Title: Assessing the role of long-noncoding RNA in nucleus accumbens in subjects with alcohol dependence

Running Title: Integrative analysis of long noncoding RNA in Nucleus Accumbens

Gowon O. McMichael¹*, John Drake²*, Eric Sean Vornholt¹, Kellen Cresswell², Vernel Williamson⁴, Chris Chatzinakos¹, Mohammed Mamdani¹, Siddharth Hariharan³, Kenneth S. Kendler¹,⁷,⁸, Michael F. Miles¹,⁹, Gursharan Kalsi⁶, Brien P. Riley¹,⁷,⁸, Mikhail Dozmorov², Silviu-Alin Bacanu¹,⁷, Vladimir I. Vladimirov¹,⁷,¹⁰,¹¹*

¹Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA
²Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA
³Summer Research Fellowship, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA
⁴Department of Pathology, Virginia Commonwealth University, Richmond, VA, USA
⁵Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA
⁶Department of Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, London SE5 8AF, UK
⁷Department Psychiatry, Virginia Commonwealth University, Richmond, VA, US
⁸Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, VA, US
⁹Department of Pharmacology & Toxicology, Virginia Commonwealth University, Richmond, VA, US
¹⁰Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, VA, USA
¹¹Lieber Institute for Brain Development, Johns Hopkins University, Baltimore, MD, USA.
Corresponding Author: Vladimir Vladimirov, MD, PhD

E-mail: vladimir.vladimirov@vcuhealth.org

**Key Words:** long noncoding RNA, gene expression, eQTLs, postmortem brain, Nucleus Accumbens, Alcohol Dependence
Abstract

Recently, long noncoding RNA (lncRNA) were implicated in the etiology of alcohol dependence (AD). As lncRNA provide another layer of complexity to the transcriptome, assessing their expression in the brain is the first critical step towards understanding lncRNA functions in AD. To that end, we profiled the expression of lncRNA and protein coding genes (PCG) in nucleus accumbens (NAc) from 41 subjects with AD and 41 controls. At false discovery rate (FDR) of 5%, we identified 69 and 309 differentially expressed lncRNA and PCG genes, respectively. Using co-expression network analyses, we identified three lncRNA and five PCG modules significantly correlated with AD at Bonferroni adj. p≤0.05. To better understand lncRNA functions, we integrated the lncRNA and PCG hubs from the significant AD modules; at FDR of 5%, we identified 3 151 positive and 2 255 negative correlations supporting the functional role of lncRNA in the development of AD. Gene enrichment analysis revealed that PCG significantly correlated with lncRNA are, among others, enriched for neuronal and immune related processes. To highlight the mechanisms, by which genetic variants contribute to AD, we integrated lncRNA and PCG hubs with genome-wide SNP data. At FDR≤0.3, we identified 276 expression quantitative trait loci (eQTL), affecting the expression of 20 and 256 lncRNA and PCG hubs, respectively. Our study is the first to profile lncRNA expression in nucleus accumbens in a large postmortem alcohol brain sample and our results may provide novel insights into the regulation of the brain transcriptome across disease.

Introduction
Alcohol Use Disorder (AUD) is a chronic and debilitating disease with an estimated heritability around 50% (1). Previous postmortem brain expression studies have shown that chronic alcohol consumption leads to broad transcriptional changes in different brain regions (2;3). Gene expression studies in prefrontal cortex (PFC), have identified genes encoding to GABA_A receptor subunits or related to mitochondrial function (4;5) as well as in functions related to myelination, cell cycling, oxidative stress, and transcription (2;6-11). Expression studies in nucleus accumbens (NAc) and ventral tegmental area (VTA), have also revealed gene expression changes related to cell architecture, signaling, vesicle formation and synaptic transmission (7). These findings suggest that there are region-specific susceptibilities and adaptations to chronic alcohol consumption in the brain that are likely to have a distinct effect on the behavioral phenotypes comprising AUD (12).

Evaluation of the regulatory mechanisms underlying genetic differentiation is necessary to better understand the neurobiology of alcohol addiction (13). The recent emergence of long noncoding RNA (lncRNA) has provided an additional layer of transcriptional and translational control that could highlight important neurobiological mechanisms underlying AUD that could be missed if only the protein coding genome was studied. LncRNA are longer than 200 base pairs (bp) with limited or no protein coding potential (14) and while they have not been functionally fully characterized yet, they were shown to participate in chromatin remodeling (15), transcriptional and post-transcriptional regulation (16), and in sequestering miRNA (17;18). LncRNA have also been implicated in the plasticity of neuronal circuity (19) and in neurodegenerative and neuropsychiatric disorders including substance abuse and alcohol dependence (20;21); the potential role of lncRNA in the etiology of AUD was also further supported by several recent alcohol-related genome-wide association scans (GWAS) (22;23). Due to the complex nature of lncRNA functions, the relationship between lncRNA and alcohol consumption has been understudied; while miRNA roles in the control of gene expression and functions in the brains of subjects with AUD have been reported (2;24-27), to our knowledge the interactions between lncRNA and protein coding genes (PCG) are still largely undescribed.

Gene expression alone cannot explain the complex etiology of AUD and assessing lncRNA...
and PCG expression in the context of available genetic data is necessary to discern the genetic basis of AUD susceptibility. This approach models associations between genetic variants and gene expression as quantitative traits, i.e. expression quantitative trait loci (eQTL)(28-30). EQTLs can help discover unknown AUD risk loci or offer specific, testable, hypotheses for the genetic impact of polymorphisms associated with AUD(31-33). Linkage disequilibrium (LD) of eQTLs with genetic variants implicated in AUD can further establish a biological mechanism for disease-associated variants with no apparent functions; there is empirical evidence suggesting that eQTLs are over-represented among GWAS signals(34;35).

We hypothesize that IncRNA are involved in the neuropathology of AUD and that, on a molecular level, this involvement is manifested by differential patterns of expression between cases with alcohol dependence (AD) and controls. To better understand the IncRNA disease functions in subjects with AD we further performed a weighted gene co-expression network analysis (WGCNA) that led to identification of IncRNA and PCG modules significantly correlated with AD. Key genes from these IncRNA and PCG modules (i.e., ‘hubs’) were then correlated with each other to identify a set of interacting IncRNA and PCG modules. Since IncRNA are not yet annotated, we used system approaches to assign biological functions for the PCG hubs interacting with the IncRNA hubs. By integrating our genome-wide SNP and expression data, we also identified expression quantitative trait loci (eQTL) affecting the expression of IncRNA and PCG in NAc. In this study, our overall goal is to profile IncRNA and PCG expressions in NAc and test whether these are under the control of specific genetic elements. To the best of our knowledge this is the only study to comprehensively assess the IncRNA expression between subjects with AD and controls in postmortem AD case and control brain tissues.

