Sex chromosome-wide association analysis suggested male-specific risk genes for alcohol dependence
Lingjun Zuo\textsuperscript{a}, Kesheng Wang\textsuperscript{e}, Xiangyang Zhang\textsuperscript{f}, Xinghua Pan\textsuperscript{b}, Guilin Wang\textsuperscript{c}, John H. Krystal\textsuperscript{a}, Heping Zhang\textsuperscript{d} and Xingguang Luo\textsuperscript{a}

Introduction
Studies in many cultures have found that the prevalence of alcohol dependence and heavy drinking generally is higher among men than among women (Heath \textit{et al.}, 1998). Potential explanations for this include the role of genes in sex chromosomes (X and Y). In the present study, we scanned the entire Y chromosome and its homologs on the X chromosome in men to identify male-specific risk genes for alcohol dependence.

The Y chromosome comprises two pseudoautosomal regions (PARs) at both ends (5%) and a male-specific region (MSY) in between (95%). Sequences in PARs are the only portions on the Y chromosome that are identical (\textasciitilde 100\%) to the corresponding PARs on the X chromosome, reflecting the exchange of DNA between X and Y (crossover) that occurs during sperm production. The PAR1 at Yp11.32/Xp22.33 (2.7 Mb) is the result of an obligatory recombination in male meiosis; gene loci in the PAR1 are present in two copies in both men and women and are not subject to dosage compensation by X chromosome inactivation (Helena Mangs and Morris, 2007). These genes in the PAR1 include \textit{ASMT}, \textit{ASMTL}, \textit{CD99}, \textit{CRLF2}, \textit{CSF2RA}, \textit{SFRS17A}, \textit{DHRS}, \textit{GTPBP6}, \textit{IL3RA}, \textit{P2RY8}, \textit{PLCD1}, \textit{PPP2R3B}, \textit{SHOX}, \textit{SLC25A6}, \textit{XG}, and \textit{ZBED1}. The PAR2 at Yq12/Xq28 (330 kb) was created by duplication of material from X to Y since the divergence of human and chimpanzee lineages about 6 million years ago. The genes in the PAR2 include \textit{SPRY3}, \textit{VAMP7}, \textit{SYBL1}, and \textit{IL9R}, some of which are subject to X chromosome inactivation (Helena Mangs and Morris, 2007). MSY consists of heterochromatin and euchromatin. Heterochromatin hugs the centromere and the Yq12

\textit{Background} Alcohol dependence is more common among men than among women. Potential explanations for this include the role of genes in sex chromosomes (X and Y). In the present study, we scanned the entire Y chromosome and its homologs on the X chromosome in men to identify male-specific risk genes for alcohol dependence.

\textit{Methods} Two thousand nine hundred and twenty-seven individuals in two independent cohorts were analyzed. The European-American male cohort (883 cases with alcohol dependence and 445 controls) served as the discovery cohort and the European-American female cohort (526 cases and 1073 controls) served as a contrast group. All individuals were genotyped on the Illumina Human 1M beadchip. Two thousand two hundred and twenty-four single nucleotide polymorphisms (SNPs) on the Y chromosome or in the homologs on the X chromosome were analyzed. The allele frequencies were compared between cases and controls within each cohort using logistic regression analysis.

\textit{Results} We found that, after experiment-wide correction, two SNPs on the X chromosome were associated significantly with alcohol dependence in European-American men ($P=1.0 \times 10^{-4}$ for rs5916144 and $P=5.5 \times 10^{-5}$ for rs5961794 at 3' UTR of \textit{NLGN4X}), but not in the women. A total of 26 SNPs at 3' UTR of or within \textit{NLGN4X} were nominally associated with alcohol dependence in men ($5.5 \times 10^{-5} \leq P \leq 0.05$), all of which were not statistically significant in women.

