

SCIENTIFIC REPORTS

OPEN

Genome-Wide Association Study of Male Sexual Orientation

Alan R. Sanders^{1,2}, Gary W. Beecham³, Shengru Guo³, Khytam Dawood⁴, Gerulf Rieger⁵, Judith A. Badner⁶, Elliot S. Gershon², Ritesha S. Krishnappa⁷, Alana B. Kolundzija⁸, Jubao Duan^{1,2}, MGS Collaboration*, Pablo V. Gejman^{1,2}, J. Michael Bailey⁹ & Eden R. Martin³

Received: 1 June 2017

Accepted: 31 October 2017

Published online: 07 December 2017

Family and twin studies suggest that genes play a role in male sexual orientation. We conducted a genome-wide association study (GWAS) of male sexual orientation on a primarily European ancestry sample of 1,077 homosexual men and 1,231 heterosexual men using Affymetrix single nucleotide polymorphism (SNP) arrays. We identified several SNPs with $p < 10^{-5}$, including regions of multiple supporting SNPs on chromosomes 13 (minimum $p = 7.5 \times 10^{-7}$) and 14 ($p = 4.7 \times 10^{-7}$). The genes nearest to these peaks have functions plausibly relevant to the development of sexual orientation. On chromosome 13, *SLITRK6* is a neurodevelopmental gene mostly expressed in the diencephalon, which contains a region previously reported as differing in size in men by sexual orientation. On chromosome 14, *TSHR* genetic variants in intron 1 could conceivably help explain past findings relating familial atypical thyroid function and male homosexuality. Furthermore, skewed X chromosome inactivation has been found in the thyroid condition, Graves' disease, as well as in mothers of homosexual men. On pericentromeric chromosome 8 within our previously reported linkage peak, we found support ($p = 4.1 \times 10^{-3}$) for a SNP association previously reported (rs77013977, $p = 7.1 \times 10^{-8}$), with the combined analysis yielding $p = 6.7 \times 10^{-9}$, i.e., a genome-wide significant association.

While the usual combination of sex chromosomes (XX or XY) predicts sexual orientation and behavior for the vast majority of humans (as heterosexual), variation exists: a stable minority of men (3~4%) are homosexual¹ and male sexual orientation appears to be bimodally distributed with most men rating themselves as predominantly heterosexual (Kinsey scale 0–1) or homosexual (Kinsey scale 5–6)^{1–5}. Male sexual orientation is moderately heritable (30~40%), but is multifactorial, with evidence of multiple genetic and environmental contributions, via family studies^{6–11}, twin studies^{4,12–16}, and segregation analyses^{8,10,11,17}. Although focused (i.e., chromosome X) and genome-wide linkage studies (GWLS) of affected (i.e., concordant) sibling pairs have been applied to the trait^{8,18–22}, these have been in relatively small samples prior to our own GWLS²³, which showed genome-wide significant linkage to the pericentromeric region of chromosome 8 and strong support for linkage to the previously reported Xq28 region. Although genetic variant(s) contributing to development of male sexual orientation have been mapped to pericentromeric chromosome 8 and chromosome Xq28, the linkage peaks are large and specific trait genes have not been identified.

Genetic association studies for male sexual orientation have been sparse, small, and used either a convenience sample with proxy markers (e.g., blood type) or usually a candidate gene approach, and have yielded negative findings (strongest finding being a nominal $p = 0.03$)^{24–27}. No genome-wide association studies (GWAS) for male sexual orientation have been heretofore published in the peer-reviewed literature. One meeting report (2012 ASHG) from the company 23andMe analyzed male sexual orientation (N = 13,733 European ancestry men) as a

