HHS Public Access

Author manuscript

Alcohol Clin Exp Res. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Alcohol Clin Exp Res. 2015 August ; 39(8): 1396–1405. doi:10.1111/acer.12790.

Combined whole methylome and genome-wide association study implicates *CNTN4* in alcohol use

Shaunna L. Clark, Ph.D.[a], **Karolina A. Aberg, Ph.D.**[a], **Srilaxmi Nerella, M.S.**[a], **Gaurav Kumar, Ph.D.**[a], **Joseph L. McClay, Ph.D.**[a], **Wenan Chen, Ph.D.**[b], **Linying Y. Xie, M.S.**[a], **Aki Harada, Ph.D.**[a], **Andrey A. Shabalin, Ph.D.**[a], **Guimin Gao, Ph.D.**[b], **Sarah E. Bergen, Ph.D.**[c,d], **Christina M. Hultman, Ph.D.**[e], **Patrik K.E. Magnusson, Ph.D.**[e], **Patrick F Sullivan, M.D., FRANZCP**[e,f], and **Edwin JCG 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 Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA

[c]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

[d]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

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

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

Abstract

Background—Methylome-wide association studies present a new way to advance the search for biological correlates for alcohol use. A challenge with methylation studies of alcohol involves the causal direction of significant methylation-alcohol associations. One way to address this issue is to combine methylome-wide association study (MWAS) data with genome-wide association study (GWAS) data.

Methods—Here, we combined MWAS and GWAS results for alcohol use from 619 individuals. Our MWAS data was generated by next generation sequencing of the methylated genomic DNA fraction, producing over 60 million reads per subject to interrogate methylation levels at ~27 million autosomal CpG sites in the human genome. Our GWAS included 5,571,786 SNPs imputed with 1000 Genomes.

Results—When combining the MWAS and GWAS data, our top finding was a region in an intron of *CNTN4* ($p = 2.55\text{x}10^{-8}$), located between chr3:2,555,403–2,555,524, encompassing SNPs rs1382874 and rs1382875. This finding was then replicated in an independent sample of 730 individuals. We used bisulfite pyrosequencing to measure methylation and found significant

Corresponding author: Dr. Shaunna L. Clark. Center for Biomarker Research and Personalized 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.

association with regular alcohol use in the same direction as the MWAS (p= 0.021). Rs1382874 and rs1382875 were genotyped and found to be associated in the same direction as the GWAS (p = 0.008 and p = 0.009). After integrating the MWAS and GWAS findings from the replication sample, we replicated our combined analysis finding (p=0.0017) in *CNTN4*.

Conclusions—Through combining methylation and SNP data, we have identified *CNTN4* as a risk factor for regular alcohol use.

Keywords

methylation; next-generation sequencing; GWAS; CNTN4

Introduction

Alcoholism is a disorder characterized by compulsive and uncontrolled consumption of alcohol despite its negative effects on the drinker's health, relationships and social standing. Alcohol use has a clear genetic component (Goldman et al., 2005) and genome-wide association studies (GWAS) have identified a number of putative risk loci (ex: (Edenberg et al., 2010, Schumann et al., 2011)). Most of these genetic variants, however, have modest predictive power, with the exception of Gelernter and colleagues whose study found variants in *ADH1B* to have reasonable predictive power (Gelernter et al., 2014). Current research is turning to epigenetic studies to supplement the search for biological correlates of alcohol use.

One of the most commonly studied epigenetic modifications is DNA methylation. Methylation occurs when a methyl group is attached at the carbon 5 position of a cytosine, and is most often, although not exclusively, found in the sequence context CpG. DNA methylation studies are a promising complement to genetic studies focusing on sequence variation. First, because methylation at critical sites can affect gene expression (Jones, 2012), epimutations are more proximal to the outcome of interest than sequence variants and can therefore have higher predictive power. Second, methylation studies can improve disease understanding as they can, in principle, explain a variety of clinical disease phenomena such as genotype-environment interactions. Third, methylation is potentially reversible. One way to modify methylation sites, either reversing or causing methylation, is through pharmacological interventions (Fuks et al., 2000), which makes methylation sites potential drug targets. Fourth, the translational potential is profound. As they involve the stable methyl-cytosine bond that can be measured cost-effectively in "naked" (histone-free) DNA, these marks are potentially excellent biomarkers for eventual use in clinical settings to improve prognosis, diagnosis, and treatment.

Previous studies have investigated methylation and alcohol abuse or dependence. Studies of global methylation have found that alcoholic patients have higher levels of methylation compared with healthy controls (Bonsch et al., 2004, Thapar et al., 2012). Other studies have found associations between alcohol dependence and CpG methylation in known candidate genes for alcohol addiction, for example: *SLC6A3* (Hillemacher et al., 2009), and *OPRM1* (Zhang et al., 2012). More recent studies have used arrays to show that methylation

of many CpGs across the methylome are associated with alcohol dependence (Philibert et al., 2012, Zhao et al., 2013, Zhou et al., 2011).

These previous studies, however, have several limitations. First, there are over 27 million autosomal CpG sites in the human reference genome. The largest commonly used array, the Infinium HumanMethylation450 Bead Chip by Illumina, only covers 450K CpGs, mostly in gene promoters, suggesting that these array studies only surveyed a small fraction of the potentially methylated portion of the genome. Furthermore, much of the methylation relevant for individual differences may occur outside gene promoters (Irizarry et al., 2009). Second, the sample sizes in the previous studies are small and therefore lack adequate statistical power. For example, the largest sample size reported in the extant literature was 20 sibling pairs (n=40) (Zhao et al., 2013). Third, instead of using primary human tissue, one of these studies used cell lines (Zhou et al., 2011), and another used DNA from Epstein Barr virus (EBV) transformed lymphocytes (Philibert et al., 2012), both of which may show different methylation profiles than primary human tissue. That is, when culturing cells, the conditions cannot fully mimic the human body and therefore DNA from cell lines may not accurately reflect the methylation occurring in humans. This is particularly true for EBV transformed cells, where the transformation itself affects methylation (Aberg et al., 2012a). Fourth, none of these previous studies considers the causal direction of significant methylation-alcohol associations. Specifically, they do not consider whether methylation is the result of alcohol exposure or if it is methylation is contributing to disruptive alcohol behaviors such as alcohol dependence. It is important to disentangle the two because it has implications for how significant methylation-alcohol sites can be utilized. For example, alterations in methylation levels as a result of exposure could potentially be useful early biomarkers of disease and may predict the risk of negative health or behavioral consequences. However, these might be poor drug targets for treating alcohol dependence as changing the methylation at the site will not affect drinking behaviors.

