Published in final edited form as:

Genes Genomics. 2015 December 1; 37(12): 1007–1016. doi:10.1007/s13258-015-0332-z.

Sex difference of autosomal alleles in populations of European and African descent

Lingjun Zuo^{1,2}, Tong Wang³, Xiandong Lin⁴, Jijun Wang⁵, Yunlong Tan⁶, Xiaoping Wang^{7,*}, Xueqing Yu⁸, and Xingguang Luo^{1,2,6,*}

¹Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

²VA Connecticut Healthcare System, West Haven Campus, CT, USA

³Department of Cellular and Molecular Physiology, Yale School of Medicine, New Haven, CT, USA

⁴Provincial Key Laboratory of Translational Cancer Medicine, Fujian Provincial Cancer Hospital, Fuzhou, Fujian, China

⁵Department of EEG & Neuroimaging, Shanghai Mental Health Center, Shanghai, China

⁶Biological Psychiatry Research Center, Beijing Huilongguan Hospital, Beijing, China

⁷Department of Neurology, Shanghai First People's Hospital, Shanghai Jiao-Tong University, China

⁸Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

Abstract

In the present study, we aimed to report the individual sex-different genetic markers across autosomes in European- and African-origin populations. A total of 8,400 females and 8,081 males in 19 independent cohorts were genotyped across genomes using Illumina or Affymetrix arrays. The allele frequencies were compared between females and males in 9 non-clean cohorts (with some human disease traits) using genome-wide logistic regression and then the nominally significant associations were replicated across 10 clean cohorts (without disease traits). Meta-analysis was performed to derive the combined p values across all cohorts. We found 13 markers that were genome-wide significant ($p < 5 \times 10^{-8}$) between females and males in the meta-analysis of all cohorts of European descent, including rs7740449 at *SYNE1*, rs7531151 at *PLD5*, rs697455 at *PPP1R12B*, rs6745746 at *LOC100128413*, rs17000079 at *PARM1*, rs11948070 at *PDE4D*, rs7801825 at *INSIG1*, rs9551642 at *MTUS2*, rs2932174 at *TPTE2*, rs1961597 at *SALL3*, rs4117529 at *METTL4*, rs6021473 at *SALL4* and rs6092466 at *RAE1*, and one marker, i.e., rs10145208 at *PCNX*, that was genome-wide significant in the meta-analysis of all cohorts of

*Corresponding authors and reprints: Xingguang Luo, PhD and Xiaoping Wang, MD; Yale University School of Medicine, West Haven, CT 06516. Xingguang.Luo@yale.edu and wangxp@ustc.edu; Tel: 203-932-5711 ext 3390; Fax: 203-937-4741.

Bioethical statement: All subjects gave written informed consent to participating in protocols approved by the relevant institutional review boards (IRBs). All subjects were de-identified in this study that was approved by Yale IRB.

Conflict of interest: The authors have declared that no competing interests exist.

African descent. The most robust finding was rs7740449 at *SYNE1*, next to *ESR1*. We conclude that there are many sex-different markers on autosomes. These markers may be informative in differentiating females and males.

Keywords

Sex-difference; autosome; allele; European; African; meta-analysis

Introduction

Growing numbers of studies report that outcomes in human health are sex-sensitive; that is, they are stronger for either males or females; others are sex-specific; that is, they apply only to males or to females; and others are opposite in males and females. These sex-differences have been widely demonstrated by human and animal pharmacokinetic studies (Zhu et al. 2007), pharmacogenetic studies (Wang and Huang 2007), behavioral studies (Keeley et al. 2013), and neuroimaging studies (Cahill 2014). Nevertheless, many gaps remain in our understanding of these sex-differences. It is hypothesized that genetic factors contribute to these biological sex-differences.

The most straightforward difference in genomes between females and males resides in the sex chromosomes; that is, males have Y chromosomes but females do not. The Y-specific genes on Y chromosome determine the male-specific characters. For example, the sex-determining region Y gene (*SRY*) in an ampliconic region on Y chromosome is a testis-determining gene. In addition to the Y-specific genes, there are also many genes on Y chromosome that are homogenous to their counterparts on X chromosome. For example, genes in the pseudoautosomal regions (PARs) on Y chromosome are identical (~100%) to the corresponding PARs on X chromosome; genes in the X-transposed regions (XTRs) on Yp11.2 is 99% identical to the genes on Xq21.3; and genes in the X-degenerate regions (XDRs) on Y are 60%-96% identical to their counterparts on X (reviewed by (Zuo et al. 2013b)). Although these genes are homogenous between X and Y chromosomes, their genotype frequency distributions are still significantly different between males and females. For example, *AMELY* in an XDR on Y chromosome has been used for gender determination of forensic DNA (Pouchkarev et al. 1998). These homogenous genes may be associated with the traits that exist in both sexes but are significantly different between females and males. For example, the *NLGN4Y* in an XDR on Y chromosome has been related to autism that has a higher prevalence (4:1) in males than females (Jamain et al. 2003).

