

RESEARCH

Open Access



# Statistical methods for testing X chromosome variant associations: application to sex-specific characteristics of bipolar disorder

William A. Jons<sup>1</sup>, Colin L. Colby<sup>1</sup>, Susan L. McElroy<sup>2</sup>, Mark A. Frye<sup>3</sup>, Joanna M. Biernacka<sup>1,3</sup> and Stacey J. Winham<sup>1\*</sup> 

## Abstract

**Background:** Bipolar disorder (BD) affects both sexes, but important sex differences exist with respect to its symptoms and comorbidities. For example, rapid cycling (RC) is more prevalent in females, and alcohol use disorder (AUD) is more prevalent in males. We hypothesize that X chromosome variants may be associated with sex-specific characteristics of BD. Few studies have explored the role of the X chromosome in BD, which is complicated by X chromosome inactivation (XCI). This process achieves “dosage compensation” for many X chromosome genes by silencing one of the two copies in females, and most statistical methods either ignore that XCI occurs or falsely assume that one copy is inactivated at all loci. We introduce new statistical methods that do not make these assumptions.

**Methods:** We investigated this hypothesis in 1001 BD patients from the Genetic Association Information Network (GAIN) and 957 BD patients from the Mayo Clinic Bipolar Disorder Biobank. We examined the association of over 14,000 X chromosome single nucleotide polymorphisms (SNPs) with sex-associated BD traits using two statistical approaches that account for whether a SNP may be undergoing or escaping XCI. In the “XCI-informed approach,” we fit a sex-adjusted logistic regression model assuming additive genetic effects where we coded the SNP either assuming one copy is expressed or two copies are expressed based on prior knowledge about which regions are inactivated. In the “XCI-robust approach,” we fit a logistic regression model with sex, SNP, and SNP-sex interaction effects that is flexible to whether the region is inactivated or escaping XCI.

**Results:** Using the “XCI-informed approach,” which considers only the main effect of SNP and does not allow the SNP effect to differ by sex, no significant associations were identified for any of the phenotypes. Using the “XCI-robust approach,” intergenic SNP rs5932307 was associated with BD ( $P = 8.3 \times 10^{-8}$ ), with a stronger effect in females (odds ratio in males ( $OR_M$ ) = 1.13, odds ratio in females for a change of two allele copies ( $OR_{W2}$ ) = 3.86).

**Conclusion:** X chromosome association studies should employ methods which account for its unique biology. Future work is needed to validate the identified associations with BD, to formally assess the performance of both approaches under different true genetic architectures, and to apply these approaches to study sex differences in other conditions.

**Keywords:** Bipolar disorder, X chromosome, Genetic association, Rapid cycling, Binge eating, Alcohol use disorder, Suicidality, X chromosome inactivation, Sex differences, X chromosome statistical analysis

\* Correspondence: [Winham.Stacey@mayo.edu](mailto:Winham.Stacey@mayo.edu)

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

## Background

Although multiple genome-wide association studies have examined the genetic contributions to the risk of bipolar disorder (BD) [1, 2], few studies have examined the genetics of specific symptoms or comorbidities of BD. We previously identified several symptoms and comorbidities of BD that differ in prevalence by sex [3]. We found that rapid cycling (RC) and a lifetime history of a suicide attempt were more common for women than men and that men more frequently had a substance use disorder. Women are also more likely to have a comorbid eating disorder, particularly binge eating behavior (BE) [4]. The reason for these sex-specific differences in BD characteristics is unclear. However, many biological sex differences are thought to arise from either hormonal differences or from genetic differences (e.g., sex chromosomes). Brain development and function as well as psychiatric traits are influenced by sex hormone levels [5] and genetic factors [2]. For example, expression of the gene *BDNF* is influenced by estradiol, and a SNP within *BDNF* Val66Met has been shown to be associated with BD and other psychiatric traits [6]. The X chromosome contains many sex and reproductive genes influencing hormone levels, such as the androgen receptor (*AR*) [7]. Patients with X chromosome aneuploidies experience higher rates of various psychiatric disorders, including mood disorders [8]. Furthermore, X chromosome dosage and dosage compensation may be relevant for polygenic complex traits, such as BD [9].

Because males and females have different numbers of copies of the X chromosome, we hypothesize that X chromosome genetics might play a role in observed sex differences in BD. In particular, females carry two X chromosomes, while males carry only one, and the X chromosome in females (but not males) undergoes a process called X chromosome inactivation (XCI). This is an epigenetic process initiated by the long non-coding RNA *XIST* that triggers silencing of the inactive X, which results in males and females expressing similar levels of many X chromosome genes [10, 11]. The identity of the inactive X is random in humans [12], and the process is also tissue- and cell-specific [13, 14]. Furthermore, XCI does not affect all loci on the X chromosome. In fact, approximately 15% of X chromosome loci escape from XCI and are expressed from both X chromosomes in females [15], although these genes are not fully expressed from the inactive X. Escape genes include genes in the pseudoautosomal regions at the ends of the chromosome (PAR1 and PAR2), as well as gametologs (genes with homologous copies on X and Y, for which females have two copies on the X and males have one copy on X and one copy on the Y), and other genes escape variably [10]. The unique biology of the X chromosome means that applying approaches for analyzing autosomal genetic variants is not appropriate.

In this work, we develop a new approach for analyzing X chromosome genetic variants, which incorporates prior biological information on XCI status of various genes, and apply the approach to examine the role of X chromosome genetic variation in sex-specific symptoms of BD. Our approach combines existing approaches for testing marginal genetic associations within a logistic regression framework. We also consider a test that accounts for single nucleotide polymorphisms (SNP)-sex interactions to allow for different effects of X chromosome variants in males and females. We compare results across methods to enable assessment of potential strengths and limitations of each approach and report on our findings regarding the association of X chromosome variants with sex-specific symptoms and comorbidities of BD.

## Methods

In this study, we examined whether X chromosome variants are associated with sex-associated symptoms and comorbidities of BD. We utilized two cohorts of individuals with BD, one from the Mayo Clinic Bipolar Disorder Biobank [16] and one from the Genetic Association Information Network (GAIN) Study of BD [17], and we employed two different X chromosome-specific statistical approaches to assess associations between SNP genotypes and phenotypes. Rather than using a discovery-validation approach, a meta-analysis was conducted in order to boost sample size and reproducibility by combining the results derived from both cohorts (GAIN and Mayo).

### Mayo cohort

#### Subject selection

Individuals with BD ( $N = 969$ ) from the Mayo Clinic Bipolar Disorder Biobank [16] (Mayo Bipolar Biobank) that had previously undergone genome-wide genotyping on the Illumina® Human OmniExpress BeadChip (Illumina®, Inc. San Diego, CA, USA) were included in this study. Control subjects ( $N = 777$ ) that did not have BD or a psychiatric illness themselves or a first-degree relative with BD were selected from the Mayo Clinic Biobank [18]. This case/control set was previously analyzed [19] and was included in a large genome-wide association study conducted by the Psychiatric Genomics Consortium [2].

