Characterizing the roles of long non-coding RNA in rat alcohol preference

Ao Zhou
Bioinformatics Program
Indiana University School of Informatics
Indianapolis, IN 46202, USA
aozhou@iu.edu

Weixing Feng
College of Automation
Harbin Engineering University
Harbin, Heilongjiang, China
fengweixing@hrbeu.edu.cn

Yadong Wang
Department of Bioengineering
University of Pittsburgh
Pittsburgh, PA 15213, USA
yaw20@pitt.edu

Howard J. Edenberg
Department of Biochemistry and Molecular Biology
Indiana University School of Medicine
Indianapolis, IN 46202, USA
edenberg@iu.edu

Yunlong Liu*
Department of Medical and Molecular Biology
Indiana University School of Medicine
Indianapolis, IN 46202, USA
yunliu@iu.edu

* Corresponding author

***Abstract*—Alcohol is one of the major threats to health in United States. With the emerging of next-generation sequencing technology, the association between alcohol preference and the variants and expression of genes has been investigated. However, the roles of long non-coding RNAs (lncRNA) in alcohol preference remains unclear. In this study, we identified 37 novel lncRNAs that differentially expressed across alcohol preferring (P) and non-preferring (NP) rats. The functional study on these lncRNAs demonstrates that they are associated with gene regulation, as well as neural functions. This suggests that these lncRNAs may contribute to the alcohol preference behaviors.**

Keywords—alcohol preference; long non-coding RNA; next generation sequencing

I. Introduction

Alcohol is the third leading cause of preventable death in the United States [1]. Alcohol misuse negatively affects the quality of life for millions of Americans, and has profound sociological and economic impacts. The neurobiological basis underlying alcohol dependence is not fully understood, but extensive evidence indicates that genetic factors play key roles in influencing the risk of alcohol dependence [2-9]. Over the past decades, several specific genes have been implicated in the risk of alcoholism [4, 10-20]. In addition, recent studies suggest that epigenetic processes play a critical role in affecting the risks of alcohol dependence [21-23].

Deep sequencing data from the Encyclopedia of DNA Elements Consortium (ENCODE) suggests that over 90% of the human genome can be transcribed, and non-protein-coding RNAs (ncRNA) exceed the number of protein-coding genes [24]. The recent discovery of over 200 ncRNAs significantly enriches the portfolio of potential genetic factors [25]. Rather than being transcriptional noise, many ncRNAs serve as master regulators that affect expression levels of dozens or even hundreds of target genes [26, 27]. These regulatory RNAs integrate signals from both genetic and environmental factors, and therefore can play major roles in controlling alcohol preference. Most notably, a strong association of epigenetic marks with long non-coding RNAs (lncRNAs, >200 nucleotides) in humans and mice was recently described [28]. Many lncRNAs contain conserved elements and show spatiotemporally restricted expression patterns, implying that they are functional and regulated [29]. These lncRNAs are reported to regulate dosage compensation, imprinting, and development by establishing chromatin domains in an allele- and cell-type specific manner [30-32]. It is also reported that lncRNAs are involved in post-transcriptional regulations [33].

It is now possible to identify novel lncRNAs from the high-throughput sequencing data. Guttman et al [34] found that genes being transcribed by RNA polymerase II (Pol II) are marked by trimethylation of lysine 4 of histone 3 (H3K4me3) in the promoters and trimethylation of lysine 36 of histone 3 (H3K36me3) along the transcribed regions. They defined such structure as "K4-K36 domains" and identified more than 1600 previously unknown K4-K36 domains from mouse by CHIP-sequencing; these transcription active regions represent either protein coding genes or lncRNAs.