Sex at birth depends on whether the sperm bears a Y chromosome. Whether a Y-bearing sperm fertilizes an egg depends on the development, motility and survival time of this sperm and more other complicated environmental factors. However, the development, motility and survival time of a sperm might not be necessary to be regulated only by the proteins encoded by genes on sex chromosomes, but possibly also by genes on autosomes. The genetic variants both on autosomes and sex chromosomes can influence natural fertility in humans (Kosova et al. 2010). Furthermore, some loci on sex chromosomes were transposed from autosomes during primate evolution (Saxena et al. 1996); X-autosome or Y-autosome

translocations may occur frequently during recombination (Pabst et al. 2002); and some genes on autosomes, e.g., estrogen receptor 1 gene (*ESR1*) on 6q25.1, are related to the function of sex-related hormones or intrauterine or postnatal hormonal influences. Most of the Sex Determining Region Y - Box (SOX) genes are located on autosomes, e.g., *SOX9* on 17q23. Therefore, theoretically, it is likely that some genes on autosomes might also be significantly different between two sexes. However, this is rarely investigated until recently (Boraska et al. 2012). In the present study, we scanned the whole autosomes with a large sample size, to report the individual sex-different genetic markers across autosomes.

Materials and Methods

Subjects

A total of 16,481 European- or African-origin subjects from 12 studies were included. Demographic data including dbGaP access numbers are presented in Table 1. They were selected because they were available to us for a previous study on neuropsychiatric diseases (Zuo et al. 2013a). These subjects were separated into 19 independent cohorts for analysis based on their ascertainment. Among them, 7 independent cohorts with European descent and 2 independent cohorts with African descent were non-clean, each with a diagnosis of one neuropsychiatric disease such as Parkinson's disease, schizophrenia, bipolar disorder, Alzheimer's disease or ALS; 8 independent cohorts with European descent and 2 independent cohorts with African descent were clean, without these diseases. Subjects of European descent included Caucasians, European-Americans, Italians and Canadians, and subjects with African descent were African-Americans. 8,400 subjects were females and 8,081 were males. All self-reported sex information was confirmed using genetic markers in *AMELX/Y* gene in an X-degenerate region (XDR) on X and Y chromosomes. All unmatched sexes were discarded. The age distributions generally matched between females and males within each cohort.

Genotyping

The subjects were genotyped using Illumina or Affymetrix microarrays with 250,000 to 1 million of SNPs. Illumina platforms included HumanCNV370v1 (for cohorts #2 and #12), HumanHap550v3.0 (for cohorts #1, #7, #8, #9 and #15), and Human_1M (for cohort #13). Affymetrix platforms included Mapping250K_Nsp/Sty (for cohorts #6 and #14) and AFFY_6.0 (1.8M) (for cohorts #3, #4, #5, #10, #11, #16, #17, #18 and #19) (Table 1).

Data cleaning

Before association analysis was conducted, we stringently cleaned the phenotype and genotype data within each cohort. Excluded were subjects with poor genotypic data and questionable diagnostic information, subjects with allele discordance, duplicated IDs, potential sample misidentification, sample relatedness, sample misspecification, gender anomalies, chromosome anomalies (such as aneuploidy and mosaic cell populations), missing race, non-European and non-African ethnicity, population group outliers, subjects with a mismatch between self-identified and genetically-inferred ethnicity, and subjects with a missing genotype call rate $\geq 2\%$ across all SNPs. Furthermore, we filtered out the monomorphic SNPs and the SNPs with allele discordance, chromosomal anomalies, batch

effect, Mendelian errors (in family samples), an overall missing genotype call rate 2%, and MAFs < 0.01. The SNPs that deviated from Hardy-Weinberg Equilibrium (HWE) ($p < 10^{-4}$) both in males and females within each cohort were excluded. Finally, the autosomal SNPs that map to multiple genomic regions were also excluded, e.g., rs55755291 on chromosomes 1, 2, 3, and X. Many technical issues such as strand flipping have also been carefully solved before association analysis. The cleaned sample sizes and SNP sizes are listed in Table 1.

Genome-wide association analysis

The allele frequencies were compared between females and males in the target cohorts using logistic regression as implemented in the program PLINK (Purcell et al. 2007). Sex status served as the dependent variable, alleles served as the independent variables, and ancestry proportions served as the covariates. The ancestry proportions for each individual were estimated by integrating the ancestry information content of 6,436 independent ancestry-informative SNPs using the program STRUCTURE (Pritchard et al. 2000). Each of these 6,436 SNPs had significant difference in allele frequency between Europeans and Africans ($p < 10^{-8}$), and was in HWE ($p > 0.05$) in each cohort. They were not associated with any disease examined ($p > 0.05$) and were not correlated ($D' = 0$). The selection procedure for these 6,436 SNPs was detailed elsewhere (Zuo et al. 2012). Ancestry proportions as covariates were included in the regression model to control for population stratification and admixture effects.

Study design and data analytic strategy

(1) Subjects with European descent and African descent are heterogeneous. Allele frequency distributions were quite different between these two populations. Thus, all association analysis and meta-analysis were performed within the same ethnicity, i.e., separately in Europeans and Africans. (2) The three Illumina platforms are different in SNP density but completely overlap. After imputation using the same reference panels (i.e., 1,000 Genome project and HapMap 3), these three platforms were highly consistent in marker sets and thus were analyzed together in meta-analysis. The two Affymetrix platforms are different in SNP density and partially overlap. After imputation, the marker set consistency increased but was still not high (<60%), thus these two platforms were separately meta-analyzed first, and then jointly meta-analyzed to detect more informative markers. (3) We performed genome-wide association analysis on sex difference in each non-clean cohort first. Next, we extracted those replicable associations that were nominally significant ($p < 0.05$) across all non-clean cohorts. We meta-analyzed these replicable associations to drive the first global p_1 value and only the associations with global $p_1 < 10^{-3}$ in meta-analysis were retained for further analysis. Then, to extract those associations that were independent of the disease status, we replicated these replicable associations (with $p_1 < 10^{-3}$) in the clean cohorts. Only those associations nominally significant ($p < 0.05$) across all clean cohorts were retained. Finally, these replicated associations were meta-analyzed across all non-clean and clean cohorts to derive the second p_2 values. Only those associations with $p_2 < 5 \times 10^{-8}$ were treated as genome-wide significant ones and presented in Tables.