Methods and Materials

Postmortem tissue
Brain tissue from 41 cases with AD and 41 controls was received from the Australian Brain Donor Program, New South Wales Tissue Resource Centre, which is supported by The University of Sydney, National Health and Medical Research Council of Australia, Schizophrenia Research Institute, National Institute of Alcohol Abuse and Alcoholism, and the New South Wales Department of Health (http://sydney.edu.au/medicine/pathology/trc/). (Suppl. Methods; Suppl. Table 1).

**RNA isolation and sample selection**

Total RNA was isolated from 50mg of frozen tissue from NAc using the mirVana-PARIS kit (Thermo Fisher, Carlsbad, CA), following manufacturer’s protocols. RNA concentration was measured using the Quant-iT Broad Range RNA Assay kit (Life Technologies), and the RNA Integrity Number (RIN) was measured on the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA).

**Expression arrays and data normalization**

The RNA samples were assayed using the Arraystar Human IncRNA Array v3.0 (Rockville, MD, USA) that is designed to profile both IncRNA and PCG. The expression values were calculated and normalized using the 75th percentile value for each microarray in the Agilent Feature Extraction text file, followed by Log2 transformation and batch effect removal in the Partek® Genomics Suite® (PGS) software, v6.6 (St. Louis, MO, USA). Probes with expression values outside the detectable ‘spike-in’ range in more than 50% of the sample, or if they were flagged as “gIsWellAboveBG=0” and “gIsPosAndSignif=0” in more than 75% of the arrays were excluded from the downstream analyses. Upon removal of low expressed probes, the final number of probes that were used in the subsequent analyses was 19,458. The efficiency of batch effect removal and overall array quality were assessed using principal components analysis (PCA) on the expression values in which each array is plotted along the first three principle components (PCs) to identify potential outliers. Of the 82 samples, 9 samples were excluded due to low RINs (i.e. RINs≤4), leaving 73 samples for the microarray run. Of
the 73 samples run on the array, 8 samples (5 cases and 3 controls) did not load on two of the first three PCs and were removed from the subsequent analysis, leaving a final sample size of N=65.

**Statistical Analyses**

**Gene Expression Analysis**

The univariate gene expression analyses were performed in the Number Cruncher Statistical Software (NCSS) v11, using a robust multiple regression model(2). Prior to the main analyses, technical (e.g. RIN, PMI, RI) and biological (e.g. smoking, medication, age) confounds and the first 5 principle components (PCs) were used as covariates in the regression analysis. Microarray reliability was validated by assessing the expression of 5 genes at the Arraystar facilities using quantitative PCR (qPCR). The assessed genes showed very high correlation between the two platforms (Suppl. Methods and Suppl. Fig. 1).

**Weighted Gene Co-expression Network Analysis (WGCNA)**

Gene co-expression networks were constructed separately for the IncRNA and PCG using the WGCNA v1.36 package in R environment (v3.5.2). While, WGCNA allows for the inclusion of additional variables, it does not correct for their effects, therefore the gene networks were built using the residuals from the regression model. WGCNA relies on pair-wise Pearson correlations to generate a signed similarity matrix, selecting for positive correlations only. The signed similarity matrix of the IncRNA and PCG expression data was raised to the lowest power (β=8 and β=11, respectively) that approximated a scale-free network topology ($R^2>0.90$), to generate an adjacency matrix. A topological overlap measure (TOM) was calculated to assess transcript interconnectedness. A dissimilarity measure was calculated from the TOM and was subsequently used for average linkage hierarchical clustering. Module definition parameters included a minimum module size of 35 genes, DeepSplit 4, and a minimum module merge height of 0.8.
Following module definition, the first principal component of each module – the module eigengene (ME) – was calculated as a synthetic gene representing the expression profile of all genes within a given module. Modules are named by a conventional color scheme and then correlated to AD case-status, matching demographics and relevant covariates. Statistical significance was assessed at Bonferroni-adj. $p \leq 0.05$ (corrected for number of tested modules).

To identify the set of genes with a high module membership and phenotype association, (i.e. the hub genes), we applied two selection criteria: 1) genes with the high intra-modular connectivity ($r^2 \geq 0.7$) and 2) significantly correlated with AD(36).

**Gene Set Enrichment Analysis**

Gene set enrichment analysis (GSEA) was used to detect known biological processes and pathways enriched within the PCG modules using GSEA v2.2.3 software from the Broad Institute as previously described(37;38). Individual gene lists for each of the PCG modules significantly correlated with AD were generated by rank-ordering all differentially expressed PCG by their module membership (MM) to each of the AD-associated modules. Prior to running GSEA, the transcript IDs were converted to HUGO Gene Nomenclature Committee (HGNC)(39). We derived the a priori gene sets from the Molecular Signatures Database v6.2 (MSigDB; http://www.broadinstitute.org/gsea/msigdb) from the Broad Institute. A total of 1329 gene sets from the Canonical Pathways subset of the C2: Curated Pathways collection of MSigDB were assessed. Default parameters were then applied to give a minimum and maximum a priori gene set size between 15 and 500 genes, respectively. The identified enriched pathways were further adjusted for multiple testing at FDR of 25%.

**Correlation Analyses**

The expression values of IncRNA and PCG hubs from the significant modules, were correlated (Pearson r) with each and adjusted at FDR of 5%, separately for the negative and positive correlations. To better understand the biological significance of these correlations, we performed
gene enrichment analysis, in the Co-IncRNA webtool (http://bio-bigdata.hrmbu.edu.cn/Co-LncRNA/) using KEGG pathways. The significance of the enrichment analysis was assessed by hypergeometric test and further adjusted for multiple testing at FDR of 10%(40).

**eQTL Detection**

The genotype calls were generated as part of a larger meta-GWAS study(22) and were integrated with the IncRNA and PCG hubs expressions to identify expression quantitative trait loci (eQTLs) for the IncRNA and PCG. Following a previously described approach and to retain power only local, cis-eQTLs, were assessed(2). Specifically, SNPs located 1 mega base pairs (Mbp) upstream and downstream from the selected IncRNA and PCG hubs were filtered with Plink v1.07 to exclude variants in LD ($R^2 \geq 0.7$)(41). To reliably estimate the eQTL effects in a sample size of 65, we further considered only SNPs with a minor allele frequency (MAF) $\geq 10\%$. After LD pruning we retained 58 317 SNPs which were tested as eQTLs. SNP impact on gene expression was detected by MatrixEQTL software package in R in using a linear regression framework and adjusting for the potential effects of the first 5 PCs(42). The significant eQTL results were adjusted for multiple testing at FDR of $q \leq 0.3$.

