



Genome-Wide Linkage Study Meta-Analysis of Male Sexual Orientation

Alan R. Sanders^{1,2} · Gary W. Beecham^{3,4} · Shengru Guo³ · Judith A. Badner⁵ · Sven Bocklandt⁶ · Brian S. Mustanski⁷ · Dean H. Hamer⁶ · Eden R. Martin^{3,4}

Received: 30 August 2019 / Revised: 1 May 2021 / Accepted: 4 May 2021 / Published online: 2 June 2021
© The Author(s) 2021

Abstract

Male sexual orientation is a scientifically and socially important trait shown by family and twin studies to be influenced by environmental and complex genetic factors. Individual genome-wide linkage studies (GWLS) have been conducted, but not jointly analyzed. Two main datasets account for > 90% of the published GWLS concordant sibling pairs on the trait and are jointly analyzed here: MGSOSO (Molecular Genetic Study of Sexual Orientation; 409 concordant sibling pairs in 384 families, Sanders et al. (2015)) and Hamer (155 concordant sibling pairs in 145 families, Mustanski et al. (2005)). We conducted multipoint linkage analyses with Merlin on the datasets separately since they were genotyped differently, integrated genetic marker positions, and combined the resultant LOD (logarithm of the odds) scores at each 1 cM grid position. We continue to find the strongest linkage support at pericentromeric chromosome 8 and chromosome Xq28. We also incorporated the remaining published GWLS dataset (on 55 families) by using meta-analytic approaches on published summary statistics. The meta-analysis has maximized the positional information from GWLS of currently available family resources and can help prioritize findings from genome-wide association studies (GWAS) and other approaches. Although increasing evidence highlights genetic contributions to male sexual orientation, our current understanding of contributory loci is still limited, consistent with the complexity of the trait. Further increasing genetic knowledge about male sexual orientation, especially via large GWAS, should help advance our understanding of the biology of this important trait.

Keywords Chromosome 8 · Chromosome Xq28 · Complex trait · Genome-wide linkage scan · Sexual orientation

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10508-021-02035-3>.

✉ Alan R. Sanders
alan.sanders.md@gmail.com

¹ Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, 1001 University Place, Evanston, IL 60201, USA

² Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA

³ John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA

⁴ Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA

⁵ Department of Psychiatry, Rush University, Chicago, IL, USA

⁶ National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

⁷ Department of Medical Social Sciences, Northwestern University, Evanston, IL, USA

Introduction

Male homosexuality runs in families, and twin studies have shown that genetic contributions appear to account for a moderate proportion of the variation in male sexual orientation with heritability estimated at ~ 32% (for review, see Bailey et al., 2016). Three genome-wide linkage studies (GWLS) have been conducted on male sexual orientation, all focusing on concordant sibling pairs (2010homosexual brothers)—we refer here to these GWLS datasets as Hamer (Mustanski et al., 2005), MGSOSO (Molecular Genetic Study of Sexual Orientation) (Sanders et al., 2015), and Canadian (Ramagopal et al.,). The Hamer GWLS combined samples from two earlier studies (Hamer et al., 1993; Hu et al., 1995) with newly collected families (Mustanski et al., 2005) to total 155 independent concordant sibling pairs in 145 families. While linkage to chromosome Xq28 was prominent in the earlier linkage studies focusing on chromosome X (Hamer et al., 1993; Hu et al., 1995), the Hamer GWLS instead had its

strongest finding of suggestive linkage at chromosome 7q36 (Mustanski et al., 2005). Another research group collected 55 families in Canada and performed a GWLS, with the strongest (albeit not significant) linkage reported at chromosome 14q32 (Ramagopalan et al., 2010). The MGSOSO performed a GWLS on 409 independent concordant sibling pairs in 384 families, making its strongest finding of significant (Lander & Kruglyak, 1995) linkage at pericentromeric chromosome 8 and also detecting suggestive (Lander & Kruglyak, 1995) linkage (supportive evidence of previous findings) at chromosome Xq28 (Sanders et al., 2015). In order to extract the maximal positional information from GWLS of currently available family resources, we jointly analyzed the Hamer and MGSOSO datasets (and included the Canadian dataset by meta-analyzing published summary statistics).