Our study aims to address these limitations. First, we interrogate all ~27 million autosomal CpG sites in a sample size of over 600 individuals. This study constitutes the largest alcohol methylation study to date in terms of the number of CpGs investigated and sample size. To increase confidence in our findings, we used a different platform to generate the methylation data for the replication, specifically the highly quantitative targeted pyrosequencing of bisulfite converted DNA.

The most comprehensive method for ascertaining methylation is whole genome sequencing of bisulfite converted DNA. However, due to the high costs of sequencing entire genomes and the large number of samples required for adequate statistical power in methylome-wide association studies (MWAS), this approach is currently not economically feasible for MWAS. A cost-effective alternative aims to sequence only the methylated part of the genome. Here, DNA is fragmented, the methylated fragments are captured, and then sequenced. The methyl-CpG binding domain (MBD) protein used in this study to capture the methylated genomic fraction strictly binds to methylated cytosines in the sequence context CpG (i.e. it will only bind to CpGs and not just a methylated cytosine). This is not a limitation as, with few exceptions (Lister et al., 2013), in most mammalian somatic tissues, DNA methylation occurs >99.75% of the time at CpG dinucleotides (Bernstein et al., 2007).

MBD-seq, next-generation sequencing combined with the use of the MBD protein, has already been demonstrated to be sensitive and capable of identifying differentially methylated regions (DMRs) (Serre et al., 2010), detecting previously reported robust associations (McClay et al., 2013) and producing findings that replicate when using “gold standard” technologies (Aberg et al., 2014). MBD-seq is quantitative, some subjects will have zero reads covering a CpG and others may not, but does not yield absolute methylation levels. This is, however, not critical for MWAS because we focus on differences between cases and controls or correlations with outcome variables. Absolute methylation levels are subtracted out from these differences and will not affect correlations that remain unchanged by linear transformations. The resolution of MBD-seq is approximately equal to the fragment size which is high enough to pinpoint a limited number of CpGs where the association occurs. Because short range correlations in the methylation status of CpGs seem pervasive across the methylome (Aberg et al., 2012b), in many cases it may not even be possible to improve the resolution obtained with MBD-seq as neighboring CpGs may have similar methylation statuses. All these properties make MBD-seq an efficient screening tool for large-scale MWAS to indentify DMRs that can then be investigated with more targeted assays.

Our approach to gain traction on the causal direction of effect is to combine MWAS data with GWAS data from the same samples. Because alcohol cannot alter sequence variation, the direction of effect in GWAS is assumed to be from the single nucleotide polymorphism (SNP) to alcohol use. Thus, for loci showing overlapping results between the MWAS and GWAS results, it is less likely that alcohol exposure effects would be driving the results.

MATERIALS AND METHODS

Discovery and Replication Sample

Table 1 describes the discovery and replication samples. Subjects from both samples were selected at random from national population registers in Sweden (Bergen et al., 2012). The subjects are essentially a population sample except that no one has a diagnosis of schizophrenia, which because of the low prevalence rate of schizophrenia, should not substantially alter the results. Current or past substance use was not an inclusion\exclusion criterion for study participation. All procedures were approved by ethical committees in Sweden and in the USA, and all subjects provided written informed consent.

The discovery sample included 619 individuals who responded to the binary alcohol use question which asked whether participants had ever considered themselves to have consumed alcohol regularly during their lifetime. This question does not distinguish between regular use and alcohol dependency. Alcohol dependency information was not collected on these participants. In the discovery sample, 94% of subjects self-identified as ever using alcohol regularly. The replication sample of 730 independent subjects, of whom 93% self-identified as ever using alcohol regularly, was used to replicate key findings. DNA was extracted from the buffy coat of whole blood donated at local medical facilities.

Stage 1: Discovery

MWAS: MBD-seq—Our MWAS sequencing approach (Aberg et al., 2012b) and computational analysis methods (Chen et al., 2013, van den Oord et al., 2013) have been described previously. Briefly, genomic DNA was fragmented with ultrasonication (Covaris) to a median length of ~150 bp, we used MethylMiner (Invitrogen, Carlsbad, CA) which employs MBD protein-based enrichment of the methylated DNA fraction, followed by single-end sequencing (50 bp reads) on the SOLiD platform (Life Technologies). As binding is better in CpG dense regions (Harris et al., 2010), we use an existing protocol variant that elutes the captured methylated fraction with 0.5 M NaCl to increase the relative number of fragments from CpG poor regions (Aberg et al., 2012b), which otherwise would not be as well covered (Bock et al., 2010), and improve coverage of the methylome. Library construction and next-generation sequencing was performed following the standard protocol for bar-coded fragment libraries (Life Technologies). Prior to alignment, but after deleting poor quality reads (> 2 missing color calls), we obtained an average of 67.3 million (SD=26.9 million) reads per sample.

MBD-seq can be used to obtain a relative quantitative methylation measure by summing the number of fragments covering each CpG site. However, methylation of any CpG in the entire fragment, not just the sequenced 50bp, could lead to its capture by MBD protein. Therefore, we estimated the fragment size distribution for each sample from the sequencing data, based on the distribution of reads around isolated CpGs. This non-parametric method has been validated against paired-end libraries where fragment size is known (van den Oord et al., 2013). The sample-specific estimated fragment size distribution was then used to calculate the probability for each read that the fragment it is tagging covers the CpG under consideration. Coverage estimates for each of the autosomal CpGs in the reference genome were then calculated for each subject by taking the sum of the probabilities that all fragments in its neighborhood cover the CpG (van den Oord et al., 2013).

To reduce the size of the data set, the CpGs were adaptively combined by collapsing highly inter-correlated coverage estimates at adjacent CpG sites into a single mean coverage estimate (Aberg et al., 2012a). This resulted in 5,074,560 CpG “blocks”. Prior to association testing, we identified 800,140 blocks with very low (< 99% of noise) levels of coverage. Eliminating these likely non-methylated sites left 4,274,420 blocks for the MWAS.