**Test for GWAS association signals enrichment**

The assessment of enrichment for low p-values was performed by using the SST test(43). Briefly, for the analysis we used the latest data from the Psychiatric Genomic Consortium (PGC) GWAS of AUD(44). As the postmortem brain sample is composed of subjects with European (EU) ancestry, for our analysis we used the EU sample of 46K subjects (Suppl. Methods).

**Results**

**Chronic alcohol consumption leads to generalized gene expression changes between cases and controls**
At the nominal p≤0.05 we identified a total of 2,859 probes differentially expressed between AD cases and controls, representing 676 IncRNA and 2,135 PCG that compromise 3.4% and 11% of the total gene pool assayed, respectively. Similar to previous studies, the overall number of differentially expressed genes was greater than expected by chance (hypergeometric p=1.3x10⁻⁵) suggesting that chronic alcohol intake leads to generalized changes in the brain transcriptome(2;45). At FDR of 5% we identified 69 IncRNA and 309 PCG. Interestingly among these we also identified a few pseudogenes with evidence for differential expression. An unsupervised hierarchical clustering of the standardized expression values (i.e., values shift to a mean of 0 and standard deviation of 1) of all differentially expressed PCG and IncRNA genes at the nominal p≤0.05 showed general patterns of association between gene expression and diagnosis (Suppl. Fig.2A and 2B). The significant results from the univariate gene expression analyses are provided in supplementary table (Suppl. Table 2).

**IncRNA and PCG show a disease relevant network pattern**

While the gene expression analysis identified differentially expressed genes in NAc of subjects with AD, these do not provide an integrative view of the interaction between the IncRNA and mRNA genes. Therefore, we performed a weighted gene co-expression network analysis (WGCNA)(46;47) separately on the nominally expressed IncRNA and PCG at p≤0.05. We chose this threshold in order to: i) include genes with smaller effect size (which otherwise would be excluded with more stringent statistical criteria), ii) enrich for genes likely to play role in AD, and iii) retain a sufficient number of genes for building of gene co-expression networks. In the IncRNA network analysis, we identified a total of 5 modules (Fig. 1)(46), of which three modules (\(M_{\text{brown}}, M_{\text{blue}},\) and \(M_{\text{grey}}\)) were significantly correlated with AD at a Bonferroni adj. p≤0.01. In the PCG network analysis, we identified 11 modules, five of which were significantly correlated with AD at a Bonferroni adj. p≤0.005 (Fig.2). In addition to the main dichotomous case/control status, we also tested additional quantitative alcohol related phenotypes such as age of initiation of alcohol drinking, amount of daily alcohol consumption and years of drinking. We identified one module (\(M_{\text{green}}\)) and two modules (\(M_{\text{brown}}\) and \(M_{\text{green}}\)) as significantly correlated with amount of daily alcohol consumption at a Bonferroni adj. p≤0.05 in the
IncRNA and PCG analyses, respectively. Interestingly, two of the five PCG modules were significantly correlated with both AD and amount of daily alcohol consumption, while in the IncRNA analysis, we did not observe overlap between modules associated with AD (i.e. \( M_{\text{brown}} \) and \( M_{\text{blue}} \)) and daily alcohol consumption (\( M_{\text{green}} \)). Full tables with module size, correlations and p-values for all PCG and IncRNA modules correlated with AD and daily alcohol consumption are provided in supplementary tables (Suppl. Tables 3 and 4).

To ensure the robustness of our gene co-expression networks and reduce the potential influence of outlier samples on network structure, we used the robust ‘bootstrap’ version of WGCNA (rWGNCA). We performed 100 iterations in which networks were created after randomly subsetting 2/3 of the total samples as previously suggested(48). The resulting 100 networks were then merged into one large, final consensus network; the individual sub-networks showed reasonably high consistency with the final IncRNA and mRNA networks (Suppl. Fig.3A and 3B).

In a scale-free network topology, ‘hubs’ are the most highly connected genes (of which there are relatively few among all the nodes within a network); thus, we also identified the hub genes for the significant modules. Of the three IncRNA modules associated with AD, \( M_{\text{blue}} \) was the largest containing 200 genes, \( M_{\text{brown}} \) contained 105 genes and \( M_{\text{grey}} \) 73 genes. As the genes in \( M_{\text{grey}} \) are unassigned we attempted to identify hubs only in \( M_{\text{blue}} \) and \( M_{\text{brown}} \). Of the 305 transcripts clustered in the two modules, 70 transcripts were considered as candidate hubs, while among the five PCG modules we identified a larger number of hubs (N=368).

**AD gene modules are enriched in alcohol related processes**

Previous studies have shown that co-expressed genes are enriched for biologically relevant functions [49-51]. We used a gene set enrichment analysis (GSEA) to assess for enrichment of cellular process and biological functional categories in modules associated with AD. The gene lists for each module were generated by ranking all 2185 differentially expressed transcripts according to their MM to each of the five significant modules as previously described(49). Using the default parameters...
in GSEA, at FDR of 25%, we identified 38 a priori gene sets significantly enriched in four of the five PCG modules associated with AD (Table 1). Among these, \( M_{\text{turquoise}} \) and \( M_{\text{greenyellow}} \), showed the highest and \( M_{\text{green}} \) the lowest numbers of enriched GO terms, respectively. The 4 modules showed distinct patterns of enrichment, with \( M_{\text{turquoise}} \) enriched for gene sets involved in neurodegenerative disorders (i.e., Alzheimer, Parkinsons and Huntingtons diseases) as well as gene sets enriched for neuronal system, neurotrophins signaling and oxidative stress, while \( M_{\text{tan}} \) and \( M_{\text{greenyellow}} \) were enriched for immune related processes and for genes involved in metabolic processes among others.