Meta-analysis

Meta-analysis was performed to derive the combined p values across cohorts using the program METAL (Willer et al. 2010). An overall z-statistic and a p-value for each SNP were calculated from a weighted sum of the individual statistics. Weights were proportional to the square-root of the number of individuals examined in each sample and selected such that the squared weights sum to 1.0. Meta-analysis was conducted across non-clean cohorts to derive the first p_1 value, then across all cohorts including non-clean and clean ones to derive the second p_2 value. The genome-wide Bonferroni-corrected α for p_2 was set at 5×10^{-8} (by one million markers).

Prediction of sex by sex-different markers

The cumulative Positive Predictive Values (PV^+) and the cumulative Positive Likelihood Ratios (LR^+) of the significant sex-different alleles were calculated using Bayesian formula.

That is, $PV^+ = P[T | \bigcap_{i=1}^n A_i]$ and $LR^+ = \frac{P[\bigcap_{i=1}^n A_i | T]}{P[\bigcap_{i=1}^n A_i | \bar{T}]}$, where P is probability; A_i is the i^{th} independent contributory allele; T is the trait, i.e., female or male; is male or female; n is the number of independent alleles; and $\cap$ denotes the Intersection approach.

Results

We checked and confirmed the quality of the association analysis results first. (1) QQ plots showed that the population stratification effects in the association findings in all cohorts had already been successfully controlled ($|\lambda_s| < 1.1$). (2) The allele frequencies were highly consistent across different cohorts within the same sex and same ethnicity. (3) The association patterns and Hardy-Weinberg Equilibrium (HWE) patterns were highly consistent across different cohorts within the same sex and same ethnicity. (4) The raw intensity data for the significant sex-different markers have been checked, which confirmed that there were no significant genotyping or calling errors. (5) Finally, the real sex-allele associations and the genotyping errors that caused Hardy-Weinberg Disequilibrium (HWD) in these sex-different markers have been carefully differentiated by multiple approaches including a model-fitting method implemented in the Mathematica Notebooks written by Wittke-Thompson et al. (2005) (Wittke-Thompson et al. 2005), which showed that genotyping errors were unlikely (see Supplementary Discussion). These pieces of evidence demonstrated that our data had high quality for association analysis. However, due to the heterogeneity of genotyping platforms, not every genetic marker was fully genotyped across all cohorts even after imputation, which is a limitation of the present study.

After quality check, we found some significant sex-allele associations. The frequencies of minor allele T of rs7740449 at *SYNE1* (Spectrin Repeat Containing, Nuclear Envelope 1 gene) on Chr6 were 0.305–0.321 in female non-clean subjects and 0.406–0.522 in male non-clean subjects in three independent cohorts of European descent with Parkinson disease or amyotrophic lateral sclerosis (ALS). They all were nominally significantly different between females and males in these three cohorts [odd ratio (OR): 0.43–0.64; 1.1×10^{-5} p 2.7×10^{-4}]. The genotype frequency distributions were in significant Hardy-

Weinberg Disequilibrium (HWD) in females in all of the three cohorts (2.5×10^{-9} p 0.015), but in Hardy-Weinberg Equilibrium (HWE) in males ($p > 0.05$). Meta-analysis of these three cohorts derived a combined $p_1 = 2.9 \times 10^{-10}$ ($Z = -7.06$). Furthermore, the allele frequencies of this allele in other five independent clean cohorts of European descent [minor allele frequencies (MAFs) = 0.289–0.342 in females and 0.357–0.471 in males] were similar to the above three non-clean cohorts. They were also nominally significantly different between females and males (3.8×10^{-6} p 0.029) in these five clean cohorts (OR: 0.58–0.80). The genotype frequency distributions were in significant HWD in females in four cohorts (4.6×10^{-13} p 1.3×10^{-3}) and in suggestive HWD in females in the 5th cohort ($p = 0.057$), but in HWE in males in all cohorts ($p > 0.05$). Meta-analysis of all of these eight European cohorts derived a combined $p_2 = 8.2 \times 10^{-22}$ ($Z = -9.60$) (Table 2). λ s from QQ plots for these eight cohorts were 0.99–1.07 (plots not shown). Other 11 cohorts we analyzed were not genotyped on this SNP, so they were not included in the meta-analysis.

Rs10145208 at *PCNX* (Pecanex Homolog gene) on Chr14 was only genotyped in the four cohorts of African descent. The frequencies of minor allele T of rs10145208 were similar across these four cohorts in disregard of their disease status. They were 0.020–0.025 in females and 0.039–0.063 in males. They were nominally significantly different between females and males (OR: 0.31–0.41; 6.5×10^{-5} p 2.9×10^{-4}) in three cohorts. They were not significantly different in the fourth cohort (#17), probably because of the small sample size of males of this cohort. The genotype frequency distributions were in HWE in both females and males in all of the four cohorts ($p > 0.05$). Meta-analysis of these four cohorts derived a combined $p_2 = 7.5 \times 10^{-11}$ ($Z = -6.51$) (Table 3). λ s from QQ plots for these four cohorts were 1.04–1.09 (plots not shown).