¹Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, Evanston, Illinois, 60201, United States of America. ²Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois, 60637, United States of America. ³Department of Human Genetics, University of Miami, Miami, Florida, 33136, United States of America. ⁴Department of Psychology, Pennsylvania State University, University Park, Pennsylvania, 16802, United States of America. ⁵Department of Psychology, University of Essex, Colchester, England, CO4 3SQ, United Kingdom. ⁶Department of Psychiatry, Rush University Medical Center, Chicago, Illinois, 60612, United States of America. ⁷Department of Psychiatry, Icahn School of Medicine at Mount Sinai, Elmhurst, New York, 11373, United States of America. ⁸Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, New York, 10027, United States of America. ⁹Department of Psychology, Northwestern University, Evanston, Illinois, 60208, United States of America. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to A.R.S. (email: alan.sanders.md@gmail.com)

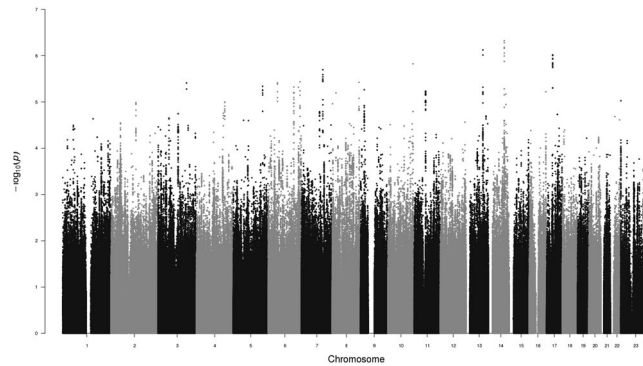


Figure 1. Manhattan Plot of GWAS for Male Sexual Orientation. Plot of negative \log_{10} of the p -values for the single SNP association analysis of 1,077 homosexual and 1,231 heterosexual men, ordered along the x-axis for each chromosome by chromosomal position.

continuous variable²⁸ and identified its strongest association ($p = 7.1 \times 10^{-8}$, direction unstated) in rs77013977²⁹, an intronic SNP in *NKAIN3* (*Na⁺/K⁺ transporting ATPase interacting 3*), which lies in a multipoint linkage peak on the pericentromeric region of chromosome 8 identified by our lab²³. To extend our gene mapping efforts for the trait, we report here the results from analyzing 1,109 homosexual and 1,231 heterosexual primarily European ancestry men in the first published GWAS on the trait.

Results/Discussion

We detected several promising regions of multiple SNPs in the 10^{-5} – 10^{-7} p -value range, as seen in the Manhattan plot (Fig. 1), though no SNP reached genome-wide significance (5×10^{-8}). The most prominent of these regions were on chromosomes 13 (minimum $p = 7.5 \times 10^{-7}$, rs9547443) and 14 ($p = 4.7 \times 10^{-7}$, rs1035144), where some SNPs had $10^{-7} < p < 10^{-6}$ (each region with 9 to 10 SNPs with $p < 10^{-5}$, Table S1). There are a number of genes of relevance to the trait in and around these regions, which we describe below. We further note that the most significant SNP (rs77013977, $p = 7.1 \times 10^{-8}$) in the 23andMe male GWAS²⁹ was nominally associated ($p = 4.1 \times 10^{-3}$) in our own GWAS. We used a meta-analytic statistic that did not need direction of effect, Fisher's combined probability test³⁰, which yielded $p = 6.7 \times 10^{-9}$ for this SNP, which is the first reported genome-wide significant association for the trait. As previously noted²⁹, rs77013977 is an intronic SNP in *NKAIN3*, which is one of a family of four proteins (*NKAIN1–4*) suggested to be critical for neuronal function³¹.

