Genome-wide scan demonstrates significant linkage for male sexual orientation

A. R. Sanders[1,2]*, E. R. Martin[3], G. W. Beecham[3], S. Guo[3], K. Dawood[4], G. Rieger[5], J. A. Badner[2], E. S. Gershon[2], R. S. Krishnappa[6], A. B. Kolundzija[7], J. Duan[1,2], P. V. Gejman[1,2] and J. M. Bailey[8]

[1] *Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, Evanston, IL, USA*

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

[3] *Department of Human Genetics, University of Miami, Miami, FL, USA*

[4] *Department of Psychology, Pennsylvania State University, University Park, PA, USA*

[5] *Department of Psychology, University of Essex, Colchester, England, UK*

[6] *Department of Psychiatry, Icahn School of Medicine at Mount Sinai, Elmhurst, NY, USA*

[7] *Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA*

[8] *Department of Psychology, Northwestern University, Evanston, IL, USA*

Background. Findings from family and twin studies support a genetic contribution to the development of sexual orientation in men. However, previous studies have yielded conflicting evidence for linkage to chromosome Xq28.

Method. We conducted a genome-wide linkage scan on 409 independent pairs of homosexual brothers (908 analyzed individuals in 384 families), by far the largest study of its kind to date.

Results. We identified two regions of linkage: the pericentromeric region on chromosome 8 (maximum two-point LOD = 4.08, maximum multipoint LOD = 2.59), which overlaps with the second strongest region from a previous separate linkage scan of 155 brother pairs; and Xq28 (maximum two-point LOD = 2.99, maximum multipoint LOD = 2.76), which was also implicated in prior research.

Conclusions. Results, especially in the context of past studies, support the existence of genes on pericentromeric chromosome 8 and chromosome Xq28 influencing development of male sexual orientation.

Introduction

Sexual orientation, the sustained erotic attraction to members of one's own sex, the opposite sex, or both – homosexuality, heterosexuality, or bisexuality, respectively – is a primary component of human sexuality. Although Kinsey proposed that sexual orientation exists upon a continuum reflected by his scale (Kinsey *et al.* 1948), male sexual orientation tends to be bimodally distributed, with most men rating themselves as predominantly heterosexual (Kinsey scale 0–1) or homosexual (Kinsey scale 5–6) (Laumann *et al.* 1994*b*; Pillard & Bailey, 1998; Bailey *et al.* 2000). In contrast, women have a more continuous distribution across the non-heterosexual orientations, lower rates of homosexuality, higher rates of bisexuality, and less temporal stability (Laumann *et al.* 1994*b*; Bailey *et al.* 2000; Diamond, 2008*a*, *b*). Although our focus is male sexual orientation, we note that female sexual orientation also appears moderately heritable (Bailey *et al.* 2000; Alanko *et al.* 2010; Langstrom *et al.* 2010; Burri *et al.* 2011). Accurate prevalence estimates are currently difficult to obtain, but the following two conclusions are uncontroversial: homosexual orientation is much less common than heterosexual orientation among men, but it is not rare in respondents in Western industrialized nations. For example, in a large and careful survey of American males approximately 2% rated their sexual orientation as homosexual when defined as identity, 3–4% when defined psychologically (i.e. attraction and fantasy), and a few more did so when defined behaviorally (Laumann *et al.* 1994*a*). Partially independent familial aggregation of male and female homosexuality (Pillard & Weinrich, 1986; Bailey & Bell, 1993; Bailey & Benishay, 1993; Bailey *et al.* 1993, 1999; Pattatucci & Hamer, 1995; Schwartz *et al.* 2010)

* Address for correspondence: A. R. Sanders, M.D., Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, 1001 University Place, Evanston, IL 60201, USA.

(Email: alan.sanders.md@gmail.com)

further suggests the utility of separate analyses by sex. Genetic epidemiological studies (family, twin, and segregation analyses; see reviews by Mustanski *et al.* 2002; Schwartz *et al.* 2010) of males have shown (1) homosexuality to be more common in biological relatives of homosexual men compared to relatives of heterosexual men (or than expected from population surveys), with the sibling recurrence ratios falling between ~2 and ~4 (Pillard & Weinrich, 1986; Bailey & Pillard, 1991; Bailey & Bell, 1993; Hamer *et al.* 1993; Bailey *et al.* 1999; Schwartz *et al.* 2010); (2) genetic contributions suggested by higher concordances of sexual orientation for identical twins compared with same-sex fraternal twins (Bailey *et al.* 2000; Kendler *et al.* 2000; Kirk *et al.* 2000; Santtila *et al.* 2008; Alanko *et al.* 2010; Langstrom *et al.* 2010); (3) environmental contributions evidenced by monozygotic concordances far less than unity, with the best demonstrated specific environmental influence being the fraternal birth order effect (Blanchard & Bogaert, 1996); and (4) non-Mendelian segregation patterns (Bailey *et al.* 1999; Rice *et al.* 1999*a*; Schwartz *et al.* 2010). Furthermore, genetic studies show some support for excess maternal transmission (Hamer *et al.* 1993), whose presence would be consistent with X-linkage and which motivated focusing initial linkage approaches to the trait on chromosome X, although most studies have not found excess maternal transmission (Bailey *et al.* 1999; Rice *et al.* 1999*b*; Schwartz *et al.* 2010). The aggregate evidence suggests multifactorial causation of male sexual orientation, both genetic and environmental (i.e. a complex genetics scenario).