\textit{Conclusion} We conclude that \textit{NLGN4X} was a significant male-specific risk gene for alcohol dependence in European-Americans. \textit{NLGN4X} might harbor a causal variant(s) for alcohol dependence. A defect of synaptogenesis in neuronal circuitry caused by \textit{NLGN4X} mutations is believed to play a role in alcohol dependence. \textit{Psychiatr Genet} 23:233–238 \textcopyright 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins.

\textbf{Keywords:} alcohol dependence, homolog, male specificity, \textit{NLGN4X}, synaptogenesis, Y chromosome

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
region (without PAR2) at the telomere, which are tightly coiled and enriched with repetitive DNA sequences. Heterochromatin maintains the chromosome’s structural integrity. It is interspersed with euchromatin along the rest of MSY. All active genes are located in the euchromatin. Euchromatin is classified to an X-transposed region (XTR), an X-degenerate region (XDR), an ampliconic region, and others. (a) The XTR (3.38 Mb) is located at Yp11.2, next to the PAR1. Its sequences are 99% identical to DNA sequences in Xq21.3 (3.91 Mb), a band in the midst of the Xq chromosome. This XTR was transposed from X to Y about 6 million years ago, after the divergence of the human and chimpanzee lineages. It escapes X chromosome inactivation and does not require dosage compensation. This XTR contains two known pairs of functional genes, that is \textit{PCDH11X} and \textit{PCDH11Y}, a pair of protocadherin genes expressed significantly in the adult human brain, and \textit{TGIF2LX} and \textit{TGIF2LY}; a pair of transforming growth factor-beta-induced factor 2-like genes expressed exclusively in testes. This region is hominid specific and absent from the Y chromosome of nonhuman primates (Blanco et al., 2000; Williams et al., 2006; Wilson et al., 2006). It might thus play a role in hominid-specific characteristics including the cerebral asymmetry (brain lateralization), a basis of human-specific cognitive functions such as language (adaptive traits). (b) The X-degenerate segments on Y are not concentrated in one region. They are separated into 12 blocks by some single-copy genes or pseudogenes. However, the homologs of these 12 blocks on the X chromosome are concentrated in a 6 Mb continuous region around Xp22.32, next to the PAR1 (Xp22.33). XDRs show 60–96% of sequence identity between X and Y. They are surviving relics of ancient autosomes from which the X and Y chromosome coevolved. XDRs hug \textit{SOX3/SRY}, \textit{RPS4Y1/Y}, \textit{ZFX/Y}, \textit{AMELXY}, \textit{TBL1XY/Y}, \textit{PRKX/Y}, \textit{USP9XY/Y}, \textit{DBX/Y}, \textit{UTX/Y}, \textit{TMSB4XY/Y}, \textit{NLGN4XY/Y}, \textit{CXorf15/CXorf15A}, \textit{CXorf15/CXorf15B}, \textit{SMCX/Y}, \textit{NLGN4X/Y}, and \textit{RPS4XY/Y}. Most of these genes are ubiquitous expressed, but only \textit{NLGN4XY} are significantly expressed in the adult human brain (Lin et al., 2012). (c) The ampliconic regions are nonrecombinant and Y-specific, that is, they are single-copy and do not have homologous regions on X. The genes in these regions are expressed exclusively in testes. Deletions in these regions are a recognized cause of male infertility. That is, the PAR1, XTR, and XDR are more likely to be relevant to brain disorders.

In summary, a total of 54 functional homologous genes have been identified so far, including 24 in PAR1, five in PAR2, three in XTR, 15 in XDR, and seven in other regions. It is estimated that \textasciitilde{15} protein-coding genes on the Y chromosome are Y-specific and have no detectable X chromosome homolog. In the present study, all homologous genes on both X and Y chromosomes, in addition to other Y-specific genes on Y, were tested. These homologs are believed to balance their dosage between men and women. Therefore, as a contrast, these homologs were also tested in women to determine whether the risk genes among those homologs are male specific or shared by women. Because the phasing information of these homologs was not available in the microarray platforms, we focused on individual variants but not multilocus haplotypes in the present study.