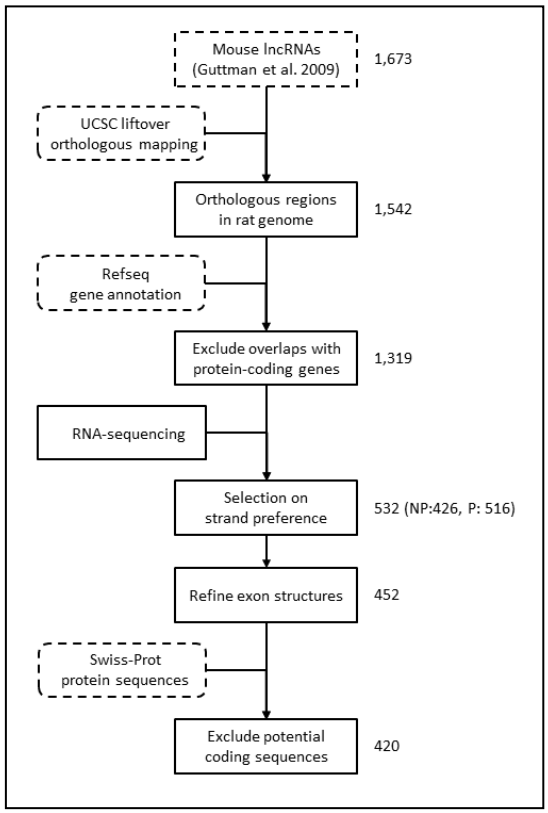
This is the author's manuscript of the article published in final edited form as:

Zhou, A., Wang, Y., Liu, Y., Feng, W., & Edenberg, H. J. (2016). Characterizing the roles of long non-coding RNA in rat alcohol preference. In 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 167–173). https://doi.org/10.1109/BIBM.2016.7822513

In the current study, we designed an RNA-sequencing experiment and a computer approach to identify and characterize novel lncRNAs that are actively transcribed and correlated with alcohol preference in rat. We conducted a scan on the transcriptional intensities within the rat orthologous regions of the mouse K4-K36 domains published by Guttman et al [34], and focused on "intergenic" lncRNAs, i.e., lncRNAs residing outside all known protein-coding genes. We identified 420 novel lncRNAs, among which 37 were differentially expressed between P (alcohol preferring) and NP (alcohol non-preferring) rats. Our pathway analysis on the differentially expressed lncRNAs demonstrated that many of them had shown significant association with neural functions. Our method is also applicable to other diseases and species.

Fig. 1. The workflow of lncRNA annotation.

The dashed boxes indicate extern data source, and solid boxes indicate results generated in our analysis. The numbers at the right of the boxes are the number of putative lncRNAs after each step of filtering. The filtering begins from 1673 K4-K36 domain in mouse and ends up with 420 putative lncRNA regions in rat.



II. Results

In order to understand the role of lncRNA in alcohol preference, we conducted RNA sequencing and bioinformatics analysis on P (alcohol preferring) and NP (alcohol non-preferring) rat strains. Our analysis includes four major steps: (i) deriving the rat orthologous regions of the K4-K36 domains in mouse; (ii) acquiring the transcriptome from the hippocampus of P and NP rats by means of next-generation sequencing; (iii) identifying potential regulatory lncRNAs associated with alcohol consumption, based on the RNA sequencing and epigenetic marker information; and (iv) inferring the functions of lncRNAs differentially expressed in P and NP rat strains (Figure 1).

A. Identifying the rat genomic regions orthologous to the K4-K36 domains in mouse

Guttman et al [34] reported 1673 K4-K36 domains in the mouse genome that may include lncRNAs [34]. To identify rat lncRNAs, we mapped these K4-K36 domains to the rat genome with UCSC LiftOver [35]; 1542 putative lncRNA regions were identified.

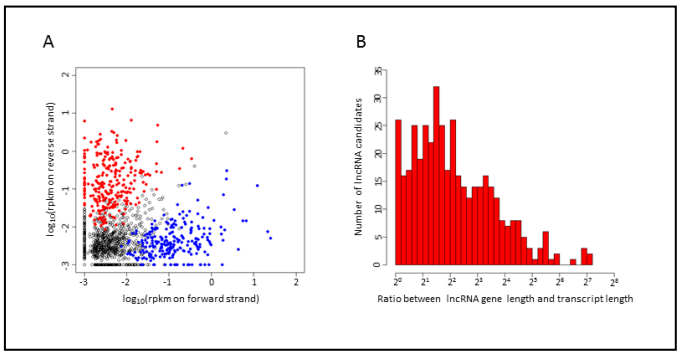
We discarded or truncated the rat orthologous domains to eliminate overlaps with (i) known protein-coding genes in rat, or (ii) orthologous regions of known protein-coding genes in mouse and human. We focused on the remaining 1319 putative lncRNA regions, in which all the known protein coding sequences were excluded.

