Genome-Wide Association Study of Sexual Orientation in a Large, Web-based Cohort

E.M. Drabant[1], A. K. Kiefer[1], N. Eriksson[1], J. L. Mountain[1], U. Francke[1], J. Y. Tung[1], D. A. Hinds[1], C. B. Do[1]

[1]23andMe, Inc, Mountain View, CA

Introduction

There is considerable variation in human sexual orientation. Heritability studies have differed on the precise contribution of genetics to this variation in sexual orientation, but it appears that both genetics and environment play a role. Though a few linkage studies have indicated a possible role for certain genes on the X chromosome (1, 2), the strength of that evidence is limited due to the conflicting nature of the reports and small sample sizes. We sought to clarify some of the questions surrounding the possible genetic underpinnings of sexual orientation by releasing a web-based survey to the large 23andMe database of over 180,000 individuals and conducting the first ever genome-wide association study (GWAS) on sexual orientation.

In parallel, we also sought to explore other non-genetic phenotypes associated with sexual orientation. By leveraging data collected from dozens of 23andMe surveys on ~1000 conditions and traits, we aimed to understand the relationship between sexual orientation and non-genetic phenotypes, including previous findings on substance use and mental illness. Research has shown that lesbians are more likely than heterosexual women to have alcoholism and alcohol-related problems (3). A number of studies have also found that women with same sex partners are more likely to have psychiatric disorders, including major depression (4) and men with same sex partners are more likely to have anxiety and mood disorders (5).

Methods

Survey

23andMe customers who were over 18 and provided consent were eligible to complete the Sexual Orientation Survey. All survey respondents were genotyped on 23andMe's custom version of an Illumina array with between 500,000 and 1 million single nucleotide polymorphisms (SNPs).

The 33-item Sexual Orientation Survey included a subset of 7 questions from the Klein Grid (6), as well as questions covering the following topics:

- Gender definition (e.g. male, male to female transgender)
- Age of awareness, sexual behavior, coming out
- Birth order, gender of siblings, adoption, family history of homosexuality
- Diagnosis of adrenal disorders
- Attitudes towards homosexuality

Information about other phenotypes was gathered through other existing 23andMe online surveys that were also filled out by the participants.

Table 1. Questions Selected from the Klein Grid (6)

Sexual Identity	How do you label, identify, or think of yourself?
Sexual Behavior	With whom have you actually had sex?
Sexual Fantasies	Whom do you have sexual fantasies about?
Sexual Attraction	To whom are you sexually attracted?
Emotional Attraction	Whom do you feel more drawn to or close to emotionally?
Social Preference	Which gender do you socialize with?
Lifestyle Preference	In which community do you like to spend your time? In which do you feel most comfortable?

Answers were provided on a 7 point scale ranging from heterosexual only to homosexual only, with an additional option "I'd rather not say".

Table 2. Questions on Attitudes Towards Homosexuality

Sex between two men is just plain wrong.
I think male homosexuals are disgusting.
Male homosexuality is a natural expression of sexuality in men.
Sex between two women is just plain wrong.
I think female homosexuals are disgusting.
Female homosexuality is a natural expression of sexuality in women.

Answers were provided on a 5 point scale ranging from "strongly agree" to "strongly disagree".

Analyses

All analyses focused on the "sexual identity" item from the Klein Grid. Individuals who responded "I'd rather not say" were not included in the analyses. Individuals who defined their gender as "transgender" (male to female, female to male) or "other" were not included in the analyses (n=109). Analyses were conducted separately for men and women.

Phenotype analyses were conducted using linear or logistic regression. The reported betas are the change on the sexual identity scale per unit of the other phenotype.

Both the phenotype analyses and GWAS analyses controlled for age, the first five principal components, and attitudes towards homosexuality (as collected in the Sexual Orientation Survey). GWAS analyses were conducted in individuals of European descent.