MWAS association testing—To test for association, we regressed the coverage of the remaining “blocks” on regular alcohol use. In MWAS, there are multiple differences (e.g. diet and medication use) between individuals that are unrelated to alcohol use but may affect the methylome. A variety of efforts were made to control for such confounders which are described in the Supplemental Material.

To account for multiple testing, we controlled the false discovery rate (FDR) at the 0.1 level (van den Oord and Sullivan, 2003). This means that on average 10% of the methylation sites declared significant are expected to be false discoveries. Operationally, the FDR was controlled using *q*-values that are FDRs calculated using the *p*-values of the individual tests as thresholds for declaring significance (Storey and Tibshirani, 2003).

GWAS—Details of the genotyping and quality control methods have been previously described (Bergen et al., 2012). Briefly, most subjects were genotyped with Affymetrix 6.0 chips (Affymetrix, Santa Clara, CA, USA) with the remainder genotyped with the Affymetrix 5.0. SNPs with minor allele frequencies less than 1%, or departure from Hardy-Weinberg equilibrium ($p < 1x10^{-6}$) were excluded. Individuals with genotype call rates less than 95% and a randomly selected member of any pair of subjects with high relatedness ($\hat{\pi} > 0.20$) were excluded.

Imputation was conducted using MaCH (Li et al., 2010) for the genotype phasing and Minimac (Howie et al., 2012) for the imputation. Our genotypes were imputed against autosomal data from 1000 Genomes (Version 3) and we retained only SNP dosages imputed with high confidence (INFO > 0.50). After imputation and quality control, there were 5,571,786 SNPs included in the analysis. Association testing was conducted by regressing regular alcohol use on the imputed SNP values with one multidimensional scaling (MDS) component, age and sex included as covariates to control for ancestry, age effects and gender differences. As in the MWAS, we controlled the FDR at the 0.1 level.

Combined Analysis—We used Fisher's method (Fisher, 1948) to combine the MWAS and GWAS results. Because there are multiple mechanisms that may explain the associations for both CpGs and SNPs, we chose to use Fisher's method because it does not require any specific mechanism and can therefore capture all mechanisms that may explain the associations. Fisher's method assumes that the data being combined are tested under the same null hypothesis of no effect of the locus. Under the null hypothesis, the tests of the MWAS and GWAS data are independent unless their errors are correlated, which is unlikely because the data were produced using two different platforms. The independence of error terms would not hold if the alternative were true, which is not relevant for the current analyses. To show that when testing under the null the *p*-values are independent, we correlated the *p*-values from the MWAS and GWAS and obtained a Pearson correlation of 0.0167 which suggests that the *p*-values are uncorrelated. Additionally, a plot of the MWAS vs. the GWAS *p*-values (Figure S3) showed no discernible relationship between the two. Both pieces of evidence demonstrate that under the null, the tests are independent.

Fisher's method first takes the natural logarithm of each *p*-value, multiplies each result by -2, and then sums them. The resulting sum is distributed as a chi-square statistic with 2L degrees of freedom, where L is the number of *p*-values. Only SNPs that were located within the boundaries of a methylation block were used. In order to be included in the combined analysis, a site must have *p*-value < 0.01 on both the MWAS and GWAS to avoid an extreme *p*-value on only one platform dominating the overall results. We tested several other potential p-value thresholds for inclusion into the combined analysis, specifically 0.10, 0.05, 0.01, and 0.001, and found the results to be robust against different thresholds.

Step 2: Replication

Pyrosequencing—For replication purposes we used pyrosequencing, which allows for targeted sequencing of bisulfite converted DNA with high quantitative accuracy (Reed et al., 2010). Genomic DNA was bisulfite converted using Epitect 96 (Qiagen) and reactions were

carried out using the PyroMark system according to standard protocols (Qiagen). Table S4 provides primer sequences. Controls including five DNA samples with known methylation levels (0%, 25%, 50% 75% and 100% methylation, created using methylated (#59665) and unmethylated (#59655) EpiTect Control DNA) were run for each assay and at least two controls of known methylation levels were included on each plate. We used logistic regression to test for association between regular alcohol use and methylation in the pyrosequencing data. Age, sex and plate indicator variables were included in the logistic regression to control for age and sex differences, and potential batch effects.

RESULTS

Discovery stage

MWAS with alcohol use—We used the quantitative coverage estimates for ~4.3 million CpG blocks as input for multiple regression analyses to test for association with regular alcohol use. A variety of efforts were made to control for potential confounders including controlling for age and sex, regressing out potential assay-related technical artifacts, and incorporating principal component analysis (PCA) to capture any remaining unmeasured potential confounders. The quantile-quantile (QQ) plot for this analysis (Figure 1A) indicates an enrichment of small *p*-values with little inflation (λ=1.04). Our test statistic inflation parameter λ of 1.04 was higher compared to what is commonly observed in GWAS. This lambda is unlikely an artifact. After performing a square root transformation to normalize the data and mitigate the effects of possible outliers, λ did not change. Instead, this λ reflects that methylation studies are more akin to gene expression studies that typically show many correlated effects with relatively large effect sizes.

Thirty-three DMRs passed our FDR threshold of *q*-value < 0.1 for genome-wide significance ensuring that only 10% of the findings are expected to be false discoveries. Table S2 provides details for all 33 DMRs below the FDR. The top finding ($p = 1.93\text{x}10^{-9}$, $q = 0.008$) was a block with two CpGs located in an intergenic region of chromosome 20.

GWAS with alcohol use—We performed a GWAS for regular alcohol use with 5,571,786 SNPs, imputed using 1000 Genomes, with one MDS component covariate to control for population stratification. A QQ plot (Figure 1B) and λ of 1.00 showed no inflation of the test statistics. There were 144 SNPs that reached our genome-wide significance threshold of $q < 0.1$, with a minimum *p*-value of $1.22\text{x}10^{-9}$ (a full list of results with $q < 0.1$ is given in the Supplementary Material). As 1000 Genomes imputation results in a dense SNP panel, many of the significant SNPs were in high linkage disequilibrium (LD) with each other meaning that the number of unique associated loci is less than the number of associated SNPs. For example, there were 20 SNPs in high LD ($0.96 < r^2 < 0.99$) located in an intron of *CNTN4* on chromosome 3 and 8 SNPs on chromosome 12 found in an intron of *CHPT1*.

