Published in final edited form as:
Alcohol Clin Exp Res. 2012 September ; 36(9): 1512–1518. doi:10.1111/j.1530-0277.2012.01758.x.

Copy number variations in 6q14.1 and 5q13.2 are associated with alcohol dependence

Peng Lin, PhD[1], **Sarah M. Hartz, MD/PhD**[1], **Jen-Chyong Wang, PhD**[1], **Arpana Agrawal, PhD**[1], **Tian-Xiao Zhang, MSc**[1], **Nicholas McKenna, BS**[1], **Kathleen Bucholz, PhD**[1], **Andrew I. Brooks, PhD**[2], **Jay A. Tischfield, PhD**[2], **Howard J. Edenberg, PhD**[3], **Victor M. Hesselbrock, PhD**[4], **John R. Kramer, PhD**[5], **Samuel Kuperman, MD**[5], **Marc A. Schuckit, MD**[6], **Alison M. Goate, PhD**[1], **Laura J. Bierut, MD**[1], **John P. Rice, PhD**[1], and **COGA Collaborators, COGEND Collaborators, GENEVA**

[1]Department of Psychiatry, Washington University, St. Louis, Missouri, United States of America

[2]Department of Genetics, Rutgers University, Piscataway, New Jersey, United States of America

[3]School of Medicine, Indiana University, Indianapolis, Indiana, United States of America

[4]Department of Psychiatry, University of Connecticut School of Medicine, Farmington, Connecticut, United States of America

[5]Department of Psychiatry, University of Iowa School of Medicine, Iowa City, Iowa, United States of America

[6]Department of Psychiatry, University of California, San Diego, California, United States of America

Abstract

Background—Excessive alcohol use is the third leading cause of preventable death and is highly correlated with alcohol dependence, a heritable phenotype. Many genetic factors for alcohol dependence have been found, but many remain unknown. In search of additional genetic factors, we examined the association between DSM-IV alcohol dependence and all common copy number variations (CNV) with good reliability in the Study of Addiction: Genetics and Environment (SAGE).

Methods—All participants in SAGE were interviewed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), as a part of three contributing studies. 2,610 non-Hispanic European American samples were genotyped on the Illumina Human 1M array. We performed CNV calling by CNVpartition, PennCNV and QuantiSNP and only CNVs identified by all three software programs were examined. Association was conducted with the CNV (as a deletion/duplication) as well as with probes in the CNV region. Quantitative polymerase chain reaction (qPCR) was used to validate the CNVs in the laboratory.

Results—CNVs in 6q14.1 ($P = 1.04 \times 10^{-6}$) and 5q13.2 ($P = 3.37 \times 10^{-4}$) were significantly associated with alcohol dependence after adjusting multiple tests. On chromosome 5q13.2 there were multiple candidate genes previously associated with various neurological disorders. The region on chromosome 6q14.1 is a gene desert that has been associated with mental retardation, and language delay. The CNV in 5q13.2 was validated whereas only a component of the CNV on 6q14.1 was validated by qPCR. Thus, the CNV on 6q14.1 should be viewed with caution.

Corresponding author: John P. Rice, Office Phone: 314-286-2572, FAX: 314-286-2577, 660 S. Euclid Ave, Campus box 8134, St. Louis MO 63110, john@zork.wustl.edu.

Conclusion—This is the first study to show an association between DSM-IV alcohol dependence and CNVs. CNVs in regions previously associated with neurological disorders may be associated with alcohol dependence.

Keywords

Copy Number Variations; Alcohol dependence; CNV Accuracy

Introduction

During 2001–2005, excessive alcohol use contributed to about 79,000 deaths and 2.3 million years of potential life lost in the United States (Kanny et al., 2011). Excessive alcohol consumption, the third leading cause of preventable death in the United States, can cause damage to the central and peripheral nervous system, and to nearly every organ system in the body (Caan and De Belleroche, 2002; Testino, 2008). It is also strongly correlated with alcohol dependence, a serious psychiatric disorder that affects about 12% of American adults across their lifetime (Hasin et al., 2007). Alcohol dependence is a common, complex disease characterized by compulsive and uncontrolled alcohol consumption despite its negative effects on the drinker's health, relationships, productivity, and social standing.

