HHS Public Access

Author manuscript

Alcohol Clin Exp Res. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Alcohol Clin Exp Res. 2017 April ; 41(4): 711–718. doi:10.1111/acer.13352.

Deep sequencing of 71 candidate genes to characterize variation associated with alcohol dependence

Shaunna L. Clark, Ph.D.[a], **Daniel E. Adkins, Ph.D.**[a], **Gaurav Kumar, Ph.D.**[a], **Karolina A. Aberg, Ph.D.**[a], **Sri Nerella, M.S.**[a], **Linying Xie, M.S.**[a], **Ann L. Collins, Ph.D.**[b], **James J. Crowley, Ph.D.**[b], **Corey R. Quackenbush**[b], **Christopher E. Hilliard, Ph.D.**[b], **Andrey A. Shabalin, Ph.D.**[a], **Scott I. Vrieze, Ph.D.**[f,g], **Roseann E. Peterson, Ph.D.**[d], **William E. Copeland, Ph.D.**[c], **Judy L. Silberg, Ph.D.**[d], **Matt McGue, Ph.D.**[g], **Hermine Maes, Ph.D.**[d], **William G. Iacono, Ph.D.**[g], **Patrick F. Sullivan, M.D., FRANZCP**[b,e], **Elizabeth J. Costello, Ph.D.**[c], and **Edwin J. van den Oord, Ph.D.**[a]

[a]Center for Biomarker Research and Precision Medicine, School of Pharmacy, Virginia Commonwealth University, Richmond,VA, USA

[b]Department of Genetics, University of North Carolina at Chapel Hill, NC, USA

[c]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

[d]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

[e]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[f]Department of Psychology and Neuroscience, Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA

[g]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

Abstract

Background—Previous genome-wide association studies (GWASs) have identified a number of putative risk loci for alcohol dependence (AD). However, only a few loci have replicated and these replicated variants only explain a small proportion of AD risk. Using an innovative approach, the goal of this study was to generate hypotheses about potentially causal variants for AD that can be explored further through functional studies.

Methods—We employed targeted capture of 71 candidate loci and flanking regions followed by next-generation deep sequencing (mean coverage 78X) in 806 European Americans. Regions included in our targeted capture library were genes identified through published GWAS of alcohol, all human alcohol and aldehyde dehydrogenases, reward system genes including dopaminegic and opioid receptors, prioritized candidate genes based on previous associations, and genes involved in the absorption, distribution, metabolism and excretion of drugs. We performed single locus tests to

Corresponding author: Dr. Shaunna L. Clark. Center for Biomarker Research and Precision Medicine, School of Pharmacy, McGuire Hall, Room 216A. P. O. Box 980533. Richmond, VA 23298-0581: slclark2@vcu.edu Phone: +1 (804) 628-3231 Fax: +1 (804) 628-3991.

determine if any single variant was associated with AD symptom count. Sets of variants that overlapped with biologically meaningful annotations were tested for association in aggregate.

Results—No single, common variant was significantly associated with AD in our study. We did, however, find evidence for association with several variant sets. Two variant sets were significant at the q-value < 0.10 level: a genic enhancer for *ADHFE1*($p=1.47\times10^{-05}$; q=0.019), an alcohol dehydrogenase, and *ADORA1*($p=5.29\times10^{-05}$;q=0.035), an adenosine receptor that belongs to a G-protein coupled receptor gene family.

Conclusions—To our knowledge, this is the first sequencing study of AD to examine variants in entire genes, including flanking and regulatory regions. We found that in addition to protein coding variant sets, regulatory variant sets may play a role in AD. From these findings, we have generated initial functional hypotheses about how these sets may influence AD.

Keywords

alcohol dependence; aldehyde dehydrogenase; genetics; next-generation sequencing; serotonin; SNP

INTRODUCTION

Alcohol dependence (AD) is a disorder characterized by compulsive and uncontrolled consumption of alcohol despite its negative effects on the drinker's health, relationships and social standing. Family, adoption, twin and sibling studies have consistently shown that AD is moderately heritable (Goldman et al., 2005). Genome-wide association studies (GWAS) have identified a number of putative risk loci (Edenberg et al., 2010, Gelernter et al., 2014, Schumann et al., 2011, Zuo et al., 2013b, Frank et al., 2012, Treutlein et al., 2009, Adkins et al., 2015) for populations of European ancestry, the focus of this study. However, only a few loci have replicated (Gelernter et al., 2014, Schumann et al., 2011) and these replicated variants only explain a small proportion of AD risk.

One explanation is that instead of following the common-disease common variant hypothesis, which says that the disease causing variants are common in the population of study and have large effect sizes (Visscher et al., 2012), there may be multiple rare causal variants for AD (Edenberg and Foroud, 2014). GWAS studies would have been unable to detect these rare variants because the analysis methods and technology used are designed to target common variants (Wagner, 2013). There is precedent for rare variants contributing to complex diseases ranging from breast cancer (Easton et al., 2007) to neuropsychiatric disorders (Heinzen et al., 2015). In addition, evolutionary theory predicts that disease-causing variants may be rare owing to negative selection and reduced fitness (Eyre-Walker, 2010). Indeed, it has been demonstrated that genes involved in the absorption, distribution, metabolism, and excretion of endogenous compounds, such as alcohol, have higher rates of selection because of the role these genes play in defending against foreign chemical substances (Li et al., 2011).