We found 12 SNPs that were significantly different between females and males in the meta-analysis of two Canadian cohorts (#6 and #14). They were rs7531151 at *PLD5* (Phospholipase D Family, Member 5 gene) ($p = 3.2 \times 10^{-43}$) and rs697455 at *PPP1R12B* (Protein Phosphatase 1, Regulatory Subunit 12B gene) ($p = 1.8 \times 10^{-8}$) on Chr1, rs6745746 at *LOC100128413* on Chr2 ($p = 4.0 \times 10^{-13}$), rs17000079 at *PARM1* (Prostate Androgen-Regulated Mucin-Like Protein 1 gene) on Chr4 ($p = 6.9 \times 10^{-13}$), rs11948070 at *PDE4D* (CAMP-Specific Phosphodiesterase 4D gene) on Chr5 ($p = 4.5 \times 10^{-15}$), rs7801825 at *INSIG1* (Insulin Induced Gene 1 gene) on Chr7 ($p = 3.7 \times 10^{-9}$), rs9551642 at *MTUS2* (Microtubule Associated Tumor Suppressor Candidate 2 gene) ($p = 4.6 \times 10^{-9}$) and rs2932174 at *TPTE2* (Transmembrane Phosphoinositide 3-Phosphatase And Tensin Homolog 2 gene) ($p = 4.4 \times 10^{-9}$) on Chr13, rs1961597 at *SALL3* (Spalt-Like Transcription Factor 3 gene) ($p = 6.8 \times 10^{-10}$) and rs4117529 at *METTL4* (Methyltransferase Like 4 gene) ($p = 1.5 \times 10^{-8}$) on Chr18, rs6021473 at *SALL4* (Spalt-Like Transcription Factor 4 gene) ($p = 1.3 \times 10^{-17}$) and rs6092466 at *RAE1* (Ribonucleic Acid Export 1 gene) ($p = 6.3 \times 10^{-9}$) on Chr20. The minor allele frequencies of these markers were similar between the two cohorts, regardless of disease status. They all were significantly different between females and males (4.2×10^{-23} p 2.5×10^{-3}) in these two cohorts. The genotype frequency distributions of rs7531151, rs6745746, rs17000079 and rs9551642 were in significant HWD in females in the two cohorts (5.8×10^{-18} p 5.5×10^{-3}) but in HWE in males; the other markers were all in HWE in both females and males ($p > 0.05$). The frequencies of the minor allele A of

rs7531151 were equal or close to 0 in males, but larger than 10% in females; it was a female-specific allele (Table 4). Both λ s from QQ plots were 1.02 for these two cohorts (Figure 1). These markers were not genotyped in the other 17 cohorts.

The cumulative Positive Predictive Values (PV^+) of alleles T-A-G-G-G-A-A-C-A-A-G of rs7740449-rs697455-rs6745746-rs17000079-rs11948070-rs7801825-rs9551642-rs2932174-rs1961597-rs4117529-rs6021473-rs6092466, which were dominant in European males at genome-wide significance level (Tables 2 and 4), was 0.983, with a cumulative Positive Likelihood Ratio (LR^+) of 57.36. The cumulative PV^+ of alleles C-G-A-A-C-G-T-C-G-G-G-C of these 12 markers, which were dominant in European females, was 0.999, with a cumulative LR^+ of 2939.50. The PV^+ of allele A of rs7531151, which was dominant in European females, was close to 1.000. The PV^+ of allele T of rs10145208, which was dominant in African males at genome-wide significance level (Table 3), was 0.733, with an LR^+ of 2.75.

Discussion

In the present study, we found 14 markers that were significantly different between females and males at genome-wide significance level in meta-analysis; among them, 13 were genome-wide significant in Europeans and one in Africans. All of these markers were not significantly associated with any disease examined. The allele frequencies and the association findings for these markers were replicable across at least two independent cohorts. The most robust finding was rs7740449 at *SYNE1* on Chr6, next to *ESR1*. The allele frequencies of this marker were highly consistent across eight independent cohorts of European descent genotyped on different platforms. Its associations with sex were highly replicable across eight cohorts with highly similar OR values and the same OR direction. The genotype distributions of this marker were in significant HWD in females across all eight cohorts, which made the HWD impossible to be attributed to systematic genotyping errors or chance (see Supplementary Discussion). The second most robust finding was rs10145208 at *PCNX* on Chr14 in Africans because of the similar reasons to rs7740449. The third most robust findings were those 12 SNPs in Canadian cohorts; among which, allele A of rs7531151 at *PLD5* was a female-specific allele. An individual of European descent carrying alleles T-A-G-G-G-A-A-C-A-A-G of rs7740449-rs697455-rs6745746-rs17000079-rs11948070-rs7801825-rs9551642-rs2932174-rs1961597-rs4117529-rs6021473-rs6092466 had 98.3% of probability to be male but only 1.7% to be female. An individual of European descent carrying alleles C-G-A-A-C-G-T-C-G-G-G-C of these 12 markers had 99.9% of probability to be female but only 0.1% to be male. An individual of European descent carrying allele A of rs7531151 had approximately 100% of probability to be female. An individual of African descent carrying allele T of rs10145208 had 73.3% of probability to be male but only 26.7% to be female. These markers may provide helpful informative content for differentiating females and males in the presence of DNA in the specimen as the Y-specific signal does. They may also provide a direction for the investigation on sex-different traits. For example, alcoholism and autism are sex-different diseases in prevalence; they might be associated with these sex-specific markers, which is worthy of future investigation.