The strongest associated region on chromosome 13 (rs9547443, $p = 7.5 \times 10^{-7}$) was located between *SLITRK6* (*SLIT and NTRK like family member 6*, ~60 kb centromeric to region) and *SLITRK5* (~1.8 Mb telomeric), with *SLITRK1* located ~2.0 Mb centromeric. Members of the *SLITRK* protein family are brain-expressed neuronal transmembrane proteins that regulate neuronal outgrowth, survival, and synapse formation; *SLITRKs* have significant homology to the secreted axonal growth-controlling *SLIT* family of proteins and also homology to the neurotrophic tyrosine kinase receptor (*NTRK*) family^{32–34}. *SLITRK6* is expressed especially in the diencephalon (which contains a region previously reported as differing in size in men by sexual orientation³⁵), and *SLITRK1* and *SLITRK5* have their highest expression in the cerebral cortex^{32–34}. Gene families, such as the *SLITRK* family that are important for neurodevelopment and are implicated as candidate genes for various neuropsychiatric phenotypes³⁴, are also of potential relevance to behavioral phenotypes such as sexual orientation.

On chromosome 14, *TSHR* (*thyroid stimulating hormone receptor*) spans the region around our most significant SNP (rs1035144, $p = 4.7 \times 10^{-7}$), and includes a cluster of SNPs with association $p < 10^{-5}$ in intron 1. *TSHR* encodes a G protein-coupled transmembrane receptor for thyrotropin (thyroid stimulating hormone) and thyrostimulin, manifests some constitutive activity (i.e., ligand independent), and is a major controller of thyroid cell metabolism^{36–38}. While the main tissue of interest and expression for *TSHR* is the thyroid gland, *TSHR* is expressed in other tissues including brain especially in neuron-rich areas (e.g., hippocampus)³⁹. *TSHR* codes for the major autoantigen in the autoimmune hyperthyroidism of Graves' disease, which is associated ($p < 10^{-20}$ with ORs 1.4–1.5) with intron 1 polymorphisms^{40–48}. A recent population-based study found that 5,351 same-sex married men among the assayed population of 2,252,751 Danish men had an elevated rate ratio of Graves' disease (RR = 1.88; 95% CI = 1.08–3.01), a finding which held when excluding men with HIV/AIDS⁴⁹. The authors⁴⁹ speculate on the possibility that a genetic (or other prenatal) factor might tie together this increased risk for a type of hyperthyroidism (Graves' disease) with separate observations of lower body weight for homosexual versus heterosexual men (independent of diet or exercise)^{50–52}. Females with Graves' disease have been reported to manifest biased X chromosome inactivation^{53–55}, and skewed X chromosome inactivation has also been reported in mothers of homosexual men compared to age-matched mothers of heterosexual men⁵⁶. Furthermore, a recent retrospective chart review of 790 adolescents (8 to 17 years) previously admitted to a child psychiatry service found 15 mothers with a history of thyroid dysfunction during pregnancy, 16 adolescents with a history of same-sex attraction and/or gender nonconformity, and 12 overlapping mother-offspring pairs with both ($p < 0.0001$), suggestive of a possible relationship⁵⁷. Thus converging findings, including suggestive evidence from the current study, point to a possible connection between thyroid function and sexual orientation in men.

The main limitations of the current study include an exclusive focus on males, sampling primarily from one ancestral group (European), combination of two datasets, and most notably the modest sample size for a GWAS

on a trait with complex genetics. Additional and larger sample sizes would be required to assess which loci might breach genome-wide significance for association in a single study, and to increase the number of such loci (as typically is the case with phenotypes with manifesting complex genetics^{58,59}). Nevertheless, our study provides support for the top finding from a previous meeting report of a GWAS on the trait²⁹, reaching genome-wide significance for the combined analysis of rs77013977 ($p = 6.7 \times 10^{-9}$) on pericentromeric chromosome 8. In addition, the current study's top two association peaks ($p < 10^{-5}$; Fig. 1) provide interesting and perhaps trait-relevant examples of their closest genes on chromosomes 13 (*SLITRK6*) and 14 (*TSHR*), though these potential connections are best characterized as speculative. The continued genetic study of male sexual orientation should help open a gateway to other studies focusing on genetic and environmental mechanisms of sexual orientation and development. Detectable genetic variants predisposing to homosexuality would have alternative alleles, which would necessarily predispose to heterosexuality, thus contributing to understanding of both typical heterosexual and minority homosexual orientations.