Exome sequencing studies (Ng et al., 2009), which interrogate the protein coding portion of the genome, can detect rare variants, but the AD studies performed to date used arrays and have not detected any significant rare variants (Vrieze et al., 2014, Zuo et al., 2013a). One

possibility for this finding is that the risk variants for AD may lie outside exons, potentially in regions of the genome which regulate gene expression (Buhler et al., 2015). Indeed, the vast majority (~88%) of GWAS findings for complex diseases are not in exons (Hindorff). The implication is that entire genes will need to be sequenced, thereby capturing variants both within and outside exons. In the past, it has been difficult to understand how risk variants located outside exons could influence disease. However, large-scale efforts have been undertaken to map the location of and better understand the role of regulatory elements in the human genome, such as the Roadmap Epigenomics Project(Kundaje et al., 2015) and the Encyclopedia of DNA Elements (ENCODE) (2012).

The goal of this paper was to generate hypotheses about potentially causal variants for AD that can be explored further through functional studies. Using an innovative approach, we first deep sequenced entire genes plus flanking regions so that we could explore both common and rare variants and variants located within and outside of exons for association with AD. Next, we formed sets of variants based on biological features that affect protein coding or regulate gene expression, in order to test whether variants with specific features are associated with AD.

To achieve our goal, we employ targeted capture (Gnirke et al., 2009) of 71 AD candidate genes in combination with deep, massively-parallel next-generation sequencing (McKernan et al., 2009). This approach is technically similar to exome sequencing, but captures the entire gene and its flanking regions, including putative regulatory regions. This method can identify several types of genetic variants, including common and rare variants, which allowed us to investigate the role these variants may play in AD. To our knowledge this is the first sequencing study of AD to examine entire genes, including introns and gene promoters, regions upstream of genes that assist with gene transcription, and regulatory regions; areas not previously covered by alcohol exome studies. Publically available resources, such as the Roadmap Epigenomics Project (Kundaje et al., 2015), will be used to aid in the interpretation of AD findings outside exons (Edwards et al., 2013).

As there may be a large number of causal risk variants that have small effects (Manolio et al., 2009), we also tested sets of variants for association with AD. These risk variants may be too small to detect individually because of low statistical power, but may be detectable when their effects are aggregated (Wang et al., 2011). Recently, methods have been developed that test the association of a set of variants with a phenotype of interest, including, but not limited to, risk profile scoring and gene-set testing. One disadvantage of these methods is that the sets are often formed based on nominal significance with AD, which can make the results difficult to interpret as there is no link to how these sets affect gene function or expression. Instead, our sets were formed based on biological function (e.g. likely to affect regulatory function or protein coding). By grouping variants in this manner, our results are more interpretable and allowed generation of functional hypotheses about how the variant set may influence AD.

MATERIALS & METHODS

Samples

The sequenced sample comprises 363 subjects from the Virginia Twin Study on Adolescent Behavioral Development (VTSABD)(Meyer et al., 1996) and 443 subjects from the Minnesota Twin Family Study (MTFS)(Iacono et al., 1999, Miller et al., 2012). Both of these studies are twin samples, however, based on random selection, only one subject per twin pair was sequenced. Table 1 provides summary descriptives data for each of the samples. The sequenced MTFS samples had a higher composition of males and monozygotic twins when compared to the VTSABD. All subjects were of European ancestry. Blood samples were collected from which DNA was extracted. Ethical committees in the USA approved all procedures, and all subjects provided written informed consent.

Measures

The phenotype considered in this study is a count of the number of alcohol dependence (AD) symptoms as measured by the Diagnostic and Statistical Manual of Mental Disorders version 4 (DSM-IV). For both samples, the AD symptoms were assessed by in-person interviews. In the VTSABD, the DSM-IV was assessed when the subjects were 18 to 31 years old. The MTFS sample was assessed on AD at multiple time points, but to be comparable to VTSABD, only measures from when the subjects were 18 to 31 years old were used. Longitudinal measures, if available, were collapsed across subject by taking the maximum count of AD symptoms endorsed between 18 and 31 years of age. The age at which this maximum occurred was used as a covariate in the analyses. AD symptom count was treated as a quantitative trait in the following analyses.

Descriptive data for each sample are shown in Table 1. The samples had comparable rates of individuals reporting using alcohol ever, ever receiving an AD diagnosis, and endorsing an AD symptom.

Sequencing

We used the solution-based hybridization targeted capture technology (SureSelectXT, Agilent) to target entire genes and ±5kb of their flanking regions. In this method, a library of synthetic oligonucleotides (baits) complementary to the sequence of interest is custom designed and manufactured (Gnirke et al., 2009). These baits are then used to pull down the desired genomic regions from fragmented genomic DNA samples. Library design and bait tiling were carried out using Agilent eArray.

The loci included in our targeted capture library were as follows: genes identified through published GWAS of alcohol (Schumann et al., 2011); human alcohol dehydrogenases (N=8) and aldehyde dehydrogenases (N=20), the key enzymes involved in alcohol metabolism(Edenberg, 2007); reward system genes, including dopaminegic and opioids receptors and related metabolic genes (N=14); and the remaining space on the array was filled through prioritization of remaining candidate genes as described in the supplementary material (N=29). After collapsing neighboring genes into single loci and removing difficult

to align repetitive elements, our selection encompassed 71 unique loci, covering a total of 5.5 Mb.