There is robust evidence for heritable influences on the liability to alcohol dependence (Bierut et al., 1998). Siblings of alcohol dependent individuals have a 3–8 fold increased risk of developing alcoholism (Reich et al., 1998) with twin studies revealing the heritability of alcohol dependence to be ≈ 50% (Heath et al., 1997; Kendler et al., 1994; Knopik et al., 2004). Given its serious public health impact (Braillon and Dubois, 2005) and the strong evidence for its biological underpinnings, numerous linkage and association studies have been targeted at gene identification for alcohol dependence (Bierut et al., 2010; Edenberg and Foroud, 2006; Hill et al., 2004; Long et al., 1998; Reich et al., 1998). Recently, several genome wide association studies (GWAS) queried the genome for association (Bierut et al., 2010; Edenberg et al., 2010; Heath et al., 2011; Treutlein et al., 2009). Results surpassed genome-wide significance in one study of early-onset male alcoholics (Treutlein et al., 2009), but across the multiple efforts, effect sizes were small and did not replicate. This has generated considerable interest in the examination of other possible contributors to the "missing heritability" for alcohol dependence. One such contributor is copy number variations (CNVs).

CNVs are duplications or deletions of a particular segment of an individual's genome and reflect inherent structural instability in the architecture of the genome. They are prevalent forms of common genetic variation and can have a substantial influence on gene expression levels (Perry et al., 2007). For instance, Mendelian disorders such as Williams-Beuren Syndrome (due to a deletion at chromosome region 7q11.23) and Charcot-Marie-Tooth neuropathy Type 1A (caused by duplication at chromosome region 17p11.2 (Krajewski et al., 2000; Martens et al., 2008) are attributable to CNVs. Despite the deleterious effects of CNVs and their links to disease, few studies have examined CNVs in the context of psychiatric illness, particularly alcohol dependence. This is primarily due to the inherent challenges involved in identification of what constitutes a CNV. While traditional methods of CNV identification involve laboratory-based experiments, they can also be identified (or "called") using GWAS data where a series of single nucleotide polymorphisms (SNPs) or "intensity" probes are interrogated for their occurrence in a state other than the expected disomic (i.e. 2 copy) state. Typically, the intensity of the probe signal that is expected when two copies of the probe are present is compared with the observed intensity, which is expected to be enhanced for duplications, or suppressed for deletions. These probes are routinely included in GWAS chips and thus, as GWAS technology became more accessible,

there was an up-swell in CNV identification efforts. However, this method of CNV calling from GWAS microarrays can be associated with relatively high error rates. For instance, in a previous study, we demonstrated the relatively modest concordance in CNV detection using three widely utilized software packages with varying algorithms. In that study, we implemented statistical measures that enhance the reliability of the detected CNVs using multiple algorithms and further, validated the CNVs identified using statistical programs by quantitative Polymerase Chain Reaction (qPCR) in the laboratory (Lin et al., 2011).

Other challenges of CNV detection include (a) size of the CNV, with smaller CNVs (<10 kb) being harder to detect, (b) number of CNV probes in the region of the CNV, with fewer probes resulting in greater noise, (c) the general quality of the data (including artifacts in the SNP data) and genomic waves (intensity variations in normalized GWAS data), (d) ethnic variations and (e) source of the sample that was genotyped – for instance, it now well known that deletions and duplications can arise in DNA drawn from cell lines (i.e. extracting cells from a DNA source and maintaining them in laboratory cultures to enhance longevity) and, thus CNV detection using cell cultures requires caution. Yet, if attention is paid to these challenges, CNVs represent a unique route for enquiry into the genetic architecture of alcohol dependence.

There continues to be a great deal of progress in statistical methods for CNV detection. In tandem, there is growing excitement about the association between these CNVs and human behavior and the extent to which these intriguing variations in the human genome may contribute to that elusive “missing heritability” in complex behavioral phenotypes and psychiatric illness. While there has been some promise in studies of autism, and intellectual disabilities (Glessner et al., 2009; Pinto et al.), as well as schizophrenia and bipolar disorder (Lachman et al., 2007; Stefansson et al., 2008), research on CNVs in studies of addiction, particularly alcohol dependence, is lacking. In this study, we examine the CNVs for DSM-IV alcohol dependence in a large sample of European-American subjects.