#### Phenotyping

Symptoms and comorbidities of BD were assessed through patient and clinical questionnaires [16]. Variables analyzed in this study included the symptom of rapid cycling (RC), comorbidities of binge eating behavior (BE), lifetime history of suicide attempt, and whether the individual had an alcohol use disorder (AUD),

defined in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSMIV) as a diagnosis of alcohol dependence or abuse [20]. Rapid cycling was defined as having four or more mood episodes within a year. Binge eating behavior was defined as an affirmative response to questions 5 and 6 of the Eating Disorders Diagnostic Scale [21]. These questions read “During the past 6 months have there been times when you felt you have eaten what other people would regard as an unusually large amount of food (e.g. a quart of ice cream) given the circumstances?” and “During the times when you ate an unusually large amount of food, did you experience a loss of control (feel you couldn’t stop eating or control what or how much you were eating)?” [21].

### Genotyping

Quality control (QC) and imputation of genotyping data were performed using standard procedures as previously described [22]. Genetic ancestry was estimated with STRUCTURE [23, 24] using 1000 Genomes Project reference panels and used to exclude individuals of non-European ancestry. Genome-wide principal components were calculated to allow for adjustment for population substructure. X chromosome SNPs were imputed using IMPUTE 2.2.2 [25] with the 1000 Genomes Project reference panel (phase 1 data, all populations). Analyses were limited to X chromosome SNPs that had minor allele frequency above 0.05 and imputation  $R^2$  above 0.8. SNPs in the pseudoautosomal region (PAR) defined by GrCh37 were excluded due to low genotyping call rate.

### GAIN cohort

#### Subject selection

Cases with BD and controls without BD were recruited to the GAIN study and underwent phenotyping and genotyping as previously described [17] with data deposited in dbGaP [26] (accession number: phs000017.v3.p1). We used data from the subjects of European ancestry that passed genetic data QC ( $N = 1001$  cases and  $N = 1034$  controls).

### Phenotyping

A history of BE, RC, suicide attempt, or an AUD was assessed in cases using the Diagnostic Interview of Genetic Studies (DIGS) (versions 2–4) [27]. Binge eating behavior was defined based on having affirmative responses to questions that addressed overeating and loss of control: “Has there ever been a time in your life when you went on food binges (i.e., rapid consumption of a large amount of food in a discrete period of time, usually less than two hours)?” and “During these binges were you afraid you could not stop eating, or that your eating was out of control?”. The presence of AUD was determined from the presence of any ICD 9 codes indicating

DSMIII-R or DSMIV diagnoses of alcohol abuse (305.00; ICD-10 = F10.10) or alcohol dependence (303.90; ICD-10 = F10.20). Rapid cycling was defined as the presence of at least four mood episodes in a year.

### Genotyping

Genotyping was performed using an Affymetrix™ Genome-Wide Human SNP Array 6.0 (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Quality control was performed as previously described [17]. Imputation was performed as previously described [28]. SNPs analyzed were limited to those with MAF above 0.05 and imputation  $R^2$  above 0.8. SNPs in the PAR (defined by GrCh37) were excluded.

### Association testing

Because of the unique biology of the X chromosome, testing associations between X chromosome genetic variants and phenotypes requires different approaches than for autosomes. Previous work has used a logistic regression framework but coded the SNP variable differently depending upon the approach applied (Table 1). The coding approach historically implemented in the PLINK software [29] codes female genotypes as 0, 1, or 2 copies of the alternate allele and male genotypes as 0 or 1 copies of the alternate allele. This genotype coding ignores that XCI occurs and assumes that variants on both copies of the X chromosome are expressed in females (i.e., escape from XCI); this implicitly assumes that the effect of a change of a single allele has the same effect in females and males. As this is not true when a SNP is in a region that is inactivated, an alternate approach is to treat all SNPs as subject to XCI, using an approach originally proposed by Clayton [30]. Male genotypes are coded as 0 or 2 copies of the alternate allele, assuming that these male genotypes have the same effect as the respective homozygotes in females. Assuming that XCI is random across cells within a woman and random across women, female heterozygotes are viewed as an intermediate genotype, coded as 1. However, this also may not be optimal as 15% of X chromosome genes are expressed from both the active and inactive X chromosome. Given prior information regarding whether a region undergoes X chromosome inactivation, it is reasonable to consider this biological information when evaluating X chromosome associations.

In this study, we employed two X chromosome-specific approaches that allow for modeling SNP effects depending on XCI status (inactivation vs. escape). In the first approach, we used biological data on which regions are likely to experience XCI to model SNP effects differently for regions subject to and escaping from XCI; this approach assumes that under a given coding scheme, the SNP effect is the same in males and females.

**Table 1** Different coding schemes for the SNP variable reflect different assumptions regarding XCI status

Assumed XCI status	Coding scheme	Coding of the SNP variable (Eq. 1)				Calculation of ORs <sup>a</sup>		
		Sex	# copies of effect allele			OR <sub>M</sub>	OR <sub>W1</sub>	OR <sub>W2</sub>
Subject	Clayton	M	0	2	NA	$e^{2\beta_1}$	$e^{\beta_1}$	$e^{2\beta_1}$
		F	0	1	2			
Escape	PLINK	M	0	1	NA	$e^{\beta_1}$	$e^{\beta_1}$	$e^{2\beta_1}$
		F	0	1	2			

<sup>a</sup>Expressions for calculating the OR are shown for the effect of one allele copy in men (OR<sub>M</sub>) and one or two copies in women (OR<sub>W1</sub> and OR<sub>W2</sub>). For men, but not women, the OR is calculated differently based on whether the SNP is believed to be subject to XCI (and hence the coding scheme, PLINK or Clayton, that is chosen)

Specifically, in regions believed to undergo XCI, we used the Clayton coding of male genotypes (0/2) and test the SNP effect while assuming that the minor allele in males has the same effect as two copies of the minor allele in females (OR<sub>M</sub> = OR<sub>W2</sub>). On the other hand, in regions believed to escape XCI, we used the PLINK coding of male genotypes (0/1) and test the SNP effect while assuming that the minor allele in males has the same effect as one copy of the minor allele in females (OR<sub>M</sub> = OR<sub>W1</sub>; Table 1). In the second approach, we fit a more flexible regression model that can model SNPs that are either subject to or escaping from XCI, without the need for prior biological knowledge of the XCI status. This approach also allows for SNP effects to differ in males and females. These approaches are compared in the context of investigating the genetics of BD-related traits.