Results

Prevalence

Prevalence of homosexuality is difficult to estimate due to sample bias and participants' unwillingness to divulge their sexual orientation, among other things. Estimates range from 0.5% of the population to 7% of men and 8% of women identifying as gay, lesbian or bisexual (7,8). Over 23,000 individuals in the 23andMe database completed the survey. 77% of participants identified as being heterosexual only while 6% of participants identified as being homosexual only. Among men, 9% identified as being homosexual only; among women 2% identified as being homosexual only.

Table 3. Survey Counts from Klein Grid Sexual Identity Question

	Men	Women	Combined
Total	13,733	10,141	23,874
Heterosexual only	10,679 (78%)	7,599 (75%)	18,278 (77%)
Heterosexual mostly	1,068 (8%)	1,543 (15%)	2,611 (11%)
Heterosexual somewhat more	150 (1%)	266 (3%)	416 (2%)
Bisexual	173 (1%)	337 (3%)	510 (2%)
Homosexual somewhat more	95 (1%)	56 (1%)	151 (1%)
Homosexual mostly	387 (3%)	148 (1%)	535 (2%)
Homosexual only	1,181 (9%)	192 (2%)	1,373 (6%)

Preliminary Phenotype Associations

We examined the correlation between sexual identity and ~1000 phenotypes already characterized in the 23andMe database through other surveys. These analyses are preliminary; we have not checked for outliers or confounders beyond what is listed in the methods. We replicated previous findings showing a positive association between lesbians and alcoholism, and between lesbians and gay men and several psychiatric conditions.