B. Hippocampus transcriptomes of P and NP rats

To examine the transcription activity of these regions in alcohol-preference, we implemented an RNA sequencing experiment on P and NP rats. P and NP rats [36] are a pair of model animals developed for alcohol dependence research, traits other than alcohol preference were strictly controlled. Total RNA was extracted from the hippocampus of 8 non-inbred P and 8 NP rats, poly-adenylated RNA was selected and reverse transcribed. The resulting cDNA was sequenced using the Illumina Genome Analyzer IIe, with the strand of the RNA transcripts restrained. RNA from each individual rat was sequenced in one Illumina lane that produced 2.8 to 12.8 million mappable reads.

Fig. 2. Features of identified lncRNAs.

(A) Strand preference. The horizontal and vertical axes denote transcription activity on forward and reverse strand, respectively. The circles denotes the lncRNA candidates not showing significant strand preference, while the blue and red dots denotes the lncRNA candidates that are transcribed on the forward and reverse strand, respectively. (B) Distribution of the ratio between lncRNA gene length and exonic region length.



C. Determining potential regulatory lncRNA regions in P and NP rats

Among the regions that were transcribed, we assumed that the strand preference for each transcript should be consistent

across all the samples, and discarded those that were not. With this filtering, 516 and 426 transcripts were derived from the putative rat lncRNA regions from P and NP rats, respectively (Figure 2A). By uniting these two sets of transcripts and removing duplicates, we derived 532 putative lncRNA transcripts with strand specificity.

We used a computational algorithm to annotate exons in the putative lncRNA regions based on the transcriptional intensity. Within each exon, we required at least 8 reads, with a maximum distance between two reads of 25 nucleotides. By discarding the putative lncRNA regions of which the total exonic lengths were less than 200 nucleotides, the candidate pool was reduced to 452 putative lncRNA regions.

We aligned the exonic sequences of putative lncRNAs and known proteins with BlastX [37, 38], and then eliminated a small portion (≈7%) of putative lncRNAs that included exons showing protein-coding capacity (Methods). Eventually, we derived 420 novel lncRNAs with significant transcriptional activity and no significant potential to code for proteins.

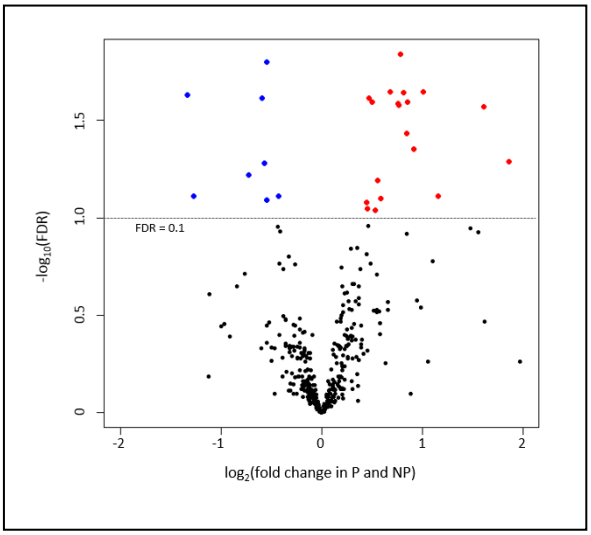
TABLE I. **STATISTICS OF PREDICTED LNCRNA, KNOWN LNCRNA AND PROTEIN-CODING GENES.**

Novel lncRNA indicates the lncRNAs identified by our pipeline; known lncRNAs include known lncRNAs in both mouse and rat; protein-coding genes refers to rat protein-coding genes only.