Methods

Study Sample. We obtained institutional review board approval from NorthShore University HealthSystem, and after a study description all enrolled subjects gave informed consent. All methods were performed in accordance with the relevant guidelines and regulations. Our GWAS analyzed sample of primarily European ancestry men included a total of 1,077 homosexual and 1,231 heterosexual men, comprised as follows. In our GWLS on 409 pairs of homosexual brothers in 384 multiplex families²³, we classified men as homosexual based on both their self-reported sexual identity and sexual feelings (Kinsey 5–6). Isolated DNA samples were genotyped (Affymetrix 5.0 SNP array) at Vanderbilt Microarray Shared Resource. For this GWAS, after removing 12 families (two homosexual brothers each) for being ancestry outliers via principal components analysis (PCA), we included all remaining homosexual males ($N = 769$) and all heterosexual males ($N = 33$) from the GWLS dataset (372 families, each with two or more homosexual brothers, Supplementary Table S2). In addition, we genotyped (same Affymetrix 5.0 platform) additional males (same phenotypic definitions), with the following subjects being retained after quality control (QC, below): 221 homosexual males and 13 heterosexual males (from 227 partially completed linkage families, i.e., reported to contain two or more homosexual brothers, Supplementary Table S2), and 51 sporadic homosexual males (i.e., without homosexual brothers). The aforementioned subjects were collected as previously described²³, largely from community festival venues. We also incorporated our Molecular Genetics of Schizophrenia (MGS) collaboration controls dataset (the male, European ancestry portion retained after QC: 36 homosexual and 1,185 heterosexual) that was genotyped with Affymetrix 6.0 at the Broad Institute, and was queried regarding sexual orientation (identity)^{60,61}.

Statistical analysis. As part of the QC design to help minimize errors due to platform-specific genotype calling differences, we genotyped 34 duplicate subjects on both platforms (Affymetrix 5.0 and 6.0), removing SNPs discrepant for any of the 34 inter-platform duplicates (Supplementary Table S3). Our further standard GWAS QC^{62,63} filters included removal of SNPs (minor allele frequency, $MAF < 0.05$; missingness $\geq 1\%$; Hardy-Weinberg equilibrium [HWE] deviation $p < 10^{-6}$; Supplementary Table S4) and removal of samples (missingness $> 5\%$; failing checks for duplications and relatedness; ancestry outliers via PCA; Supplementary Table S3, and also see Supplementary Figure S1 for plotting of PC1 and PC2), resulting in $\lambda_{1000} = 1.029$ (i.e., a low genomic inflation factor). For the X chromosome, we followed the same QC protocol, but estimated HWE only in females (mothers in multiplex families). The pseudoautosomal regions (PAR) were removed prior to analysis. We imputed to 1,000 Genomes⁶⁴ using the IMPUTE2 software⁶⁵ (removing SNPs with an information score < 0.6 , $MAF < 0.05$) prior to performing GWAS analyses. The final QC'd SNP dataset contained a total of 5,642,880 retained SNPs (200,367 typed and 5,442,513 imputed). Association analysis was conducted using logistic regression with typed and imputed data in the R package, Genome-Wide Association analyses with Family (GWAFF)⁶⁶, with PC1 and PC2 as covariates. Regional association plots for the two regions highlighted in Supplementary Table S1 are displayed in Supplementary Figure S2.

Data Availability. The MGS collaboration controls dataset has been previously deposited into the database of Genotypes and Phenotypes (dbGaP, dbgap.ncbi.nlm.nih.gov, phs000021 and phs000167). The GWLS dataset generation was previously described²³. For all studied subjects, additional information is included within the accompanying supporting information (Supplementary Tables S1–S4 and Supplementary Figures S1–S2). The results from all data analyzed during this study are included or displayed in this published article (and its Supplementary Information files, such as the Manhattan plot and PCA plot), though only top individual SNP results for the two strongest regions are tabulated (Table S1) and regionally displayed (Figure S2). However, other individual SNP results may be made available from the corresponding author on reasonable request.