Method

Joint Linkage Analyses

The two jointly analyzed datasets used very similar phenotype definitions for homosexual men from their questionnaire data: Hamer used “Kinsey 5–6” for several questions (attraction, fantasy, behavior, and self-identification) (Mustanski et al., 2005), and MGSOSO used “Kinsey 5–6” for fantasy along with homosexual identity (Sanders et al., 2015). The Hamer dataset consisted of 441 individuals in 145 families genotyped with 408 short tandem repeat polymorphism genetic markers (STRPs) (Mustanski et al., 2005), and the MGSOSO dataset consisted of 908 individuals in 384 families and genotyped with 45,387 single-nucleotide polymorphism genetic markers (SNPs) (Sanders et al., 2015). Various quality control steps had already been performed in the respective GWLS as previously detailed (Mustanski et al., 2005; Sanders et al., 2015). After obtaining collaborative access to genotypes for each dataset, we conducted multipoint nonparametric linkage analyses with Merlin v1.1.2 (Abecasis et al., 2002) on the Hamer (Mustanski et al., 2005) and MGSOSO (Sanders et al., 2015) datasets separately since they were genotyped differently (STRPs vs. SNPs). To integrate, we found the genetic positions of the respective markers in the Rutgers Map v.3 (hg19 build) (Nato et al., 2018) and then used the nonparametric S-pairs and grid 1 cM options to perform multipoint linkage on both data sets, followed

by combining LOD scores at each grid position across the marker sets.

Meta-Analyses of Summary Statistics

For phenotype definitions for homosexual men, the Canadian dataset used an interview approach based on identity and corroboration by sibling, and on a sub-sample all also had Kinsey 5–6 for several questions (attraction, fantasy, and behavior) (Rice et al., 1999a, b). As we were unable to access genotypes for the Canadian dataset (accounting for < 10% of the families in GWLS on the trait), we were only able to incorporate the Canadian GWLS by meta-analyzing summary statistics. Thus, we used the plotted multipoint Canadian GWLS Fig. 1 (Ramagopalan et al., 2010) and interpolated into cM bins enabling use of GWLS meta-analytic methods not needing genotypes, namely the multi-scan probability (MSP) approach utilizing regional *p*-values (Badner & Gershon, 2002), and the rank-based genome scan meta-analysis (GSMA) approach (Levinson et al., 2003; Wise & Lewis, 1999).

Results

The multipoint plots for the Hamer and the MGSOSO datasets for the current analyses (Supplementary Figs. 1 and 2, respectively) line up very well with the original GWLS manuscripts’ multipoint plots—Fig. 1a (Mustanski et al., 2005) and Fig. 1 (Sanders et al., 2015), respectively. This overlap of multipoint findings was found despite some differences between the original reports (Mustanski et al., 2005; Sanders et al., 2015) and the current manuscript in statistical analysis software (Aspex vs. Merlin for the Hamer dataset) and genetic map used (deCode vs. Rutgers for both the Hamer and MGSOSO datasets). The joint analysis of the combined Hamer and MGSOSO datasets is shown in Fig. 1, with zoomed-in plots of the top two multipoint linkage peaks from this joint GWLS depicted for chromosomes 8 and X in Fig. 2. The results of the meta-analyses of summary statistics from Hamer, MGSOSO, and Canadian GWLS datasets are presented in Supplementary Tables 1 (MSP) and 2 (GSMA).