### Approach 1: XCI-informed approach

#### Deriving a presumed XCI status for each X chromosome SNP

Previous work by Balaton et al. [31] derived a “consensus” inactivation status across multiple studies and multiple tissue types for approximately 400 genes on the X chromosome. To infer XCI status at a SNP level, we used the presumed XCI status for each gene (as given in “Additional file 1: Table S1.” from Balaton et al. [31]). Start and stop positions of all genes are per the transcription start and stop sites. Any SNPs overlapped by only “subject” genes (category: Subject) or only “escape” genes (categories: PAR and escape) were assigned the corresponding XCI status (“subject” or “escape”); SNPs lying between genes of the same type were also assigned the corresponding XCI status. SNPs between “subject” and “escape” genes or overlapping both “subject” and “escape” genes were assigned an XCI status of “unknown.”

#### Using an XCI status informed approach for testing associations between X chromosome SNPs and phenotype

To test association with each phenotype, a sex-adjusted logistic regression model (Eq. 1) was used:

$$\text{logit}(\text{pheno}) = \beta_0 + \beta_1 (\text{SNP}) + \beta_2 (\text{sex}) \quad (1)$$

Sex was coded as 0 for females and 1 for males. Irrespective of presumed XCI status, the SNP variable in females was set equal to the number of copies of the minor allele. However, in males, the SNP variable’s coding depended on the presumed XCI status and hence the coding scheme chosen (Clayton or PLINK coding; Table 1). SNPs of unknown XCI status were modeled under both coding schemes (Clayton and PLINK), and Akaike information criterion (AIC) was used to determine which XCI status led to the better fitting model in each cohort (lower AIC indicates better model fit).

When XCI status at a SNP was unknown and the cohorts gave discordant presumed XCI statuses, the coding used for generating the cohort-specific summary statistics for the meta-analysis was the Clayton coding, since most of the X chromosome is subject to XCI.

### Approach 2: XCI-robust approach

In this second approach, a logistic regression model with a SNP-sex interaction term (Eq. 2) was employed, where the SNP variable was the count of copies of the minor allele, and a likelihood ratio test with two degrees-of-freedom (df) was used to jointly assess the significance of the SNP and SNP-sex interaction terms. Sex was coded as 1 for males and 0 for females.

$$\text{logit}(\text{pheno}) = \beta_0 + \beta_1 (\text{SNP}) + \beta_2 (\text{sex}) + \beta_3 (\text{SNP})(\text{sex}) \quad (2)$$

To facilitate the interpretation of the top SNP effects in males and females, sex-stratified logistic regression analyses were conducted in Mayo and GAIN.

For all analyses, a chromosome-wide Bonferroni-corrected significance threshold was set by dividing 0.05 by the number of SNPs passing QC in the GAIN set prior to imputation ( $P = 0.05/26,662 = 1.88 \times 10^{-6}$ ). Regression analyses were performed in R using the “glm” function. Analyses incorporated additional covariates for genetic ancestry as assessed by principal components,



DIGS questionnaire version (for GAIN), and enrollment site (Mayo Clinic cohort only) when necessary. For the XCI-informed approach, meta-analysis of results from the Mayo and GAIN cohorts was conducted in METAL by weighting observations from each study inversely proportional to their standard errors [32]. For the XCI-robust approach, the  $P$  values from the 2df test in the Mayo and GAIN cohorts were combined by Fisher's method to derive a joint  $P$  value implemented in R [29]. Meta-analyses of sex-stratified results from the Mayo and GAIN cohorts were performed using inverse-variance weighting using METAL [32] to estimate SNP effects in men and women separately for each phenotype.

### Candidate SNP study

Previously, Jancic et al. [33] analyzed the association of X chromosome SNPs with risk of suicide attempt in individuals with BD (983 suicide attempters, 1143 non-attempters), which included the individuals from the GAIN sample analyzed here. We attempted to replicate the top ten SNPs from that paper in the independent Mayo sample. The original work used the PLINK coding and sex-adjusted logistic regression to identify top SNPs. We applied the two X chromosome-specific approaches described here to the Mayo data. As all ten SNPs reported by Jancic lay in a region subject to XCI, the "XCI-informed approach" used Clayton coding for all of these SNPs.

### Annotation of lead SNPs

All lead SNPs reported in this paper were annotated to the nearest gene (not counting pseudo-genes or lncRNAs) using BioR [34] and Gr37Chp5 or by visual inspection in the University of California Santa Cruz (UCSC) Genome Browser. The GTEx database [35] was used to verify if any of the top SNPs are expression quantitative trait locus (eQTLs) in any tissue ( $FDR < 0.05$ ) or are splice quantitative trait locus (sQTLs) ( $FDR < 0.05$ ).

### Results

All characteristics of BD examined (RC, suicide attempt, BE, and AUD) were relatively common in both Mayo

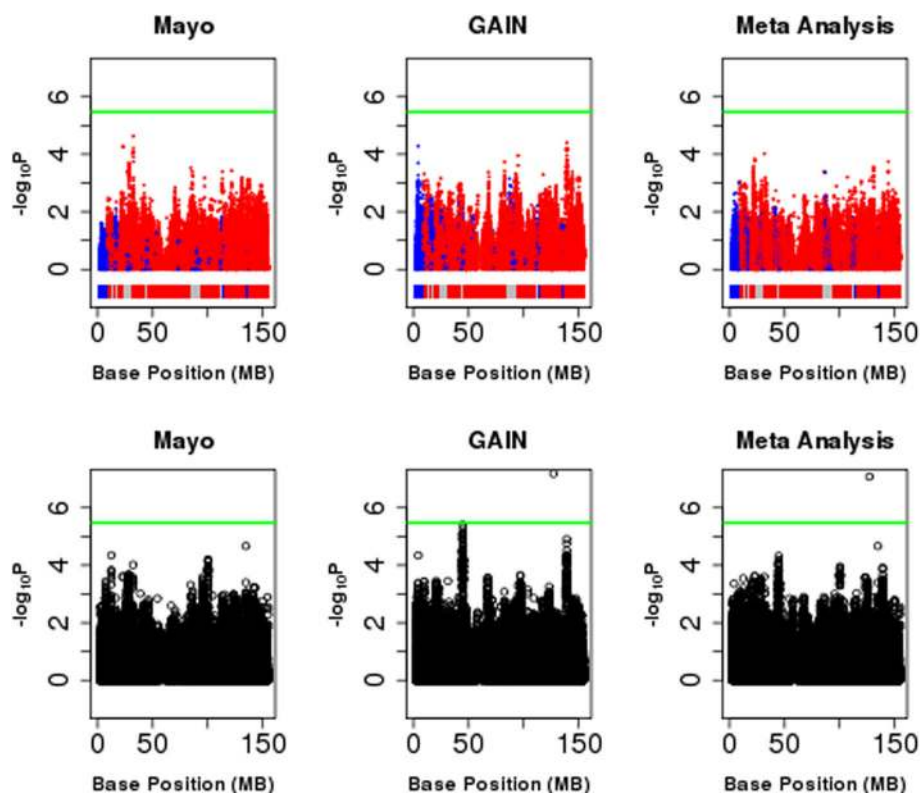
Clinic and GAIN datasets (Table 2). For both Mayo Clinic and GAIN, women were more likely than men to engage in BE or to have attempted suicide, and men were at greater risk for having an AUD. Additionally, RC was significantly more common for female cases ( $P = 0.004$ ) for Mayo, although this was not true for GAIN ( $P = 0.580$ ).