References

1. Laumann, E. O., Gagnon, J. H., Michael, R. T. & Michaels, S. *The Social Organization of Sexuality: Sexual Practices in the United States* (University of Chicago Press, 1994).
2. Kinsey, A. C., Pomeroy, W. B. & Martin, C. E. *Sexual Behavior in the Human Male* (W. B. Saunders Company, 1948).
3. Pillard, R. C. & Bailey, J. M. Human sexual orientation has a heritable component. *Hum. Biol.* **70**, 347–365 (1998).
4. Bailey, J. M., Dunne, M. P. & Martin, N. G. Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. *J. Pers. Soc. Psychol.* **78**, 524–536 (2000).
5. Bailey, J. M. *et al.* Sexual Orientation, Controversy, and Science. *Psychol. Sci. Public Interest.* **17**, 45–101 (2016).
6. Pillard, R. C. & Weinrich, J. D. Evidence of familial nature of male homosexuality. *Arch. Gen. Psychiatry.* **43**, 808–812 (1986).
7. Bailey, J. M. & Pillard, R. C. A genetic study of male sexual orientation. *Arch. Gen. Psychiatry.* **48**, 1089–1096 (1991).
8. Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N. & Pattatucci, A. M. A linkage between DNA markers on the X chromosome and male sexual orientation. *Science.* **261**, 321–327 (1993).
9. Bailey, J. M. & Bell, A. P. Familiality of female and male homosexuality. *Behav. Genet.* **23**, 313–322 (1993).