The frequency of major allele C of rs7740449 at *SYNE1* in females was about 0.05–0.20 higher than males of European descent. This marker was in significant HWD in females in all cohorts examined. The minor allele frequency of rs10145208 at *PCNX* in males was about 1.6–3.2 times of that in females of African descent. The minor allele frequencies of rs697455 at *PPP1R12B*, rs6745746 at *LOC100128413*, rs17000079 at *PARM1*, rs11948070 at *PDE4D*, rs9551642 at *MTUS2*, and rs2932174 at *TPTE2*, rs1961597 at *SALL3* and rs6092466 at *RAE1* in Canadian females were about 1.7–7.6 times of those in Canadian males. The minor allele frequencies of rs7801825 at *INSIG1*, rs4117529 at *METTL4* and rs6021473 at *SALL4* in Canadian males were about 2.0–3.2 times of those in Canadian females. The minor allele A of rs7531151 at *PLD5* was female-specific in Canadians. Among these markers, rs6745746 at *LOC100128413*, rs17000079 at *PARM1*, rs9551642 at *MTUS2*, and rs7531151 at *PLD5* were in significant HWD in females in both Canadian cohorts. A female-associated allele of a SNP in females should be present at a higher rate than the general population (a natural selection), which may violate HWE of that SNP in females, as observed for the major allele C of rs7740449 and the minor alleles of several other markers. Thus, HWD of a SNP in females could be an indicator of an allele-female association and a possible recessive inheritance mode (Luo et al. 2008) (see Supplementary Discussion). The female-specific alleles of these 5 SNPs may contribute to female-specific traits. Other 9 SNPs were in HWE in both females and males. Their alleles may contribute to sex-different traits. Six of the above protein-coding genes including *SYNE1*, *PPP1R12B*, *PDE4D*, *TPTE2*, *RAE1* and *PLD5* are related to phosphorylation and two genes including *LOC100128413* and *PARM1* are related to apoptosis. Both protein phosphorylation and apoptosis are important physiological processes involving in the development, motility and survival time of X-bearing and Y-bearing sperm (Chan et al. 2009; Navara 2010). They have been reported to contribute to the gender-difference of many traits (Lagranha et al. 2010; Siegel et al. 2011). Furthermore, one gene (*PARM1*) is androgen-related gene; one gene (*PDE4D*) is next to the androgen-related gene *PART1*; and other two genes (*SYNE1* and *TPTE2*) are next to estrogen-related genes, consistent with their roles in sex-difference. Finally, the major limitation of this study is that many markers were not consistently genotyped across all cohorts, which might result in information loss. More sex-different markers are anticipated to be identified if using denser and more consistent marker sets.

Identifying and understanding sex differences can improve our understanding of the nature and etiology of human diseases and have implications for tailoring prevention and treatment interventions to maximize outcomes for both males and females. Despite some important progresses have been made in laboratory sex-based research and the present study also contributes to these progresses, it is in an early stage in this field. More potential biological mechanisms leading to sex-difference of markers on autosomes and the biological implications of these markers in human health are warranted by future investigations.

We conclude that there are many sex-different markers on autosomes. These markers may provide an alternative approach to differentiate females and males in the absence of sex chromosomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported in part by National Institute on Drug Abuse (NIDA) grant K01 DA029643, National Institute on Alcohol Abuse and Alcoholism (NIAAA) grants R21 AA021380, R21 AA020319 and R21 AA023237, the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award 17616 (L.Z.) and ABMRF/The Foundation for Alcohol Research (L.Z.). We thank NIH GWAS Data Repository, the Contributing Investigator(s) who contributed the phenotype and genotype data from his/her original study (e.g., Drs. Bierut, Edenberg, Heath, Singleton, Hardy, Foroud, Myers, Gejman, Faraone, Sonuga-Barke, Sullivan, Nurnberger, Devlin, Monaco, etc.), and the primary funding organization that supported the contributing study. Funding and other supports for phenotype and genotype data were provided through the National Institutes of Health (NIH) Genes, Environment and Health Initiative (GEI) (U01HG004422, U01HG004436 and U01HG004438); the GENEVA Coordinating Center (U01HG004446); the NIAAA (U10AA008401, R01AA013320, P60AA011998); the NIDA (R01DA013423); the National Cancer Institute (P01 CA089392); the Division of Neuroscience, the NIA National Institute of Neurological Disorders and Stroke (NINDS); the NINDS Human Genetics Resource Center DNA and Cell Line Repository; the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C); the Center for Inherited Disease Research (CIDR); a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and Prevention; the NIH Office of Research on Women's Health (ORWH) (R01NS45012); the Department of Veterans Affairs; the University of Maryland General Clinical Research Center (M01RR165001), the National Center for Research Resources, NIH; the National Institute of Mental Health (R01MH059160, R01MH59565, R01MH59566, R01MH59571, R01MH59586, R01MH59587, R01MH59588, R01MH60870, R01MH60879, R01MH61675, R01MH62873, R01MH081803, R01MH67257, R01MH81800, U01MH46276, U01MH46282, U01MH46289, U01MH46318, U01MH79469, U01MH79470 and R01MH67257); the NIMH Genetics Initiative for Bipolar Disorder; the Genetic Association Information Network (GAIN); the Genetic Consortium for Late Onset Alzheimer's Disease; the Autism Genome Project, the MARC: Risk Mechanisms in Alcoholism and Comorbidity; the Molecular Genetics of Schizophrenia Collaboration; the Medical Research Council (G0601030) and the Wellcome Trust (075491/Z/04), University of Oxford; the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, NWO Genomics, SPI 56-464-1419) the Centre for Neurogenomics and Cognitive Research (CNCR-VU); Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR); and the European Union (EU/WLRT-2001-01254), ZonMW (geestkracht program, 10-000-1002). Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Genetic Consortium for Late Onset Alzheimer's Disease, the GENEVA Coordinating Center (U01 HG004446), and the National Center for Biotechnology Information. Genotyping was performed at the Johns Hopkins University Center for Inherited Disease Research, and GlaxoSmithKline, R&D Limited. The datasets used for the analyses described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap>. XL had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The financial sponsors had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