All previous genetic linkage studies of male sexual orientation were conducted on smaller sample sets. The two initial studies reported linkage to Xq28: first in 40 pairs of homosexual brothers (Hamer *et al.* 1993) and then further supported in 33 additional such pairs (Hu *et al.* 1995). However, Xq28 linkage was not supported in three subsequent studies of 54 US pairs of homosexual brothers (Sanders *et al.* 1998), 52 Canadian pairs of homosexual brothers (Rice *et al.* 1999*a*), or 73 additional pairs of homosexual brothers studied by the original research group (Mustanski *et al.* 2005). Extended family samples [55 Canadian families (Ramagopalan *et al.* 2010) and 146 US families with 155 independent pairs of homosexual brothers (Mustanski *et al.* 2005)] that underwent genome-wide linkage scans (GWLS) found neither genome-wide significant linkage nor any further support for Xq28 linkage. The strongest findings from the larger of the reported GWLS were at chromosomes 7qtel (~7q35–q36), pericentromeric 8 (~8p21–p11), and 10qtel (~10q26) (Mustanski *et al.* 2005). We chose to study male sexual orientation with an independent sample in a GWLS to seek more consistent findings by improvements in statistical power with a larger sample.

Method

Sample collection

We recruited families with two or more homosexual brothers in several primarily English-speaking counties, especially the United States, resulting in completed families from the United States (98.2%), Canada (1.6%), and the United Kingdom (0.2%). We excluded families known to have previously participated in other linkage studies of sexual orientation. Our primary source for identifying probands was through booths at community festivals, especially Gay Pride and related festivals. We also publicized our study through advertisements and articles in homophile media, liaisons with analogous groups (e.g. chapters of PFLAG; Parents, Families and Friends of Lesbians and Gays), and through an educationally oriented internet site (gaybros.com). We enrolled other interested members (available parents and any brothers, regardless of sexual orientation) in a family through the proband. The large majority (97.9%) of the studied families were of European ancestry, 1.6% were African American, and 0.5% were Asian; 95.1% were non-Hispanic and 4.9% were Hispanic. The average age of the studied brothers was 44.3 years [range 18.7–88.9, standard deviation (S.D.) = 10.7], and the sample was recruited from 2004 to 2008. Brothers completed a self-report questionnaire, and all subjects provided a DNA sample via blood (EDTA lavender top tubes, with venipuncture performed primarily through a vendor: LabCorp is the Laboratory Corporation of America, LabCorp, www.labcorp. com) or saliva (Oragene DNA Self-Collection kits, DNA Genotek, Canada). We obtained institutional review board approval from NorthShore University HealthSystem, and also convened and utilized a community advisory board. After complete description of the study to the subjects, all enrolled subjects provided written informed consent.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Measurement of sexual orientation

We assessed male sexual orientation as a primarily bimodal, psychological trait. Using data from standard

Kinsey self-report questionnaires obtained on all participating brothers, we classified the sexual orientation of subjects as homosexual (homosexual identity and past year Kinsey 5–6 for feelings, i.e. sexual fantasies) or heterosexual (heterosexual identity and past year Kinsey 0–1 for feelings), or otherwise unknown (we collected no questionnaires on parents, and they were all classified as unknown phenotype, but nevertheless made their studied families more genetically informative). For the linkage analysis families, which contained a minimum of two homosexual brothers, no subjects were ultimately classified as unknown, intermediate, or bisexual. Note that the primary linkage analyses employed 'affecteds-only analysis' (i.e. focusing on the less common phenotypic variant), which in this case means that only the homosexual brothers' phenotypes were considered certain (i.e. the heterosexual brothers were considered unknown by the analytical programs).

Genotyping

DNA samples were mailed (saliva) or couriered (blood via FedEx) to our laboratory. We isolated DNA following manufacturers' recommendations from saliva (Oragene kits) or whole blood (QIAamp® DNA Blood Mini kit, Qiagen, USA). We measured DNA concentrations with PicoGreen, adjusting to 50 ng/μl. We performed initial identity, sex, and relationship screens using AmpFlSTR® Profiler Plus® [Applied Biosystems, USA; 10 simple tandem repeat polymorphism (STRP) markers] prior to shipping plates of DNA to the Vanderbilt Microarray Shared Resource (VMSR, now known as Vanderbilt Technologies for Advanced Genomics; USA). After receipt of samples provided as 15 μl aliquots in 96-well plates and performance of quality control (QC) procedures (primarily DNA concentration and A_{260}/A_{280} ratios via Nanodrop Spectrophotometer, Thermo Scientific, USA), VMSR genotyped the sample on the Affymetrix 5.0 Genotyping Array [440 793 single nucleotide polymorphisms (SNPs)]. The initial linkage dataset consisted of 959 linkage and QC-related samples with BRLMM-P call rates $\geqslant$98%. We conducted further QC iteratively at the sample and SNP level, incorporating inter- and intra-plate QC duplicates, following standard QC procedures for genome-wide association studies (GWAS) (Laurie *et al.* 2010; Turner *et al.* 2011), but further adapted for genome-wide linkage analyses such as the current study.

Sample QC

All 959 samples passed Affymetrix Power Tools QC (APTqc) and had call rates $\geqslant$95%. A single HapMap sample (NA10857) was duplicated 13 times across the plates, and 12 internal duplicate pairs were placed within each plate (one pair per plate). Mean concordance between duplicate pairs was >99% for both interplate duplicates and intra plate duplicates. Checks for Mendelian errors, sex concordance, and relationship confirmation identified eight pairs of monozygotic (MZ) twins, four other unintentionally duplicated samples, and one additional error (possible cryptic adoptee within a brother pair), all of which were removed. Two of the eight MZ twin families had a third non-MZ homosexual brother, thus we removed the one of the MZ twins with the lower genotyping rate, but retained these families in the analysis. The other six MZ twin families (comprising seven individuals besides the proband) had no additional siblings and so were removed from the analysis. For the four non-MZ unintentionally duplicated samples, one family still had a genotyped homosexual sibling pair and thus the remainder of the family was retained. The other three families no longer had a sibling pair and thus were entirely dropped (i.e. the remaining sibling) from the linkage analyses. The remaining samples passed the heterozygosity filter (0.273–0.309 retained). We removed one member (the one with the lower genotyping rate) of each of 12 pairs of intra-plate QC duplicates, and for two of these QC duplicates removed the other pair member (because the family did not have a sibling pair for linkage analyses). Finally, after then removing the 13 HapMap inter-plate duplicates, 908 genotyped family members in 384 independent multiplex families remained for linkage analysis.