10. Bailey, J. M. *et al.* A family history study of male sexual orientation using three independent samples. *Behav. Genet.* **29**, 79–86 (1999).
11. Schwartz, G., Kim, R. M., Kolundzija, A. B., Rieger, G. & Sanders, A. R. Biodemographic and physical correlates of sexual orientation in men. *Arch. Sex. Behav.* **39**, 93–109 (2010).
12. Kirk, K. M., Bailey, J. M., Dunne, M. P. & Martin, N. G. Measurement models for sexual orientation in a community twin sample. *Behav. Genet.* **30**, 345–356 (2000).
13. Kendler, K. S., Thornton, L. M., Gilman, S. E. & Kessler, R. C. Sexual orientation in a U.S. national sample of twin and nontwin sibling pairs. *Am. J. Psychiatry.* **157**, 1843–1846 (2000).
14. Santtila, P. *et al.* Potential for homosexual response is prevalent and genetic. *Biol. Psychol.* **77**, 102–105 (2008).
15. Langstrom, N., Rahman, Q., Carlstrom, E. & Lichtenstein, P. Genetic and environmental effects on same-sex sexual behavior: a population study of twins in Sweden. *Arch. Sex. Behav.* **39**, 75–80 (2010).
16. Alanko, K. *et al.* Common genetic effects of gender atypical behavior in childhood and sexual orientation in adulthood: a study of Finnish twins. *Arch. Sex. Behav.* **39**, 81–92 (2010).
17. Rice, G., Risch, N. & Ebers, G. Genetics and male sexual orientation. *Science.* **285**, 803 (1999).
18. Hu, S. *et al.* Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nat. Genet.* **11**, 248–256 (1995).
19. Sanders, A. R. *et al.* Genetic linkage study of male homosexual orientation. 151st Annual Meeting (American Psychiatric Association, New Research Abstracts NR149, Toronto, Ontario, Canada, 1998).
20. Rice, G., Anderson, C., Risch, N. & Ebers, G. Male homosexuality: absence of linkage to microsatellite markers at Xq28. *Science.* **284**, 665–667 (1999).
21. Mustanski, B. S. *et al.* A genomewide scan of male sexual orientation. *Hum. Genet.* **116**, 272–278 (2005).
22. Ramagopalan, S. V., Dymont, D. A., Handunnethi, L., Rice, G. P. & Ebers, G. C. A genome-wide scan of male sexual orientation. *J. Hum. Genet.* **55**, 131–132 (2010).
23. Sanders, A. R. *et al.* Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychol. Med.* **45**, 1379–1388 (2015).
24. DuPree, M. G., Mustanski, B. S., Bocklandt, S., Nievergelt, C. & Hamer, D. H. A candidate gene study of CYP19 (aromatase) and male sexual orientation. *Behav. Genet.* **34**, 243–250 (2004).
25. Wang, B. *et al.* Association analysis between the tag SNP for sonic hedgehog rs9333613 polymorphism and male sexual orientation. *J. Androl.* **33**, 951–954 (2012).
26. Macke, J. P. *et al.* Sequence variation in the androgen receptor gene is not a common determinant of male sexual orientation. *Am. J. Hum. Genet.* **53**, 844–852 (1993).
27. Ellis, L., Ficek, C., Burke, D. & Das, S. Eye color, hair color, blood type, and the rhesus factor: Exploring possible genetic links to sexual orientation. *Arch. Sex. Behav.* **37**, 145–149 (2008).
28. Klein, F. *The Bisexual Option* (The Haworth Press, 1993).
29. Drabant, E. M. *et al.* Genome Wide Association Study of Sexual Orientation in a Large, Web-based Cohort (<https://blog.23andme.com/wp-content/uploads/2012/11/Drabant-Poster-v7.pdf>). 62nd Annual Meeting (American Society of Human Genetics, Poster 2100W, San Francisco, California, 2012).
30. Fisher, R. A. *Statistical Methods for Research Workers* (Oliver and Boyd, 1925).
31. Gorokhova, S., Bibert, S., Geering, K. & Heintz, N. A novel family of transmembrane proteins interacting with beta subunits of the Na,K-ATPase. *Hum. Mol. Genet.* **16**, 2394–2410 (2007).
32. Aruga, J. & Mikoshiba, K. Identification and characterization of Slitrk, a novel neuronal transmembrane protein family controlling neurite outgrowth. *Mol. Cell. Neurosci.* **24**, 117–129 (2003).
33. Aruga, J., Yokota, N. & Mikoshiba, K. Human SLITRK family genes: genomic organization and expression profiling in normal brain and brain tumor tissue. *Gene.* **315**, 87–94 (2003).
34. Proenca, C. C., Gao, K. P., Shmelkov, S. V., Rafii, S. & Lee, F. S. Slitrks as emerging candidate genes involved in neuropsychiatric disorders. *Trends Neurosci.* **34**, 143–153 (2011).
35. LeVay, S. A difference in hypothalamic structure between heterosexual and homosexual men. *Science.* **253**, 1034–1037 (1991).
36. Kleinau, G. & Biebermann, H. Constitutive activities in the thyrotropin receptor: regulation and significance. *Adv. Pharmacol.* **70**, 81–119 (2014).
37. Huth, S., Jaeschke, H., Schaarschmidt, J. & Paschke, R. Controversial constitutive TSHR activity: patients, physiology, and *in vitro* characterization. *Horm. Metab. Res.* **46**, 453–461 (2014).
38. Kleinau, G., Neumann, S., Gruters, A., Krude, H. & Biebermann, H. Novel insights on thyroid-stimulating hormone receptor signal transduction. *Endocr. Rev.* **34**, 691–724 (2013).
39. Crisanti, P. *et al.* The expression of thyrotropin receptor in the brain. *Endocrinology.* **142**, 812–822 (2001).
40. Davies, T. F., Yin, X. & Latif, R. The genetics of the thyroid stimulating hormone receptor: history and relevance. *Thyroid.* **20**, 727–736 (2010).
41. Ho, S. C., Goh, S. S. & Khoo, D. H. Association of Graves' disease with intragenic polymorphism of the thyrotropin receptor gene in a cohort of Singapore patients of multi-ethnic origins. *Thyroid.* **13**, 523–528 (2003).
42. Dechairo, B. M. *et al.* Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur. J. Hum. Genet.* **13**, 1223–1230 (2005).
43. Yin, X., Latif, R., Bahn, R., Tomer, Y. & Davies, T. F. Influence of the TSH receptor gene on susceptibility to Graves' disease and Graves' ophthalmopathy. *Thyroid.* **18**, 1201–1206 (2008).
44. Brand, O. J. *et al.* Association of the thyroid stimulating hormone receptor gene (TSHR) with Graves' disease. *Hum. Mol. Genet.* **18**, 1704–113 (2009).
45. Ploski, R. *et al.* Thyroid stimulating hormone receptor (TSHR) intron 1 variants are major risk factors for Graves' disease in three European Caucasian cohorts. *PLoS One.* **5**, e15512 (2010).
46. Pujol-Borrell, R., Gimenez-Barcons, M., Marin-Sanchez, A. & Colobran, R. Genetics of Graves' Disease: Special Focus on the Role of TSHR Gene. *Horm. Metab. Res.* **47**, 753–766 (2015).
47. Chu, X. *et al.* A genome-wide association study identifies two new risk loci for Graves' disease. *Nat. Genet.* **43**, 897–901 (2011).
48. Liu, B. L. *et al.* Refined association of TSH receptor susceptibility locus to Graves' disease in the Chinese Han population. *Eur. J. Endocrinol.* **170**, 109–119 (2014).
49. Frisch, M., Nielsen, N. M. & Pedersen, B. V. Same-sex marriage, autoimmune thyroid gland dysfunction and other autoimmune diseases in Denmark 1989–2008. *Eur. J. Epidemiol.* **29**, 63–71 (2014).
50. Blanchard, R. & Bogaert, A. F. Biodemographic comparisons of homosexual and heterosexual men in the Kinsey interview data. *Arch. Sex. Behav.* **25**, 551–579 (1996).
51. Frisch, M. & Zdravkovic, S. Body size at birth and same-sex marriage in young adulthood. *Arch. Sex. Behav.* **39**, 117–123 (2010).
52. Deputy, N. P. & Boehmer, U. Determinants of body weight among men of different sexual orientation. *Prev. Med.* **51**, 129–131 (2010).
53. Brix, T. H. *et al.* High frequency of skewed X-chromosome inactivation in females with autoimmune thyroid disease: a possible explanation for the female predisposition to thyroid autoimmunity. *J. Clin. Endocrinol. Metab.* **90**, 5949–5653 (2005).
54. Ozelik, T. *et al.* Evidence from autoimmune thyroiditis of skewed X-chromosome inactivation in female predisposition to autoimmunity. *Eur. J. Hum. Genet.* **14**, 791–797 (2006).