References

- Boraska V, Jeroncic A, Colonna V, Southam L, Nyholt DR, Rayner NW, Perry JR, Toniolo D, Albrecht E, Ang W, et al. Genome-wide meta-analysis of common variant differences between men and women. *Hum Mol Genet.* 2012; 21:4805–4815. [PubMed: 22843499]
- Cahill L. Fundamental sex difference in human brain architecture. *Proc Natl Acad Sci U S A.* 2014; 111:577–578. [PubMed: 24381154]
- Chan CC, Shui HA, Wu CH, Wang CY, Sun GH, Chen HM, Wu GJ. Motility and protein phosphorylation in healthy and asthenozoospermic sperm. *J Proteome Res.* 2009; 8:5382–5386. [PubMed: 19678645]
- Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet.* 2003; 34:27–29. [PubMed: 12669065]
- Keeley RJ, Tyndall AV, Scott GA, Saucier DM. Sex difference in cue strategy in a modified version of the Morris water task: correlations between brain and behaviour. *PLoS One.* 2013; 8:e69727. [PubMed: 23874990]

- Kosova G, Abney M, Ober C. Colloquium papers: Heritability of reproductive fitness traits in a human population. *Proc Natl Acad Sci U S A*. 2010; 107(Suppl 1):1772–1778. [PubMed: 19822755]
- Lagranha CJ, Deschamps A, Aponte A, Steenbergen C, Murphy E. Sex differences in the phosphorylation of mitochondrial proteins result in reduced production of reactive oxygen species and cardioprotection in females. *Circ Res*. 2010; 106:1681–1691. [PubMed: 20413785]
- Luo X, Zuo L, Kranzler HR, Wang S, Anton RF and Gelernter J. Recessive genetic mode of an ADH4 variant in substance dependence in African-Americans: A model of utility of the HWD test. *Behav Brain Funct*. 2008; 4:42. [PubMed: 18801187]
- Navara KJ. Programming of offspring sex ratios by maternal stress in humans: assessment of physiological mechanisms using a comparative approach. *J Comp Physiol B*. 2010; 180:785–796. [PubMed: 20544204]
- Pabst B, Glaubitz R, Schalk T, Schneider U, Schulze W, Miller K. Reciprocal translocation between Y chromosome long arm euchromatin and the short arm of chromosome 1. *Ann Genet*. 2002; 45:5–8. [PubMed: 11934382]
- Pouchkarev VP, Shved EF, Novikov PI. Sex determination of forensic samples by polymerase chain reaction of the amelogenin gene and analysis by capillary electrophoresis with polymer matrix. *Electrophoresis*. 1998; 19:76–79. [PubMed: 9511866]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
- Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, Rozen S, Dinulos MB, Disteché CM, et al. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet*. 1996; 14:292–299. [PubMed: 8896558]
- Siegel C, Li J, Liu F, Benashski SE, McCullough LD. miR-23a regulation of X-linked inhibitor of apoptosis (XIAP) contributes to sex differences in the response to cerebral ischemia. *Proc Natl Acad Sci U S A*. 2011; 108:11662–11667. [PubMed: 21709246]
- Wang J, Huang Y. Pharmacogenomics of sex difference in chemotherapeutic toxicity. *Curr Drug Discov Technol*. 2007; 4:59–68. [PubMed: 17630929]
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–2191. [PubMed: 20616382]
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet*. 2005; 76:967–986. [PubMed: 15834813]
- Zhu X, Lee DY, Shin WG. Gender difference in the pharmacokinetic interaction between oral warfarin and oxolamine in rats: inhibition of CYP2B1 by oxolamine in male rats. *Biopharm Drug Dispos*. 2007; 28:125–133. [PubMed: 17295362]
- Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, Malison RT, Li CR, Wang F, Zhang XY, et al. Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacology*. 2012; 37:557–566. [PubMed: 21956439]
- Zuo L, Wang K, Zhang X JHK, Li CR, Zhang F, Zhang H, Luo X. NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk region specific for alcohol dependence in subjects of European descent. *Drug Alcohol Depend*. 2013a; 129:254–264. [PubMed: 23455491]
- Zuo L, Wang K, Zhang X, Pan X, Wang G, Krystal JH, Zhang H and Luo X. Sex chromosome-wide association analysis suggested male-specific risk genes for alcohol dependence. *Psychiatr Genet*. 2013b; 23:233–238. [PubMed: 23907288]