Family structures

We confirmed full sibling relationships for the 351 families containing $\geqslant$1 full sibling pair, and identified half-sibling relationships for the 35 families containing $\geqslant$1 half-sibling pair. The latter group included 21 reported half-sibling families (already described in the questionnaires) and 14 previously undeclared half-sibling families (for which we corrected family structures). For all half-sibling pair families, we specified as paternal (11 families, which did not contribute linkage information for X chromosome analyses) if explicitly noted as sharing the father in the questionnaire, or otherwise as maternal (shared mother, 24 families). We also annotated one family for which the relative pair of homosexual brothers was explicitly described in the questionnaire (and consistent with the genotypes) as double first cousins (here, with the mothers being full sisters to each other, and the fathers being full brothers to each other). The 908 genotyped family members included 793 homosexual brothers (708/793 = 89% were Kinsey 6, the rest Kinsey 5), 33 heterosexual brothers (31/33 = 94% were Kinsey 0, the rest Kinsey 1), 49 mothers, and 33 fathers. Most families

Table 1. *SNP genotyped families and individuals analyzed for linkage*

Number of homosexual brothers/family	Family count	Independent homosexual brother pairs	Number of homosexual brothers	Number of heterosexual brothers	Number of parents
2	361	361	722	32	75
3	21	42	63	1	5
4	2	6	8	0	2
Totals	384	409	793	33	82

All counts reflect the 908 (793 + 33 + 82) SNP genotyped individuals analyzed for linkage. There were 11 paternal half-sibling pairs and 24 maternal half-sibling pairs in these families.

(361/384) had two homosexual brothers, but 21 families had three and two families had four, yielding 409 independent homosexual brother pairs ($n - 1$ method, see Table 1).

SNP QC

While not analyzed in the current study (except as a further indication of QC), we note that 34 duplicated samples genotyped with two related genotypic arrays at different times and sites [Affymetrix 5.0 at VMSR *v.* Affymetrix 6.0 at the Broad Institute (Shi *et al.* 2009)] revealed ⩾98.8% SNP concordance. After this concordance check, we removed 22 928 SNPs that were discrepant for any of the 34 platform-duplicated (Affymetrix 5.0 and 6.0) samples. We also removed 497 Affymetrix 5.0 SNPs not mapped to chromosomes 1–23 (i.e. those mapped to chromosomes 0, 24, or 25 per Affymetrix information). We then evaluated SNPs for genotyping call rates, deviations from Hardy–Weinberg equilibrium (HWE), and minor allele frequency (MAF) using PLINK (Purcell *et al.* 2007). We removed 78 112 SNPs (29 033 had call rate <95%, an additional six SNPs showed significant deviations from HWE at $p < 10^{-6}$, and an additional 49 073 had MAF < 0.01), leaving 362 681 SNPs for further analysis.

Linkage analysis

We conducted two-point and multipoint analysis for linkage using Merlin non-parametric linkage (Abecasis *et al.* 2002) using the Kong and Cox linear model (Kong & Cox, 1997) and the S-pairs option. We chose S-pairs to assess independent sibling pairs rather than all possible sibling pairs (S-all), although the results for the latter were almost identical to the former as expected in a sample largely consisting of families genotyped for a single sibling pair. We calculated allele frequency estimates within the dataset (using founding parents when available, otherwise one brother per family), and used the deCODE genetic

map (Kong *et al.* 2010). For multipoint analysis, we pruned SNPs for linkage disequilibrium (LD, $r^2 > 0.16$) using PLINK (Purcell *et al.* 2007) and MAF < 0.1, which left 44 766 autosomal SNPs and 621 chromosome X SNPs for multipoint analysis. As a further check of results, we examined the top results (two-point LOD > 2.2) and found no enrichment (compared to the remaining analyzed SNPs) of QC classes such as lower call rate, lower MAF, or lower *p* values for HWE deviations (online Supplementary Table S1). We estimated power for our GWLS using the formulas for non-parametric linkage analysis (Risch, 1990) in 409 independent homosexual sibling pairs (online Supplementary Fig. S1). We found good power (>0.8) to detect suggestive (LOD > 2.2) linkage (Lander & Kruglyak, 1995) down to locus-specific sibling recurrence risk ratios of $\lambda_{sibs} \sim 1.39$ and to detect significant (LOD > 3.6) linkage (Lander & Kruglyak, 1995) down to $\lambda_{sibs} \sim 1.52$. Given that the overall sibling recurrence risk ratio is estimated to be ~2 to ~4, this suggests that we were able to detect moderate to major loci contributing to the trait.