55. Yin, X., Latif, R., Tomer, Y. & Davies, T. F. Thyroid epigenetics: X chromosome inactivation in patients with autoimmune thyroid disease. *Ann. N. Y. Acad. Sci.* **1110**, 193–200 (2007).
56. Bocklandt, S., Horvath, S., Vilain, E. & Hamer, D. H. Extreme skewing of X chromosome inactivation in mothers of homosexual men. *Hum. Genet.* **118**, 691–694 (2006).
57. Sabuncuoglu, O. High rates of same-sex attraction/gender nonconformity in the offspring of mothers with thyroid dysfunction during pregnancy: Proposal of prenatal thyroid model. *Ment. Illn.* **7**, 5810 (2015).
58. Robinson, M. R., Wray, N. R. & Visscher, P. M. Explaining additional genetic variation in complex traits. *Trends Genet.* **30**, 124–132 (2014).
59. Levinson, D. F. *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biol. Psychiatry.* **76**, 510–512 (2014).
60. Sanders, A. R. *et al.* The internet-based MGS2 control sample: Self report of mental illness. *Am. J. Psychiatry.* **167**, 854–865 (2010).
61. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature.* **460**, 753–757 (2009).
62. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).
63. Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Curr. Protoc. Hum. Genet.* Chapter 1, Unit119 (2011).
64. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature.* **467**, 1061–1073 (2010).
65. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
66. Chen, M. H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics.* **26**, 580–581 (2010).