The libraries were paired-end sequenced (75bp + 35bp reads) on the SOLiD 5500xl (Life Technologies). The sequence reads were aligned to the human genome (build hg19/GRCh37) using Bioscope 1.3 (Life Technologies) that aligns in color-space. After alignment, quality control measures were implemented including dropping subjects with low mapped reads (<1 million) and fold enrichment (<365). Mean coverage across the targeted regions for each individual was 78X, with at least 10X coverage for 88.2% of the targeted regions, an average fold enrichment of 393.8, and 97.9% of baits covered. This level of coverage is very high for color space data, where two color call errors must occur by chance in adjacent positions before a single-nucleotide polymorphism (SNP) is incorrectly called and therefore should result in fewer base calling errors relative to equivalent coverage on other sequencing platforms (McKernan et al., 2009).

Variant Calling and Annotation

The variants were called and filtered using GATK (McKenna et al., 2010) according to best practices recommendations (Van der Auwera et al., 2013, DePristo et al., 2011). Specifically, we employed variant quality score recalibration to filter variants. We defined rare variants as having a minor allele frequency (MAF) < 0.01 and common variants as MAF ≥ 0.01.

All variants passing quality control (QC) were annotated to examine if variants overlapped with bioinformatic features from the following databases: UCSC Genome Browser and GENCODE (for a full list of annotations see Table S1). To determine their novelty, identified variants were compared with dbSNP v141 (Phillips, 2007) and the 1000 Genomes Database (Clarke et al., 2012) (1KG). Variants were also annotated for overlap with 15 chromatin states, elements that regulate gene transcription (Baker, 2011), in liver and brain tissue from the anterior caudate, hippocampus, mid-frontal lobe, and substantia nigra regions, which are all known to be involved with alcohol addiction (Ozsoy et al., 2013, Wrase et al., 2008, Bassareo et al., 2003, George et al., 2001). See Supplemental Material for a description of the chromatin states and how they were generated by the Roadmap Epigenomics Project (Kundaje et al., 2015).

Statistical Analyses

Individual Variants—To determine if AD symptom count was associated with any single, common variant, we performed single locus tests on all common variants passing QC filters using a linear regression model of additive effects in PLINK (Purcell et al., 2007) for the VTASBD and MTFS separately. Sex, age and 10 ancestry principal components (PCs; see Supplemental Material) were included as covariates to control for sex differences and ancestry. The results from each dataset were then meta-analyzed using a fixed effects model in PLINK (Purcell et al., 2007). We used a false discovery rate (FDR) based approach to declare significance (Supplemental Material). Briefly, we set an FDR threshold of 0.10 (van den Oord and Sullivan, 2003) for declaring genome-wide significance that was implemented using *q*-values (Black, 2004), which are FDRs calculated using the *p*-value of the markers as thresholds for declaring significance.

Sets of Variants—Variant set-based association tests were performed using SKAT (Lee et al., 2012b) to identify sets of variants associated with AD symptom count. To form the sets, we used the annotations described in 2.2. Taking the POLYPHEN – Damaging annotation as an example, within a given gene, all variants that had a POLYPHEN – Damaging annotation were used to form the variant set. Those variants that did not have the annotation were not included in the set. If a set had only one variant that overlapped with a specific annotation in a given gene, then it was excluded from testing as the test would be equivalent to the individual variant test described above. By forming variant sets based on their overlap with a specific annotation, we are able to test whether variants that cause damaging amino acid changes or variants involved in regulatory processes or some other function could potentially be causal.

The VTSABD and MTFS were analyzed together as a mega-analysis to increase power and ensure that the same variants were used to form each set. An indicator variable that identified which study a subject originated from was included as a covariate to address any potential study-related effects. Sex, age and 10 PCs were also included as covariates in this analysis. We tested the combined effects of common and rare variation in SKAT as described in Ionita-Lanza et al. (2013). Briefly, the effect of common and rare variants are assessed separately and then the corresponding test statistics are combined using weights such that common and rare variants contribute equally to the overall test statistic. We required that all variants be polymorphic in both the VTSABD and MTFS to avoid singleton or rare variants in one data set overly influencing the results. In total, 1,312 sets were tested for association with AD.

Previous studies have shown that under certain conditions SKAT can be biased when calculating *p*-values (Lee et al., 2012a, Ionita-Laza et al., 2013). To check the significance of the results with q-values < 0.30, we conducted 100,000 permutations. Specifically, we permuted the phenotypes in each sample separately to preserve the genetic architecture in each study. The analyses were then conducted on the combined permuted data.

RESULTS

Variant Calling

In the 71 loci investigated, we identified 25,922 variants of which 18,442 (71.4%) were rare and 12,417 (47.9%) were novel. Of these novel variants, 73.1% were singleton variants. After removing all singleton variants and requiring that variants be polymorphic in both samples, there were 18,028 variants remaining for analysis. For a complete list of variants investigated see Table S2.