Linkage models accounting for fraternal birth order effect

We incorporated the known environmental contribution of more older biological brothers from the same mother increasing the chance later born men would be homosexual, i.e. the fraternal birth order effect (Blanchard & Bogaert, 1996), into our linkage analyses for the two strongest peaks. We did this by means of parametric models (dominant, recessive, and X-linked) using variable penetrance estimates to simulate increasing phenocopy rates (increasing importance of the number of older brothers relative to genetic contributions) as the number of older brothers increased. For the 826 genotyped brothers, we had the number of older brothers for 819 of them, with the number of older brothers ranging from 0 to 7 (online Supplementary Table S2). For the baseline penetrance (for no older brothers), we

used 0.005 and increased this by 33% for each older brother a subject had, since it has been estimated that in demographically similar populations to our sample that each additional older brother increases the odds of male homosexuality by ~33% (Cantor *et al.* 2002) (we assigned the baseline penetrance for the seven brothers with an unknown number of older brothers). Thus, the penetrances were: 0.005 for 0 older brother, and 0.0067, 0.0088, 0.012, 0.016, 0.021, 0.028, and 0.037 for 1–7 older brothers, respectively. Under the dominant model all carrier penetrances were set at 0.75 for all liability classes; under the recessive model heterozygote carriers matched the baseline (non-carrier) penetrances, with homozygote carriers having penetrance of 0.75 for all liability classes. All parents and heterosexual brothers were classified as phenotype unknown, and the population frequency of the homosexuality-associated allele was set to 0.02 in these parametric analyses (as in Hamer *et al.* 1993). We compared the results from these variable penetrance models to those where all subjects' penetrances were set to 0.005, i.e. as if there were no fraternal birth order effect. We used Merlin (Abecasis *et al.* 2002) to conduct multipoint linkage analyses with these parameters, and compared to the previous non-parametric linkage results on chromosomes 8 and X.

Supplementary chromosome X analysis dataset

For chromosome X, in addition to the dataset described above genotyped with SNPs, we had an independent set of 50 families (146 genotyped individuals, online Supplementary Table S3) that had been previously genotyped for 30 STRPs spanning chromosome X. This additional dataset included 56 independent homosexual brother pairs (same phenotyping as SNP-genotyped subjects, i.e. Kinsey 5–6), one of which was a maternal half-sibling pair. To integrate the STRP and SNP maps, we found the physical positions of the 30 STRPs on the hg19 build, and then translated these positions to the deCODE genetic map using the Affymetrix 5.0 annotation file. We ran multipoint linkage analysis with Merlin (minx) using the S-pairs option for the SNP-genotyped families and the STRP-genotyped families separately using the grid option of 1 cM (--grid 0.01). We then summed LOD scores at each grid position across all SNP and STRP-genotyped families to achieve total LOD scores.

Results

We detected suggestive two-point linkage (LOD ≥ 2.2) (Lander & Kruglyak, 1995) for 352 SNPs (online Supplementary Table S1) analyzed in the 409 SNP-genotyped homosexual brother pairs (Table 1), with five SNPs exceeding the threshold for genome-wide significance (LOD ≥ 3.6) (Lander & Kruglyak, 1995). One of the genome-wide significant SNPs was intergenic (rs13212974) with the nearest RefSeq gene being *FRK* (*fyn-related kinase*). The other four significant SNPs were located within introns of RefSeq genes: rs6990254 at *CLVS1* (*clavesin 1*), rs2498600 at *PTPRD* (*protein tyrosine phosphatase, receptor type, D*), rs2221108 at *GRM5* (*glutamate receptor, metabotropic 5*), and rs7964186 at *DNAH10* (*dynein, axonemal, heavy chain 10*). However, we note that linkage signals are imprecise (especially for traits manifesting complex genetics) and thus larger regions containing additional genes are implicated. The two-point LOD maxima were 4.08 (rs6990254) at 8q12 for autosomes, and 2.99 (rs5925403) at Xq28 for chromosome X. Our genome-wide multipoint non-parametric linkage results are presented in Fig. 1, which shows the two strongest peaks (both with multipoint LOD > 2.5) to be at the pericentromeric region of chromosome 8 (~60–90 cM, based on multipoint drop-1 LOD support interval) and on Xqtel (~160–180 cM). These chromosomes are plotted in Fig. 2a (chromosome 8) and Fig. 2b (chromosome X), along with their two-point LOD results. We note that our strongest autosomal finding (based on two-point findings and supported by multipoint results) at pericentromeric chromosome 8 (8p11-q21, Figs 1 and 2a) overlapped with the second strongest linkage peak in the next largest reported GWLS on 155 homosexual brother pairs which found a peak multipoint LOD = 1.96 at 8p12 (Mustanski *et al.* 2005). We found little support (Fig. 1) for the 7q35–q36 (Mustanski *et al.* 2005) or the 14q32 (Ramagopalan *et al.* 2010) regions highlighted in previous smaller GWLS. However, our second strongest linkage region (Xq27–q28, Figs 1 and 2b) overlaid the previously reported Xq28 linkage (Hamer *et al.* 1993; Hu *et al.* 1995). After taking into account the fraternal birth-order effect by incorporating the number of older brothers into the linkage analysis using variable penetrance parametric models, the multipoint linkage peaks on chromosomes 8 and X remained largely unchanged, both compared to the non-parametric linkage results and compared to the parametric linkage results without such birth-order adjustment (online Supplementary Fig. S2). For the X-chromosome multipoint linkage analyses, we merged an independent set (i.e. different families) of 56 homosexual brother pairs (online Supplementary Table S3) previously genotyped for STRP markers (Sanders *et al.* 1998) into the same genetic map (deCODE) to assess the effect on the Xq28 linkage peak. There was minimal influence from the smaller dataset on the multipoint curve except a small (~0.5
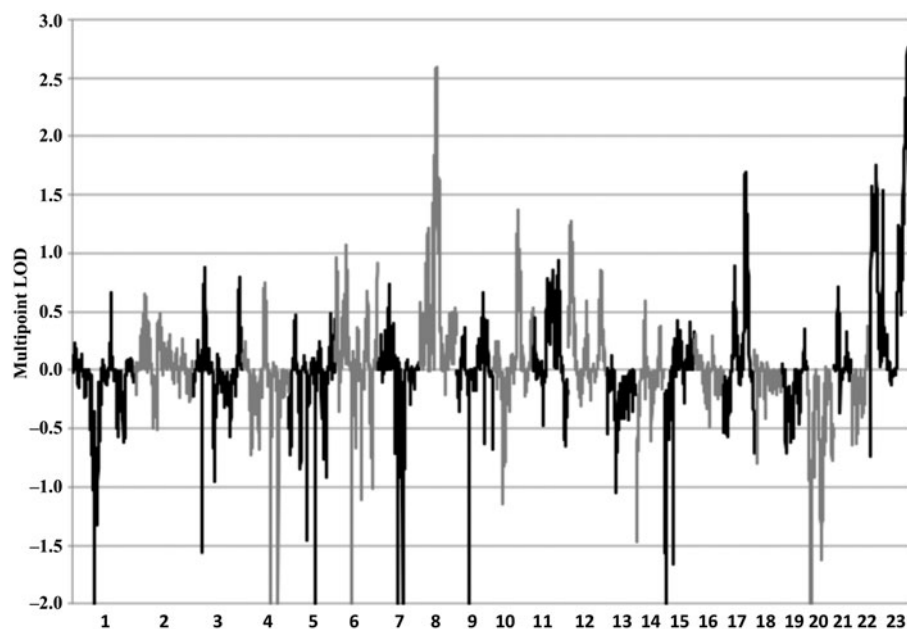
Fig. 1. Multipoint LOD scores are plotted *v.* the chromosomal positions for the non-parametric linkage analysis. Adjacent chromosomes are separated by alternating black and gray lines. Maximum multipoint LOD scores are ~74 cM on chromosome 8q12 and ~175 cM on chromosome Xq28.