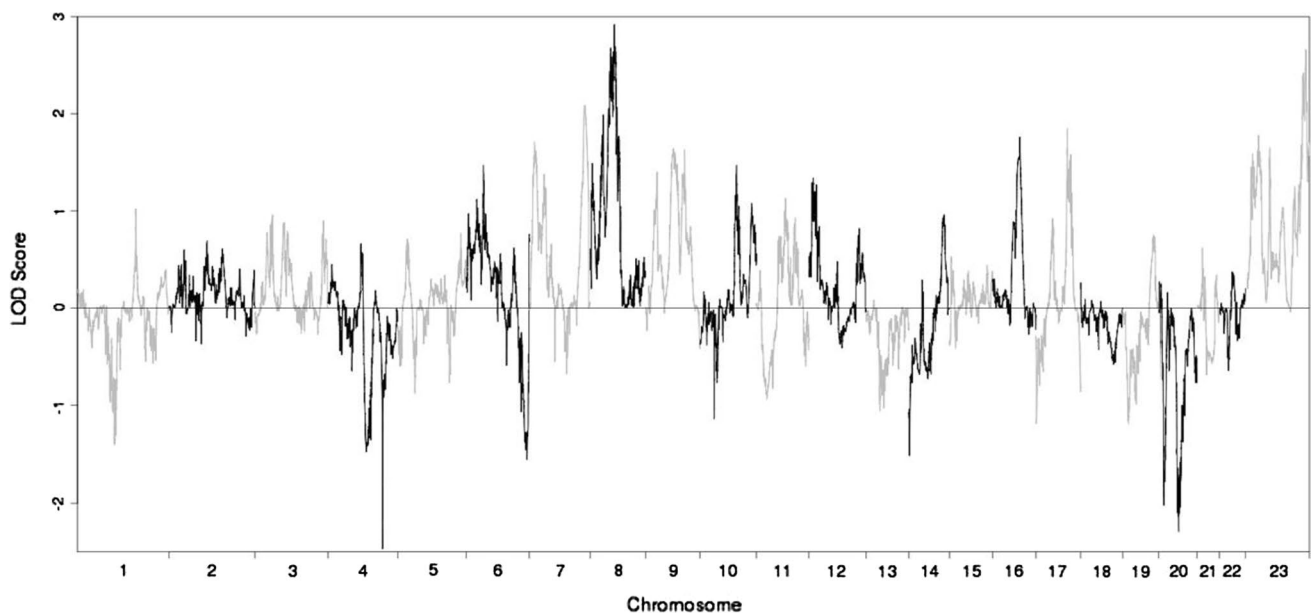


Fig. 1 Joint multipoint linkage analysis of the combined Hamer and MGSOSO datasets. Multipoint LOD scores are plotted v. the chromosomal positions for the nonparametric linkage analysis. Adjacent chromosomes are separated by alternating black and gray lines

Discussion

Our primary analysis for this investigation was the joint analysis of multipoint linkage from the Hamer and MGSOSO datasets (Mustanski et al., 2005; Sanders et al., 2015), to which each dataset contributed some peaks (Fig. 1, Supplementary Figs. 1 and 2). Overall, the maximum multipoint peaks increased little in height, though the pericentromeric chromosome 8 peak was broadened (Fig. 2). Chromosomes 8 and X retained the highest multipoint peaks genome-wide, mostly arising from the larger (MGSOSO) dataset (Fig. 2). The joint analysis gives a more comprehensive picture of shared and heterogeneous linkage regions (e.g., at pericentromeric chromosome 8), the studies share overlapping peaks (possibly suggesting heterogeneity, perhaps with different genes involved in the different datasets), and the evidence broadens the search. The secondary analyses on summary statistics using MSP and GSMA to incorporate all three (Hamer, MGSOSO, Canadian) GWLS datasets showed no genome-wide significant results though suggestive findings remained present. The joint analysis of multipoint linkage (Fig. 1) extracted the available positional information

from collaborating GWLS, though previous GWLS findings were not much further strengthened in these analyses. Nevertheless, this provides information to complement other approaches, such as helping prioritize findings from GWAS. Linkage and association studies measure different genetic properties (i.e., segregation of a region within families, vs. correlation of alleles in a population), both of which provide clues about underlying trait genetics. Thus, since GWLS are different from GWAS, we were unable to directly combine any GWAS (e.g., Ganna et al., 2019) with the studied GWLS in our GWLS meta-analysis. Limitations include those inherent to linkage (as opposed to GWAS) of traits with complex genetics (e.g., their limited utility for phenotypes with contributions from more than one or a few genes); on the other hand, linkage retains some advantages over association approaches, such as being robust to allelic heterogeneity (Lipner & Greenberg, 2018). Accumulating genetic studies of the trait such as by much enlarged GWAS (e.g., Ganna et al., 2019) will be especially useful, given its successful application in the study of other phenotypes manifesting complex genetics (e.g., Fig. 3b in Sullivan et al. (2018)).