Figure 1. QQ plots of p-values
[left: in Canadian without diseases; right: in Canadians with Alzheimer disease]

Table 1

Demographic data of all cohorts

#	Dataset name	Ethnicity	Disease status	Genotyping platform	dbGaP #	SNP Size	Sample Size		Age (years)	
							Female	Male	Female	Male
1	Ing_coriell_pd	Caucasian	PD	ILLUMINA HumanHap550v3.0	phs000089.v3.p2	542,325	465	336	56.0±17.2	62.3±14.4
2	PDRD+GRU	Caucasian	PD	ILLUMINA HumanCNV370v1	phs000126.v1.p1	338,807	363	537	70.2±10.0	70.5±9.3
3	nonGAIN	European-American	Schz	AFFYMETRIX AFFY_6.0	phs000021.v3.p2	898,519	404	947	45.1±11.2	42.5±11.3
4	BDO+GRU	European-American	BP	AFFYMETRIX AFFY_6.0	phs000017.v3.p1	729,450	178	190	45.4±10.0	43.1±8.0
5	BARD+GRU	European-American	BP	AFFYMETRIX AFFY_6.0	phs000017.v3.p1	729,455	331	322	44.4±9.7	42.1±8.3
6	GenADA	Canadian	AD	AFFYMETRIX Mapping250K_Nsp/Sty	phs000219.v1.p1	483,727	466	340	78.3±8.8	77.6±8.3
7	Italian	Italian	ALS	ILLUMINA HumanHap550v3.0	phs000101.v3.p1	545,366	123	138	60.7±8.5	60.3±8.1
8	US_Control	European-American	Clean	ILLUMINA HumanHap550v3.0	phs000101.v3.p1	553,077	323	204	69.8±8.9	69.6±8.6
9	Ing_coriell_pd	Caucasian	Clean	ILLUMINA HumanHap550v3.0	phs000089.v3.p2	540,204	465	336	56.0±17.2	62.3±14.4
10	nonGAIN	European-American	Clean	AFFYMETRIX AFFY_6.0	phs000021.v3.p2	890,997	744	634	49.2±16.7	49.2±16.7
11	BDO+GRU	European-American	Clean	AFFYMETRIX AFFY_6.0	phs000017.v3.p1	729,455	502	532	50.1±17.6	54.7±17.3
12	PDRD+GRU	Caucasian	Clean	ILLUMINA HumanCNV370v1	phs000126.v1.p1	338,587	521	346	55.2±11.2	53.5±15.7
13	AGP	European-American	Clean	ILLUMINA_Human_IM	phs000267.v1.p1	842,645	1360	1385	37.1±8.0	35.2±8.2
14	GenADA	Canadian	Clean	AFFYMETRIX Mapping250K_Nsp/Sty	phs000219.v1.p1	483,228	503	279	72.8±8.1	74.4±7.7
15	Italian	Italian	Clean	ILLUMINA HumanHap550v3.0	phs000101.v3.p1	546,695	110	136	66.8±8.7	67.0±8.4
16	nonGAIN	African-American	Schz	AFFYMETRIX AFFY_6.0	phs000021.v3.p2	900,220	449	746	43.0±9.8	41.9±10.8
17	BARD+GRU	African-American	BP	AFFYMETRIX AFFY_6.0	phs000017.v3.p1	845,571	102	39	42.0±7.8	42.4±7.9
18	nonGAIN	African-American	Clean	AFFYMETRIX AFFY_6.0	phs000021.v3.p2	858,472	592	362	45.0±12.9	46.2±13.7
19	BARD+GRU	African-American	Clean	AFFYMETRIX AFFY_6.0	phs000017.v3.p1	845,815	399	272	45.7±13.5	46.0±14.0

PD, Parkinson's disease. Schz, Schizophrenia. BP, Bipolar disorder. AD, Alzheimer's disease. ALS, Amyotrophic Lateral Sclerosis. Clean, without diseases.

Table 2

Sex difference of rs7740449 at SYNE1 intron on Chr6

Cohort #	Ethnicity	Disease status	Frequency (Count) of Minor Allele T				OR/Z	P	HWE
			Female	Male	N				
1	Caucasian	Parkinson disease	0.305 (148/486)	0.406 (325/800)	643		0.64	2.7×10^{-4}	2.5×10^{-9} in females
2	Caucasian	Parkinson disease	0.309 (205/664)	0.414 (444/1072)	868		0.63	1.1×10^{-5}	0.015 in females
7	Italian	ALS	0.321 (59/184)	0.522 (144/278)	231		0.43	2.4×10^{-5}	2.7×10^{-4} in females
Meta	White	Clean					-7.06	2.9×10^{-10}	
8	European-American	Clean	0.289 (155/536)	0.357 (145/406)	471		0.73	0.029	2.9×10^{-10} in females
9	Caucasian	Clean	0.292 (172/590)	0.372 (169/454)	522		0.69	6.4×10^{-3}	4.6×10^{-13} in females
12	Caucasian	Clean	0.320 (302/944)	0.432 (299/692)	818		0.62	3.8×10^{-6}	0.057 in females
13	European-American	Clean	0.342 (859/2512)	0.394 (1092/2770)	2641		0.80	9.2×10^{-5}	4.4×10^{-5} in females
15	Italian	Clean	0.341 (58/170)	0.471 (128/272)	221		0.58	7.6×10^{-3}	1.3×10^{-3} in females
Meta	White	Clean+non-clean					-9.60	8.2×10^{-22}	

N, sample size. OR/Z, odd ratios or z-scores (meta). Non-clean, with one disease. HWE, Hardy-Weinberg Equilibrium.