Materials and methods

Samples

Data were drawn from the Study of Addiction: Genetics and Environment (SAGE) (Bierut et al.). SAGE is one study of the Gene Environment Association Studies (GENEVA) project (Cornelis et al., 2010). Unrelated cases and controls for the SAGE sample were drawn from 3 contributing projects: the Collaborative Study on the Genetics of Alcoholism (COGA), the Collaborative Study on the Genetics of Nicotine Dependence (COGEND) and the Family Study of Cocaine Dependence (FSCD). While the contributing studies originally ascertained subjects for alcohol dependence (COGA), nicotine dependence (COGEND: based on an FTND score of 4 or greater in current smokers, controls being smokers) and for cocaine dependence (FSCD), the subset of cases selected for genotyping in SAGE were uniformly defined as those meeting criteria for DSM-IV alcohol dependence (N=1899) while controls (N=1946) were individuals who reported drinking alcohol but did not meet criteria, during their lifetime, for alcohol dependence. Note that even though the parent projects were family-based, the series of cases and controls for SAGE, and consequently for these analyses, were selected to be unrelated individuals. Of these, 1,186 cases and 1,397 controls are of self-reported non-Hispanic European-American descent. All participants agreed to share their DNA and phenotypic information for research purposes and provided written informed consent following instructions from institutional review boards at all data collection sites.

Measures: A lifetime diagnosis of DSM-IV alcohol dependence was determined via self-reported interview information collected using the Semi-Structured Assessment for the

Genetics of Alcoholism (SSAGA). Controls were individuals who had drunk alcohol at least once in their lifetime but did not meet criteria for alcohol dependence.

Genotyping and quality control—The Center for Inherited Disease Research (CIDR) at Johns Hopkins University genotyped all samples on the Illumina Human 1M array. An extensive data cleaning effort had been made to ensure data quality. These procedures included, but not limited to, using HapMap controls, detection of gender mis-annotation and chromosomal anomalies, cryptic relatedness, population structure, batch effects, Mendelian error detection, and duplication error detection. A detailed description of data cleaning effort is described elsewhere (Bierut et al., 2010; Laurie et al.).

CNV calling—The Illumina 1M array has a total of 1,072,820 probes, predominantly indexed by polymorphic SNPs. 23,812 of these probes are non-SNP "intensity-only" markers for CNV detection. All of the 1,072,820 probes were used for the CNV analyses. Three common programs were used to call CNVs: CNVPartition, PennCNV (Wang et al., 2007), and QuantiSNP (Colella et al., 2007). Genomic waves were also adjusted when we called CNVs by PennCNV and QuantiSNP (Diskin et al., 2008). Both PennCNV and QuantiSNP report a metric score for quality control purposes. As recommended by QuantiSNP documentation, we removed all CNV calls that had Log Bayes Factor (LBF) less than 10, as well as poor quality samples based on quality control measures for CNV analysis, following the approaches described in our previous work (Lin et al., 2011). In total, we genotyped 2,583 non-Hispanic European American samples in SAGE and among them 95 samples failed to pass quality controls for CNV analysis.

Comparative statistics—The CNV calls from different programs were compared against each other. In our previous work, we have demonstrated that a CNV that is confirmed by all three CNV calling programs has a higher reproducibility rate, and thus, a higher reliability (Lin et al., 2011). Therefore, we required that only CNVs detected by all three programs would be studied.

Association analysis—Logistic regressions were performed on all CNV regions. After identifying potential regions, individual dummy variables for duplications and deletions were created to dissect the association signal with DSM-IV alcohol dependence. Several covariates were included in the model based on the previous GWAS of these data (Bierut et al., 2010), including sex, age, and two principal components indexing continuous ancestral genetic variation. We also included a dummy variable to indicate the source of DNA (cell line versus whole blood). In addition, we ensured that these potentially confounding variables were not directly associated with the identified CNVs.