Combined MWAS and GWAS Results—We combined our MWAS and GWAS results using Fisher's method (Fisher, 1948). The QQ-plot for the combined analysis (Figure 1C) shows that there is an enrichment of small *p*-values and there is little evidence of inflation (λ =1.008). There were 17 signals with a $q < 0.1$ (Table 2; Figure S2). The top two results from

the combined analysis are two SNPs overlapping with a methylation block located in an intron of *CNTN4*. Only one of the top 17 findings was located in a CpG island or shore, 2000 bp flanking region of a CpG island. Other top findings were located in introns or DNAse clusters (Table 2).

Replication of CNTN4

Given the converging evidence between the MWAS and GWAS results for *CNTN4* on chromosome 3, we conducted further assays and analyses of this locus in an independent sample of 730 individuals (Table 1). First, we generated the methylation data using pyrosequencing of bisulfite converted DNA. The region implicated in *CNTN4* on chromosome 3 by the combined analysis had two CpGs at coordinates 2,555,483 and 2,555,524 that were highly correlated with each other. As it is practically impossible to design a high quality pyrosequencing assay that would target both these CpG sites, we instead used a high quality assay that targeted only the CpG at 2,555,524. Given that the two sites are highly correlated, the methylation measure from this assay should give a good estimate of the methylation levels at both sites. The regression results indicated that the CpG site was significantly associated with regular alcohol use ($p = 0.021$) and in the same direction as the discovery MWAS ($\beta_{discovery} = -2.71$; $\beta_{replication} = -0.06$). These results remained significant (Table S5) even when regressing out other substance use covariates that are often comorbid with alcohol use (i.e. smoking and narcotic use).

We also replicated our GWAS findings. GWAS data from Affymetrix 6.0 chips were available for the replication sample and replication was carried out in the same manner as the discovery GWAS. Of the 144 significant SNPs in the discovery GWAS, only 16 were considered to be replicated (Table S3) as they had a p-value < 0.05 and the same direction of effect in both the discovery and replication samples. All 16 of these SNPs were located in *CNTN4* and are in high LD with one another. Considering specifically rs1382874 and rs1382875, the two SNPs implicated in the combined analysis *CNTN4* finding, both SNPs were associated with regular alcohol use ($p = 0.008$ and $p = 0.009$, respectively) in the replication even after controlling for multiple testing. Two additional SNPs on chromosome 14 had p-values < 0.05, but the direction of effect was not the same in the replication as in the discovery GWAS, and therefore we considered these SNPs to not have replicated.

Following the same procedure as in the discovery sample, we then replicated our combined GWAS and MWAS finding, by using Fisher's method to combine the results. The combined one-sided *p*-values were 0.0018 and 0.0017 for rs1382874 and rs1382875, respectively. In contrast to the GWAS which implicated 16 SNPs, the combined analysis only implicated two SNPS. This is likely due to correlations among methylation sites being much more localized when compared to LD between SNPs (Kumar et al., 2015).

Bioinformatic investigations of these two SNPs showed that they were located in an intron of *CNTN4*. These SNPs, rs1382874 and rs1382875, are 102 and 41 base pairs away, respectively, from the CpG site that was pyrosequenced. The functional relevance of these two SNPs was investigated using the SNP Function Prediction (FuncPred) (Xu and Taylor, 2009) (http://snpinfo.niehs.nih.gov). The results indicated that both SNPs may have regulatory potential as predicted by the ESPERR (evolutionary and sequence pattern

extraction through reduced representations) method (Taylor et al., 2006). We also examined functional potential by seeing if the two CpGs in the region were encompassed by transcription factor (TF) binding sites. We used empirical data for TF binding from the ENCODE project, in addition to conserved TF recognition sequences across humans and rodents from the TRANSFAC database (Matys et al., 2006). We included a flanking region of ± 250 bp to account for the maximum length of sequenced fragments in our MBD-seq experiments around the two CpGs that bounded the region at chr3:2,555,403 and chr3: 2,555,524. This region encompassed several TFBS according to TRANSFAC. These included myocyte-enhancer factor 2 (MEF2), TATA-binding protein (TBP), octamer transcription factor (OCT1) and signal transducer and activator of transcription 3 (STAT3). These sites were not detected with ENCODE empirical data.

We then investigated possible mechanisms through which *CNTN4* may affect regular alcohol use. Given that both methylation and SNPs are involved, one possibility is that there is a CpG-SNP, which is a SNP that creates or destroys CpGs. Rs1382875 directly overlaps with a CpG and can therefore be considered a CpG-SNP. We tested whether rs1382875 was operating under a CpG-SNP mechanism by interrogating whether methylation was associated with the number of copies of the G allele, which creates a CpG site, an individual had. The regression indicated that the CpG-SNP is potentially regulating methylation in the region ($p = 0.047$).

DISCUSSION

By integrating MWAS and GWAS data were able to identify two SNPs in *CNTN4* that influence regular alcohol use. Even without considering the methylation data, the GWAS results implicated 16 SNPs in *CNTN4* spanning a 7Kb region. However, using both GWAS and methylation data adds additional confidence to the robustness of our findings as the two different technologies converging on the same solution reduces platform specific errors (Niculescu et al., 2000). Furthermore, it enabled us to fine-map the location of the putative casual sites from 16 SNPs implicated by the GWAS to two potential SNPs, because in contrast to LD between SNPs, correlations among methylation sites tend to be much more localized (Kumar et al., 2015).

CNTN4 (contactin 4) locus on chromosome 3p26.3 is a member of the contactin family of immunoglobulins that are expressed predominantly in the central nervous system (Shimoda and Watanabe, 2009). The genes in this family function as cell adhesion molecules, guiding axon growth during development, and have roles in modulating synaptic transmission and plasticity (Dityatev et al., 2008). Recently, two SNPs at *CNTN4* were associated with measures of human brain network connectivity (Jahanshad et al., 2013) indicating that genetic variants at *CNTN4* have the potential to affect brain function. Indeed, evidence has shown that *CNTN4* is associated with developmental disorders including 3p deletion syndrome (Fernandez et al., 2004) and autism (Roohi et al., 2009). In summary, evidence indicates that *CNTN4* is of critical importance for human brain development and function and, when disrupted, can lead to neurodevelopmental disorders.