Materials and methods

Participants
Two thousand nine hundred and twenty-seven individuals in two independent cohorts were analyzed. The European-American male cohort [883 cases with alcohol dependence \textit{Diagnostic and Statistical Manual of Mental Disorders}, 4th ed. (DSM-IV) and 445 controls] served as the discovery cohort and the European-American female cohort (526 cases and 1073 controls) served as a contrast group. In addition, two smaller African-American cohorts including a male cohort (428 cases and 169 controls) and a female cohort (253 cases and 339 controls) were also analyzed as contrast groups on an exploratory basis. Affected individuals fulfilled the lifetime DSM-IV criteria (American Psychiatric Association, 1994) for alcohol dependence. Controls were defined as individuals who had been exposed to alcohol (and possibly to other drugs), but had never become addicted to alcohol or other illicit substances (lifetime diagnoses). Controls were also screened to exclude individuals with major axis I disorders, including schizophrenia, mood disorders, and anxiety disorders. More detailed demographic data have been published previously (Bierut et al., 2010; Edenberg et al., 2010; Heath et al., 2011; Zuo et al., 2011a, 2011b, 2012, 2013a, 2013b). All participants were deidentified in this study. All participants were genotyped on the Illumina Human 1M beadchip (Illumina Inc., San Diego, California, USA).

Data cleaning
Individuals with poor genotypic data and questionable diagnostic information, those with allele discordance, duplicated IDs, potential sample misidentification, sample relatedness, sample misspecification, sex anomalies, chromosome anomalies (such as aneuploidy and mosaic cell populations), missing race, non-European and non-African ethnicity, population group outliers, individuals with a mismatch between self-identified and genetically inferred ethnicity, and those with a missing genotype call rate of at least 2% across all SNPs examined were excluded step by step (Zuo et al., 2011a). Furthermore, only the SNPs on the Y chromosome and the SNPs in the homologous genes on the X chromosome were extracted from the 1M beadchip for association analysis. Because some homologous genes, especially in XDR, were not completely identical in sequence between X and Y chromosomes, some of the SNPs in these genes appeared only on the X chromosome. Among these extracted SNPs, those with allele discordance or chromosomal anomalies were excluded. We then filtered out the SNPs with an
overall missing genotype call rate of at least 2% and the SNPs with minor allele frequencies less than 0.01 in either European-Americans or African-Americans. The SNPs on the X chromosome in women and the SNPs in the homologous regions in men that deviated from the Hardy–Weinberg equilibrium ($P < 10^{-4}$) within European-American or African-American controls were also excluded. This selection process yielded a total of 2224 SNPs for analysis, including 1481 in 17 homologous genes on the X chromosome only, 480 SNPs in 34 genes on the Y chromosome only, and 263 SNPs in the recombinant regions (i.e. XY regions) of both X and Y chromosomes including 16 genes in the PAR1 (ASMT, ASMTL, CD99, CRLF2, CSF2RA, CXorf2, DHR5XY, DXYS155E, GTPBP6, IL3RA, P2RY8, PLCD1, PPP2R3B, SHOX, SL2C5A6, XG, and ZBED1), three genes in the PAR2 (IL9R, SPRY3, and SYBL1), two genes in the XTR (PCDH11X/Y and TGF21X/Y), 12 genes in XDR (AMELXY, DDX3XY, EFIA1XY, NLGN4XY, PRKXXY, RPS4X/Y, SMCYX, TBL1XY, TMBS4XY, USP9XY, UTX/Y, and ZFX/Y), and one gene in another region (NXY2).