CNVs with different starting and ending point—Even when all three programs detect a CNV, they often report different starting and ending points for the same CNV segment, which leads to computational challenges in combining CNV reports. There is a lack of consensus in the research community regarding this issue and therefore, in addition to studying the CNV as a deletion or duplication, we adopted an additional straightforward approach for association. First, SNP probes and intensity-only probes were used to detect CNVs by multiple programs. Second, a change of copy number at a particular probe was considered detected when all CNV programs reported CNV segments that cover the probe. Third, association between alcohol dependence and each probe (assigned the same copy number as the CNV) was examined. For instance, if a CNV (duplication or deletion) was detected in region X, using probes (SNP or intensity probes) A, B C, D, E, F and G by three different programs, then the results from the three programs for each probe were compared against each other (See Supplementary Figure 1). If agreement was reached among three

programs, then the CNV for these probes (Probe D and E) were confirmed and would be used in the following analysis. If there was disagreement among the three programs, then a missing value was assigned to these probes (Probe B, C, and F).

Validation—CNVs identified by 3 independent programs were validated in subjects carrying the variant with quantitative PCR (qPCR). We selected a TaqMan CNV probe in the target region. The probe was predesigned by Applied Biosystems (Applied Biosytems, Foster City, CA, USA). Genomic DNA was analyzed with real-time PCR using an ABI-7900 Fast PCR system. Each real-time PCR run included within-plate duplicates. Correction for sample-to-sample variation was done by simultaneously amplifying a standard CNV reference assay, RNAse P. Real-time data were analyzed using the comparative C_t method (Muller et al., 2002). The C_t values of each sample were normalized with the C_t value for the RNAse P assay. Only the samples with a standard error <0.15 were analyzed. Copy numbers were calculated using ABI CopyCaller™ Software v1.0.

Results

Socio-demographic characteristics

Of the 2,583 non-Hispanic European American samples from SAGE, 95 failed to pass quality control for CNV analysis, leaving 1,140 cases and 1,348 controls. The mean age among subjects with alcohol dependence was 38.2 [SD=10.0], and for controls was 39.0 [SD=9.5]. Sixty percent of the cases and 29.2% of the controls were male. As shown in Table 1, cases were more likely than controls to be dependent on nicotine and illicit drugs, including nicotine, cocaine, and marijuana. They were also more likely to meet criteria for a lifetime history of conduct disorder and major depression.

Alcohol history

Cases also reported an earlier age of heavy and regular alcohol use, and, by definition, reported more alcohol symptoms (Table 2).

CNV detection

Of the samples that passed quality control, we identified 1,139 CNV regions with length greater than 50 kb and number of probes not less than 10 (Lin et al., 2011). Among them, only 141 CNV regions have frequency higher than 1%. All of these CNV regions had previously been documented in the database of genetic variants (Zhang et al., 2006). Level of agreement across the various algorithms was moderate (see Supplementary Table S1). Thus, after adjusting for multiple tests, our significance threshold for association analyses is $0.05/141=3.54 \times 10^{-4}$.

Association between CNVs and Alcohol Dependence

Two CNV regions were significantly associated with alcohol dependence (Table 3): chromosome 6q14.1 (OR=2.86, $P= 1.04 \times 10^{-6}$, n=121 subjects with the duplication) and chromosome 5q13.2 (OR=1.99, $P= 3.37 \times 10^{-4}$, n=59 subjects with the duplication, and n=58 subjects with deletions). The P values for each probe in these two regions are listed in Supplementary Tables 2 & 3. There were significant differences in the distribution of males and females across cases and controls. While all analyses modeled sex as a covariate, we also conducted secondary analyses separately in males and females. Significant results for SNPs in the region of the CNV on 5q13.2 and 6q14.1 was replicated with the same direction of effect (albeit at less significant p-values, see Supplemental Tables 4A, 4B, 5A and 5B).

Validation using qPCR—For the CNV at 5q13.2, over 97% of these CNVs (called by all three algorithms) were confirmed as true CNVs using qPCR. Only 85% were validated when the CNV was called by 2 of 3 algorithms and 38% when only 1 of 3 algorithms called it. Redoing analyses with successfully validated CNVs yielded identical results. However, for 6q14.1, while all deletions were confirmed, none of the duplications were reproduced via qPCR. This suggests that the result for 6q14.1 should be viewed with caution.