Table 3

Sex difference of rs10145208 at 3' to PCNX on Chr14

Cohort #	Ethnicity	Disease status	Frequency (Count) of Minor Allele T				OR/Z	P
			Female	Male	N			
16	African-American	Schizophrenia	0.021 (19/896)	0.050 (75/1490)	1193		0.41	2.9×10^{-4}
17	African-American	Bipolar disorder	0.025 (5/202)	0.039 (3/76)	139		0.62	0.688
18	African-American	Clean	0.020 (24/1180)	0.055 (40/724)	952		0.36	6.5×10^{-5}
19	African-American	Clean	0.020 (16/794)	0.063 (34/544)	669		0.31	9.6×10^{-5}
Meta	African-American	Clean+non-clean					-6.51	7.5×10^{-11}

Table 4

Sex difference in Canadian cohorts

Cohort #		#6										#14				Meta	
Disease status (sample size)		Alzheimer's disease (N=806)										Clean (N=782)				Non-clean+Clean	
		Frequency (Count)										Frequency (Count)					
SNP	CHR	Gene	Minor Allele	Female	Male	OR	P	HWE (F)	Female	Male	OR	P	HWE (F)	Z	P		
rs7531151	1	PLD5 ^{intron}	A	0.116 (108/932)	0.004 (3/680)	29.58	4.2×10 ⁻²³	1.2×10 ⁻³	0.105 (106/1006)	0.000 (0/558)	∞	8.8×10 ⁻²²	3.4×10 ⁻³	13.78	3.2×10 ⁻⁴³		
rs697455	1	PPP1R12B ^{intron}	G	0.039 (36/932)	0.009 (6/680)	4.51	1.1×10 ⁻⁴		0.038 (38/1006)	0.005 (3/558)	7.26	4.1×10 ⁻⁵		-5.63	1.8×10 ⁻⁸		
rs6745746	2	5' to LOC100128413	A	0.231 (215/932)	0.146 (99/680)	1.76	1.2×10 ⁻⁶	9.0×10 ⁻¹⁵	0.244 (245/1006)	0.143 (80/558)	1.92	6.6×10 ⁻⁸	5.8×10 ⁻¹⁸	-7.25	4.0×10 ⁻¹³		
rs17000079	4	PARM1 ^{intron}	A	0.104 (97/932)	0.041 (28/680)	2.71	2.4×10 ⁻⁶	5.5×10 ⁻³	0.122 (123/1006)	0.039 (22/558)	3.39	5.2×10 ⁻⁸	2.8×10 ⁻⁴	-7.18	6.9×10 ⁻¹³		
rs11948070	5	PDE4D ^{intron}	C	0.184 (171/932)	0.069 (47/680)	3.03	1.0×10 ⁻¹¹		0.133 (134/1006)	0.065 (36/558)	2.23	2.0×10 ⁻⁵		7.84	4.5×10 ⁻¹⁵		
rs7801825	7	5' to INSIG1	A	0.020 (19/932)	0.057 (39/680)	0.34	1.1×10 ⁻⁴		0.022 (22/1006)	0.070 (39/558)	0.30	7.2×10 ⁻⁶		-5.90	3.7×10 ⁻⁹		
rs9551642	13	MTUS2 ^{intron}	T	0.147 (137/932)	0.069 (47/680)	2.32	7.5×10 ⁻⁷	1.0×10 ⁻⁵	0.141 (142/1006)	0.084 (47/558)	1.79	8.7×10 ⁻⁴	1.2×10 ⁻⁵	-5.86	4.6×10 ⁻⁹		
rs2932174	13	TPTE2 ^{intron}	C	0.161 (150/932)	0.079 (54/680)	2.22	1.0×10 ⁻⁶		0.136 (137/1006)	0.079 (44/558)	1.84	6.8×10 ⁻⁴		-5.87	4.4×10 ⁻⁹		
rs1961597	18	5' to SALL3	G	0.059 (55/932)	0.021 (14/680)	2.98	1.5×10 ⁻⁴		0.089 (90/1006)	0.027 (19/558)	3.56	7.8×10 ⁻⁷		-6.17	6.8×10 ⁻¹⁰		
rs4117529	18	METTL4 ^{intron}	A	0.041 (38/932)	0.087 (59/680)	0.45	1.8×10 ⁻⁴		0.041 (41/1006)	0.097 (54/558)	0.40	2.0×10 ⁻⁵		-5.66	1.5×10 ⁻⁸		
rs6021473	20	SALL4 ^{intron}	A	0.097 (90/932)	0.212 (144/680)	0.40	2.5×10 ⁻¹¹		0.097 (98/1006)	0.192 (107/558)	0.45	6.4×10 ⁻⁸		-8.55	1.3×10 ⁻¹⁷		
rs6092466	20	RAE1 ^{intron}	C	0.077 (72/932)	0.021 (14/680)	3.98	2.3×10 ⁻⁷		0.059 (59/1006)	0.026 (14/558)	2.42	2.5×10 ⁻³		5.81	6.3×10 ⁻⁹		

HWE (F), Hardy-Weinberg Equilibrium in females.