X chromosome-wide results for all phenotypes under both the "XCI-informed approach" and the "XCI-robust approach" are displayed in Fig. 1 and Additional files 1, 2, 3, and 4. Using the "XCI-informed approach," which examines marginal SNP effects, no SNPs were identified that were significantly associated with BD or any of its sex-specific symptoms and comorbidities (Additional file 5: Table S1). However, using the "XCI-robust approach," which considers SNP-sex interactions, the SNP rs5932307 was significantly associated with BD ( $P = 8.31E-8$ ; Table 3). The minor A allele was associated with higher odds of BD, with a stronger effect in females ( $OR_{W2} = 3.86$  vs.  $OR_M = 1.13$ ). This SNP is downstream of the ACTRT1 gene, which has the highest gene expression in the testes [35] and encodes a beta actin-like protein that is suggested to be important for spermatid formation [36]. It has not been identified as an expression quantitative trait locus (eQTL) in any tissues or spliceQTL. However, we should note that this SNP marginally deviates from Hardy-Weinberg Equilibrium in female controls in the GAIN sample ( $P = 1.2E-4$ ), but not in the Mayo sample ( $P > 0.05$ ).

Top SNPs for suicide attempt and AUD under the "XCI-robust approach," although not significant after Bonferroni correction, were single-tissue eQTLs (Table 3). The SNP most strongly associated with suicide attempt was rs5975146, an eQTL of the gene X-prolyl aminopeptidase 2 (XPNPEP2) in both tibial nerve and adipose tissue. The meta-analysis of results from Mayo and GAIN under the "XCI-robust approach," which allows SNP effects to differ by sex, suggests that the minor A allele of rs5975146 may be associated with greater risk of suicide attempt, but only among females ( $OR_{W1} = 1.40$ ,  $P_{2df} = 1.5E-5$ ). Additionally, the SNP most associated with AUD (rs145649722) was an eQTL of CLCN5 in the skin. The results from the meta-analysis suggest

**Table 2** Characteristics of bipolar disorder cases

	Mayo			GAIN		
	Male ( <i>N</i> = 396)	Female ( <i>N</i> = 573)	<i>P</i>	Male ( <i>N</i> = 500)	Female ( <i>N</i> = 501)	<i>P</i>
Age, mean (SD)	43.48 (15.78)	42.35 (14.82)	0.255	41.96 (13.01)	42.28 (12.86)	0.697
Rapid cycling, <i>N</i> (%)	211 (53.4)	357 (62.9)	0.004	159 (38.6)	175 (40.7)	0.580
Suicide attempt, <i>N</i> (%)	87 (22.1)	222 (39.2)	< 0.001	175 (36.0)	252 (50.9)	< 0.001
BE behavior, <i>N</i> (%)	61 (21.9)	131 (31.0)	0.011	30 (6.6)	94 (20.1)	< 0.001
Alcohol use disorder, <i>N</i> (%)	170 (44.4)	199 (35.9)	0.011	257 (51.4)	208 (41.5)	0.002



**Fig. 1** Association of X chromosome variants with BD. Top row denotes results from XCI-informed approach. Bottom row denotes results from XCI-robust approach. Green line denotes the study-wide significance threshold of  $3.36 \times 10^{-6}$ . Domains as shown in the colored bars beneath the Manhattan plots for XCI-informed approach denote whether SNPs fall into regions experiencing (red) or escaping (blue) from X chromosome inactivation. Grey denotes regions for which a domain (subject or escaping) could not be assigned based on the paper by Balaton et al. [31]. SNPs are colored by the chosen XCI status used in the meta-analysis

that the minor G allele of rs145649722 may be associated with greater odds of AUD, primarily in males ( $OR_M = 3.20$ ,  $OR_{W2} = 0.55$ ,  $P_{2df} = 4.1E-4$ ).

We analyzed ten SNPs most strongly associated with suicide attempt in prior work [33] in the Mayo Clinic cohort. None of these SNPs was even nominally associated ( $P < 0.05$ ) with the risk of suicide attempt in the independent Mayo Clinic sample (Additional file 6: Table S2). When the GAIN data was analyzed using the XCI-informed and XCI-robust methods, only two SNPs were nominally associated (rs5909133,  $P_{informed} = 0.0037$ ,  $P_{robust} = 0.014$ ; rs695214,  $P_{informed} = 0.00052$ ,  $P_{robust} = 0.0013$ ); this cannot be considered an independent replication, as the prior study included the GAIN data.

## Discussion

In this study, we examined the association of X chromosome SNPs with sex-associated characteristics of BD using two different X chromosome-specific analysis approaches. These approaches consider the sex-specific nature of the X chromosome and the process of XCI and allow for a more flexible interpretation of the findings.

The sex associations of the BD characteristics are as expected based on prior work, including higher rates of RC, lifetime history of suicide attempt, and greater prevalence of BE in women, as well as greater prevalence of AUDs in men.

The SNP rs5932307 was significantly associated with BD under the “XCI-robust approach” ( $P = 8.3 \times 10^{-8}$ ), even with a conservative, Bonferroni-corrected significance threshold of  $P = 1.88 \times 10^{-6}$ . This contrasts with results from a recent GWAS that employed a two-stage methodology with independent discovery (7467 cases/27,303 controls) and replication samples (2313 cases/3489 controls); in that study, despite the larger sample size of the discovery cohort, no X chromosome SNPs passed the threshold ( $P = 1 \times 10^{-6}$ ) to advance to testing in the replication sample [1]. However, this may be because different approaches to association testing were employed. In the previous study, the association test used the Clayton coding, which assumes that the minor allele in males has the same effect as two copies of the minor allele in females. However, the approach that yielded the significant result for our analysis was the “XCI-robust approach,” which allowed the effect of the