Discussion

These analyses evaluated the association between CNVs and alcohol dependence among a relatively large sample of alcohol-dependent cases and non-dependent alcohol exposed controls. We found two regions significantly associated with alcohol dependence: Chr5: 69,916,523-70,373,564 and chr6:79,034,386-79,090,197. To our knowledge, this is the first study to connect CNVs and alcohol dependence.

The identified chromosomal regions have been previously associated with several neurological and other disorders. Chr5: 69,916,523-70,373,564 covers several genes, including *SMA4, SERF1, SERF1B, SMN2, SMA3, NAIP, GTF2H2, GTF2H2D* and the downstream *OCLN*. Among them, *SMA3, SMA4* and *SMN2* are known to be associated spinal muscular atrophy (Also-Rallo et al., 2011; Wong et al., 2007). Recent research shows that the genes in this region have a function in the nervous system (Dachs et al., 2011), including *OCLN,* another candidate in this region, which is an integral membrane protein that is required for cytokine-induced regulation of the tight junction paracellular permeability barrier. Mutations in this gene are thought to be a cause of pseudo-*TORCH* syndrome, an autosomal recessive neurologic disorder that mimics the clinical characteristics (e.g. microcephaly, seizures, spasticity) attributable to congenital infections due to **T**oxoplasmosis, **O**ther Agents, **R**ubella, **C**ytomegalovirus or **H**erpes Simplex (O'Driscoll et al., 2010). While the CNV in Chr6:79,034,386-79,090,197 is located in a gene desert, there is evidence that suggests a link between chromosome region 6q14.1 and mild mental retardation, language delay, and minor dysmorphisms (Lespinasse et al., 2009; Puppala et al., 2011). Also, it is hypothesized that non-coding intergenic regions, such as this, may contain regulatory elements, such as enhancers and chromosome scaffold components that are capable of changing gene expression.

We restricted our association tests to non-rare (>1%) CNVs for two reasons. First, the traditional genome wide association study design has little power to detect rare genetic variants (<1%), and the case control study design of this project cannot provide enough power to detect rare CNVs. Second, accuracy of CNV detection diminishes with decreasing frequency.

In addition we required that all CNVs be reproduced by 3 independent programs, a step that increases confidence in the results but that raises the potential problem of the same CNV region being detected with different starting and ending points, which results in uncertainty on how to combine these different CNV calls. In order to avoid this controversy, we adopted an intuitive method where we tested each genetic marker instead of a particular CNV segment. A CNV status is assigned to a particular genetic marker when all programs report a CNV that covers this probe (Supplementary Tables 1 & 2). We validated our findings by qPCR – while the region on chromosome 5q13.2 was successfully validated, the duplication on 6q14.1 did not. Thus, the finding on 6q14.1 should we viewed with caution and as such, it cannot be claimed as a true CNV until laboratory validation is successful. However, we cannot be certain that absence of experimental validation necessarily negates a CNV called using statistical algorithms – in some instances, lack of suitable laboratory probes in the region of the CNV can result in lack of experimental validation. Nonetheless, this

underscores the considerable importance of experimental validation of CNVs identified using software algorithms.

The frequency of the CNV deletion on 6q14.1 is high (48.5% in cases and 49.8% in controls). Supplemental Figure 2 provides the distribution of deletions and duplications from the DGV. CNVs with frequencies such as those noted for 5q13.2 are not uncommon. Additionally, based on other reports in the Database of Genomic Variants, for the 5q13.2 region, both deletions and duplications have been identified but frequencies vary dramatically by sample size and ethnicity of studied individuals. For 6q14.1, the DGV does not report on additional studies with reasonably large or diverse samples for us to validate the frequency of the observed CNV – nonetheless, the Database of Genomic Variants (DGV) notes 228 other instances where CNVs of such high frequency have been previously identified (restricting sample size in studies to be greater than 30).