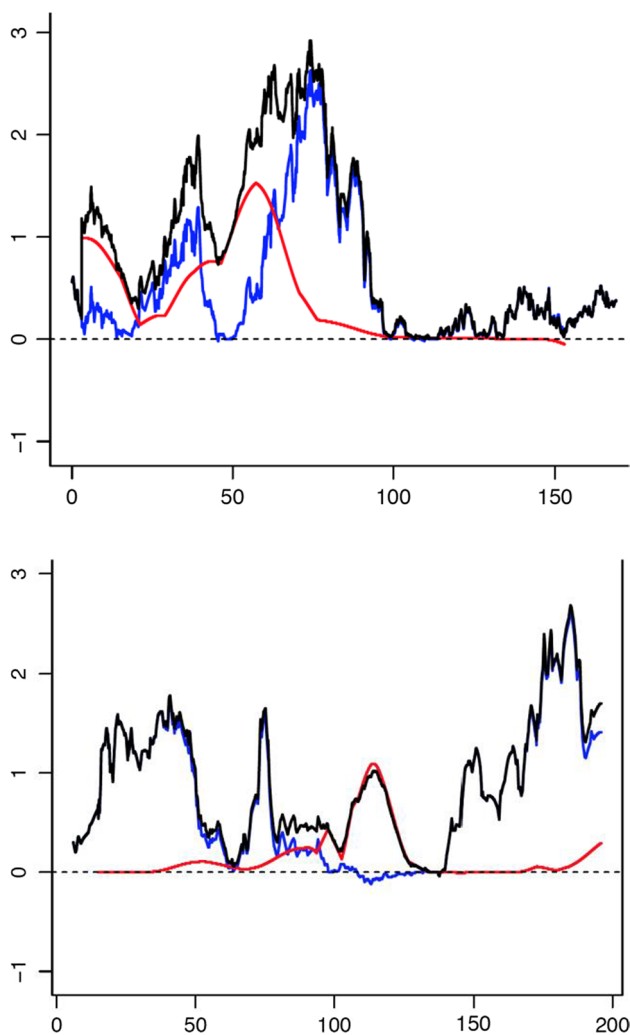


Fig. 2 Individual chromosomal plots of multipoint linkage peaks on chromosomes 8 and X. Multipoint LOD scores are plotted v. the chromosomal positions for the nonparametric linkage analysis, the red line indicating the results from the Hamer dataset, the blue line for the MGSOSO dataset, and the black line for the joint analysis. Chromosome 8 (top) has its multipoint linkage peak in the pericentromeric region, and chromosome X (bottom) has its multipoint linkage peak at Xq28

Acknowledgements This work was supported by NICHD, the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Award No. R01HD041563 for the linkage sample to Alan R. Sanders, M.D.; and Award No. R21HD080410 for meta-analyses to Alan R. Sanders, M.D. and Eden R. Martin, Ph.D.), and by intramural NIH funds (to Dean H. Hamer, Ph.D.). We thank the men and their families for their participation.

Compliance with Ethical Standards

Human and Animal Rights and Informed Consent There was no participant contact for the current study, as it was a meta-analysis of previously collected genetic data. Two of the three studied samples (MGSOSO and Hamer) describe ethical aspects in their earlier manuscripts in more detail, but we briefly summarize here. For the third sample, the Cana-

dian one (Ramagopalan et al., 2010), we only meta-analyzed published summary statistics, i.e., we used no individual level data. For MGSOSO, institutional review board (IRB) approval was obtained from North-Shore University HealthSystem, and all participants provided informed consent (Sanders et al., 2015). For the Hamer dataset, IRB approval was obtained from the National Cancer Institute, and all participants provided informed consent (Mustanski et al., 2005).