**Association analysis**

The allele frequencies were compared between cases and controls within each cohort using logistic regression analysis implemented in the program PLINK (Purcell *et al.*, 2007). Diagnosis served as the dependent variable, alleles served as the independent variables, and age and the first 10 principal components served as the covariates. The principal component scores of our samples were derived from all autosomal SNPs across the genome using principal component analysis implemented in the software package EIGENSTRAT (Price *et al.*, 2006). Each individual received scores on each principal component. The first 10 principal component scores accounted for greater than 95% of variance. These principal components reflected the population structure of our samples. These principal components served as covariates in the regression model controlled for the population stratification and admixture effects on association analysis.

The experiment-wide significance levels ($z$) were corrected for the numbers of effective markers that were calculated using the program SNPSpD (Nyholt, 2004), which is an adjusted Bonferroni correction taking the linkage disequilibrium structure into account. Approximately 379 and 494 effective SNPs that had an $r^2$ of up to 0.1 between any two of them captured most of the information of all variants examined in cohorts of European and African descent, respectively. Thus, the corrected significance levels ($z$) for association tests were set at 1.3 x 10^{-4} in cohorts of European descent and 1.0 x 10^{-4} in cohorts of African descent, respectively.

**Power analysis**

We computed the power of test using the R package ‘pwr’, which uses the effect sizes and the notations from Cohen (1988). Given the effect size ($w$), the total number of observations ($n$), the degree of freedom ($df$), and the significance level ($z$), the statistical power of test can be determined from these parameters. The effect size is $w = \sum (O - E)^2 / E$, where $O$ is the observed allele frequency of the targeted SNP and $E$ is its expected allele frequency (Table 1). The $z$ was corrected for multiple testing as described above.

**Results**

We found that 79 (62 on X, four on Y, 13 on XY) and 132 (50 on X, 73 on Y, nine on XY) SNPs were nominally associated with alcohol dependence in European-American men and African-American men, respectively ($P < 0.05$), and 51 (37 on X and 14 on XY) and 69 (62 on X and seven on XY) SNPs were nominally associated with alcohol dependence in European-American women and African-American women, respectively ($P < 0.05$). After experiment-wide correction, two SNPs on X were associated significantly with alcohol dependence in European-American men ($P = 5.0 \times 10^{-4}$ for rs5916144 and $P = 5.5 \times 10^{-5}$ for rs5916794 at 3’UTR of NLGN4X), but not in other cohorts.

Given the allele frequencies discovered in the European-American male cohort listed in Table 1, $w = 0.167$, $df = 1$, and $z = 10^{-4}$, the European-American female cohort ($n = 1599$) had 99.8% of power to detect the two most significant risk markers. However, the African-American male and female cohorts had only 68.1 and 51.4% of power to significantly detect them, respectively; thus, the findings in African-American cohorts were considered exploratory.

We found that 26 of 444 SNPs at 3’UTR or within NLGN4X were nominally associated with alcohol dependence in European-American men ($5.5 \times 10^{-5} \leq P \leq 0.05$), including those two significant ones. All of these associations were not statistically significant in European-American women, and most of them were not significant in African-Americans men and women either (Table 1).

**Discussion**

We found that NLGN4X on the X chromosome was a significant male-specific risk gene for alcohol dependence in European-Americans. NLGN4X might harbor a causal variant(s) for alcohol dependence.

NLGN4X, spanning 338 kb, maps to Xp22.32 on XDR, proximal to the PAR1 of the X chromosome. It contains six exons and codes for neurexin 4 (with 816 amino acids) (Ichtchenko *et al.*, 1996). Its homolog on the Y chromosome, that is, NLGN4Y (at Yq11.221), is an X-degenerate gene. Forty million years ago, during the process of divergence of Old World from New World Monkeys, NLGN4X and NLGN4Y were introduced into X and Y chromosomes, respectively, originating from the common ancient autosomes (De Oliveira *et al.*, 2012).
Females have two Xs and males have one X and one Y. To make the gene dosage equivalent between both sexes, females inactivate almost an entire X chromosome. However, about 15% of the genes on the inactive X chromosome, mostly situated on the short arm (Xp) of the X chromosome, and men are more likely to have X chromosomes. The male-specific associations between NLGN4X and alcohol dependence are paralleled by male-dominant inheritance for other disorders where NLGN4X is also named as autism susceptibility X-linked (AUTSX2). Given the significant phenotypic overlap with autism and the putative role for synaptic plasticity in the etiology of alcohol dependence, it is understandable that NLGN4X is also a risk gene for type 2 gene (AUTSX2). NLGN4X is also a risk gene for plasticity in the etiology of alcohol dependence, it is understandable that NLGN4X is also a risk gene for.