LOD) increase at the telomeric extreme of Xq28, slightly broadening the peak (online Supplementary Fig. S3).

Discussion

We have detected genome-wide significant linkage to pericentromeric chromosome 8 with multipoint support, and replicated linkage to chromosome Xq28. In context with the previous linkage scans, it seems likely that genes contributing to variation in male sexual orientation reside in these regions. As usual with linkage peaks for complex traits, there are a number of genes of potential relevance under each broad peak, such as transcription factors, microRNAs, and various brain-expressed genes including some with neurodevelopment, neuroendocrine, and/or neurotransmission roles (e.g. *CHRNB3, CHRNA6, SNTG1, NPBWR1, OPRK1, RGS20, PENK, CRH, TRPA1, GDAP1, SLITRK2, CNGA2, GABRE, GABRA3, GABRQ, PLXNB3, L1CAM, GDI1, PLXNA3*), and at Xq27–28 the testes-specific members of the *SPANX* (*sperm protein associated with the nucleus on chromosome X*) family (these ones being ampliconic genes independently acquired since human-mouse divergence; Mueller *et al.* 2013). We comment further on selected examples, acknowledging the speculative nature of genotype-behavior correlations across species (e.g. human and mice) and even between more closely related species (e.g. within rodents) (Baud *et al.* 2013). Arginine

vasopressin is a hormone that mediates social and affiliative behaviors (see review by Ebstein *et al.* 2012), and one of its receptors (*AVPR2, arginine vasopressin receptor 2*) is located on Xq28, although *AVPR2* is primarily expressed in the kidney. However, *neuropeptides B/W receptor 1* (*NPBWR1*) on 8q11.23 is expressed in limbic regions including the hippocampus, has been shown to modulate social interactions in mice (Nagata-Kuroiwa *et al.* 2011), and a common functional variant (Tyr[135]Phe, rs33977775, CEU MAF~10%) has been reported as associated with altered emotional responses to facial expressions in humans, especially valence (with heterozygotes perceiving expressions in more pleasant terms) (Watanabe *et al.* 2012). Another brain-expressed gene, *CNGA2* (*cyclic nucleotide gated channel alpha 2*) on Xq28, has been shown in mice to be critical for regulation of odor-evoked socio-sexual behaviors, including major histocompatibility complex (MHC)-related odors (Mandiyan *et al.* 2005; Spehr *et al.* 2006); and although more uncertain, MHC-related odors may have relevance in social communication and mate selection in humans (Wedekind *et al.* 1995; Havlicek & Roberts, 2009; Milinski *et al.* 2013). However, linkage can only indicate a region and not particular genes, a task better suited to other approaches such as resequencing approaches, GWAS, and meta-analyses thereof.