	Novel lncRNA	**Known lncRNA**	**Protein-coding genes**
Number of regions	420	99	13892
Length of longest transcript (nt)	72075	83437	17599
Length of shortest transcript (nt)	200	374	105
Mean of all transcript lengths (nt)	4053	4947	2131
Median of all transcript lengths (nt)	1939	3240	1767
Maximum exon number	244	48	106
Minimum exon number	1	1	1
Mean of exon number	14	5	9
Median of exon number	8	4	7
Length of longest exon (bp)	10454	83437	11972
Length of shortest exon (bp)	32	20	3
Mean exon length (bp)	283	913	244
Median of exon length (bp)	208	150	132
Maximum expression intensity (rpkm)	1104.17	N/A	2905.00
Minimum expression intensity (rpkm)	0	N/A	0
Mean expression intensity (rpkm)	3.73	N/A	15.44
Median of expression intensity (rpkm)	0.94	N/A	2.26

Fig. 3. Volcano plot of differential expression in P and NP samples.

The black dots denote the lncRNAs that are not differentially expressed. Red and blue dots denote lncRNAs that are significantly higher expressed in P and NP rats, respectively.



These novel lncRNAs are equally distributed along different chromosomes, with 43 and 9 on chr1 and chr12, respectively, which are the longest and shortest chromosomes of rat (Figure S3). The transcript lengths of novel lncRNAs fell between 200 and 72,075 nucleotides, and the ratio of lncRNA transcripts and lncRNA genes ranges from 1 to 165 (Figure 2B), which are similar to known lncRNAs (Table 1). The mean transcriptional intensity of novel lncRNAs is 3.7 RPKM, while the average RPKM of protein-coding genes is 15.44.

D. Inferring the functions of novel lncRNAs and their associations with alcohol preference

Among 420 lncRNAs identified in P or NP rat hippocampus, 37 were differentially expressed at a false discovery rate of 0.1 in our Friedman test (Figure 3). Among the differentially expressed lncRNAs, expression levels of 26 are higher in P rats, while 11 are higher in NP rats. This trend is significantly different from protein-coding genes ($p \leq 0.001$), where expression levels of 1401 and 2009 genes were high in P and NP rats respectively (Table 2). This is consistent with the observations that most known lncRNAs exert their functions by repressing the expression of protein-coding genes.

TABLE II. **CHI-SQUARE TEST OF LNCRNA NEGATIVE REGULATION ON GENE.**

Up-regulated indicates the lncRNA or gene is up-regulated in P rats vs NP rats; conversely, down-regulated indicates the lncRNA or gene is down-regulated in P rats vs NP rats.

	Up-regulated	**Down-regulated**
lncRNA	26	11
Protein coding gene	1401	2009

P-value=0.0006 (Chi-square)

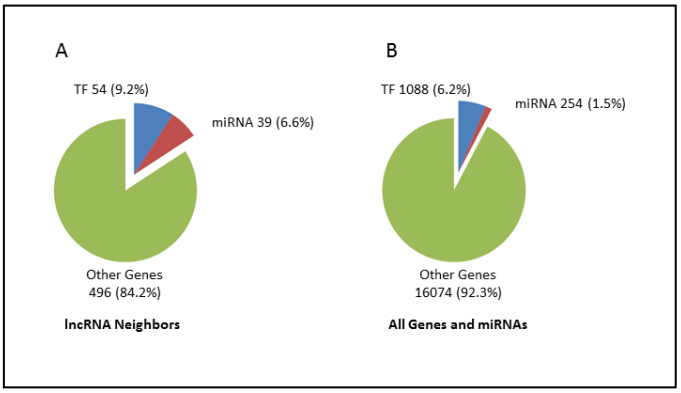
We used a generalized linear model to characterize the correlations between the transcription levels of lncRNAs and genes. Among the 2873 genes significantly correlating (FDR<0.2, p-value<0.005) with the differentially expressed lncRNAs, 120 are correlated with more than 3 lncRNAs. Several of lncRNA correlating genes are known as associated with alcohol dependence, including ALDH1A1, ALDH9A1, GABRA2, CHRM2, PDYN and CNR1. We conducted a pathway analysis on these 120 genes with Ingenuity Pathway Analysis (IPA) and found that physiological function most frequently associated to lncRNAs is nervous system development and functions.