**IncRNA and PCG show a complex pattern of interactions**

Since IncRNA can function as potential molecular “scaffolders”(50;51), i.e. by bringing different sets of genes near each other, we were interested to identify PCG hubs significantly correlated with IncRNA hubs. Limited studies have shown that IncRNA can affect gene functions by interfering with the expression of their specific gene targets(52). We integrated the IncRNA and PCG hubs by correlating the expression values of IncRNA and PCG hubs across the significant modules. At FDR of 5% we observed 5,404 correlations, of which the number of positive correlations (N=3,150) significantly exceeded the number of negative (N=2,254) correlations (Mann-Whitney U test \( p=1E^{-36} \)). In these analyses, we observed generalized as well as module specific correlations. For example, hubs from the blue and brown IncRNA modules showed distinct either negative or positive correlations, respectively with hubs from the black PCG module (Fig.3A and 3B). Similarly, hubs from the blue IncRNA module were exclusively positively correlated with the hubs from \( M_{\text{turquoise}} \) PCG module, while hubs from the brown IncRNA module were negatively correlated with \( M_{\text{turquoise}} \). We further observed that the unique module correlations also drove the strongest positive and negative individual IncRNA/PCG hubs correlations. For example, the strongest negative correlation (Pearson \( r=-0.72, \ p=1.63E^{-11}, \ q=2.07E^{-8} \)) was observed between IncRNA (G075391) a hub gene in the brown IncRNA module on chromosome 7 and VPS51 (a member of the vacuolar protein sorting-associated protein 51 family) a hub from \( M_{\text{turquoise}} \) on chromosome 11. Further exploring the negative correlation between these two hubs we observed a moderate mediating impact of G075391 on the expression of
VPS51 between cases and controls (ANCOVA, F=5.07, df=1, p=0.028). In the gene expression analysis, G075391 was overexpressed in the alcohol subjects, while VPS51 was underexpressed (Fig. 4) suggesting that one potential mechanism by which G075391 contributes to the neuropathology of AUD is by reducing the expression of VPS51. The entire list of positive and negative IncRNA/PCG hub correlations at FDR of 5% is provided in supplementary table (Suppl. Table 5).

To better understand the overall functional role of the two significant IncRNA modules, we performed a pathway enrichment analyses using the Co-IncRNA webtool(40) for the negatively and positively correlated IncRNA and PCG modules. We observed distinct pathway enrichments, in which PCG modules negatively correlated with the IncRNA modules (M_blue) were enriched for pathways related to the complement system, cell adhesion and toll-like receptors, while the positively correlated PCG modules were enriched for pathways belonging to neurodegenerative disorders such as Parkinson's, Alzheimer and Huntington disease and in pathways involved in oxidative phosphorylation, long-term potentiation among others. PCG modules positively correlated with IncRNA M_brown module were enriched for lysosome trafficking, while PCG modules negatively correlated with IncRNA M_brown were enriched for genes involved in calcium signaling, long-term potentiation and neurotrophin signaling (Suppl Figs. 4A-4O). The PCG M_brown significantly correlated with daily alcohol consumption and negatively correlated with IncRNA M_green was enriched for gene pathways involved in focal adhesion, MAPK signaling, peroxisome (Suppl. Figs. 5A-5D).

**Genetic variants affect IncRNA and PCG expression in a disease specific manner**

To understand the underlying genetic mechanisms of AD risk polymorphisms we further integrated the IncRNA and PCG hubs expression data from our significant modules with previously collected genotypic data(22). Due to the small sample size, we focused on testing cis-eQTLs (defined within 1Mbp up- and down-stream from the tested loci) only. At FDR of q≤0.3 we identified 293 eQTLs that impact hubs expression, with disproportionate number of eQTLs for the PCG (n=282) vs.
In a follow up analysis, we further tested these 293 eQTLs for an interaction between genotype and disease status; at FDR of 5% we identified 31 interactions that mediate the expression of 30 PCG hubs, and one IncRNA hub (G054549) on Chr. 3 (Table 2). The most significant interaction we observed was between FKBP prolyl isomerase 5 (FKBP5) and rs9394312 on Chr. 6 (Fig.5).

GWAS signals of AUD show suggestive enrichment in our IncRNA specific eQTLs

While we and others have shown alcohol related eQTLs are enriched for GWAS signals of AUD and other alcohol related phenotypes (ARPs) in the protein coding genome(53), no such studies have been performed to interrogate for such enrichment in IncRNA genes. To increase power of our test and be rather inclusive, we selected all eQTLs significant at the nominal p≤0.05. At this threshold level we have identified a total of 3064 eQTLs (366 IncRNA and 2698 PCG specific eQTLs). The assessment of enrichment was performed by using the SST test (Suppl. Text) and we detected only a suggestive enrichment (p=0.16) of AUD signals, most likely due to the underpowered alcohol PGC sample and our own postmortem brain sample.

Discussion

The main goal of this study is to profile the expression patterns of the non-coding and coding transcriptome of NAc in a large sample of subjects with AD and controls. The NAc is a central component of the mesocorticolimbic system (MCL) and has been shown to be involved in addictive behaviors(54-56). By integrating GWAS data previously collected on the sample(22) with our expression data, we further attempted to reveal hitherto unknown relationships between IncRNA and PCG and to assess how this relationship is modulated by genetic variants in the brain. Unsurprisingly and in agreement with previous alcohol related postmortem brain expression studies, we identified differentially expressed PCG, suggesting that chronic alcohol consumption is associated with widespread transcriptomic changes in the brain. We further identified numerous IncRNA differentially expressed between alcoholic cases and control, supporting their involvement in the neuroadaptations...
associated with chronic alcohol consumption. Among the most significant differentially expressed PCG, were serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3 (SERPINA3), interferon induced transmembrane protein 2 and 3 (IFITM2 and IFITM3) and solute carrier family 14 (urea transporter), member 1 (Kidd blood group) (SLC14A1), that were also identified from previous alcohol related postmortem brain gene expression studies(9;57). Interestingly, we also detected a few differentially expressed pseudogenes, such as the annexin A2 pseudogene 2 and 3 (ANXA2P2 and ANXA2P3) and interferon induced transmembrane protein 4 pseudogene (IFITM4P) on chromosomes 9, 10 and 6, respectively. Previous postmortem brain expression studies have shown ANXA2P2 to be involved in stress response pathways in chronic alcoholics(57), however, little is known about the functions of ANXA2P3 and IFITM4P; it appears that ANXA2P3 is involved in reducing lipids levels in response to statins(58) and in chronic obstructive pulmonary disease (COPD)(59), while no known functions were reported for IFITM4P.