Acknowledgements

This work was supported by NICHD: the Eunice Kennedy Shriver National Institute of Child Health and Human Development (A.R.S., grant no. R01HD041563 for the Affymetrix 5.0 genotyped sample and A.R.S. and E.R.M., grant no. R21HD080410 for various analyses), and by the Molecular Genetics of Schizophrenia (MGS) Collaboration for the Affymetrix 6.0 genotyped sample. The MGS Collaboration includes P.V.G., A.R.S., and J.D., in addition to the individuals in the MGS Collaboration group author list. MGS was mainly supported by R01MH059571, R01MH081800, and U01MH079469 (to P.V.G.), and other NIH grants for other MGS sites (R01MH067257 to N.G.B., R01MH059588 to B.J.M., R01MH059565 to R.F., R01MH059587 to F.A., R01MH060870 to W.F.B., R01MH059566 to D.W.B., R01MH059586 to J.M.S., R01MH061675 to D.F.L., R01MH060879 to C.R.C., U01MH046276 to C.R.C., and U01MH079470 to D.F.L.). We thank the men for their participation, Timothy F. Murphy for his work on the community advisory board and study website, and Besiana Liti for technical assistance.

Author Contributions

A.R.S. conceived the experiment. A.R.S., K.D., G.R., R.S.K., A.B.K. and J.M.B. collected the sample. A.R.S., K.D., G.R. and J.M.B. advised and contributed to the phenotypic characterization. A.R.S., J.A.B., E.S.G., J.D. and P.V.G. advised and contributed to the genotypic characterization. G.W.B., S.G. and E.R.M. conducted the statistical analyses. A.R.S., G.W.B., S.G., G.R., J.M.B., and E.R.M. wrote the main manuscript text. A.R.S., G.W.B., S.G. and E.R.M. prepared figures and tables. MGS Collaboration (J.S., D.F.L., J.D., A.R.S., B.J.M., A.O., F.A., C.R.C., D.M.S., J.M.S., N.G.B., W.F.B., D.W.B., R.F. & P.V.G.) provided the Affymetrix 6.0 genotyped sample. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-15736-4>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

Consortia MGS Collaboration

Jianxin Shi¹⁰, Douglas F. Levinson¹¹, Bryan J. Mowry¹², Ann Olincy¹³, Farooq Amin¹⁴, C. Robert Cloninger¹⁵, Dragan M. Svrakic¹⁵, Jeremy M. Silverman¹⁶, Nancy G. Buccola¹⁷, William F. Byerley¹⁸, Donald W. Black¹⁹ & Robert Freedman¹³

¹⁰Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, 20892, United States of America. ¹¹Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, 94305, United States of America. ¹²Queensland Centre for Mental Health Research, Brisbane and Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, 4072, Australia. ¹³Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, 80045, United States of America. ¹⁴Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, and Emory University, Atlanta, Georgia, 30329, United States of America. ¹⁵Department of Psychiatry, Washington University, St. Louis, Missouri, 63110, United States of America. ¹⁶Department of Psychiatry, Mount Sinai School of Medicine New York, New York, 10029, United States of America. ¹⁷School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana, 70112, United States of America. ¹⁸Department of Psychiatry, University of California at San Francisco, San Francisco, California, 94143, United States of America. ¹⁹Mental Health Clinical Research Center, and Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa, 52242, United States of America.