Individual Variants

The Quantile-Quantile (QQ) plot of the individual, common variant meta-analysis results (Figure 1) shows that the distribution of *p*-values is generally on a straight line, indicating the expected *p*-value distribution under the null hypothesis assuming no effects of the markers. If there were evidence of a true association between these variants and AD

symptom count, we would see points in the upper right corner of the plot. However, there is no evidence to support an association between these variants and AD symptom count.

Examining the top ten variants (Table S3) confirms that none of the common variants were significant at the stringent *q*-value < 0.1 level, nor at a very liberal threshold of *q*-value < 0.95.

Sets of Variants

The QQ-plot, Figure 2, shows evidence for association as indicated by the points above the confidence interval in the upper, right corner of the plot. The results show (Table 2) that there are two significant sets associated with AD symptom count at the *q*-value < 0.10 level and several suggestive sets. Of the results listed in Table 2, only one set involved variants that could potentially affect protein coding (i.e. *SLC6A3* POLYPHEN – Damaging sets). Three sets involved regulatory annotations based on chromatin activity classifications and a further six involved other annotations indicative of regulatory potential such as promoter regions, CpG shores, non-coding RNA or enhancer regions. The list of variants in each set listed in Table 2 is provided in Table S4.

Our results confirm the finding from previous studies that the results from SKAT can be biased. Specifically, we observed a λ of 0.797. Values less than 1 suggest that the observed test statistics are smaller than expected when compared with the complete null hypothesis assuming no effect for any of the sets. The permutation p-values were comparable to the observed *p*-values from SKAT (Table 2), with the permutation p-values being slightly more conservative in some cases, and unrelated to the variant set size (Figure S1).

DISCUSSION

We tested variants in 71 candidate loci plausibly associated with AD. Results showed that no single common variant was associated in any of the loci, which is unsurprising given that we do not have the large sample size needed to detect a small effect of a common variant (N > 10,000). This is also consistent with other GWASs of AD where either no variants were found to be significantly associated or the associated variants only explained a small portion of AD risk(Hart and Kranzler, 2015). We then tested if sets of variants formed on the basis of overlap with biologically meaningful annotations within a given locus were associated with AD. These analyses identified sets overlapping with potentially functional (affecting protein coding or regulatory processes) annotations that were significantly associated with AD.

Six of the top variant sets were located in alcohol or aldehyde dehydrogenase genes. These genes encode for different enzymes that are involved in alcohol metabolism. Some of the strongest and most widely replicated genetic associations for alcohol dependence have occurred with common variants in these genes. There has been limited work to date examining the role of rare variants in these genes in AD. Only one study examined rare variants and found an association with rare variants across the entire ADH gene cluster (Zuo et al., 2013c). Our variant set findings suggest a role for both common and rare variants in these genes. Indeed, a recent commentary has called for the need to examine genetic

variation across the entire allelic spectrum to identify risk variants for AD (Edenberg and Foroud, 2014). This strategy has proved successful across a number of other disorders such as, for example, nicotine dependence (Olfson et al., 2016) and schizophrenia (Chang et al., 2016), suggesting that there is the potential for a more general trend of both common and rare variants playing a role in risk.

Three of the top variant sets involved the dopamine transporter gene, *SLC6A3*, also known as *DAT1*, which has been extensively studied for association with AD because of dopamine's role in reward processes. Most previous genetic studies have focused on the 40-base-pair variable-number tandem repeat (VNTR) in the 3′-unstranslated region (3′-UTR) of the gene. There has been mixed evidence for association of genetic variants in this region with AD(Du et al., 2011), although a recent meta-analysis found a significant association with the A9 genotype(Ma et al., 2016). Our variant set findings implicate the opposite end of this gene, specifically the promoter region, suggesting that genetic variants in this region of the gene may influence AD by regulating gene expression. Bioinformatic analyses of *SLC6A3* have shown an abundance of CpGs in the promoter implicating DNA methylation as the regulatory mechanism (Shumay et al., 2010). DNA methylation (Kerkel et al., 2008) has been shown to be regulated by genetic variation and there is ample evidence suggesting that this in turn may affect gene regulation/expression. In the specific context of addiction, a recent study showed that different genotypes influence epigenetic modifications and gene transcription differently in response to cocaine (Vasiliou et al., 2012).

Only two studies have examined the exons of the loci considered here with the goal of identifying causal variants for AD in subjects of European ancestry, to our knowledge. These studies also found that single, common variants are not causal (Vrieze et al., 2014, Zuo et al., 2013a). Only the Vrieze et al. study examined if a variant set was associated with AD. The sets they examined, non-synonymous rare variants, were not significantly associated with AD. This is equivalent to our non-synonymous rare variant findings which is unsurprising given that the MTFS is a subset of the subjects in the Vrieze et al. study. As these previous studies focused only on exons, they would have missed 65 out of 73 (89.1%) variant set findings with p-value < 0.05 found outside protein coding regions. To our knowledge, no previous investigation has examined the role regulatory variant sets may play in AD.