**Table 3** Top SNPs under “XCI-robust” approach

Phenotype	SNP	Alleles (min/maj)	Nearest gene	Cohort	MAF	XCI-informed				XCI-robust			
						Chosen XCI status	OR <sub>w</sub>	OR <sub>w1</sub>	OR <sub>w2</sub>	P	OR <sub>w</sub>	OR <sub>w1</sub>	OR <sub>w2</sub>
Case status	rs5932307	A/G	ACTRT1	Mayo	0.08	S	1.59 (1.04–2.43)	1.26 (1.02–1.56)	1.59 (1.04–2.43)	0.03	1.39 (0.82–2.36)	1.42 (1.00–2.02)	2.02 (1.00–4.08)
				GAIN	0.07	S	1.60 (1.12–2.30)	1.27 (1.06–1.52)	1.60 (1.12–2.30)	0.01	1.01 (0.67–1.50)	3.54 (2.20–5.69)	12.52 (4.85–32.34)
				Meta		S	1.60 (1.21–2.10)	1.26 (1.10–1.45)	1.60 (1.21–2.10)	8.2E–04	1.13 (0.82–1.56)	1.96 (1.48–2.60)	3.86 (2.19–6.78)
Rapid cycling	rs5933727	G/C	TBL1X	Mayo	0.07	E*	1.91 (1.19–3.06)	1.91 (1.19–3.06)	3.65 (1.42–9.38)	0.01	1.31 (0.56–3.07)	2.25 (1.26–4.02)	5.08 (1.60–16.18)
				GAIN	0.08	E*	0.38 (0.23–0.62)	0.38 (0.23–0.62)	0.15 (0.05–0.39)	1.2E–04	0.76 (0.34–1.71)	0.27 (0.14–0.51)	0.07 (0.02–0.26)
				Meta		E*	0.77 (0.55–1.08)	0.77 (0.55–1.08)	0.60 (0.30–1.18)	0.14	0.99 (0.55–1.77)	0.86 (0.56–1.31)	0.73 (0.31–1.73)
Suicide attempt	rs5975146	A/G	XPNIPEP2	Mayo	0.12	S	0.47 (0.25–0.85)	0.68 (0.50–0.92)	0.47 (0.25–0.85)	0.01	0.10 (0.02–0.45)	0.99 (0.67–1.46)	0.98 (0.44–2.15)
				GAIN	0.11	S	1.77 (1.09–2.87)	1.33 (1.04–1.69)	1.77 (1.09–2.87)	0.02	1.07 (0.58–1.99)	2.15 (1.39–3.32)	4.63 (1.94–11.05)
				Meta		S	1.05 (0.72–1.53)	1.02 (0.85–1.24)	1.05 (0.72–1.53)	0.80	0.76 (0.43–1.33)	1.40 (1.05–1.88)	1.97 (1.10–3.53)
Binge eating	rs6627188	A/C	CNGA2	Mayo	0.14	S	0.17 (0.08–0.39)	0.41 (0.28–0.62)	0.17 (0.08–0.39)	2.0E–05	0.06 (0.01–0.50)	0.48 (0.30–0.78)	0.23 (0.09–0.61)
				GAIN	0.13	S	1.65 (0.86–3.14)	1.28 (0.93–1.77)	1.65 (0.86–3.14)	0.13	2.00 (0.76–5.27)	1.21 (0.78–1.86)	1.45 (0.61–3.46)
				Meta		S	0.68 (0.41–1.13)	0.83 (0.64–1.06)	0.68 (0.41–1.13)	0.14	1.08 (0.45–2.60)	0.80 (0.58–1.11)	0.64 (0.34–1.23)
Alcohol use disorder	rs145649722	G/C	CLCN5	Mayo	0.04	S	0.78 (0.28–2.15)	0.88 (0.53–1.47)	0.78 (0.28–2.15)	0.63	0.73 (0.16–3.29)	0.92 (0.46–1.83)	0.84 (0.21–3.36)
				GAIN	0.05	S	2.60 (1.24–5.47)	1.61 (1.11–2.34)	2.60 (1.24–5.47)	0.01	8.29 (2.50–27.51)	0.58 (0.29–1.18)	0.34 (0.08–1.39)
				Meta		S	1.68 (0.92–3.05)	1.30 (0.96–1.75)	1.68 (0.92–3.05)	0.09	3.20 (1.26–8.15)	0.74 (0.45–1.21)	0.55 (0.21–1.47)

Odds ratios associated with an increase of one minor allele copy in men (OR<sub>w</sub>) or women (OR<sub>w1</sub>) or an increase of two copies in women (OR<sub>w2</sub>) are reported for two different analysis approaches. The XCI-informed approach employed a sex-adjusted logistic regression model (Eq. 1), but coded the SNP variable differently dependent on presumed XCI status (listed in the “Chosen XCI status” column). SNPs were assigned a status of subject (S) or escaping from inactivation, based on prior work on which regions of the X chromosome experience inactivation. For SNPs in regions of unknown XCI status (entries with asterisk in the “Chosen XCI status” column), presumed XCI status was determined by fitting the model using both the PLINK and Clayton coding schemes and using Akaike information criterion to select the more appropriate model. The XCI-robust approach employed a sex-adjusted logistic regression model with a SNP-sex interaction term (Eq. 2). The significance of the SNP and SNP-sex terms in the model was assessed by a  $\chi^2$  test with two degrees-of-freedom (2df)

Italics denote the results of the meta-analysis

SNP to differ by sex. The potential importance of allowing SNP effects to differ by sex is highlighted by the fact that for this SNP, sex-stratified analyses suggest that the minor allele is more strongly associated with BD for females ( $OR_{W2} = 3.86$ , 95% CI 2.19–6.78) than for males ( $OR_M = 1.13$ , 95% CI 0.82–1.56). However, this result should be interpreted cautiously given that this SNP showed some deviation from Hardy Weinberg equilibrium in one of the analyzed datasets.

Although not significant after multiple testing correction, the SNP most strongly associated with suicide attempt (rs5975146) was an eQTL of the X-prolyl aminopeptidase 2 (XPNPEP2) in both tibial nerve and adipose tissue, and the SNP most associated with AUD (rs145649722) was an eQTL of CLCN5 in the skin. The gene CLCN5 encodes the protein chloride channel 5 (Clc-5), and one study found the gene CLCN5 to be differentially methylated in brain tissue from obsessive-compulsive disorder subjects and controls [34].

Candidate SNPs most significantly associated with risk of suicide attempt in a prior study in a BD population of which the GAIN data was a subset [33] were not significantly associated with suicide attempt within our Mayo cohort, regardless of coding or approach, with most OR estimates close to one. This may have been due to differences in methodology, as most of these SNPs were also not associated in our analysis of the GAIN data, with the exception of rs695214.

Importantly, correct interpretation of X chromosome association results depends on the statistical model that was fit and genotype coding that was used, which reflect assumptions that were made. When interpreting effect size for X chromosome SNPs, multiple ORs are informative. Whereas for autosomes ORs are commonly reported for the change of one allele copy (assuming an additive model for allele effects), it is less clear what is most appropriate to report for X chromosome variants, because the effect of the SNP varies with sex. Under the “XCI-informed approach,” for SNPs lying in regions that escape from XCI, the assumption is that the OR in males ( $OR_M$ ) is the same as females for a change of one allele copy ( $OR_{W1}$ ). However, for SNPs lying in regions experiencing XCI, the effect of a change of one allele copy in males ( $OR_M$ ) is expected to be comparable to a change of two copies in females ( $OR_{W2}$ ). These assumptions are implicit in the “XCI-informed approach,” which assumes a log-additive effect of SNP in females.