A major limitation of the single gene expression analysis to identify and prioritize a set of genes associated with a given phenotype, is its inability to consider the complex molecular interactions between such identified genes(60); this is especially relevant in studying the non-coding transcriptome, which is poorly annotated, with many IncRNA genes still of unknown function. To address this limitation, we performed a weighted gene co-expression analysis (WGCNA), which allowed us to identify a set of co-regulated PCG and IncRNA. We identified three IncRNA modules and four PCG modules that were significantly correlated with AD. Among the significant IncRNA modules correlated with AD status was M_{grey}. Genes in the grey module are frequently considered as ‘noise’, i.e. genes whose expression does not correlate with genes in the other modules, although there are suggestions that this interpretation may not always be accurate(61). In general, the genes in the grey module are usually ubiquitously expressed, exhibiting oscillating or highly variable patterns of expression; of course these could also be simply assigned to the wrong module(62). Regardless, one potential explanation for our observation could be that, at least some IncRNA, may impact the neuropathology of AD in isolation and not as a part of gene network. For example, other studies have
also reported similar observation, where despite of their unassigned status, genes in M\text{grey} show significant disease associations in certain phenotypes(63;64). Using the available quantitative measures of alcohol consumption, we further identified one lncRNA module and two PCG modules significantly correlated with amount of daily alcohol consumption after correction for multiple testing. Interestingly, while we observed same PCG modules to be significantly correlated with both AD status and quantitative alcohol phenotype, no such relationship was revealed for the lncRNA module, suggesting that different sets of lncRNA genes may impact these related, yet distinct, measures of alcohol related phenotypes. To further understand the potential regulatory functions of lncRNA, we performed a series of correlation analyses between the hub genes identified from the significant lncRNA and PCG modules. In these analyses we observed highly significant positive and negative correlations, in which the positive correlations significantly outnumbered the negative correlations; this is not entirely surprising, considering that the positive correlation can reflect both a genuine regulatory function as well as co-expression patterns between lncRNA and PCG. We further observed that the module specific correlations were driven exclusively by lncRNA belonging to individual modules, suggesting that individual modules are likely to contain a set of lncRNA with different and non-overlapping functions. Indeed, in our pathway enrichment analyses we observed the PCG modules negatively or positively correlated with the lncRNA modules are enriched for distinct and mostly non-overlapping biological processes, further corroborating the suggestion that lncRNA involved in the neuroadaptation to alcohol consumption likely have non-overlapping functions. From our GSEA analysis, we observed that similarly to other studies(2;45;65), some of the PCG modules are also enriched for immune related or neurodegenerative processes. Immune related responses to alcohol consumption in the brain has been well documented(66;67). Likewise, the neurodegenerative processes observed for some of our modules may reflect the nature of our postmortem sample composed exclusively from chronic alcoholics. This is not surprising considering that our sample is composed of subjects that have been exposed to alcohol for over two decades. Both animal and human genetic models have demonstrated that prolonged exposure to alcohol leads
to activation of the microglial population with the concurrent activation of immune related processes in the brain(68;69). It has been well known, for over several decades now, that prolonged and excessive alcohol drinking leads to loss of a brain matter(70;71), and consequently activation of genes involved in neurodegenerative disorders.

Integrating our expression and genetic data led to identification of eQTLs that affect lncRNA and PCG expression. Having an integrated view on the expression of lncRNA and PCG we were also interested to identify eQTL that can impact both lncRNA and PCG expressions; we, however, failed to observe eQTLs that have a shared impact on lncRNA and PCG expressions. We further observed disproportionate numbers of eQTLs for PCG vs. lncRNA genes. This could reflect either a genuine biological mechanism, or conversely can be due to a lack of sufficient power to reliably identify eQTLs for lncRNA whose expression is associated with AD. Interestingly, among the eQTLs affecting PCG expression, certain polymorphisms showed a strong interaction effect between PCG expression and disease status. Among these, the strongest interaction effect was observed for FKBP5 and rs9394312. Several studies have implicated FKBP5 in the severity of alcohol withdrawal(72), alcohol drinking patterns in rodents(73), problematic drinking(74), as well as direct regulation of FKBP5 expression and phosphorylation by acute ethanol in mouse prefrontal cortex(75;76).

The importance of our postmortem brain expression study to assess lncRNA expression in nucleus accumbens is further supported by three recent alcohol related GWAS reporting genome-wide signals near lncRNA genes: i) a study conducted by Gelernter at al. (2014)(77) that implicated the ADH gene cluster on chromosome 4 and the LOC100507053 locus, ii) a study conducted by the COGA group(78) that identified a polymorphism near LOC151121 on chromosome 2, and iii) a study conducted by our own group that have reported several polymorphisms in LOC339975(22). However, in our study none of these three loci showed evidence for differential expression, which also agrees with our previous study reporting no evidence for differential expression of these loci, although we observed the expression of LOC339975 to be affected by the most significantly associated polymorphism (rs11726136) in NAc.
In summary, the main goal of this study was to test the hypothesis that lncRNA, as a novel class of none coding RNAs, may contribute to the neuropathology of alcohol dependence. In this study we identified a set of lncRNA and PCG differentially expressed between cases and controls. Our gene network analyses identified lncRNA and PCG modules significantly correlated with AD in NAc, which were shown to be enriched for immune related processes and neurodegeneration. We further identified distinct patterns of correlation between the lncRNA and PCG hubs with minimal overlap between modules, i.e. either positive or negative correlations, but not mixed. Gene pathway enrichment revealed that of genes whose functions were related to immune, neurological and neurodegenerative processes. Finally, we identified eQTLs that affect the expression of lncRNAs and PCG hubs, with some of these eQTLs showing a clear mediating effect on the PCG expression between cases and controls. Unlike previous studies though, here we saw only suggestive enrichment for association signals among the eQTLs, which could be attributed to either a limited importance of lncRNAs in the etiology of AD, or most likely due to our limited sample size.

While our study is novel in its approach to integrate genetic and molecular data in postmortem alcoholic brains as well as addressing important questions regarding lncRNA involvement in the etiology of AD, it is not without limitations. First, postmortem brain studies are observational as manipulation of living human subjects is not possible. Although the cross-sectional nature of these studies limits the causal inference we can make, we believe the eQTL analysis is a major step toward clarifying the directionality of these observations. Secondly, although our sample size (N=65) is by far the largest postmortem alcohol study to date, in comparison with other publically available postmortem brain expression samples(79) used to study neuropsychiatric disorders such as schizophrenia, bipolar disorder or major depression, is still prohibitively small. However, we believe that with our careful experimental design and implementation of integrative multivariate approaches, we can circumvent some of these limitations and further broaden our understanding of alcohol addiction processes. We hope that our study will provide additional molecular targets that will help translate these advances into effective therapeutic strategies for patients suffering with substance use
disorders. Thus, in conclusion, we believe our study to be important in the field, as to the best of our knowledge, it is among the first such studies to address the role of IncRNA in the neuropathology of alcohol addiction paving the road for future such studies.
Reference List


(28) Farris SP, Wolen AR, Miles MF. Using expression genetics to study the neurobiology of ethanol and alcoholism. International review of neurobiology 2010;91.


(35) Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010;6(4).


(41) Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007 Sep;81(3):559-75.