Some previous evidence also indicates a role for *CNTN4* in alcohol use. An early GWAS reported an association of a pathway containing *CNTN4* and level of response to alcohol, a measure inversely correlated with problem drinking behavior (Joslyn et al., 2010). A later GWAS found SNPs in close proximity to *CNTN4* to be associated with co-morbid alcohol dependence in bipolar patients (Kerner et al., 2011). Studies have linked *CNTN4* to olfaction and development of olfactory neurons (Mimmack et al., 1997). Therefore, taste preference for alcohol could be speculated as a mechanism by which *CNTN4* could influence drinking. Although further research is needed to define specific mechanisms by which *CNTN4* influences regular alcohol use, it is a plausible candidate.

There are several scenarios that can explain the relationship between genetic variation, methylation and alcohol use we see in our study. One mechanism involves SNPs that create or destroy CpGs, called CpG-SNPs. The signal in *CNTN4* was produced by CpG-SNP which suggests that allele frequency differences between cases and controls at the CpG-SNP may also influence methylation differences there. The statistical power is generally low to detect methylation differences caused by CpG-SNPs. In the case of rs1382875 in the replication study, with minor allele frequencies of 0.25 and 0.37 in cases and controls respectively and 0.12% of the variance in methylation explained by the SNP, a sample of just over 1300 is required to achieve 80% power to detect methylation differences. Our sample size of 619 makes it possible that we have detected methylation differences related to allele frequency differences between cases and controls. However, we cannot rule out an alternative mechanism in which methylation plays an active role in regular alcohol use. For example, similar to SNPs, methylation in critical sites can inhibit the binding of TFs to their recognition elements (Prendergast and Ziff, 1991), resulting in gene silencing.

Assuming that the methylation signal in the region is tagging a causal SNP, this motivates the investigation for potential functional SNPs in the region. We examined the functional relevance of SNPs within the suggested region and found that two SNPs, rs1382874 and rs1382875, may have regulatory potential, as predicted using the ESPERR (Taylor et al., 2006). This method allows for classification of functional versus neutral genomic regions, based on cross-species sequence homology, GC content and other metrics. It has been shown to accurately classify a range of functional elements, including binding sites for cis-acting factors, DNaseI hypersensitive sites and developmental enhancers.

As the signal may involve an active role of methylation, we also examined genomic features of the implicated region. We found that the methylation site encompassed several TF binding sites which is relevant because methylation in critical sites can inhibit the binding of TFs to their recognition elements (Prendergast and Ziff, 1991), resulting in gene silencing. Specifically, the methylation site overlapped with TF binding sites for MEF2, TBP, OCT1, and STAT3, according to TRANSFAC (Matys et al., 2006). MEF2 is a regulator of cellular differentiation and plays a critical role in embryonic development, in addition to involvement in neuronal differentiation and regulation of stress response in adults (Potthoff and Olson, 2007). TBP is a generic TF that is part of the basic transcriptional machinery, while OCT1 is also widely expressed but necessary for cellular stress response. STAT3 has a role in regulating synaptic plasticity in the central nervous system (Nicolas et al., 2013). Notably, reduced expression of STAT3 has been shown in peripheral blood of non-human

primates following chronic ethanol consumption (Asquith et al., 2014). The overlap of the region of interest with TF binding sites suggests a potential mechanism by which methylation may influence alcohol use.

While the methylation finding in blood is interesting because of the potential to indicate liability for alcohol use or other alcohol-related health outcomes such as inflammation, it is important to examine how methylation may play a role in other tissues which might be more relevant for addiction, such as brain. Whether blood is the most appropriate tissue for methylation addiction studies has been a discussion point (Wong et al., 2011) because methylation patterns can be tissue specific. If the methylation signal is merely tagging differences in allele frequencies the issue of using blood is less relevant in the context of our *CNTN4* finding because methylation would play no causal role and merely tags a SNP effect that is not tissue specific. If, however, there is a causal role of the methylation at the CpG-SNP site the issue is relevant. Several studies have found a high degree of overlap between methylation in blood and brain tissue (Aberg et al., 2013, Davies et al., 2012). These studies, however, considered all CpGs, not specifically CpG-SNPs. We therefore conducted a study to examine the overlap in methylation status of CpG-SNPs in blood and brain. The blood methylation data came from an MBD-seq study of 1459 blood samples (Aberg et al., 2014). The 75 brain samples from the prefrontal cortex, Broadmann Area 10, on which we performed MBD-seq, were obtained from the Stanley Medical Research Institute. Using the 99th percentile of coverage at non-CpGs sites (i.e. MBD-seq captures only CpG methylation, as these sites are at least 400 bp away from the nearest CpG they cannot be methylated and any reads covering them represent “noise”) as a threshold below which we considered sites to be unmethylated, we found that 68.3% and 67.7% of CpG-SNPs where likely methylated in blood and brain, respectively. The overlap of methylated sites in blood and brain was high, as 94% of the methylated sites in blood were also methylated in brain. One caveat to these results is that methylation patterns in the prefrontal cortex may not be similar to patterns in other brain regions. Research has shown that there are some differences in individual sites between brain regions, but overall the correlation between methylation in different brain regions is as high as 0.76 (Davies et al., 2012). This evidence suggests that the methylation of CpG-SNPs in blood likely mirror the methylation of the same site in brain. Thus, it is not unreasonable to hypothesize about casual processes in brain based on our findings in blood.

In addition to the MWAS results that overlapped with the GWAS results, there were several MWAS results that did not overlap. We further investigated these top non-overlapping MWAS results using pathway analyses, but found that they did not organize into significant pathways. This is consistent with the idea of alcohol exposure causing methylation changes in blood that may not result in coordinated, functional biological processes. Even though these significant methylation-alcohol associations may reflect alcohol exposure, they have the potential to be useful as biomarkers of exposure to alcohol, and potential predictors of health and risk outcomes associated with detrimental alcohol use. Further investigation of these signals is therefore warranted.

Through combining methylation and GWAS data, we have identified *CNTN4* as a risk factor for alcohol use. Strengths of the methylation aspect of this study included the utilization of a

much larger sample size compared to previous studies with arrays. Also, our use of next-generation sequencing results in an improvement in the number of sites interrogated over arrays. In addition to replicating our separate MWAS and GWAS results, we were able to replicate our combined *CNTN4* finding in an independent sample using a different technology to measure methylation, providing further evidence that this is a true finding. There may be two possible mechanisms which explain the *CNTN4* finding: the methylation differences are tagging allele frequency differences at a CpG-SNP or there is an active role for methylation influencing alcohol use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the National Institute on Alcohol Abuse and Alcoholism (grant K01AA021266) and the National Institute of Mental Health (grants RC2MH089996 & 1R01MH097283). The present study is part of a larger project entitled '*A Large-Scale Schizophrenia Association Study in Sweden'* that is supported by grants from the NIMH and the Stanley Medical Research Institute. Institutions involved in this project are: Karolinska Institutet, Icahn School of Medicine at Mount Sinai, University of North Carolina at Chapel Hill, Virginia Commonwealth University, Broad Institute, and the US National Institute of Mental Health.