Table 4. Preliminary Phenotype Associations

	Men				Women			
SPORTS	**sig**	**p value**	**beta**	**sample**	**sig**	**p value**	**beta**	**sample**
Dislocated Joint(s)	****	1.00E-10	-0.23	9735	**	2.30E-06	0.14	7302
Sprained Ankle	****	2.20E-08	-0.31	4507	.	7.10E-03	0.10	4028
Played Baseball	****	1.90E-12	-0.48	5862	.	3.00E-03	0.22	4933
Played Soccer	****	1.90E-17	-0.63	5880	.	5.50E-01	0.05	5009
Played Golf	****	1.20E-09	-0.43	5949	.	6.20E-01	-0.04	5044
Played Basketball	****	2.90E-15	-0.58	5835	.	1.80E-03	0.29	4940
Downhill Skiing	****	2.50E-08	-0.38	6029	.	8.90E-01	-0.01	4989
Running	***	4.80E-08	-0.33	5632	.	2.80E-02	-0.15	4909
Backpacking	***	1.90E-07	-0.43	5981	.	1.60E-01	0.13	5015
DISEASE								
Seasonal Allergies	****	2.00E-08	0.25	6020	.	9.30E-01	0.00	4469
AIDS	***	2.40E-07	4.18	797	.	NA	NA	559
Hepatitis	****	1.10E-30	1.01	9241	.	4.20E-04	0.26	7121
Hemorrhoids	****	4.60E-10	0.24	9179	.	8.30E-01	0.01	7099
High Blood Pressure	**	5.30E-07	0.21	9205	.	8.90E-04	0.11	7124
Reflux	****	1.80E-08	0.23	9115	.	3.10E-03	0.09	7056
Elevated Liver Test	****	8.30E-13	0.42	9156	.	1.40E-04	0.18	7009
ADDICTION								
Alcoholism	.	1.30E-04	0.26	8967	****	5.70E-09	0.37	6927
Ever Used Tobacco	.	2.70E-03	0.11	10641	***	2.80E-07	0.14	7845
MENTAL ILLNESS								
Panic Symptoms (current)	****	1.20E-08	0.37	6148	.	2.00E-04	0.15	5221
Anxiety	****	7.40E-15	0.40	9183	****	3.50E-10	0.20	7056
Bipolar	*	1.10E-05	0.45	10031	*	4.30E-06	0.34	7528
Depression	****	5.30E-21	0.39	9155	****	1.70E-11	0.18	7062
OCD	****	9.40E-14	0.85	9218	.	1.60E-02	0.18	7109
Panic Disorder	**	1.10E-06	0.44	4645	.	3.60E-03	0.15	4058
PTSD	*	1.30E-05	0.59	4709	***	5.90E-08	0.31	4085
PERSONALITY								
Cries Easily	****	3.40E-11	0.30	7130	.	8.00E-01	-0.01	5473
Adventurous	****	1.40E-08	-0.14	3265	.	1.90E-01	0.02	2992
Optimistic	*	4.60E-06	-0.02	6137	.	7.70E-03	-0.01	5128
Conscientiousness	.	8.50E-01	0.00	7323	****	2.80E-08	-0.01	5917
Neuroticism	****	5.30E-30	0.03	7329	.	8.60E-03	0.01	5923
Perceived Stress Scale	****	3.20E-10	0.02	6234	.	3.50E-04	0.01	5208
LIFESTYLE								
Atheist or Agnostic	****	1.20E-15	-0.53	3179	.	7.50E-02	0.08	2892
Ever Been Pregnant	.	-	-	-	****	3.40E-27	-0.34	6955
Number of Children	****	4.30E-41	-0.22	6606	****	4.80E-15	-0.10	5498
Single	****	2.30E-38	0.54	7823	.	7.60E-01	0.01	6278
Monogamous	****	2.60E-20	-0.22	3143	***	3.90E-08	0.13	3016
Ever Cheated on Partner	***	1.70E-07	0.31	3857	.	7.50E-04	0.14	3635
Spontaneous Traveler	***	1.30E-07	-0.10	4301	.	1.90E-01	0.02	3678
Served in Military	.	8.40E-03	-0.16	4749	***	6.10E-08	0.48	4155
COSMETIC								
BMI	.	3.50E-01	0.00	10662	****	1.60E-10	0.01	7919
Stretch Marks	****	4.10E-09	0.29	7056	.	1.30E-02	0.08	5579
Shaves Legs	.	NA	NA	11	****	4.10E-10	-0.36	2991
Tanning Bed Use	****	6.00E-24	0.16	3285	.	3.20E-01	-0.01	2995
Sunbathing Frequency	.	NA	NA	9	***	2.10E-07	-0.10	3596
Liposuction	****	2.30E-14	1.22	9177	.	1.10E-02	-0.16	7098
Cosmetic Surgery	****	7.70E-09	0.48	9167	.	1.90E-02	-0.09	7096
MEDICATIONS								
Currently on SSRI	**	2.50E-06	0.39	4607	.	1.40E-03	0.15	4012
Sudafed Causes Jitters	***	6.80E-08	0.30	4588	.	9.60E-03	0.09	4154
Taken Finasteride	****	1.20E-13	0.65	5276	.	7.90E-01	-0.12	4366
Taken Sleep Medications	****	5.70E-11	0.44	4713	.	6.30E-03	0.11	4161
Taken Psych Meds	****	5.10E-28	0.41	9706	****	4.90E-09	0.16	7176
MISC								
Ability to Smell	****	1.90E-08	0.14	4486	.	4.40E-01	-0.01	4066
Blood Donation Frequency	****	2.30E-28	-0.29	4652	.	8.60E-01	0.00	4054
Low HDL	**	9.60E-07	0.26	5718	.	1.00E-02	0.11	4781
Dry Eye Treatment	**	3.60E-07	0.12	3754	.	6.50E-03	0.04	3317
Short of Breath from Stairs	****	3.80E-10	0.35	6551	.	8.20E-05	0.14	5094
Taken Advanced Math	****	1.80E-09	-0.10	3188	.	2.20E-01	-0.01	2842
Self Rated Math Ability	****	6.50E-11	-0.21	3337	.	6.80E-01	-0.01	3034

Threshold corrected for multiple comparisons.
. = not significant
Bonferroni corrected p values: * = p < .05 ** = p < .01 *** = p < .001 **** = p < .0001

Results (continued)

GWAS

Our GWAS results did not identify any genetic loci reaching genome-wide significance at $p < 5 \times 10^{-8}$ among men or women. Among men, the peak (non-significant) hit was in chromosome 8q12.3 (chr8:63532921 in *NKAIN3*, $p = 7.1 \times 10^{-8}$).