Our finding of the association between these CNVs and alcohol dependence is encouraging because it identifies regions previously associated with neurological disorders, however these findings will require replication. Furthermore, whether these CNVs co-segregate with alcoholism in families remains to be explored – unfortunately, despite the cases and controls being drawn from family-based data, GWAS data and biological samples on parents are currently not available. Nonetheless, our study is amongst the first to examine the role of CNVs in the etiology of alcohol dependence. This reflects the exciting phase of the post-GWAS genomics era where the quest to articulate the genetic architecture of serious psychiatric problems like alcohol dependence moves beyond single SNP association to new frontiers, such as CNVs and rare variants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423, R01 DA019963). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract “High throughput genotyping for studying the genetic contributions to human disease” (HHSN268200782096C). Arpana Agrawal also acknowledges support from ABMRF/The Foundation for Alcohol Research.

References

Also-Rallo E, Alias L, Martinez-Hernandez R, Caselles L, Barcelo MJ, Baiget M, Bernal S, Tizzano EF. Treatment of spinal muscular atrophy cells with drugs that upregulate SMN expression reveals inter-and intra-patient variability. Eur J Hum Genet. 2011

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A. 107(11):5082–7. [PubMed: 20202923]

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A. 2010; 107(11): 5082–7. [PubMed: 20202923]

Bierut LJ, Dinwiddie SH, Begleiter H, Crowe RR, Hesselbrock V, Nurnberger JI Jr, Porjesz B, Schuckit MA, Reich T. Familial transmission of substance dependence: alcohol, marijuana, cocaine, and habitual smoking: a report from the Collaborative Study on the Genetics of Alcoholism. Arch Gen Psychiatry. 1998; 55(11):982–8. [PubMed: 9819066]

Braillon A, Dubois G. Alcohol and public health. Lancet. 2005; 365(9468):1387. [PubMed: 15836885]

Caan, W.; De Belleroche, J. Drink, drugs and dependence : from science to clinical practice. Routledge; London ; New York, NY: 2002.

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007; 35(6):2013–25. [PubMed: 17341461]

Cornelis MC, Agrawal A, Cole JW, Hansel NN, Barnes KC, Beaty TH, Bennett SN, Bierut LJ, Boerwinkle E, Doheny KF, Feenstra B, Feingold E, Fornage M, Haiman CA, Harris EL, Hayes MG, Heit JA, Hu FB, Kang JH, Laurie CC, Ling H, Manolio TA, Marazita ML, Mathias RA, Mirel DB, Paschall J, Pasquale LR, Pugh EW, Rice JP, Udren J, van Dam RM, Wang X, Wiggs JL, Williams K, Yu K. The gene, environment association studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. Genet Epidemiol. 2010:364–372. [PubMed: 20091798]

Dachs E, Hereu M, Piedrafita L, Casanovas A, Caldero J, Esquerda JE. Defective neuromuscular junction organization and postnatal myogenesis in mice with severe spinal muscular atrophy. J Neuropathol Exp Neurol. 2011; 70(6):444–61. [PubMed: 21572339]

Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris JM, Wang K. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008; 36(19):e126. [PubMed: 18784189]

Edenberg HJ, Foroud T. The genetics of alcoholism: identifying specific genes through family studies. Addict Biol. 2006; 11(3–4):386–96. [PubMed: 16961766]

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI Jr, Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res. 2010; 34(5):840–52. [PubMed: 20201924]

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PM, Chiavacci R, Annaiah K, Thomas K, Hou C, Glaberson W, Flory J, Otieno F, Garris M, Soorya L, Klei L, Piven J, Meyer KJ, Anagnostou E, Sakurai T, Game RM, Rudd DS, Zurawiecki D, McDougle CJ, Davis LK, Miller J, Posey DJ, Michaels S, Kolevzon A, Silverman JM, Bernier R, Levy SE, Schultz RT, Dawson G, Owley T, McMahon WM, Wassink TH, Sweeney JA, Nurnberger JI, Coon H, Sutcliffe JS, Minshew NJ, Grant SF, Bucan M, Cook EH, Buxbaum JD, Devlin B, Schellenberg GD, Hakonarson H. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009; 459(7246):569–73. [PubMed: 19404257]

Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry. 2007; 64(7): 830–42. [PubMed: 17606817]

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychol Med. 1997; 27(6): 1381–96. [PubMed: 9403910]

Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A, Nyholt DR, Bucholz KK, Madden PA, Montgomery GW. A Quantitative-Trait Genome-Wide Association Study of Alcoholism Risk in the Community: Findings and Implications. Biol Psychiatry. 2011

Hill SY, Shen S, Zezza N, Hoffman EK, Perlin M, Allan W. A genome wide search for alcoholism susceptibility genes. Am J Med Genet B Neuropsychiatr Genet. 2004; 128B(1):102–13. [PubMed: 15211641]

Kanny D, Liu Y, Brewer RD. Binge drinking-United States, 2009. MMWR Surveill Summ. 2011; 60(Suppl):101–4. [PubMed: 21430633]

Kendler KS, Neale MC, Heath AC, Kessler RC, Eaves LJ. A twin-family study of alcoholism in women. Am J Psychiatry. 1994; 151(5):707–15. [PubMed: 8166312]

Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, Statham D, Whitfield JB, Martin NG. Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. Psychol Med. 2004; 34(8):1519–30. [PubMed: 15724882]

Krajewski KM, Lewis RA, Fuerst DR, Turansky C, Hinderer SR, Garbern J, Kamholz J, Shy ME. Neurological dysfunction and axonal degeneration in Charcot-Marie-Tooth disease type 1A. Brain. 2000; 123 (Pt 7):1516–27. [PubMed: 10869062]

Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T, Papolos DF, Stopkova P. Increase in GSK3beta gene copy number variation in bipolar disorder. Am J Med Genet B Neuropsychiatr Genet. 2007; 144B(3):259–65. [PubMed: 17357145]

Laurie C, Bierut L, Bhangale T, Boehm F, Caporaso N, Doheny K, Gabriel S, Harris E, Hu F, Jacobs K, Kraft P, Landi M, Manolio T, McHugh C, Mirel D, Pugh E, Rice J, Weir BS. Investigators TG Genotype data cleaning for whole-genome association studies. In preparation.

Lespinasse J, Gimelli S, Bena F, Antonarakis SE, Ansermet F, Paoloni-Giacobino A. Characterization of an interstitial deletion 6q13-q14.1 in a female with mild mental retardation, language delay and minor dysmorphisms. Eur J Med Genet. 2009; 52(1):49–52. [PubMed: 18992376]

Lin P, Hartz SM, Wang J-C, Krueger RF, Foroud TM, Edenberg HJ Jr, JIN, Brooks AI, Tischfield JA, Almasy L, Webb BT, Hesselbrock VM, Porjesz B, Goate AM, Bierut LJ, Rice JP. Collaborators C, Collaborators C, GENEVA. Copy Number Variation Accuracy in Genome Wide Studies. human heredity. 2011 In Press.

Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, Bennett PH, Goldman D. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. Am J Med Genet. 1998; 81(3):216–21. [PubMed: 9603607]

Martens MA, Wilson SJ, Reutens DC. Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. J Child Psychol Psychiatry. 2008; 49(6): 576–608. [PubMed: 18489677]

Muller PY, Janovjak H, Miserez AR, Dobbie Z. Processing of gene expression data generated by quantitative real-time RT-PCR. Biotechniques. 2002; 32(6):1372-4, 1376, 1378–9. [PubMed: 12074169]

O'Driscoll MC, Daly SB, Urquhart JE, Black GC, Pilz DT, Brockmann K, McEntagart M, Abdel-Salam G, Zaki M, Wolf NI, Ladda RL, Sell S, D'Arrigo S, Squier W, Dobyns WB, Livingston JH, Crow YJ. Recessive mutations in the gene encoding the tight junction protein occludin cause band-like calcification with simplified gyration and polymicrogyria. Am J Hum Genet. 2010; 87(3):354–64. [PubMed: 20727516]

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007; 39(10):1256–60. [PubMed: 17828263]

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez

BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 466(7304):368–72. [PubMed: 20531469]