Library construction and next generation sequencing was performed by EdgeBio, Gaithersburg, MD.

References

Aberg K, Khachane AN, Rudolf G, Nerella S, Fugman DA, Tischfield JA, Van Den Oord EJ. Methylome-wide comparison of human genomic DNA extracted from whole blood and from EBV-transformed lymphocyte cell lines. Eur J Hum Genet. 2012a; 20:953–5. [PubMed: 22378283]

Aberg KA, Mcclay JL, Nerella S, Clark S, Kumar G, Chen W, Khachane AN, Xie L, Hudson A, Gao G, Harada A, Hultman CM, Sullivan PF, Magnusson PK, Van Den Oord EJ. Methylome-Wide Association Study of Schizophrenia: Identifying Blood Biomarker Signatures of Environmental Insults. JAMA Psychiatry. 2014

Aberg KA, Mcclay JL, Nerella S, Xie LY, Clark SL, Hudson AD, Bukszar J, Adkins D, Consortium SS, Hultman CM, Sullivan PF, Magnusson PK, Van Den Oord EJ. MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case--control samples. Epigenomics. 2012b; 4:605–21. [PubMed: 23244307]

Aberg KA, Xie LY, Mcclay JL, Nerella S, Vunck S, Snider S, Beardsley PM, Van Den Oord EJ. Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. Epigenomics. 2013; 5:367–77. [PubMed: 23895651]

Asquith M, Pasala S, Engelmann F, Haberthur K, Meyer C, Park B, Grant KA, Messaoudi I. Chronic ethanol consumption modulates growth factor release, mucosal cytokine production, and microRNA expression in nonhuman primates. Alcohol Clin Exp Res. 2014; 38:980–93. [PubMed: 24329418]

Bergen SE, O'dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, Moran JL, Chambert KD, Handsaker RE, Backlund L, Osby U, Mccarroll S, Landen M, Scolnick EM, Magnusson PK, Lichtenstein P, Hultman CM, Purcell SM, Sklar P, Sullivan PF. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry. 2012; 17:880–6. [PubMed: 22688191]

Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell. 2007; 128:669–81. [PubMed: 17320505]

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol. 2010; 28:1106–14. [PubMed: 20852634]

Bonsch D, Lenz B, Reulbach U, Kornhuber J, Bleich S. Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. J Neural Transm. 2004; 111:1611–6. [PubMed: 15565495]

Chen W, Gao G, Nerella S, Hultman CM, Magnusson PK, Sullivan PF, Aberg KA, Van Den Oord EJ. MethylPCA: a toolkit to control for confounders in methylome-wide association studies. BMC Bioinformatics. 2013; 14:74. [PubMed: 23452721]

Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, Al-Sarraj S, Dobson R, Schalkwyk LC, Mill J. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 2012; 13:R43. [PubMed: 22703893]

Dityatev A, Bukalo O, Schachner M. Modulation of synaptic transmission and plasticity by cell adhesion and repulsion molecules. Neuron Glia Biol. 2008; 4:197–209. [PubMed: 19674506]

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]

Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, Lifton RP, State MW. Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet. 2004; 74:1286–93. [PubMed: 15106122]

Fisher RA. Combining independent tests of significance. American Statistician. 1948; 2:30.

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet. 2000; 24:88–91. [PubMed: 10615135]

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]

Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. Nat Rev Genet. 2005; 6:521–32. [PubMed: 15995696]

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, Ecker JR, Li W, Farnham PJ, Waterland RA, Meissner A, Marra MA, Hirst M, Milosavljevic A, Costello JF. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010; 28:1097–105. [PubMed: 20852635]

Hillemacher T, Frieling H, Hartl T, Wilhelm J, Kornhuber J, Bleich S. Promoter specific methylation of the dopamine transporter gene is altered in alcohol dependence and associated with craving. J Psychiatr Res. 2009; 43:388–92. [PubMed: 18504048]

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012; 44:955–9. [PubMed: 22820512]

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, Feinberg AP. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009; 41:178–86. [PubMed: 19151715]

Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, Jack CR Jr, Saykin AJ, Green RC, Weiner MW, Medland SE, Montgomery GW, Hansell NK, Mcmahon KL, De Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. Proc Natl Acad Sci U S A. 2013; 110:4768–73. [PubMed: 23471985]

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484–92. [PubMed: 22641018]

Joslyn G, Ravindranathan A, Brush G, Schuckit M, White RL. Human variation in alcohol response is influenced by variation in neuronal signaling genes. Alcohol Clin Exp Res. 2010; 34:800–12. [PubMed: 20201926]

Kerner B, Lambert CG, Muthen BO. Genome-wide association study in bipolar patients stratified by co-morbidity. PLoS One. 2011; 6:e28477. [PubMed: 22205951]

Kumar G, Clark SL, Mcclay JL, Shabalin AA, Adkins DE, Xie L, Chan R, Nerella S, Kim Y, Sullivan PF, Hultman CM, Magnusson PK, Aberg KA, Van Den Oord EJ. Refinement of schizophrenia GWAS loci using methylome-wide association data. Hum Genet. 2015; 134:77–87. [PubMed: 25284466]

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR. Global epigenomic reconfiguration during mammalian brain development. Science. 2013; 341:1237905. [PubMed: 23828890]

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–10. [PubMed: 16381825]

Mcclay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, Hudson AD, Harada A, Hultman CM, Magnusson PK, Sullivan PF, Van Den Oord EJ. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. Hum Mol Genet. 2013

Mimmack ML, Saito H, Evans G, Bresler M, Keverne EB, Emson PC. A novel splice variant of the cell adhesion molecule BIG-2 is expressed in the olfactory and vomeronasal neuroepithelia. Brain Res Mol Brain Res. 1997; 47:345–50. [PubMed: 9221934]