Our findings may also begin to provide a genetic basis for the considerable evolutionary paradox that
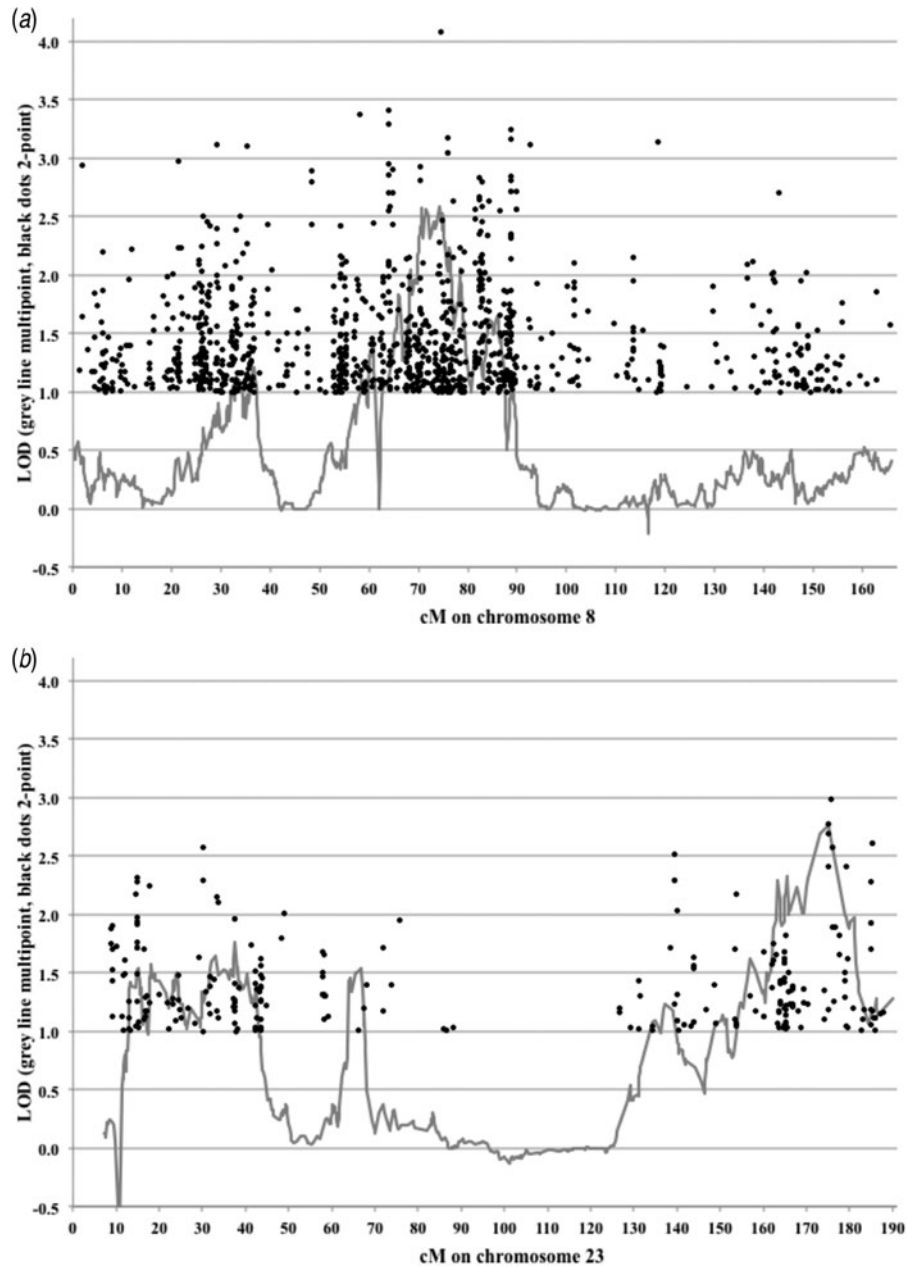
Fig. 2. Individual chromosomal plots for the strongest multipoint linkage peaks on chromosomes 8 and X. LOD scores are plotted *v.* the chromosomal positions for the non-parametric linkage analysis, the gray line indicating the multipoint values and the black dots indicating the two-point linkage results (with any two-point LOD < 1.0 undisplayed to enable visualization of the multipoint line). (*a*) Chromosome 8 has its multipoint linkage peak in the pericentromeric region containing several two-point LOD > 3.0, the highest being LOD = 4.08 at 8q12. (*b*) Chromosome X has its multipoint linkage peak in the pericentromeric region containing several two-point LOD > 2.5, the highest being LOD = 2.99 at Xq28.

homosexual men are less motivated than heterosexual men to have procreative sex and yet exist as a stable non-trivial minority of the population (Wilson, 1975; Bell & Weinberg, 1978). Linkage to Xq28 is especially relevant to the X-linked sexually antagonistic selection hypothesis that women with genetic variant/s predisposing to homosexuality in men have a reproductive advantage compared with other women, i.e. that

fertility costs of variants that increase the likelihood of a man's homosexuality are balanced by increased fecundity when expressed in a woman (Miller, 2000; Gavrilets & Rice, 2006; Camperio Ciani *et al.* 2008; Rahman *et al.* 2008). Modeling of the sexually antagonistic hypothesis (Gavrilets & Rice, 2006) predicts such loci to be strongly over-represented on chromosome X with 1–2 loci there, and allows for strong asymmetries

in effect size (i.e. large decrease in reproductive fitness in males can be balanced by smaller increase in females). The putative effects of sexually antagonistic genes are necessarily even more speculative and uncertain than the existence of such genes. Speculation has included the possibility that such genes increase sexual attraction to men in both homosexual men and their heterosexual female relatives, leading to increased reproduction in the females (Camperio Ciani *et al.* 2008). An empirical study focusing on female maternal relatives of homosexual men found several personality and self-reported health differences (Camperio Ciani *et al.* 2012); it is unclear if any of the differences were related to increased fecundity. A different possibility, overdominance (male heterozygote advantage) (MacIntyre & Estep, 1993), could conceivably help explain the paradox of autosomal genes for male homosexuality, including our findings of linkage to the pericentromeric region of chromosome 8.

Limitations of this research include statistical power considerations (despite the sample size) and challenges inherent to linkage mapping of traits with complex genetics. However, the larger size of the currently studied sample should provide a more stable estimate of the degree of linkage, such as at chromosome Xq28, compared to earlier smaller samples, in part by reducing the influence of stochastic variation. We note that while not studied here, female sexual orientation merits its own scientific study. Future investigations of sexual orientation using GWAS may shed further light on the development of human sexual orientation. Regarding any scenario that research in this area will result in a prenatal genetic test for homosexuality, the small magnitude of effects suggested herein are inconsistent with a test that those motivated to influence their children's sexual orientation would find useful. Furthermore, we agree with Murphy (1997, 2012) that fear of this unlikely scenario should not prevent further research. Indeed, factual information about sexual orientation can help prevent distorted and hostile views (reviewed in Murphy, 1997). Finding genetic linkage contributes to the overall societal debate by extending support for genetic influence on variation in male sexual orientation from the epidemiological into the molecular realm. While our study results provide further evidence for early (prenatal) biological influences on variation in male sexual orientation, we also emphasize that genetic contributions are far from determinant but instead represent a part of the trait's multifactorial causation, both genetic and environmental.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291714002451.