Puppala S, Coletta DK, Schneider J, Hu SL, Farook VS, Dyer TD, Arya R, Blangero J, Duggirala R, Defronzo RA, Jenkinson CP. Genome-Wide Linkage Screen for Systolic Blood Pressure in the Veterans Administration Genetic Epidemiology Study (VAGES) of Mexican-Americans and Confirmation of a Major Susceptibility Locus on Chromosome 6q14.1. Hum Hered. 2011; 71(1): 1–10. [PubMed: 21293138]

Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally PM, Nurnberger JI Jr, Tischfield JA, Crowe RR, Cloninger CR, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet. 1998; 81(3): 207–15. [PubMed: 9603606]

Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. Nature. 2008; 455(7210):232–6. [PubMed: 18668039]

Testino G. Alcoholic diseases in hepato-gastroenterology: a point of view. Hepatogastroenterology. 2008; 55(82–83):371–7. [PubMed: 18613369]

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry. 2009; 66(7):773–84. [PubMed: 19581569]

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17(11):1665–74. [PubMed: 17921354]

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet. 2007; 80(1):91–104. [PubMed: 17160897]

Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet Genome Res. 2006; 115(3–4):205–14. [PubMed: 17124402]

Table 1

Socio-demographic characteristics

	Total No.	Gender (male %)	Ave. Age	Comorbidity with other addictions*			
				Nicotine dependence	Cocaine dependence	Marijuana dependence	Opiates dependence
Alcohol dependence	1,140	60.0%	38.2	70.4%	38.3%	34.2%	14.7%
No dependence	1,348	29.8%	39.0	22.5%	0.0 %	0.0 %	0.0 %
Total	2,488	43.7%	38.7	44.5%	17.5%	15.6%	6.7 %

*All Psychiatric diagnoses were categorized by Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

Alcohol Clin Exp Res. Author manuscript; available in PMC 2013 September 01.

Table 2

Alcohol history

	Age of onset of regular drinking[a]	Age when first got drunk	Maximum drinks per day[b]	Maximum drinks per week	Number of alcohol symptoms endorsed
Alcohol dependence	17.8	15.2	30.1	54.7	5.3
No dependence	20.8	18.2	9.9	10.7	0.7
Total	19.3	16.7	19.1	27.5	2.8

[a] Regular drinking is defined as drinking once a month for 6 months or more

[b] Largest number of alcoholic drinks consumed in 24 hours

Alcohol Clin Exp Res. Author manuscript; available in PMC 2013 September 01.

Table 3

Associations of CNVs and alcohol dependence

Locus[a]	CNV type	CNV frequency			Beta [b]	P value [c]	Genes
			Cases	Controls			
Chr5: 68,921,426-70,412,247	Copy number/dosage[e]	Duplication	43 (4.97%)	16 (1.59%)	0.69	3.37×10^{-4}	*SMA4, SERF1, SERF1B, SMN2, SMA3, NAIP, GTF2H2, GTF2H2D, OCLN* [d]
		Deletion	22 (2.54%)	36 (3.57%)			
		Non-missing	865 (100%)	1008 (100%)			
chr6:79,034,386-79,090,197	Duplication	Duplication	83 (8.99%)	38 (3.49%)	1.05	1.04×10^{-6}	Gene desert
		Deletion	448 (48.54%)	542 (49.77%)			
		Non-missing	923 (100%)	1089 (100%)			

[a]The starting and the ending point are defined by the probes whose P values have a clear deviation from the rest.

[b.c]The P value and beta of this region is annotated by the most significant probe in this CNV.

[d]*OCLN* is located less than 30 kb downstream of this region.

[e]This CNV was analyzed as ≤1= deletion, 2=normal diploid, ≥3=duplication. For cases: Deletions: 4/22 had a dosage score of 0.667 (18/22 had a score of 1, or heterozygous for deletion); Duplications: 39/43 had a dosage score of 3 (heterozygous) and the remainder had a mean score of 3.3. For controls: Deletions: 11/36 had a dosage score of 0.667, remainder scoring 1; Duplications: 13/22 had a dosage score of 3, 2/22 had a score of 3.3 and 1/22 had a score of 4 (homozygous duplication).