While the “XCI-robust approach” that includes SNP-sex interactions also assumes that the effects of SNPs are log-additive in females, it is more flexible because the effect of a SNP can vary by sex.  $OR_M$  is not constrained to equal the effect of the SNP in females ( $OR_{W1}$  or  $OR_{W2}$ ), which even allows for a SNP to exhibit a protective effect in one sex and to be a risk factor for the

other sex. It is worth noting that the “XCI-informed approach” and the “XCI-robust approach” are designed to detect different genetic effects on the phenotype. The “XCI-informed approach” examines the main effect of the SNP variable on the phenotype, whereas the “XCI-robust approach” with the 2df test reflects the joint importance of the SNP and SNP-sex interaction terms, and hence is sensitive not only to the main effects but also to differences in the SNP effect between sexes.

The importance of allowing for this flexibility in the model can be seen by looking at top SNPs for each phenotype under the more restrictive “XCI-informed approach.” All of these SNPs are in a region subject to XCI, which would lead one to predict those SNPs have the same effect for one allele in males as two copies in females (i.e.,  $OR_M = OR_{W2}$ ). However, examining the sex-stratified ORs for those SNPs (Additional file 5: Table S1) shows that many of those SNPs potentially have SNP effects that do not follow the expected theoretical pattern. For example, the top SNP for AUD under the “XCI-informed approach,” rs62587381, has an estimated OR in males that is much greater than in females ( $OR_M = 4.32$  versus  $OR_{W2} = 1.85$ ).

One might be concerned that the increased flexibility of the model might come at the expense of reduced power to detect genetic differences. However, this does not appear to be a major concern, at least in our study. For three of the five top SNPs for each phenotype under the “XCI-informed approach,” we observed a *P* value within an order of magnitude for the “XCI-robust approach.” Additionally, only the “XCI-robust approach” resulted in a significant finding for any of the phenotypes studied. However, a disadvantage of the “XCI-robust approach” when used across datasets that are subsequently meta-analyzed is that it relies on a two degrees-of-freedom likelihood ratio test statistic that does not retain the directionality of the SNP effect, which can lead to difficulties in interpreting meta-analysis results.

Selection of prior gene-level XCI states is necessary for the “XCI-informed approach.” We used the XCI consensus states described in Balaton et al. [31], because they were assessed across multiple studies and multiple tissue types and could be considered generally applicable, and it is unclear which tissue type might best inform BD risk. Because XCI patterns are known to be tissue-specific, a tissue-specific XCI source could be used for conditions with clearly defined normal tissue types, if it exists [13]. Failing to properly account for tissue-specific patterns could possibly lead to a reduction in power for the “XCI-informed approach” if the wrong XCI state is modeled. An advantage of the “XCI-robust approach” is that it does not rely on specification of the tissue-specific XCI pattern. Furthermore, the “XCI-robust approach”



can also accommodate the phenomenon of partial or incomplete escape from XCI, which is not accounted for in the “XCI-informed approach.”

Neither the “XCI-informed” or the “XCI-robust” approaches directly account for genes that are homologous across the X and Y chromosome (gametologs), as they do not incorporate Y chromosome data from males. The “XCI-informed” approach treats SNPs within these genes as escaping from XCI, whereas the “XCI-robust” approach does not make any assumptions about XCI status. This suggests that development of methods that incorporate X and Y data for studying these regions would be valuable.

Strengths of our work include the investigation of the role of X chromosome genetic variants to multiple symptoms and comorbidities of BD with known sex-differences in prevalence, and the use of two methods of analysis that can model the effect of SNPs both subject to and escaping from XCI. Importantly, we developed a new approach for analyzing X chromosome genetic variants that incorporates prior biological information on XCI status. However, our study also has limitations. The biological relevance of our observed associations is unclear, and laboratory validation required to establish biological associations is beyond the scope of this work, as is a comparison of the genetic versus hormonal influences on sex differences in BD. The relatively small sample size limited statistical power and makes interpretation of the significance of our findings difficult. Additionally, our cohorts were composed solely of individuals of European ancestry. Future work in more ethnically diverse cohorts or larger cohorts such as the Psychiatric Genomics Consortium might allow us to discover new X chromosome genetic variants that are important to BD risk and allow for findings with greater generalizability.

This work provides a basis for future methodological studies. Future work should extend both approaches to incorporate data from the Y chromosome in males for the XY gametolog genes. The relative merits of the two approaches should be more rigorously assessed by simulation studies assessing type I error and statistical power, as well as comparison to other existing approaches [37]. Alternate approaches could be explored, such as prioritizing SNPs in sex-biased genes or using Bayesian methods or model averaging [38], which could reflect the uncertainty that exists about a locus’ XCI status. In addition, statistical approaches to determine the likely genetic architecture by which genotypes alter phenotypes (e.g., additivity vs. dominance of allelic effects) could also be pursued; additionally, information about the genetic architecture may also imply XCI status. Finally, the versatility and relative ease of implementation of our approach should encourage its broad application,

particularly in conditions where X chromosome involvement is suggested, but few if any specific genes have been identified.

## Perspectives and significance

In conclusion, we employed two different approaches to the analysis of X chromosome genetic variants that are able to model SNPs both subject to and escaping from XCI. In the “XCI-informed approach,” we used biological information regarding what regions of the X chromosome undergo XCI to code the SNP variable differently for regions believed to undergo versus escape from inactivation. In the “XCI-robust approach,” a more flexible model with a SNP-sex interaction term was fit that allowed for SNPs both in regions of inactivation and escape, without the need for prior knowledge as to the true XCI status. We also describe how the SNP effect sizes can be interpreted for each sex based on the model that was fit.

Neither approach identified SNPs that were significantly associated with sex-specific symptoms of BD, although the interaction approach identified a SNP (rs5932307) associated with risk of BD ( $P = 8.31 \times 10^{-8}$ ). Future work in larger, independent cohorts is needed to replicate this finding, but our work highlights the importance of applying X chromosome-specific methods and careful interpretation of the results when analyzing phenotypes with known sex differences.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13293-019-0272-4>.

**Additional file 1: Figure S1.** Association of X chromosome genetic variants with RC. Top row denotes results from XCI-Informed Approach. Bottom row denotes results from XCI-Robust Approach. Green line denotes the study wide significance threshold of  $3.36 \times 10^{-6}$ . Domains as shown in the colored bars beneath the Manhattan plots for XCI-Informed Approach denote whether SNPs fall into regions experiencing (red) or escaping (blue) from X chromosome inactivation. Grey denotes regions for which a domain (subject or escaping) could not be assigned based on the paper by Balaton et al [31]. SNPs are colored by the chosen XCI status used in the meta-analysis.

**Additional file 2: Figure S2.** Association of X chromosome genetic variants with attempted suicide. Top row denotes results from XCI-Informed Approach. Bottom row denotes results from XCI-Robust Approach. Green line denotes the study wide significance threshold of  $3.36 \times 10^{-6}$ . Domains as shown in the colored bars beneath the Manhattan plots for XCI-Informed Approach denote whether SNPs fall into regions experiencing (red) or escaping (blue) from X chromosome inactivation. Grey denotes regions for which a domain (subject or escaping) could not be assigned based on the paper by Balaton et al [31]. SNPs are colored by the chosen XCI status used in the meta-analysis.