Declaration of Interest

None.

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002). Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* **30**, 97–101.

Alanko K, Santtila P, Harlaar N, Witting K, Varjonen M, Jern P, Johansson A, von der Pahlen B, Sandnabba NK (2010). Common genetic effects of gender atypical behavior in childhood and sexual orientation in adulthood: a study of Finnish twins. *Archives of Sexual Behavior* **39**, 81–92.

Bailey JM, Bell AP (1993). Familiality of female and male homosexuality. *Behavior Genetics* **23**, 313–322.

Bailey JM, Benishay DS (1993). Familial aggregation of female sexual orientation. *American Journal of Psychiatry* **150**, 272–277.

Bailey JM, Dunne MP, Martin NG (2000). Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. *Journal of Personality and Social Psychology* **78**, 524–536.

Bailey JM, Pillard RC (1991). A genetic study of male sexual orientation. *Archives of General Psychiatry* **48**, 1089–1096.

Bailey JM, Pillard RC, Dawood K, Miller MB, Trivedi S, Farrer LA, Murphy RL (1999). A family history study of male sexual orientation using three independent samples. *Behavior Genetics* **29**, 79–86.

Bailey JM, Pillard RC, Neale MC, Agyei Y (1993). Heritable factors influence sexual orientation in women. *Archives of General Psychiatry* **50**, 217–223.

Baud A, Hermsen R, Guryev V, Stridh P, Graham D, McBride MW, Foroud T, Calderari S, Diez M, Ockinger J, Beyeen AD, Gillett A, Abdelmagid N, Guerreiro-Cacais AO, Jagodic M, Tuncel J, Norin U, Beattie E, Huynh N, Miller WH, Koller DL, Alam I, Falak S, Osborne-Pellegrin M, Martinez-Membrives E, Canete T, Blazquez G, Vicens-Costa E, Mont-Cardona C, Diaz-Moran S, Tobena A, Hummel O, Zelenika D, Saar

K, Patone G, Bauerfeind A, Bihoreau MT, Heinig M, Lee YA, Rintisch C, Schulz H, Wheeler DA, Worley KC, Muzny DM, Gibbs RA, Lathrop M, Lansu N, Toonen P, Ruzius FP, de Bruijn E, Hauser H, Adams DJ, Keane T, Atanur SS, Aitman TJ, Flicek P, Malinauskas T, Jones EY, Ekman D, Lopez-Aumatell R, Dominiczak AF, Johannesson M, Holmdahl R, Olsson T, Gauguier D, Hubner N, Fernandez-Teruel A, Cuppen E, Mott R, Flint J (2013). Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nature Genetics* **45**, 767–775.

Bell AP, Weinberg MS (1978). *Homosexualities: a Study of Diversity Among Men and Women*. Simon and Schuster: New York.

Blanchard R, Bogaert AF (1996). Biodemographic comparisons of homosexual and heterosexual men in the Kinsey interview data. *Archives of Sexual Behavior* **25**, 551–579.

Burri A, Cherkas L, Spector T, Rahman Q (2011). Genetic and environmental influences on female sexual orientation, childhood gender typicality and adult gender identity. *PLoS ONE* **6**, e21982.

Camperio Ciani A, Cermelli P, Zanzotto G (2008). Sexually antagonistic selection in human male homosexuality. *PLoS ONE* **3**, e2282.

Camperio Ciani AS, Fontanesi L, Iemmola F, Giannella E, Ferron C, Lombardi L (2012). Factors associated with higher fecundity in female maternal relatives of homosexual men. *Journal of Sexual Medicine* **9**, 2878–2887.

Cantor JM, Blanchard R, Paterson AD, Bogaert AF (2002). How many gay men owe their sexual orientation to fraternal birth order? *Archives of Sexual Behavior* **31**, 63–71.

Diamond LM (2008a). Female bisexuality from adolescence to adulthood: results from a 10-year longitudinal study. *Developmental Psychology* **44**, 5–14.

Diamond LM (2008b). *Sexual Fluidity: Understanding Women's Love and Desire*. Harvard University Press: Cambridge, MA.

Ebstein RP, Knafo A, Mankuta D, Chew SH, Lai PS (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. *Hormones and Behavior* **61**, 359–379.

Gavrilets S, Rice WR (2006). Genetic models of homosexuality: generating testable predictions. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **273**, 3031–3038.

Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AM (1993). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* **261**, 321–327.

Havlicek J, Roberts SC (2009). MHC-correlated mate choice in humans: a review. *Psychoneuroendocrinology* **34**, 497–512.

Hu S, Pattatucci AM, Patterson C, Li L, Fulker DW, Cherny SS, Kruglyak L, Hamer DH (1995). Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nature Genetics* **11**, 248–256.

Kendler KS, Thornton LM, Gilman SE, Kessler RC (2000). Sexual orientation in a U.S. national sample of twin and nontwin sibling pairs. *American Journal of Psychiatry* **157**, 1843–1846.

Kinsey AC, Pomeroy WB, Martin CE (1948). *Sexual Behavior in the Human Male*. W. B. Saunders Company: Philadelphia.

Kirk KM, Bailey JM, Dunne MP, Martin NG (2000). Measurement models for sexual orientation in a community twin sample. *Behavior Genetics* **30**, 345–356.

Kong A, Cox NJ (1997). Allele-sharing models: LOD scores and accurate linkage tests. *American Journal of Human Genetics* **61**, 1179–1188.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Gylfason A, Kristinsson KT, Gudjonsson SA, Frigge ML, Helgason A, Thorsteinsdottir U, Stefansson K (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103.