The male-specific associations between NLGN4X and alcohol dependence are paralleled by male-dominant inheritance for other disorders where NLGN4X is implicated, particularly autism (Jamain et al., 2003). NLGN4X is also named as autism susceptibility X-linked type 2 gene (AUTSX2). Given the significant phenotypic overlap with autism and the putative role for synaptic plasticity in the etiology of alcohol dependence, it is understandable that NLGN4X is also a risk gene for type 2 gene (AUTSX2).

Two copies of NLGN4X in women escape the inactivation and thus both express protein product. When one of the pair has a defective effect, another copy might compensate the defective effect of the risk copy, and thus might protect against diseases. Similarly, when the variants are paired between NLGN4X and NLGN4Y in men, the defective effects of risk alleles could also be compensated. Only when the defective effects come from the single-copy of X (i.e. in men) could these effects be invariably expressed, increasing the risk for diseases, which was exactly what we observed in this study. This is why many male-dominant diseases are caused by genes on the X chromosome, and men are more likely to have X chromosome-related diseases (Zinn et al., 2007).

The male-specific associations between NLGN4X and alcohol dependence are paralleled by male-dominant inheritance for other disorders where NLGN4X is implicated, particularly autism (Jamain et al., 2003). NLGN4X is also named as autism susceptibility X-linked type 2 gene (AUTSX2). Given the significant phenotypic overlap with autism and the putative role for synaptic plasticity in the etiology of alcohol dependence, it is understandable that NLGN4X is also a risk gene for type 2 gene (AUTSX2).

Two copies of NLGN4X in women escape the inactivation and thus both express protein product. When one of the pair has a defective effect, another copy might compensate the defective effect of the risk copy, and thus might protect against diseases. Similarly, when the variants are paired between NLGN4X and NLGN4Y in men, the defective effects of risk alleles could also be compensated. Only when the defective effects come from the single-copy of X (i.e. in men) could these effects be invariably expressed, increasing the risk for diseases, which was exactly what we observed in this study. This is why many male-dominant diseases are caused by genes on the X chromosome, and men are more likely to have X chromosome-related diseases (Zinn et al., 2007).

The male-specific associations between NLGN4X and alcohol dependence are paralleled by male-dominant inheritance for other disorders where NLGN4X is implicated, particularly autism (Jamain et al., 2003). NLGN4X is also named as autism susceptibility X-linked type 2 gene (AUTSX2). Given the significant phenotypic overlap with autism and the putative role for synaptic plasticity in the etiology of alcohol dependence, it is understandable that NLGN4X is also a risk gene for type 2 gene (AUTSX2).
alcohol dependence. Interestingly, other genes have been linked to both, the heritable risk for alcoholism and that for autism. For example, the autism susceptibility candidate gene 2 (AUTS2) was reported to be a risk gene for alcohol consumption by a genome-wide association study (Schumann et al., 2011). In addition to autism and alcohol dependence, NLGN4X has also been associated with several other cognitive disorders including Asperger syndrome, Tourette syndrome, mental retardation, selective nonverbal deficits (e.g. impaired visual-spatial abilities), learning disability, and/or schizophrenia (Jamain et al., 2003; Laumonnier et al., 2004; Yan et al., 2005; Zinn et al., 2007; Lawson-Yuen et al., 2008; Walsh et al., 2008).