Nicolas CS, Amici M, Bortolotto ZA, Doherty A, Csaba Z, Fafouri A, Dournaud P, Gressens P, Collingridge GL, Peineau S. The role of JAK-STAT signaling within the CNS. JAKSTAT. 2013; 2:e22925. [PubMed: 24058789]

Niculescu AB 3rd, Segal DS, Kuczenski R, Barrett T, Hauger RL, Kelsoe JR. Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. Physiol Genomics. 2000; 4:83–91. [PubMed: 11074017]

Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SR. The impact of recent alcohol use on genome wide DNA methylation signatures. Front Genet. 2012; 3:54. [PubMed: 22514556]

Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. Development. 2007; 134:4131–40. [PubMed: 17959722]

Prendergast GC, Ziff EB. Methylation–sensitive sequence-specific DNA binding by the c-Myc basic region. Science. 1991; 251:186–9. [PubMed: 1987636]

Reed K, Poulin ML, Yan L, Parissenti AM. Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. Anal Biochem. 2010; 397:96–106. [PubMed: 19835834]

Roohi J, Montagna C, Tegay DH, Palmer LE, Devincent C, Pomeroy JC, Christian SL, Nowak N, Hatchwell E. Disruption of contactin 4 in three subjects with autism spectrum disorder. J Med Genet. 2009; 46:176–82. [PubMed: 18349135]

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]

Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 2010; 38:391–9. [PubMed: 19906696]

Shimoda Y, Watanabe K. Contactins: emerging key roles in the development and function of the nervous system. Cell Adh Migr. 2009; 3:64–70. [PubMed: 19262165]

Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100:9440–5. [PubMed: 12883005]

Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res. 2006; 16:1596–604. [PubMed: 17053093]

Thapar M, Covault J, Hesselbrock V, Bonkovsky HL. DNA methylation patterns in alcoholics and family controls. World J Gastrointest Oncol. 2012; 4:138–44. [PubMed: 22737275]

Van Den Oord EJ, Bukszar J, Rudolf G, Nerella S, Mcclay JL, Xie LY, Aberg KA. Estimation of CpG coverage in whole methylome next-generation sequencing studies. BMC Bioinformatics. 2013; 14:50. [PubMed: 23398781]

Van Den Oord EJ, Sullivan PF. False discoveries and models for gene discovery. Trends Genet. 2003; 19:537–42. [PubMed: 14550627]

Wong CC, Mill J, Fernandes C. Drugs and addiction: an introduction to epigenetics. Addiction. 2011; 106:480–9. [PubMed: 21205049]

Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009; 37:W600–5. [PubMed: 19417063]

Zhang H, Herman AI, Kranzler HR, Anton RF, Simen AA, Gelernter J. Hypermethylation of OPRM1 promoter region in European Americans with alcohol dependence. J Hum Genet. 2012; 57:670–5. [PubMed: 22914673]

Zhao R, Zhang R, Li W, Liao Y, Tang J, Miao Q, Hao W. Genome-wide DNA methylation patterns in discordant sib pairs with alcohol dependence. Asia Pac Psychiatry. 2013; 5:39–50. [PubMed: 23857790]

Zhou FC, Balaraman Y, Teng M, Liu Y, Singh RP, Nephew KP. Alcohol alters DNA methylation patterns and inhibits neural stem cell differentiation. Alcohol Clin Exp Res. 2011; 35:735–46. [PubMed: 21223309]