Lander E, Kruglyak L (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247.

Langstrom N, Rahman Q, Carlstrom E, Lichtenstein P (2010). Genetic and environmental effects on same-sex sexual behavior: a population study of twins in Sweden. *Archives of Sexual Behavior* **39**, 75–80.

Laumann EO, Gagnon JH, Michael RT, Michaels S (eds) (1994a). Homosexuality. In *The Social Organization of Sexuality: Sexual Practices in the United States*, pp. 283–321. The University of Chicago Press: Chicago.

Laumann EO, Gagnon JH, Michael RT, Michaels S (1994b). *The Social Organization of Sexuality: Sexual Practices in the United States*. University of Chicago Press: Chicago.

Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ, Gabriel SB, Harris EL, Hu FB, Jacobs KB, Kraft P, Landi MT, Lumley T, Manolio TA, McHugh C, Painter I, Paschall J, Rice JP, Rice KM, Zheng X, Weir BS (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* **34**, 591–602.

MacIntyre F, Estep KW (1993). Sperm competition and the persistence of genes for male homosexuality. *Biosystems* **31**, 223–233.

Mandiyan VS, Coats JK, Shah NM (2005). Deficits in sexual and aggressive behaviors in Cnga2 mutant mice. *Nature Neuroscience* **8**, 1660–1662.

Milinski M, Croy I, Hummel T, Boehm T (2013). Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **280**, 20122889.

Miller EM (2000). Homosexuality, birth order, and evolution: toward an equilibrium reproductive economics of homosexuality. *Archives of Sexual Behavior* **29**, 1–34.

Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics* **45**, 1083–1087.

Murphy TF (1997). *Gay Science: the Ethics of Sexual Orientation Research*. Columbia University Press: New York, NY.

Murphy TF (2012). *Ethics, Sexual Orientation, and Choices about Children*. The MIT Press: Cambridge, MA.

Mustanski BS, Chivers ML, Bailey JM (2002). A critical review of recent biological research on human sexual orientation. *Annual Review of Sex Research* **13**, 89–140.

Mustanski BS, Dupree MG, Nievergelt CM, Bocklandt S, Schork NJ, Hamer DH (2005). A genomewide scan of male sexual orientation. *Human Genetics* **116**, 272–278.

Nagata-Kuroiwa R, Furutani N, Hara J, Hondo M, Ishii M, Abe T, Mieda M, Tsujino N, Motoike T, Yanagawa Y, Kuwaki T, Yamamoto M, Yanagisawa M, Sakurai T (2011). Critical role of neuropeptides B/W receptor 1 signaling in social behavior and fear memory. *PLoS ONE* **6**, e16972.

Pattatucci AML, Hamer DH (1995). Development and familiality of sexual orientation in females. *Behavior Genetics* **25**, 407–420.

Pillard RC, Bailey JM (1998). Human sexual orientation has a heritable component. *Human Biology* **70**, 347–365.

Pillard RC, Weinrich JD (1986). Evidence of familial nature of male homosexuality. *Archives of General Psychiatry* **43**, 808–812.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575.

Rahman Q, Collins A, Morrison M, Orrells JC, Cadinouche K, Greenfield S, Begum S (2008). Maternal inheritance and familial fecundity factors in male homosexuality. *Archives of Sexual Behavior* **37**, 962–969.

Ramagopalan SV, Dyment DA, Handunnetthi L, Rice GP, Ebers GC (2010). A genome-wide scan of male sexual orientation. *Journal of Human Genetics* **55**, 131–132.

Rice G, Anderson C, Risch N, Ebers G (1999*a*). Male homosexuality: absence of linkage to microsatellite markers at Xq28. *Science* **284**, 665–667.

Rice G, Risch N, Ebers G (1999*b*). Genetics and male sexual orientation. *Science* **285**, 803.

Risch N (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *American Journal of Human Genetics* **46**, 229–241.

Sanders AR, Cao Q, Zhang J, Badner JA, Goldin LR, Guroff JJ, Gershon ES, Gejman PV (1998). *Genetic Linkage Study of Male Homosexual Orientation*. American Psychiatric Association: Toronto, Ontario, Canada.

Santtila P, Sandnabba NK, Harlaar N, Varjonen M, Alanko K, von der Pahlen B (2008). Potential for homosexual response is prevalent and genetic. *Biological Psychology* **77**, 102–105.

Schwartz G, Kim RM, Kolundzija AB, Rieger G, Sanders AR (2010). Biodemographic and physical correlates of sexual orientation in men. *Archives of Sexual Behavior* **39**, 93–109.

Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757.

Spehr M, Kelliher KR, Li XH, Boehm T, Leinders-Zufall T, Zufall F (2006). Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *Journal of Neuroscience*: **26**, 1961–1970.

Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, de Andrade M, Doheny KF, Haines JL, Hayes G, Jarvik G, Jiang L, Kullo IJ, Li R, Ling H, Manolio TA, Matsumoto M, McCarty CA, McDavid AN, Mirel DB, Paschall JE, Pugh EW, Rasmussen LV, Wilke RA, Zuvich RL, Ritchie MD (2011). Quality control procedures for genome-wide association studies. *Current Protocols in Human Genetics*, chapter 1, unit 19, 1–18.

Watanabe N, Wada M, Irukayama-Tomobe Y, Ogata Y, Tsujino N, Suzuki M, Furutani N, Sakurai T, Yamamoto M (2012). A single nucleotide polymorphism of the neuropeptide B/W receptor-1 gene influences the evaluation of facial expressions. *PLoS ONE* **7**, e35390.

Wedekind C, Seebeck T, Bettens F, Paepke AJ (1995). MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **260**, 245–249.

Wilson EO (1975). *Sociobiology: the New Synthesis*. Harvard University Press: Cambridge.