(54) Marty VN, Spigelman I. Effects of alcohol on the membrane excitability and synaptic transmission of medium spiny neurons in the nucleus accumbens. Alcohol 2012;46(4).


**Fig. 1** Module-trait relationships for IncRNA. The residuals of the expression values used to generate the IncRNA module MEs are correlated (Pearson) to the dichotomous AD case/control status (Diagnosis) and to quantitative alcohol measures such as daily alcohol consumptions (Alc-Cons), total amount of drinks (Tot_drinks) and initial age of drinking (Age_began). The IncRNA modules were also correlated to the first 5 PCs to assess for confounding. P-values shown are unadjusted for multiple testing. After adjusting for number of modules tested, ME_{brown}, ME_{blue}, and ME_{grey}, are significantly correlated with AD case-status (Class) and ME_{green} with daily alcohol consumption.

**Fig. 2** Module-trait relationships for protein coding genes (PCG). The PCG module MEs are correlated (Pearson) to the dichotomous AD case/control status (Diagnosis) and to quantitative alcohol measures such as daily alcohol consumptions (Alc-Cons), total amount of drinks (Tot_drinks) and initial age of drinking (Age_began). The IncRNA modules were also correlated to the first 5 PCs to assess for confounding. P-values shown are unadjusted for multiple testing. After adjusting for number of modules tested, ME_{turquoise}, ME_{black}, ME_{greenyellow}, ME_{tan} and ME_{green} are significantly correlated with AD case-status (Class) and ME_{brown} and ME_{green} also with daily alcohol consumption.

**Fig. 3** Histogram of IncRNA/PCG module correlations. (A) IncRNA ME_{brown} shows negative correlation with PCG ME_{turquoise}, while IncRNA ME_{blue} showed negative correlations with all five significant PCG MEs. (B) PCG ME_{black} shows a positive correlation only with IncRNA ME_{brown} and PCG ME_{turquoise} shows a positive correlation only with IncRNA ME_{blue}.

**Fig. 4** Plot of the mediating effect of IncRNA (G075391) on the expression of PCG Vacuolar Protein Sorting-Associated Protein 51 (VPS51) in cases and controls.

**Fig. 5** Plot showing a SNP interaction effect of disease status on the expression of FKBP5. The minor allele of rs9394312 confers a higher FKBP5 expression in controls and lower expression in cases.
Suppl. Fig. 1 Microarray expression data validation using quantitative PCR. Expression levels of five genes measured by the expression array-based approach were validated using quantitative PCR in all 65 postmortem Nucleus Accumbens (NAc) RNA samples.

Suppl. Fig. 2 Unsupervised hierarchical clustering of the lncRNA and PCG differentially expressed at \( p \leq 0.05 \) was performed on the standardized expression values, which were centered at mean of 0 and SD of 1. The genes were clustered according to the similarity of their expression profile in PGS v.6.6 using complete linkage and Euclidean distance metrics. The color scheme of the Y-axis reflects the clustering based on diagnosis, while the color scheme of the X-axis is an arbitrary identification of the two large clusters.

Suppl. Fig. 3 Robust, bootstrapped version of WGCNA (rWGCNA). The purpose of rWGCNA is to assess for potential influence of outlier samples on network structure. We performed 100 iterations in which networks were created after first randomly subsetting 2/3 of the total samples. The resulting 100 networks were then merged into one large, final consensus network. The individual sub-networks show consistent structure with each other and the final network for (A) lncRNA and (B) PCG.

