed to data acquisition and analysis of the Add Health data. S.G., G.W.B., E.R.M., and A.R.S. contributed to data acquisition and analysis of the MGSOSO data. A.G., P.L., S.L., and N.L. contributed to data acquisition and analysis of the CATSS data. K.J.H.V., M.G.N., R.M., R.W., F.R.D., A.S.B., A.A., S.G., J.F.S., G.W.B., E.R.M. contributed to secondary analyses of the data.

B.P.Z. wrote most of the manuscript; A.G., K.J.H.V., M.G.N., R.W., J.R.B.P., and B.M.N. contributed substantially to the writing of the manuscript. K.J.H.V. was in charge of writing the supplementary material. A.G., M.G.N., R.M., R.W., F.R.D., A.S.B., E.R.M., A.R.S., and B.P.Z. provided important input for the supplementary materials. A.G., K.J.H.V., R.W., F.R.D., A.R.S., B.M.N., and B.P.Z. contributed to engaging in community-based outreach with alliance and advocacy groups, and creation of the website for the general public. All authors provided critical revision of the manuscript for important intellectual content.

Competing interests. J.F.S., A.A. (Adam Auton), and members of the 23andMe Research Team are employees of 23andMe, Inc., and hold stock or stock options in 23andMe. B.M.N. is a member of the scientific advisory board at Deep Genomics and a paid consultant for Camp4 Therapeutics Corporation, Takeda Pharmaceutical and Biogen.

Data and materials availability. The code is available through GitHub (https://github.com/andgan/sexual_orientation_GWAS), archived at Zenodo *(48)* and the GWAS summary statistics of the UK-Biobank sample (and the top 10,000 independent SNPs from the meta-analysis including 23andMe data) are available at GWAS Catalog (search for this paper's PubMed ID on https://www.ebi.ac.uk/gwas/downloads/summary-statistics). Access to the full summary statistics of the 23andMe sample (i.e. for all SNPs) can be obtained by qualified researchers through a Data Transfer Agreement with 23andMe that protects the privacy of the 23andMe participants. Researchers interested in the full meta-analysis summary statistics containing 23andMe data must also apply to 23andMe. Please visit https://research.23andme.com/dataset-access/ for more information and to apply to access the data. Access to individual level data from the UK Biobank can be obtained by bona fide scientists through application with UK Biobank, see https://www.ukbiobank.ac.uk/researchers/. Summary statistics from the Neale Lab database used for the pheWAS are available at GWAS Catalog(https://www.ebi.ac.uk/gwas/downloads/summary-statistics).

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 material [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 "non-heterosexuals", while we refer to individuals who have never had sex with someone of the same sex as "heterosexuals."
- We acknowledge that the grouping phrase "non-heterosexuals" has the potential to present messages of othering (that is, undesirable marginalization of another person or group based on their sexual expression) - by defining an "outgroup" in reference to an "ingroup" and implying that "*non-heterosexual behavior*" may have a negative connotation, while "*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 drop individuals from our study whose biological sex and self-identified sex/gender do not match. This is an important limitation of our analyses, as the analyses do not include transgender persons, intersex persons, and other important persons and groups within queer community. We hope that this limitation will be addressed in future work.

Box 2. Communication and interpretation

- The topic explored in this study is complex and intersects with sexuality, identity, and attraction, and potentially has civil and political implications for sexual minority groups. Therefore, we have:
    - Engaged with science communication teams
    - Engaged with LGBTQ advocacy groups nationally and within our local institutions
    - Tried to make clear the many limitations and nuances of our study and our phenotypes
- We wish to make it clear that our results overwhelmingly point toward the richness and diversity of human sexuality. Our results do not point toward a role for discrimination based on sexual identity or attraction, nor do our results make any conclusive statements about the degree to which "nature" and "nurture" influence sexual orientation.

Figure legends

Figure 1. (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 (i.e. lower birth-year adjusted number of children) (y-axis). Vertical bars represent 95% confidence intervals.

Figure 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, while triangles represent genome wide significant signals that are female (upright, blue) or male (upside down, green) specific.

Figure 3. SNP-based vs. family-based heritability (h^2) estimates for same-sex sexual behavior (red dot) compared with a variety of other traits (grey dots); see **Table S23** for the estimates for all traits. Horizontal bars represent 95% confidence intervals for the SNP-based estimate, vertical bars represent 95% confidence intervals 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.

Figure 4. Genetic correlations of same-sex sexual behavior with various preselected traits and disorders, separately for males (green) and females (blue). Yellow asterisks denote the genetic correlations that were experiment-wise significant ($p<8.9\text{x}10^{-4}$; 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% confidence intervals.

Figure 5. (**A**) Genetic correlations between the main phenotype (same-sex sexual behavior; heterosexuals vs. non-heterosexuals) and proportion of same-sex to total sexual partners

among non-heterosexuals, 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 non-heterosexuals (y-axis) with various other traits (see **Table S21**). (**C**) Genetic correlations among different sexual orientation items in the 23andMe sample.