We did not find evidence of an association between sexual identity and SNPs on the X chromosome in men, women or the samples combined at genome-wide significance.

We calculated that the study had 80% power to detect a genome-wide significant association with an odds ratio over 1.7 in women and over 1.3 in men for a minor allele frequency of 0.3.

Conclusions

To our knowledge, with over 23,000 participants this is the largest GWAS ever conducted on sexual orientation, as well as one of the largest analyses of non-genetic phenotypes associated with sexual orientation.

Prevalence:

- Rates of reported homosexuality are in line with the range of previous studies; men (9%) and women (2%).

Phenotype Associations:

- These analyses are preliminary; we have not checked for outliers or confounders beyond what is listed in the methods.
- We replicated previous studies showing that lesbians are more likely to be alcoholic than heterosexual women.
- We replicated previous studies showing that both lesbians and gay men are more likely to have psychiatric illnesses such as anxiety and depression than their heterosexual counterparts.
- We identified many phenotypic associations with sexual identity, including relationships with athletics, lifestyle, personality, medication usage, mental illness, addiction, and cosmetic procedures.
- Many of these associations may be novel.

GWAS:

- We did not find evidence of SNPs associated with sexual identity in men or women, nor did we replicate previous findings showing an association with regions on the X chromosome.
- Interestingly, our top (non-significant) hit was in 8q12.3 for men, which was also identified as a top (non-significant) hit in a previous linkage study among a small sample of gay men using microsatellites (2) as well as a current ASHG abstract (9).

This study focused on one aspect of sexual orientation, namely sexual identity. Further examination of the role of other aspects of sexuality is needed. Additionally, although this is a large sample, the number of individuals who identify as homosexual only is still small. Further data are needed to better understand the relationship between genetics and sexual orientation.

Acknowledgments

We thank the 23andMe customers who consented to participate in research for enabling this study. We also thank the employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References and Resources

1. Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AML. A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* 1993 (261): 321-327.
2. Mustanski BS, DuPress MG, Nievergelt CM, Bocklandt S, Schork NJ, Hamer DH. A genomewide scan of male sexual orientation. *Human Genetics* 2005 (116): 272-278.
3. Drabble L, Midanik LT, Trocki K. Reports of alcohol consumption and alcohol-related problems among homosexual, bisexual and heterosexual respondents: results from the 2000 National Alcohol Survey. *Journal of Studies on Alcohol* 2005: 111-120.
4. Gilman SE, Cochran SD, Mays VM, Hughes M, Ostrow D, Kessler RC. Risk of psychiatric disorders among individuals reporting same-sex sexual partners in the National Comorbidity Survey. *American Journal of Public Health* 2001 (91): 933-939.
5. Cochran SD, Mays VM, Sullivan JG. Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay and bisexual adults in the United States. *Journal of Consulting and Clinical Psychology* 2001(71): 53-61.
6. Klein F, Sepekoff B, Wolf TJ. Sexual orientation: a multi-variable dynamic process. *Journal of Homosexuality* 1985 (11): 35-49.
7. Gates GJ. Same-sex couples and the gay, lesbian, bisexual population: new estimates from the American Community Survey. *The Williams Institute* 2006.
8. National Survey of Health and Behavior 2010 www.nationalsexstudy.indiana.edu/
9. Sanders AR, Dawood K, Rieger G, Badner JA, Gershon ES, Krishnappa RS, Kolundzija AB,Guo S, Beecham GW, Martin ER, Bailey JM. Genome-wide linkage scan of male sexual orientation. *American Society of Human Genetics* 2012 conference presentation 1957T.