Suppl. Fig. 4 Figures show the enrichment of the PCG hubs positively and negatively correlated with lncRNA hubs in KEGG pathways for AD status. The pathways were derived from the Co-lncRNA web-tool(40). (A) PCG hubs from the ME\textsuperscript{green} positively correlated with lncRNA ME\textsuperscript{blue} are enriched for lysosome functions; (B) PCG hubs from the ME\textsuperscript{greenyellow} negatively correlated with lncRNA ME\textsuperscript{blue} are enriched for cell adhesions functions; (C) PCG hubs from the ME\textsuperscript{tan} negatively correlated with lncRNA ME\textsuperscript{blue} are enriched for genes with functions in complement system; (D) PCG hubs from the ME\textsuperscript{tan} negatively correlated with lncRNA ME\textsuperscript{blue} are also enriched for genes involved in Toll-like receptors; (E, F, G) PCG hubs from the ME\textsuperscript{turquoise} positively correlated with lncRNA ME\textsuperscript{blue} are enriched for genes involved in neurodegenerative diseases, i.e. Alzheimer’s, Huntington and Parkinson; (H, I, J, and K) PCG hubs from the ME\textsuperscript{turquoise} positively correlated with lncRNA ME\textsuperscript{blue} are also enriched for genes involved oxidative stress, glioma, long term potentiation, and spliceosome; (L) PCG hubs from the ME\textsuperscript{green} positively correlated with lncRNA ME\textsuperscript{brown} are enriched for genes involved in processes related to lysosome functions; (M, N, and O) PCG hubs from the ME\textsuperscript{turquoise} negatively correlated with lncRNA ME\textsuperscript{blue} are enriched for genes involved in calcium signaling, long term potentiation and neurotrophin signaling.
Suppl. Fig. 5 Figures show the enrichment of the PCG hubs positively and negatively correlated with IncRNA hubs in KEGG pathways for daily alcohol consumption. (A, B, C, D) PCG hubs were enriched for processes related to focal adhesion, MAPK signaling, pathways in cancer, and peroxisome functions.
<table>
<thead>
<tr>
<th>Name</th>
<th>Module Color</th>
<th>SIZE</th>
<th>ES</th>
<th>NES</th>
<th>NOM p-val</th>
<th>FDR q-val</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REACTOME_GENERIC_TRANSCRIPTION_PATHWAY</strong></td>
<td>Green</td>
<td>18</td>
<td>0.32074964</td>
<td>2.0500951</td>
<td>0.00817286</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>NABA_MATRISOME</strong></td>
<td>Greenyellow</td>
<td>52</td>
<td>0.29716546</td>
<td>2.5380103</td>
<td>0.00136364</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_GENERIC_TRANSCRIPTION_PATHWAY</strong></td>
<td>Green</td>
<td>18</td>
<td>0.44453996</td>
<td>2.712143</td>
<td>0.00110024</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM</strong></td>
<td>Greenyellow</td>
<td>16</td>
<td>0.36958157</td>
<td>1.8165785</td>
<td>0.01814516</td>
<td>0.05221438</td>
</tr>
<tr>
<td><strong>REACTOME_TRANSMEMBRANE_TRANSPORT_OF_SMALL_MOLECULES</strong></td>
<td>Green</td>
<td>24</td>
<td>0.28899164</td>
<td>1.7495738</td>
<td>0.03180914</td>
<td>0.05658437</td>
</tr>
<tr>
<td><strong>KEGG_FOCAL_ADHESION</strong></td>
<td>Greenyellow</td>
<td>25</td>
<td>0.26358476</td>
<td>1.7666267</td>
<td>0.02647653</td>
<td>0.05739538</td>
</tr>
<tr>
<td><strong>NABA_MATRISOME</strong></td>
<td>Greenyellow</td>
<td>32</td>
<td>0.3983636</td>
<td>1.8274673</td>
<td>0.01153846</td>
<td>0.05828804</td>
</tr>
<tr>
<td><strong>KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION</strong></td>
<td>Greenyellow</td>
<td>15</td>
<td>0.39268157</td>
<td>1.8297838</td>
<td>0.01704545</td>
<td>0.06925771</td>
</tr>
<tr>
<td><strong>REACTOME_FATTY_ACID_TRACYLGLYCEROL_AND_KETONE_BODY_METABOLISM</strong></td>
<td>Green</td>
<td>16</td>
<td>0.39104077</td>
<td>1.8677335</td>
<td>0.01417004</td>
<td>0.07023604</td>
</tr>
<tr>
<td><strong>REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS</strong></td>
<td>Green</td>
<td>40</td>
<td>0.25898344</td>
<td>1.897424</td>
<td>0.00978473</td>
<td>0.08481742</td>
</tr>
<tr>
<td><strong>REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM</strong></td>
<td>Tan</td>
<td>30</td>
<td>0.5582678</td>
<td>3.6189888</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_INTERFERON_SIGNALING</strong></td>
<td>Tan</td>
<td>23</td>
<td>0.6121739</td>
<td>3.4482594</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_IMMUNE_SYSTEM</strong></td>
<td>Tan</td>
<td>80</td>
<td>0.31643414</td>
<td>3.1721213</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_INNATE_IMMUNE_SYSTEM</strong></td>
<td>Tan</td>
<td>20</td>
<td>0.51097137</td>
<td>2.762049</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_GENERIC_TRANSCRIPTION_PATHWAY</strong></td>
<td>Tan</td>
<td>16</td>
<td>0.49382012</td>
<td>2.3938062</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>NABA_MATRISOME</strong></td>
<td>Tan</td>
<td>52</td>
<td>0.2769824</td>
<td>2.2962203</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION</strong></td>
<td>Tan</td>
<td>15</td>
<td>0.46424875</td>
<td>2.1254985</td>
<td>0.00196437</td>
<td>0.07368225</td>
</tr>
<tr>
<td><strong>NABA_MATRISOME_ASSOCIATED</strong></td>
<td>Tan</td>
<td>36</td>
<td>0.28877153</td>
<td>2.0132291</td>
<td>0.00813008</td>
<td>0.03437438</td>
</tr>
<tr>
<td><strong>REACTOME_ADAPTIVE_IMMUNE_SYSTEM</strong></td>
<td>Tan</td>
<td>49</td>
<td>0.25886484</td>
<td>1.8398738</td>
<td>0.01386138</td>
<td>0.06948614</td>
</tr>
<tr>
<td><strong>REACTOME_HIV_INFECTION</strong></td>
<td>Tan</td>
<td>21</td>
<td>0.30567454</td>
<td>1.7100884</td>
<td>0.02687140</td>
<td>0.06993838</td>
</tr>
<tr>
<td><strong>REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM</strong></td>
<td>Tan</td>
<td>16</td>
<td>0.40826204</td>
<td>1.9951951</td>
<td>0.00581395</td>
<td>0.07636424</td>
</tr>
<tr>
<td><strong>REACTOME_HQST_INTERACTIONS_OF_HIV_FACTORS</strong></td>
<td>Tan</td>
<td>15</td>
<td>0.35319522</td>
<td>1.6167567</td>
<td>0.0323</td>
<td>0.10923666</td>
</tr>
<tr>
<td><strong>NABA_ECM_REGULATORS</strong></td>
<td>Tan</td>
<td>15</td>
<td>0.3412781</td>
<td>1.5847824</td>
<td>0.04637097</td>
<td>0.10880603</td>
</tr>
<tr>
<td><strong>REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT</strong></td>
<td>Turquoise</td>
<td>38</td>
<td>0.52046474</td>
<td>3.5809069</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_RESPIRATORY_ELECTRON_TRANSPORT</strong></td>
<td>Turquoise</td>
<td>27</td>
<td>0.58163541</td>