Our findings must be interpreted in the context of the potential limitations. One limitation is that our results suggest potential mechanisms through which the variant sets may affect AD, rather than proving the mechanism. Possible next steps to test the suggested mechanisms include testing these variants sets in an independent sample and examining the function of significant sets in targeted laboratory experiments. We explored whether some of variants included in the top variants sets influenced gene expression using GTEx (Carithers and Moore, 2015) (Table S6) and found several variants are expression quantitative trait loci (eQTL). Other potential methods to explore the functional effects of these variants include targeted genome editing where the specific variant of interest is artificially engineered against a standard cell line background (e.g. CRISPR-Cas9) and the functional effects are observed (de Souza, 2012), or targeted chromatin immunoprecipitation (ChIP) assays of

regulatory elements such as transcription factor binding sites and histone marks overlapping with the significant results.

In conclusion, we did not find a single, common variant that was significantly associated with AD in any of the 71 susceptibility loci considered. However, we identified specific sets of variants within some candidate genes as being potentially causal. We found interesting protein coding variant sets, however they do not account for all signals and it is likely that other variants also contribute via a regulatory role.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support: This work was supported by the National Institutes of Health (R01 DA024413, R01 MH045268, R01 MH068521, R01 DA036216, R01 DA05147, R01 DA024417, R01 AA09367, R25 DA026119, and K01 AA021266 to S.C.).

References

1. An integrated encyclopedia of DNA elements in the human genome. Nature. 489:57–74. [PubMed: 22955616]
2. Adkins DE, Clark SL, Copeland WE, Kennedy M, Conway K, Angold A, Maes H, Liu Y, Kumar G, Erkanli A, Patkar AA, Silberg J, Brown TH, Fergusson DM, Horwood LJ, Eaves L, Van Den Oord EJCG, Sullivan PF, Costello EJ. Genome-Wide Meta-Analysis of Longitudinal Alcohol Consumption Across Youth and Early Adulthood. Twin Research and Human Genetics. 2015; 18:335–347. [PubMed: 26081443]
3. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]
4. Baker M. Making sense of chromatin states. Nat Methods. 2011; 8:717–22. [PubMed: 21878916]
5. Bassareo V, De Luca MA, Aresu M, Aste A, Ariu T, Di Chiara G. Differential adaptive properties of accumbens shell dopamine responses to ethanol as a drug and as a motivational stimulus. Eur J Neurosci. 2003; 17:1465–72. [PubMed: 12713649]
6. Black MA. A note on the adaptive control of false discovery rates. J R Stat Soc B. 2004; 66:297–304.
7. Buhler KM, Gine E, Echeverry-Alzate V, Calleja-Conde J, De Fonseca FR, Lopez-Moreno JA. Common single nucleotide variants underlying drug addiction: more than a decade of research. Addict Biol. 2015
8. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. Biopreserv Biobank. 2015; 13:307–8. [PubMed: 26484569]
9. Chang H, Li L, Li M, Xiao X. Rare and common variants at 16p11.2 are associated with schizophrenia. Schizophr Res. 2016
10. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D, Leinonen R, Shumway M, Sherry S, Flicek P. The 1000 Genomes Project: data management and community access. Nat Methods. 2012; 9:459–62. [PubMed: 22543379]
11. De Souza N. Primer: genome editing with engineered nucleases. Nat Methods. 2012; 9:27. [PubMed: 22312638]
12. Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, Mckenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–8. [PubMed: 21478889]

13. Du Y, Nie Y, Li Y, Wan YJ. The association between the SLC6A3 VNTR 9-repeat allele and alcoholism-a meta-analysis. Alcohol Clin Exp Res. 2011; 35:1625–34. [PubMed: 21554332]
14. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet. 2007; 81:873–83. [PubMed: 17924331]
15. Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol Res Health. 2007; 30:5–13. [PubMed: 17718394]
16. Edenberg HJ, Foroud T. Genetics of alcoholism. Handb Clin Neurol. 2014; 125:561–71. [PubMed: 25307596]
17. 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; 34:840–52. [PubMed: 20201924]
18. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013; 93:779–97. [PubMed: 24210251]
19. Eyre-Walker A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci U S A. 2010; 107(Suppl 1):1752–6. [PubMed: 20133822]
20. Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P, Herms S, Wodarz N, Soyka M, Zill P, Maier W, Mossner R, Gaebel W, Dahmen N, Scherbaum N, Schmal C, Steffens M, Lucae S, Ising M, Muller-Myhsok B, Nothen MM, Mann K, Kiefer F, Rietschel M. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. Addict Biol. 2012; 17:171–80. [PubMed: 22004471]
21. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, Anton R, Preuss UW, Ridinger M, Rujescu D, Wodarz N, Zill P, Zhao H, Farrer LA. Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry. 2014; 19:41–9. [PubMed: 24166409]
22. George MS, Anton RF, Bloomer C, Teneback C, Drobes DJ, Lorberbaum JP, Nahas Z, Vincent DJ. Activation of prefrontal cortex and anterior thalamus in alcoholic subjects on exposure to alcohol-specific cues. Arch Gen Psychiatry. 2001; 58:345–52. [PubMed: 11296095]
23. Gnirke A, Melnikov A, Maguire J, Rogov P, Leproust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009; 27:182–9. [PubMed: 19182786]
24. Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. Nat Rev Genet. 2005; 6:521–32. [PubMed: 15995696]
25. Hart AB, Kranzler HR. Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. Alcohol Clin Exp Res. 2015
26. Heinzen EL, Neale BM, Traynelis SF, Allen AS, Goldstein DB. The genetics of neuropsychiatric diseases: looking in and beyond the exome. Annu Rev Neurosci. 2015; 38:47–68. [PubMed: 25840007]
27. Hindorff, LA., Macarthur, J., Morales, J., Junkins, HA., Hall, PN., Klemm, AK., Manolio, TA. A Catalog of Published Genome-Wide Association Studies. [Online]. Available: http://www.genome.gov/gwastudiesAccessed
28. Iacono WG, Carlson SR, Taylor J, Elkins IJ, Mcgue M. Behavioral disinhibition and the development of substance-use disorders: findings from the Minnesota Twin Family Study. Dev Psychopathol. 1999; 11:869–900. [PubMed: 10624730]
29. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet. 2013; 92:841–53. [PubMed: 23684009]
30. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B. Genomic surveys by methylation-sensitive SNP analysis identify