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2001). GRR: Graphical representation of relationship errors. *Bioinformatics*, *17*, 742–743. <https://doi.org/10.1093/bioinformatics/17.8.742>
- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, *30*, 97–101. <https://doi.org/10.1038/ng786>
- Badner, J. A., & Gershon, E. S. (2002). Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. *Molecular Psychiatry*, *7*, 56–66. <https://doi.org/10.1038/sj/mp/4000922>
- Bailey, J. M., Vasey, P. L., Diamond, L. M., Breedlove, S. M., Vilain, E., & Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest*, *17*, 45–101. <https://doi.org/10.1177/1529100616637616>
- Boyles, A. L., Scott, W. K., Martin, E. R., Schmidt, S., Li, Y. J., Ashley-Koch, A., Bass, M. P., Schmidt, M., Pericak-Vance, M. A., Speer, M. C., & Hauser, E. R. (2005). Linkage disequilibrium inflates type I error rates in multipoint linkage analysis when parental genotypes are missing. *Human Heredity*, *59*, 220–227. <https://doi.org/10.1159/000087122>
- Ganna, A., Verweij, K. J., Nivard, M. G., Maier, R., Wedow, R., Busch, A. S., Abdellaoui, A., Guo, S., Sathirapongsasuti, J. F., Lichtenstein, P., & Lundström, S. (2019). Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science*, *365*. <https://doi.org/10.1126/science.aat7693>
- Hamer, D. H. (1999). Genetics and male sexual orientation. *Science*, *285*, 803. <https://doi.org/10.1126/science.285.5429.803a>
- Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N., & Pattatucci, A. M. (1993). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science*, *261*, 321–327. <https://doi.org/10.1126/science.8332896>
- Hu, S., Pattatucci, A. M., Patterson, C., Li, L., Fulker, D. W., Cherny, S. S., Kruglyak, L., & Hamer, D. H. (1995). Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nature Genetics*, *11*, 248–256. <https://doi.org/10.1038/ng1195-248>
- Huang, Q., Shete, S., & Amos, C. I. (2004). Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis. *American Journal of Human Genetics*, *75*, 1106–1112. <https://doi.org/10.1086/426000>

- Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genetics*, *11*, 241–247. <https://doi.org/10.1038/ng1195-241>
- Levinson, D. F., Levinson, M. D., Segurado, R., & Lewis, C. M. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. *American Journal of Human Genetics*, *73*, 17–33. <https://doi.org/10.1086/376548>
- Lipner, E. M., & Greenberg, D. A. (2018). The rise and fall and rise of linkage analysis as a technique for finding and characterizing inherited influences on disease expression. *Methods in Molecular Biology*, *1706*, 381–397. https://doi.org/10.1007/978-1-4939-7471-9_21
- McPeck, M. S., & Sun, L. (2000). Statistical tests for detection of misspecified relationships by use of genome-screen data. *American Journal of Human Genetics*, *66*, 1076–1094. <https://doi.org/10.1086/302800>
- Mustanski, B. S., Dupree, M. G., Nievergelt, C. M., Bocklandt, S., Schork, N. J., & Hamer, D. H. (2005). A genomewide scan of male sexual orientation. *Human Genetics*, *116*, 272–278. <https://doi.org/10.1007/s00439-004-1241-4>
- Nato, A. Q., Buyske, S., & Matise, T. C. (2018). *The Rutgers map: A third-generation combined linkage-physical map of the human genome*. Retrieved from http://compgen.rutgers.edu/rutgers_maps.shtml
- Ramagopalan, S. V., Dymont, D. A., Handunnetthi, L., Rice, G. P., & Ebers, G. C. (2010). A genome-wide scan of male sexual orientation. *Journal of Human Genetics*, *55*, 131–132. <https://doi.org/10.1038/jhg.2009.135>
- Rice, G., Anderson, C., Risch, N., & Ebers, G. (1999a). Male homosexuality: Absence of linkage to microsatellite markers at Xq28. *Science*, *284*, 665–667. <https://doi.org/10.1126/science.284.5414.665>
- Rice, G., Risch, N., & Ebers, G. (1999b). Genetics and male sexual orientation. *Science*, *285*, 803. <https://doi.org/10.1126/science.285.5429.803a>
- Sanders, A. R., Martin, E. R., Beecham, G. W., Guo, S., Dawood, K., Rieger, G., Badner, J. A., Gershon, E. S., Krishnappa, R. S., Kolundzija, A. B., & Duan, J. (2015). Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychological Medicine*, *45*, 1379–1388. <https://doi.org/10.1017/S0033291714002451>
- Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børglum, A. D., Breen, G., Cichon, S., Edenberg, H. J., Faraone, S. V., Gelernter, J., & Mathews, C. A. (2018). Psychiatric genomics: An update and an agenda. *American Journal of Psychiatry*, *175*, 15–27. <https://doi.org/10.1176/appi.ajp.2017.17030283>
- Wigginton, J. E., & Abecasis, G. R. (2005). PEDSTATS: Descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics*, *21*, 3445–3447. <https://doi.org/10.1093/bioinformatics/bti529>
- Wise, L. H., & Lewis, C. M. (1999). A method for meta-analysis of genome searches: Application to simulated data. *Genetic Epidemiology*, *17*(Suppl. 1), S767–S771. <https://doi.org/10.1002/gepi.13701707126>

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