<td>3.4332826</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_Respiratory_Electron_Transport_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_PUMP</strong></td>
<td>Turquoise</td>
<td>31</td>
<td>0.51505631</td>
<td>3.3772025</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>KEGG_HUNTINGTONS_DISEASE</strong></td>
<td>Turquoise</td>
<td>43</td>
<td>0.45448094</td>
<td>3.3454773</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>KEGG_ALZHEIMERS_DISEASE</strong></td>
<td>Turquoise</td>
<td>38</td>
<td>0.42519427</td>
<td>3.100907</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>KEGG_OXIDATIVE_PHOSPHORYLATION</strong></td>
<td>Turquoise</td>
<td>36</td>
<td>0.40677297</td>
<td>2.9108124</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>KEGG_PARKINSONS_DISEASE</strong></td>
<td>Turquoise</td>
<td>36</td>
<td>0.40677297</td>
<td>2.8989124</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>KEGG_CARDIAC_MUSCLE_CONTRACTION</strong></td>
<td>Turquoise</td>
<td>15</td>
<td>0.53685029</td>
<td>2.4297204</td>
<td>0.00171172</td>
<td>0.09609686</td>
</tr>
<tr>
<td><strong>REACTOME_NGF_SIGNALING_VIA_TRAK_FROM_THE_PLASMA_MEMBRANE</strong></td>
<td>Turquoise</td>
<td>15</td>
<td>0.47475696</td>
<td>2.2123554</td>
<td>0.00213675</td>
<td>0.06058243</td>
</tr>
<tr>
<td><strong>REACTOME_NEURONAL_SYSTEM</strong></td>
<td>Turquoise</td>
<td>27</td>
<td>0.30835184</td>
<td>1.9067713</td>
<td>0.01452282</td>
<td>0.05451782</td>
</tr>
<tr>
<td><strong>REACTOME_Metabolism_of_MRNA</strong></td>
<td>Turquoise</td>
<td>17</td>
<td>0.36503627</td>
<td>1.8311599</td>
<td>0.00798128</td>
<td>0.02636573</td>
</tr>
<tr>
<td><strong>REACTOME_Signaling_by_NGF</strong></td>
<td>Turquoise</td>
<td>22</td>
<td>0.31407282</td>
<td>1.7909727</td>
<td>0.02080312</td>
<td>0.07282838</td>
</tr>
<tr>
<td><strong>REACTOME_PLATELET_ACTIVATION_SIGNALING_AND_AGGREGITION</strong></td>
<td>Turquoise</td>
<td>17</td>
<td>0.35916377</td>
<td>1.7582463</td>
<td>0.01393651</td>
<td>0.08196814</td>
</tr>
<tr>
<td><strong>REACTOME_TRANSMISSION_Across_CHEMICAL_SYNAPSES</strong></td>
<td>Turquoise</td>
<td>19</td>
<td>0.32222223</td>
<td>1.6900269</td>
<td>0.02626262</td>
<td>0.10649994</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>transcript_type</td>
<td>$b_{(i)}$</td>
<td>t-value</td>
<td>p-value</td>
<td>q-value</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>protein_coding</td>
<td>-4339.099284</td>
<td>-4.17296636</td>
<td>1.06E-04</td>
<td>0.0021200</td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td>protein_coding</td>
<td>-194.1447756</td>
<td>-3.892794888</td>
<td>2.66E-04</td>
<td>0.0026600</td>
<td></td>
</tr>
<tr>
<td>CFB</td>
<td>protein_coding</td>
<td>-816.8607338</td>
<td>-3.57516617</td>
<td>7.29E-04</td>
<td>0.0048600</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>protein_coding</td>
<td>-2674.651616</td>
<td>-3.216831651</td>
<td>0.002155381</td>
<td>0.0090439</td>
<td></td>
</tr>
<tr>
<td>VWF</td>
<td>protein_coding</td>
<td>-10483.74203</td>
<td>-3.20519733</td>
<td>0.002260962</td>
<td>0.0090439</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>protein_coding</td>
<td>-2902.166648</td>
<td>-2.992404662</td>
<td>0.00531313</td>
<td>0.0118070</td>
<td></td>
</tr>
<tr>
<td>SERPINB1</td>
<td>protein_coding</td>
<td>-469.2666009</td>
<td>-2.978236654</td>
<td>0.004278888</td>
<td>0.0118070</td>
<td></td>
</tr>
<tr>
<td>MS4A6A</td>
<td>protein_coding</td>
<td>-892.6519535</td>
<td>-2.94793914</td>
<td>0.004738888</td>
<td>0.0118070</td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td>protein_coding</td>
<td>-160.8875175</td>
<td>-2.900623299</td>
<td>0.006358287</td>
<td>0.0123182</td>
<td></td>
</tr>
<tr>
<td>PEF1</td>
<td>protein_coding</td>
<td>1051.251682</td>
<td>2.710880859</td>
<td>0.00892136</td>
<td>0.0136802</td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td>protein_coding</td>
<td>-161.1241294</td>
<td>-2.80411694</td>
<td>0.010102021</td>
<td>0.0144315</td>
<td></td>
</tr>
<tr>
<td>CFB</td>
<td>protein_coding</td>
<td>-455.8770002</td>
<td>-2.60100512</td>
<td>0.01186758</td>
<td>0.0150922</td>
<td></td>
</tr>
<tr>
<td>ETSF</td>
<td>protein_coding</td>
<td>-2998.120486</td>
<td>-2.59435938</td>
<td>0.012073783</td>
<td>0.0150922</td>
<td></td>
</tr>
<tr>
<td>CD14</td>
<td>protein_coding</td>
<td>-2937.772327</td>
<td>-2.556507693</td>
<td>0.01331174</td>
<td>0.0156609</td>
<td></td>
</tr>
<tr>
<td>KIF19</td>
<td>protein_coding</td>
<td>-176.2846059</td>
<td>-2.478319252</td>
<td>0.016240449</td>
<td>0.0176072</td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td>protein_coding</td>
<td>-135.5325182</td>
<td>-2.46453997</td>
<td>0.016812949</td>
<td>0.0176072</td>
<td></td>
</tr>
<tr>
<td>G054549</td>
<td>long noncoding</td>
<td>-76.58008707</td>
<td>-2.44596443</td>
<td>0.01761392</td>
<td>0.0176072</td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>protein_coding</td>
<td>-479.7336288</td>
<td>-2.422108455</td>
<td>0.018692651</td>
<td>0.0176072</td>
<td></td>
</tr>
<tr>
<td>SLC2A5</td>
<td>protein_coding</td>
<td>-680.8779148</td>
<td>-2.407797313</td>
<td>0.019367912</td>
<td>0.0176072</td>
<td></td>
</tr>
<tr>
<td>LIPA</td>
<td>protein_coding</td>
<td>2571.97528</td>
<td>2.377762233</td>
<td>0.02085671</td>
<td>0.0181363</td>
<td></td>
</tr>
<tr>
<td>CFB</td>
<td>protein_coding</td>
<td>-487.8252969</td>
<td>-2.343161594</td>
<td>0.02269128</td>
<td>0.0189151</td>
<td></td>
</tr>
<tr>
<td>IRF1</td>
<td>protein_coding</td>
<td>-893.4275376</td>
<td>-2.183954119</td>
<td>0.033170244</td>
<td>0.0259255</td>
<td></td>
</tr>
<tr>
<td>C1QB</td>
<td>protein_coding</td>
<td>-794.4082368</td>
<td>-2.16374536</td>
<td>0.034765474</td>
<td>0.0259255</td>
<td></td>
</tr>
<tr>
<td>FKBP5</td>
<td>protein_coding</td>
<td>-143.4466841</td>
<td>-2.160400441</td>
<td>0.035035906</td>
<td>0.0259255</td>
<td></td>
</tr>
<tr>
<td>CLIC1</td>
<td>protein_coding</td>
<td>-8449.40885</td>
<td>-1.9737918</td>
<td>0.038346681</td>
<td>0.0381048</td>
<td></td>
</tr>
<tr>
<td>SMARCD3</td>
<td>protein_coding</td>
<td>-4225.619536</td>
<td>-1.917513507</td>
<td>0.040278624</td>
<td>0.0417515</td>
<td></td>
</tr>
<tr>
<td>CLIC1</td>
<td>protein_coding</td>
<td>-6404.73533</td>
<td>-1.874201051</td>
<td>0.046122845</td>
<td>0.0440819</td>
<td></td>
</tr>
<tr>
<td>CD14</td>
<td>protein_coding</td>
<td>-3360.788427</td>
<td>-1.857196165</td>
<td>0.048544994</td>
<td>0.0442226</td>
<td></td>
</tr>
</tbody>
</table>