sequence-dependent allele-specific DNA methylation. Nat Genet. 2008; 40:904–8. [PubMed: 18568024]

31. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, Mcmanus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–30. [PubMed: 25693563]
32. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012a; 91:224–37. [PubMed: 22863193]
33. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012b; 13:762–75. [PubMed: 22699862]
34. Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. Hum Mol Genet. 2011; 20:528–40. [PubMed: 21081654]
35. Ma Y, Fan R, Li MD. Meta-Analysis Reveals Significant Association of the 3′-UTR VNTR in SLC6A3 with Alcohol Dependence. Alcohol Clin Exp Res. 2016; 40:1443–53. [PubMed: 27219321]
36. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, Mccarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, Mccarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. [PubMed: 19812666]
37. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, Depristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]
38. Mckernan KJ, Peckham HE, Costa GL, Mclaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, Coleman BE, Laptewicz MW, Sannicandro AE, Rhodes MD, Gottimukkala RK, Yang S, Bafna V, Bashir A, Macbride A, Alkan C, Kidd JM, Eichler EE, Reese MG, De La Vega FM, Blanchard AP. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res. 2009; 19:1527–41. [PubMed: 19546169]
39. Meyer JM, Silberg JL, Simonoff E, Kendler KS, Hewitt JK. The Virginia Twin-Family Study of Adolescent Behavioral Development: assessing sample biases in demographic correlates of psychopathology. Psychol Med. 1996; 26:1119–33. [PubMed: 8931158]
40. Miller MB, Basu S, Cunningham J, Eskin E, Malone SM, Oetting WS, Schork N, Sul JH, Iacono WG, Mcgue M. The Minnesota Center for Twin and Family Research genome-wide association study. Twin Res Hum Genet. 2012; 15:767–74. [PubMed: 23363460]
41. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461:272–6. [PubMed: 19684571]

42. Olfson E, Saccone NL, Johnson EO, Chen LS, Culverhouse R, Doheny K, Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Barr RG, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, Van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardia SL, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang JC, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel JA, Rice J, Goate A, Bierut LJ. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. Mol Psychiatry. 2016; 21:601–7. [PubMed: 26239294]
43. Ozsoy S, Durak AC, Esel E. Hippocampal volumes and cognitive functions in adult alcoholic patients with adolescent-onset. Alcohol. 2013; 47:9–14. [PubMed: 23063480]
44. Phillips C. Online resources for SNP analysis: a review and route map. Mol Biotechnol. 2007; 35:65–97. [PubMed: 17401150]
45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]
46. Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, Desrivieres S, Aliev FA, Khan AA, Amin N, Aulchenko YS, Bakalkin G, Bakker SJ, Balkau B, Beulens JW, Bilbao A, De Boer RA, Beury D, Bots ML, Breetvelt EJ, Cauchi S, Cavalcanti-Proenca C, Chambers JC, Clarke TK, Dahmen N, De Geus EJ, Dick D, Ducci F, Easton A, Edenberg HJ, Esk T, Fernandez-Medarde A, Foroud T, Freimer NB, Girault JA, Grobbee DE, Guarrera S, Gudbjartsson DF, Hartikainen AL, Heath AC, Hesselbrock V, Hofman A, Hottenga JJ, Isohanni MK, Kaprio J, Khaw KT, Kuehnel B, Laitinen J, Lobbens S, Luan J, Mangino M, Maroteaux M, Matullo G, Mccarthy MI, Mueller C, Navis G, Numans ME, Nunez A, Nyholt DR, Onland-Moret CN, Oostra BA, O'reilly PF, Palkovits M, Penninx BW, Polidoro S, Pouta A, Prokopenko I, Ricceri F, Santos E, Smit JH, Soranzo N, Song K, Sovio U, Stumvoll M, Surakk I, Thorgeirsson TE, Thorsteinsdottir U, Troakes C, Tyrfingsson T, Tonjes A, Uiterwaal CS, Uitterlinden AG, Van Der Harst P, Van Der Schouw YT, Staehlin O, Vogelzangs N, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Whitfield JB, Wichmann EH, Willemsen G, Witteman JC, Yuan X, Zhai G, Zhao JH, Zhang W, Martin NG, Metspalu A, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proc Natl Acad Sci U S A. 2011; 108:7119–24. [PubMed: 21471458]
47. Shumay E, Fowler JS, Volkow ND. Genomic features of the human dopamine transporter gene and its potential epigenetic States: implications for phenotypic diversity. PLoS One. 2010; 5:e11067. [PubMed: 20548783]
48. 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; 66:773–84. [PubMed: 19581569]
49. Van Den Oord EJ, Sullivan PF. False discoveries and models for gene discovery. Trends Genet. 2003; 19:537–42. [PubMed: 14550627]
50. Van Der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, Depristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 11:11 10 1–11 10 33.
51. Vasiliou SA, Ali FR, Haddley K, Cardoso MC, Bubb VJ, Quinn JP. The SLC6A4 VNTR genotype determines transcription factor binding and epigenetic variation of this gene in response to cocaine in vitro. Addict Biol. 2012; 17:156–70. [PubMed: 21309950]
52. Visscher PM, Brown MA, Mccarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012; 90:7–24. [PubMed: 22243964]
53. Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, Iacono WG, Mcgue M. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. Biol Psychiatry. 2014; 75:783–9. [PubMed: 24094508]