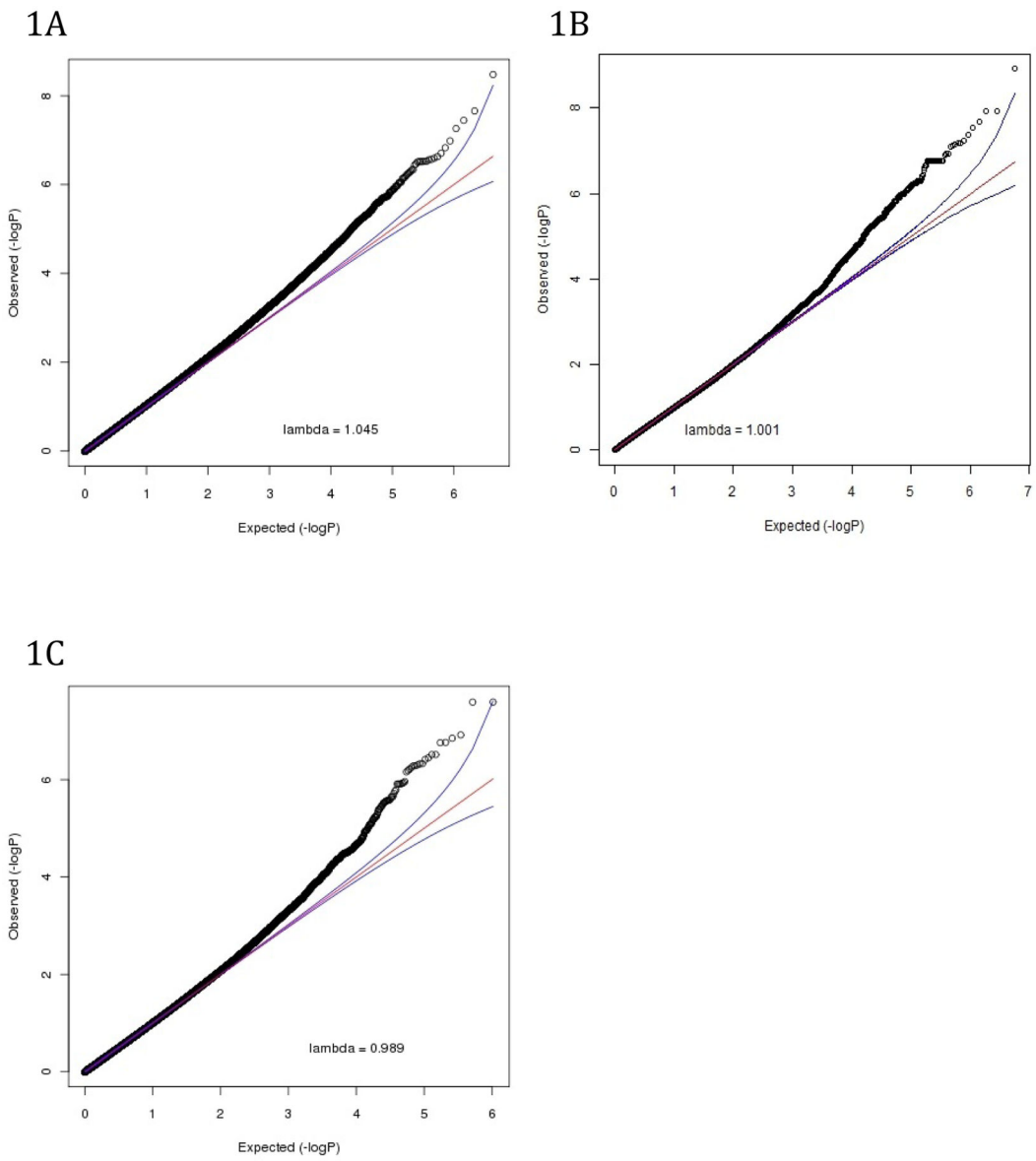

Figure 1.
QQ plots for A) MWAS, B) GWAS, and C) Combined MWAS and GWAS results

Alcohol Clin Exp Res. Author manuscript; available in PMC 2016 August 01.

Table 1

Descriptive data for the discovery and replication samples*.

	Discovery (N = 619)				Replication (N = 730)			
	Regular User (N = 580)		Never Regular User (N = 39)		Regular User (N = 689)		Never Regular User (N = 41)	
Variable	**%**	**% Complete**	**%**	**% Complete**	**%**	**%Complete**	**%**	**%Complete**
Male	53.4	100.0	56.4	100.0	59.1	100.0	58.5	100.0
Finnish Ancestors	7.51	85.0	13.3	76.9	0.0	100.0	0.0	100.0
Smokes	55.3	16.2	36.4	28.2	12.8	100.0	4.9	100.0
Narcotics	5.2	100.0	7.7	100.0	5.2	100.0	5.0	97.6
Epilepsy	0.5	99.3	0.0	100.0	0.0	100.0	0.0	100.0
Diabetes	2.9	99.1	2.6	100.0	0.0	100.0	0.0	100.0
Hyperthyroid	1.6	95.6	0.0	94.8	0.0	100.0	0.0	100.0
Hypothyroid	2.3	95.8	5.3	97.4	0.0	100.0	0.0	100.0
Autoimmune Disorder	0.2	97.4	0.0	97.4	0.1	96.8	2.7	90.2
Age: Mean (SD)	55.3 (11.4)	100.0	59.8 (15.3)	100.0	57.3 (10.9)	100.0	54.1 (10.7)	100.0

"N" indicates the number of individuals; "%" indicates the percentage of individuals to whom variable label applies; "% complete" indicates the percentage of individuals without missing information.

*Because of screening, 100% of the replication subjects have Non-Finnish Nordic Parents and 0% of the replication subjects have been diagnosed with epilepsy, diabetes, hyperthyroid or hypothyroid.

Table 2

Top results from combined MWAS and GWAS analysis with $q < 0.10$.

Chr	Start	End	SNP	Minor Allele	MAF	MWAS *p*-value	GWAS *p*-value	Combined *p*-value	Combined *q*-value	Gene name	Feature
3	2555403	2555524	rs1382874	C	0.132	6.89E-03	1.72E-07	2.55E-08	0.013	CNTN4	Intron, DNAse cluster
3	2555403	2555524	rs1382875	G	0.133	6.89E-03	1.72E-07	2.55E-08	0.013	CNTN4	Intron, DNAse cluster
3	140995064	140995209	rs6772975	A	0.053	4.90E-05	1.24E-04	1.21E-07	0.030	ACPL2	Intron, Repeat, DNAse cluster
6	130500431	130500437	rs1543727	C	0.085	2.34E-05	8.12E-04	3.57E-07	0.036	SAMD3	Intron
7	85513858	85513869	rs4541830	G	0.123	3.24E-05	6.34E-04	3.84E-07	0.036		Repeat
21	19587716	19587772	rs2824706	C	0.149	9.51E-04	2.98E-05	5.20E-07	0.036	CHODL	Intron, Repeat
1	119536705	119536735	rs1886918	T	0.191	3.46E-04	1.04E-04	6.51E-07	0.038	TBX15	Upstream 8K, CpG Shore
5	150618945	150619179	rs72794132	T	0.277	5.99E-05	1.10E-03	1.16E-06	0.050	CCDC69/GM2A	Repeat, DNAse cluster
12	102069072	102069125	rs764291	G	0.233	6.28E-03	1.50E-05	1.61E-06	0.062	MYBPC1	Exon
12	83049561	83049639	rs17722140	A	0.088	2.28E-03	4.43E-05	1.72E-06	0.064		Repeat, DNAse cluster
5	109278059	109278173	rs10066593	A	0.073	5.82E-04	2.73E-04	2.65E-06	0.076		
6	41981045	41981121	rs9349211	G	0.385	2.05E-04	7.98E-04	2.72E-06	0.076	CCND3	Intron, Repeat, DNAse cluster, DNAse cluster
3	107705304	107705365	rs35887155	A	0.366	1.49E-03	1.13E-04	2.78E-06	0.076		
3	8713597	8713795	rs72624406	A	0.058	9.78E-04	1.75E-04	2.84E-06	0.076	C3orf32	Repeat
9	108152147	108152390	rs2771040	G	0.062	2.86E-04	7.72E-04	3.60E-06	0.083	SLC44A1	Exon, DNAse cluster
20	11423547	11423674	rs1474786	G	0.198	9.83E-04	2.35E-04	3.76E-06	0.085		
7	20347535	20347594	rs3114431	G	0.380	7.52E-04	3.81E-04	4.60E-06	0.096		DNAse cluster

Chromosome ("Chr"), "Start" and "End" positions for each CpG block are given. MAF is the minor allele frequency. Also shown are the test statistic *p*-values for the MWAS and GWAS in addition to the combined *p*-value and *q*-value. "Gene name" indicates genes within 20 Kb (+/−) of the block. "Feature" describes attributes overlapping with the CpG block. "Exon" and "Intron" designate overlap with RefSeq genes; "CGI" denotes overlap a CpG island; "Shore" is +/− 2 kb flanking a CGI; "Upstream 8kb" indicates within 8kb upstream of transcription start site; "DNase cluster" indicates a genomic region hypersensitive to DNaseI; "Repeat" indicates overlap with repetitive elements from RepeatMasker (www.repeatmasker.org)