**Additional file 3: Figure S3.** Association of X chromosome genetic variants with BE. Top row denotes results from XCI-Informed Approach. Bottom row denotes results from XCI-Robust Approach. Green line denotes the study wide significance threshold of  $3.36 \times 10^{-6}$ . Domains as shown in the colored bars beneath the Manhattan plots for XCI-Informed Approach denote whether SNPs fall into regions experiencing (red) or escaping (blue) from X chromosome inactivation. Grey denotes regions

for which a domain (subject or escaping) could not be assigned based on the paper by Balaton et al [31]. SNPs are colored by the chosen XCI status used in the meta-analysis.

**Additional file 4: Figure S4.** Association of X Chromosome Genetic Variants with AUD. Top row denotes results from XCI-Informed Approach. Bottom row denotes results from XCI-Robust Approach. Green line denotes the study wide significance threshold of  $3.36 \times 10^{-6}$ . Domains as shown in the colored bars beneath the Manhattan plots for XCI-Informed Approach denote whether SNPs fall into regions experiencing (red) or escaping (blue) from X chromosome inactivation. Grey denotes regions for which a domain (subject or escaping) could not be assigned based on the paper by Balaton et al [31]. SNPs are colored by the chosen XCI status used in the meta-analysis.

**Additional file 5: Table S1.** Top SNPs under "XCI-informed" Approach.

**Additional file 6: Table S2.** Candidate SNPs for Association with Suicide Attempt.

## Abbreviations

AIC: Akaike information criterion; AUD: Alcohol use disorder; BD: Bipolar disorder; BE: Binge eating behavior; df: Degrees-of-freedom; DIGS: Diagnostic Interview of Genetic Studies; eQTL: Expression quantitative trait locus; GAIN: Genetic Association Information Network; MAF: Minor allele frequency; OR: Odds ratio; OR<sub>M</sub>: Odds ratio in males for a change of 1 allele copy; OR<sub>W</sub>: Odds ratio in females for a change of 1 allele copy; OR<sub>W2</sub>: Odds ratio in females for a change of 2 allele copies; PAR: Pseudoautosomal region; QC: Quality control; QTL: Quantitative trait locus; RC: Rapid cycling; SNP: Single nucleotide polymorphism

## Acknowledgements

Funding support for the Whole Genome Association Study of Bipolar Disorder was provided by the National Institute of Mental Health (NIMH), and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of Genotypes and Phenotypes (dbGaP) found at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000017.v3.p1. Samples and associated phenotype data for the Collaborative Genomic Study of Bipolar Disorder were provided by the NIMH Genetics Initiative for Bipolar Disorder. Data and biomaterials were collected in four projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1991 to 1998, the Principal Investigators and Co-Investigators were Indiana University, Indianapolis, IN, U01 MH46282, John Nurnberger, M.D., Ph.D., Marvin Miller, M.D., and Elizabeth Bowman, M.D.; Washington University, St. Louis, MO, U01 MH46280, Theodore Reich, M.D., Allison Goate, Ph.D., and John Rice, Ph.D.; Johns Hopkins University, Baltimore, MD U01 MH46274, J. Raymond DePaulo, Jr., M.D., Sylvia Simpson, M.D., MPH, and Colin Stine, Ph.D.; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD, Elliot Gershon, M.D., Diane Kazuba, B.A., and Elizabeth Maxwell, M.S.W. Data and biomaterials were collected as part of ten projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1999-03, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, R01 MH59545, John Nurnberger, M.D., Ph.D., Marvin J. Miller, M.D., Elizabeth S. Bowman, M.D., N. Leela Rau, M.D., P. Ryan Moe, M.D., Nalini Samavedy, M.D., Rif El-Mallakh, M.D. (at University of Louisville), Hussein Manji, M.D. (at Wayne State University), Debra A. Glitz, M.D. (at Wayne State University), Eric T. Meyer, M.S., Carrie Smiley, R.N., Tatiana Foroud, Ph.D., Leah Flury, M.S., Danielle M. Dick, Ph.D., Howard Edenberg, Ph.D.; Washington University, St. Louis, MO, R01 MH059534, John Rice, Ph.D., Theodore Reich, M.D., Allison Goate, Ph.D., Laura Bierut, M.D.; Johns Hopkins University, Baltimore, MD, R01 MH59533, Melvin McInnis, M.D., J. Raymond DePaulo, Jr., M.D., Dean F. MacKinnon, M.D., Francis M. Mondimore, M.D., James B. Potash, M.D., Peter P. Zandi, Ph.D., Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, PA, R01 MH59553, Wade Berrettini, M.D., Ph.D.; University of California at Irvine, CA, R01 MH60068, William Byerley, M.D., and Mark Vawter, M.D.; University of Iowa, IA, R01 MH059548, William Coryell, M.D., and Raymond Crowe, M.D.; University of Chicago, IL, R01 MH59535, Elliot Gershon, M.D., Judith Badner, Ph.D., Francis McMahon, M.D., Chunyu Liu, Ph.D., Alan Sanders, M.D., Maria Caserta, Steven Dinwiddie, M.D., Tu Nguyen, Donna Harakal; University of California at San Diego, CA, R01 MH59567, John Kelsoe, M.D., Rebecca McKinney, B.A.; Rush

University, IL, R01 MH059556, William Scheftner, M.D., Howard M. Kravitz, D.O., M.P.H., Diana Marta, B.S., Annette Vaughn-Brown, MSN, RN, and Laurie Bederow, MA; NIMH Intramural Research Program, Bethesda, MD, 1Z01MH002810-01, Francis J. McMahon, M.D., Layla Kassem, PsyD, Sevilla Detera-Wadleigh, Ph.D., Lisa Austin, Ph.D., Dennis L. Murphy, M.D. The Genotype-Tissue Expression (GTEx) Project was supported by the [Common Fund](#) of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 5/31/2019.

## Authors' contributions

WAJ contributed to the design of the study, conducted the analyses, interpreted the results, and wrote the manuscript. CLC manages the data collection system database, provided assistance with data preparation, and conducted imputation and QC for datasets described. SJW contributed to the conception and design of the study, contributed to the conduct of the analysis, interpreted the results, contributed to the writing of the manuscript, and supervised this work. JMB contributed to the conception and design of the study, contributed to the conduct of the analysis, interpreted the results, contributed to the writing of the manuscript, supervised this work, and also served as co-PI for Mayo Clinic Individualized Medicine Biobank for Bipolar Disorder. MAF provided oversight for Mayo Clinic Individualized Medicine Biobank for Bipolar Disorder and assisted with patient recruitment and phenotyping. SLM is the principal investigator at the Lindner Center of HOPE/University of Cincinnati and participated in and supervised the patient recruitment and phenotyping. All authors reviewed, revised, and approved the final manuscript.