54. Wagner MJ. Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits. Pharmacogenomics. 2013; 14:413–24. [PubMed: 23438888]

55. Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics. 2011; 98:1–8. [PubMed: 21565265]

56. Wrase J, Makris N, Braus DF, Mann K, Smolka MN, Kennedy DN, Caviness VS, Hodge SM, Tang L, Albaugh M, Ziegler DA, Davis OC, Kissling C, Schumann G, Breiter HC, Heinz A. Amygdala volume associated with alcohol abuse relapse and craving. Am J Psychiatry. 2008; 165:1179–84. [PubMed: 18593776]

57. Zuo L, Saba L, Wang K, Zhang X, Krystal JH, Tabakoff B, Luo X. Exome-wide association study of replicable nonsynonymous variants conferring risk for alcohol dependence. J Stud Alcohol Drugs. 2013a; 74:622–5. [PubMed: 23739027]

58. Zuo L, Wang K, Zhang XY, Krystal JH, Li CS, Zhang F, Zhang H, Luo X. NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk gene region specific for alcohol dependence in subjects of European descent. Drug Alcohol Depend. 2013b; 129:254–64. [PubMed: 23455491]

59. Zuo L, Zhang H, Malison RT, Li CS, Zhang XY, Wang F, Lu L, Wang X, Krystal JH, Zhang F, Deng HW, Luo X. Rare ADH variant constellations are specific for alcohol dependence. Alcohol Alcohol. 2013c; 48:9–14. [PubMed: 23019235]

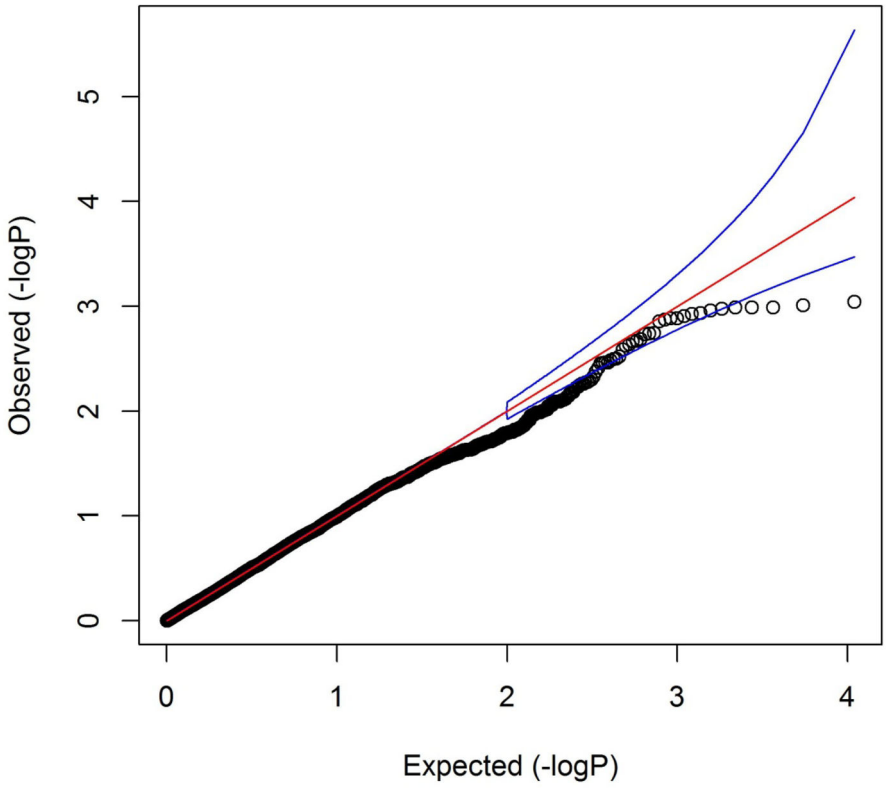

Figure 1.
QQ-Plot of the individual, common variant association meta-analysis results.