The brain processes information by transmitting signals at synapses, which connect neurons into vast networks of communicating cells. In these networks, synapses not only transmit signals but also transform and refine them. All processing of information in the brain involves synapses, and almost all abnormalities in brain function have a direct or an indirect effect on synaptic function. Neurexins and neuroligins are enriched specifically in the membrane of glutamatergic synapses, especially in the dendritic spines in the hippocampal neurons (Chih et al., 2005). They are arguably the best-characterized synaptic cell-adhesion molecules, and they are the only ones for which a specifically synaptic function has been established (Dean and Dresbach, 2006; Craig and Kang, 2007). Neuroligin–neurexin complexes connect presynaptic and postsynaptic neurons at synapses, mediate signaling across the synapse, and shape the properties of neural networks by specifying synaptic functions, playing a central role in the brain’s ability to process information during nervous system development. Neuroligin deficiency in the brain caused by NLGN4X mutations may lead to abnormal development of synaptic structures and may have adverse effects on communication processes and cognitive development; thus, neuroligin is a key target in the pathogenesis of cognitive diseases (Sudhof, 2008).

Imbalance between neuronal excitation and inhibition could lead to epilepsy; the prevalence of epilepsy among alcoholics is at least triple that in the general population (Chan, 1985). It is hypothesized that at least some forms of alcohol dependence might be caused by an imbalance of neuronal excitation/inhibition. Neuroligin 4 is localized to excitatory glutamatergic axons and plays a role in adjusting the balance between excitatory and inhibitory synapses (Graf et al., 2004; Prange et al., 2004; Chih et al., 2005). Neuroligin defects caused by NLGN4X mutations may abolish the formation, stabilization, or recognition of specific synapses essential for the communication processes, and subsequently lead to selective loss of excitatory function and abnormal excitatory/inhibitory balance in neurons (Jamain et al., 2003). Such a defect of synaptogenesis in neuronal circuitry is believed to play a role in alcohol dependence.

NLGN4X conferring a risk for alcohol dependence might be population specific (i.e. European-Americans only). However, this specificity should be confirmed in the future by expanding the sample sizes of African-American cohorts to achieve sufficient study power.

Acknowledgements
The authors thank NIH GWAS Data Repository, the Contributing Investigator(s) (Drs. Bierut, Edenberg, etc.) who contributed the phenotype and genotype data (SAGE and COGA) from his/her original study, and the primary funding organization that supported the contributing study. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Genotyping was performed at the Johns Hopkins University Center for Inherited Disease Research or at deCODE. Grant sponsor: National Alliance for Research on Schizophrenia and Depression (NARSAD) Award (L.Z.); Grant number: 17616; Grant sponsor: ABMRF/The Foundation for Alcohol Research (L.Z.); Grant sponsor: National Institute on Drug Abuse (NIDA); Grant numbers: R01 DA029643, K24 DA017899, R01 DA016750, K02 DA026990, R01 DA013423; Grant sponsor: National Institute on Alcohol Abuse and Alcoholism (NIAAA); Grant numbers: R01 AA016015, R21 AA020319, K24 AA013736, P50 AA012870, U10 AA008401; Grant sponsor: Genes, Environment and Health Initiative (GEI); Grant numbers: U01HG004422, U01HG004438; Grant sponsor: GENEVA Coordinating Center; Grant number: U01 HG004446; Grant sponsor: National Cancer Institute; Grant number: P01 CA089392; Grant sponsor: NIH contract ‘High throughput genotyping for studying the genetic contributions to human disease’; Grant number: HHSN268200782096C; Grant sponsor: Department of Veterans Affairs (the VA Alcohol Research Center, the VA National Center for PTSD, and the Depression REAP).

Conflicts of interest
There are no conflicts of interest.

References


Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320:539–543.