## Funding

This work was funded by the Marriott Foundation and Mayo Clinic Center for Individualized Medicine. WJ is supported by grant R25 GM075148 from the National Institutes of Health.

## Availability of data and materials

The datasets generated and/or analyzed for the GAIN cohort during the current study are available and were collected in previous work [17] and deposited in the dbGaP repository [26] (accession #: phs000017.v3.p1). Datasets generated and/or analyzed for the Mayo cohort contain protected health information and will not be shared to protect patient privacy.

## Ethics approval and consent to participate

The research in the Mayo cohort was approved under the title "Mayo Clinic Individualized Medicine Biobank for Bipolar Disorder" by the Mayo Clinic Institutional Review Board (IRB#: 08-008794).

## Consent for publication

Not applicable.

## Competing interests

SLMc has received research grants from Alkermes, AstraZeneca, Cephalon, Eli Lilly & Co., Forest, Marriott Foundation, Orexigen Therapeutics, Inc., Naurex, Pfizer, Shire, Takeda Pharmaceutical Company Ltd., and Transcept Pharmaceutical, Inc.; has been a consultant to or member of the scientific advisory boards of Alkermes, Bracket, Corcept, F. Hoffman La Roche, MedAvante, Naurex, Novo Nordisk, Shire, and Teva; and is also an inventor on US patent no. 6,323,236 B2, Use of Sulfamate Derivatives for Treating Impulse Control Disorders, and, along with the patent's assignee, University of Cincinnati, Cincinnati, OH, USA, has received payments from Johnson & Johnson, which has exclusive rights under the patent. MAF has received grant support from Pfizer and Myriad and has served as an unpaid consultant for Allergan, Myriad, Sunovion, and Teva Pharmaceuticals. WAJ, CLC, JMB, and SJW declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA. <sup>2</sup>Lindner Center of HOPE, University of Cincinnati College of Medicine, Mason, OH 45040, USA. <sup>3</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN 55905, USA.

Received: 2 July 2019 Accepted: 21 November 2019

Published online: 09 December 2019

# References

- Hou L, Bergen SE, Akula N, et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Human molecular genetics*. 2016;25(15):3383–94.
- Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics*. 2019; 51(5):793–803.
- Erol A, Winham SJ, McElroy SL, et al. Sex differences in the risk of rapid cycling and other indicators of adverse illness course in patients with bipolar I and II disorder. *Bipolar disorders*. 2015;17(6):670–6.
- McElroy SL, Crow S, Blom TJ, et al. Clinical features of bipolar spectrum with binge eating behaviour. *Journal of affective disorders*. 2016;201:95–8.
- Marrocco J, McEwen BS. Sex in the brain: hormones and sex differences. *Dialogues Clin Neurosci*. 2016;18(4):373–83.
- Munkholm K, Vinberg M, Kessing LV. Peripheral blood brain-derived neurotrophic factor in bipolar disorder: a comprehensive systematic review and meta-analysis. *Mol Psychiatry*. 2016;21(2):216–28.
- Saifi GM, Chandra HS. An apparent excess of sex- and reproduction-related genes on the human X chromosome. *Proc Biol Sci*. 1999;266(1415):203–9.
- Green T, Flash S, Reiss AL. Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacology*. 2019;44(1):9–21.
- Sidorenko J, Kassam I, Kemper KE, et al. The effect of X-linked dosage compensation on complex trait variation. *Nature Commun*. 2019;10(1):3009.
- Ross MT, Grafham DV, Coffey AJ, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434(7031):325–37.
- Brown CJ, Ballabio A, Rupert JL, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*. 1991;349(6304):38–44.
- Heard E, Distech CM. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev*. 2006;20(14):1848–67.
- Tukiainen T, Villani AC, Yen A, et al. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017;550(7675):244–8.
- Cotton AM, Price EM, Jones MJ, et al. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet*. 2015;24(6):1528–39.
- Carrel L, Willard HF. Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc Natl Acad Sci U S A*. 1999;96(13):7364–9.
- Frye MA, McElroy SL, Fuentes M, et al. Development of a bipolar disorder biobank: differential phenotyping for subsequent biomarker analyses. *Int J Bipolar Disord*. 2015;3(1):30.
- Smith EN, Bloss CS, Badner JA, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009;14(8):755–63.
- Olson JE, Ryu E, Johnson KJ, et al. The Mayo Clinic Biobank: a building block for individualized medicine. *Mayo Clin Proc*. 2013;88(9):952–62.
- Cuellar-Barboza AB, Winham SJ, McElroy SL, et al. Accumulating evidence for a role of TCF7L2 variants in bipolar disorder with elevated body mass index. *Bipolar Disord*. 2016;18(2):124–35.
- American Psychiatric Association, Frances A. Diagnostic and statistical manual of mental disorders DSM-IV. 4th Edition ed. American Psychiatric Association, Washington D.C. 1997. 886 p.
- Stice E, Telch CF, Rizvi SL. Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychol Assess*. 2000;12(2):123–31.
- McElroy SL, Winham SJ, Cuellar-Barboza AB, et al. Bipolar disorder with binge eating behavior: a genome-wide association study implicates PRR5-ARHGAP8. *Transl Psychiatry*. 2018;8(1):40.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945–59.
- Porrás-Hurtado L, Ruiz Y, Santos C, et al. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front Genet*. 2013;4:98.
- Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955–9.
- Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39(10):1181–6.
- Nurnberger JI Jr, Blehar MC, Kaufmann CA, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry*. 1994;51(11):849–59 discussion 63–4.
- Winham SJ, Cuellar-Barboza AB, Oliveros A, et al. Genome-wide association study of bipolar disorder accounting for effect of body mass index identifies a new risk allele in TCF7L2. *Mol Psychiatry*. 2014;19(9):1010–6.
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Human Genet*. 2007;81(3):559–75.
- Clayton D. Testing for association on the X chromosome. *Biostatistics (Oxford, England)*. 2008;9(4):593–600.
- Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol Sex Differ*. 2015;6: 35.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190–1.
- Jancic D, Seifuddin F, Zandi PP, et al. Association study of X chromosome SNPs in attempted suicide. *Psychiatry Res*. 2012;200(2–3):1044–6.
- Yue W, Cheng W, Liu Z, et al. Genome-wide DNA methylation analysis in obsessive-compulsive disorder patients. *Sci Rep*. 2016;6:31333.
- GTEX Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, NY)*. 2015;348(6235):648–60.
- O’Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733–45.
- Wang P, Xu S-Q, Wang B-Q, et al. A robust and powerful test for case-control genetic association study on X chromosome. *Stat Methods Med Res*. 2019;28(10–11):3260–72.
- Chen B, Craiu R, Sun L. Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *Biostatistics*. 2018. <https://www.ncbi.nlm.nih.gov/pubmed/30247537>.

# Publisher’s Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