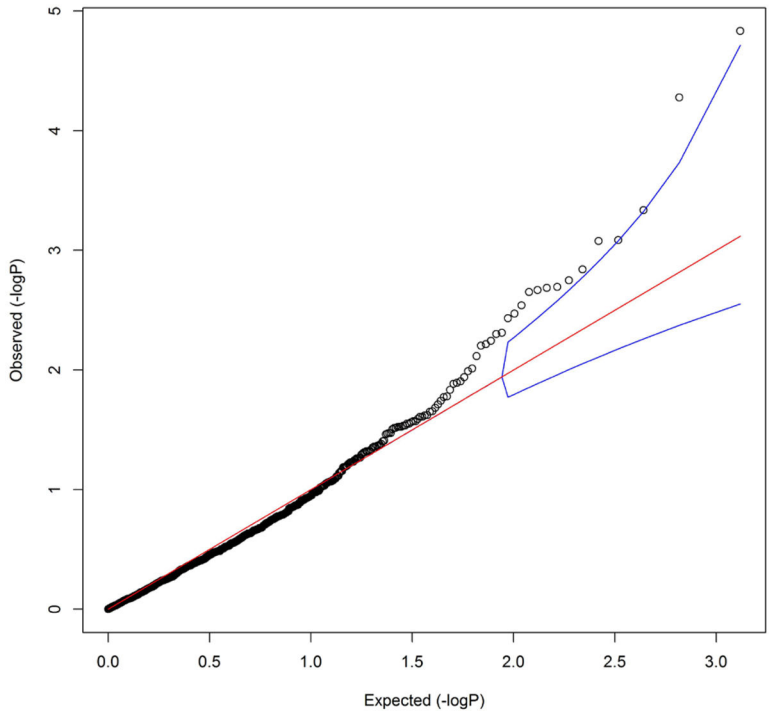

Figure 2.
QQ-Plot of the variant set association results.

Table 1

Descriptive data of sequenced samples.

	VTSABD (n=363)		MTFS (n=443)	
	n	**%**	**n**	**%**
Male	141	38.8	221	49.89
Age (Mean, SD)	24.1	2.11	23.5	1.51
Monozygotic	195	53.7	443	100.0
Caucasian	363	100.0	443	100.0
Use Alcohol Ever (Lifetime)	339	93.9	402	90.74
DSM IV Diagnosis Ever (Lifetime)	39	11.4	58	13.2
DSM IV Symptom Count				
0	252	73.5	246	55.53
1	20	5.83	64	14.45
2	26	7.58	41	9.26
3	22	6.41	40	9.03
4	11	3.21	20	4.51
5	6	1.75	17	3.84
6	3	0.87	9	2.03
7	2	0.58	4	0.90
8	0	0.0	1	0.23
9	1	0.29	1	0.23
DSM IV Symptom (Mean, SD)	0.74	1.48	1.19	1.75

Note: n = sample size; SD = standard deviation

Table 2

Genomic feature set results with observed *q*-values < 0.30

Gene	Feature	N Rare	N Common	P-value	Q-value	Perm. P-value
ADHFE1	Enhancer	1	7	1.47×10^{-05}	0.019	1.19×10^{-04}
ADORA1	TSS Flank - Brain (MFL)	3	5	5.29×10^{-05}	0.035	7.80×10^{-04}
ALDH1A2	Non Coding RNA	6	3	4.61×10^{-04}	0.202	1.32×10^{-03}
ALDH1A2	CpG Shore	5	5	8.20×10^{-04}	0.219	1.95×10^{-03}
ADH1A	TFBS	2	3	8.34×10^{-04}	0.219	7.26×10^{-03}
SLC6A3	POLYPHEN-Damaging	1	1	1.45×10^{-03}	0.266	6.98×10^{-03}
SLC6A3	CpG Shore	12	28	1.78×10^{-03}	0.266	1.95×10^{-03}
ADH5	TSS Flank - Liver	1	3	2.02×10^{-03}	0.266	6.33×10^{-03}
ADH1A	TSS - Liver	2	1	2.06×10^{-03}	0.266	8.36×10^{-03}
SLC6A3	Promoter	4	11	2.15×10^{-03}	0.266	3.64×10^{-03}

Note: “N Rare” and “N Common” is the number of rare variants (MAF < 0.05) and common variants (MAF ≥ 0.05) included in the tested set. “*P*-value” and “*Q*-value” are the observed association *p*-values and *q*-values from the test of whether the set of variants that overlap with the specified genomic feature within the given gene is associated with AD symptom count. “Perm. P-value” are the permutation p-values. “Gene” indicates that the name of the gene the variant set falls within the boundary of as defined by RefSeq. “Feature” describes genomic attributes under consideration. “Promoter” indicates the variant is within 5kb of a transcription start site of the given gene; “Non Coding RNA” indicates a functional RNA molecule that is not translated into a protein; “CpG Shore” is ± 2kb flanking a CpG Island; “POLYPHEN – Deleterious” indicates that the variant is predicted to cause a deleterious amino acid substitution by PolyPhen2(Adzhubei et al., 2010). Chromatin states are indicated by the following format: chromatin state name – tissue type (region if tissue is brain). Possible chromatin states are Active transcription start site (TSS), Flanking active TSS, Transcription at gene 5′ and 3′, Strong transcription, Weak transcription, Genic Enhancer, Enhancer, ZNF genes and repeats, Heterochromatin, Bivalent\Poised TSS, Flanking Bivalent TSS\Enhancer, Bivalent Enhancer, Repressed Polycomb, Weak Repressed Polycomb, and Quiescent. The brain regions examined for chromatin states were: AC – anterior caudate, HM – hippocampus, MFL – mid-frontal lobe, and SN – substantia nigra.