Among all the differentially expressed lncRNAs, 22 correlate with more than 30 genes each. 15 of these lncRNAs correlate with genes enriched in nervous system development and function, 8 with neurological diseases, and 6 with behaviors. Moreover, 10 of the 22 lncRNAs are associated to genetic disorder, which may reveal the hereditary nature of alcoholism.

Among genes proximal to lncRNAs differentially expressed between P and NP, 9.2% are annotated as transcription factors, and 6.6% as miRNA; for genes not proximal to differentially expressed lncRNAs, however, this proportion dropped to 6.2% and 1.5% for transcription factors and miRNAs. This difference is significant with $\chi 2$ p-value 2.2×10-16, suggesting that many of the novel lncRNAs may be associated with neighboring transcription factors and miRNAs, and work in a cis-acting manner.

Fig. 4. Potential cis-regulation of lncRNA.

The pie-chart demonstrates the percentage of transcription factors, miRNA and other genes in lncRNA neighbors and in all genes. The proportion of transcription factors and miRNAs in lncRNA neighbors is significantly higher (p=2.2×10-16) than that in all genes.



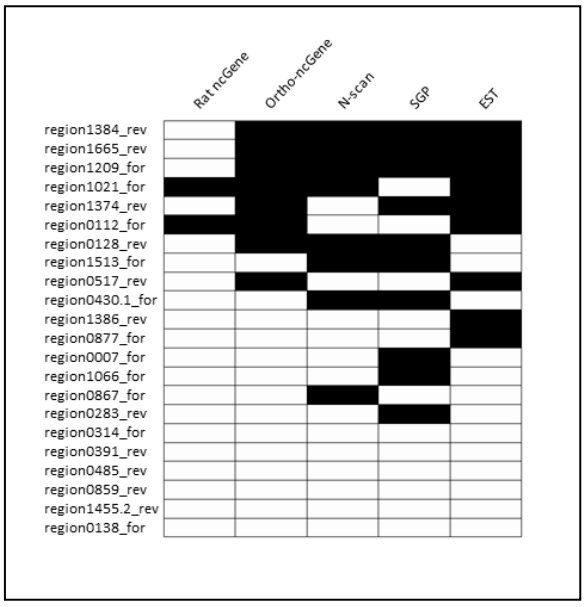
III. Discussions

We report an RNA-seq experiment on the hippocampus of P and NP rats, and a bioinformatics strategy to identify lncRNAs from the RNA-seq information and characterize their roles in alcohol preference. Our strategy includes four components, orthologous lncRNA region mapping from mouse to rat, RNA-seq on P and NP rats, lncRNA annotation and pathway characterization. We identified 420 lncRNAs, 37 of which are differentially expressed across P and NP rats. By applying a generalized linear model to differentially expressed lncRNAs and protein-coding genes, we derived 3699 significantly correlated lncRNA-gene pairs involving 2873 genes. We created a set of significantly correlated genes for each lncRNA, and inferred their functional roles by pathway analysis. The result revealed that 15 are significantly correlated with nervous system development and function. Our statistical analysis also revealed that the proportion of TF and miRNA are significantly higher among the lncRNA neighboring genes than other genes, implying a cis-acting mechanism (Figure 4).

Evidence was found supporting the existence and potential regulation functions of the differentially expressed lncRNAs. Region1384_rev is significantly correlated with 824 genes, of which 61 are significantly associated with nervous system development and function; it is located in the promoter (179 nt upstream of transcription start site) of a protein-coding gene CHD2, whose product alters gene expression by modification of chromatin structures [39]. Given the observations that many lncRNAs locating in the promoter of protein-coding genes possess regulation functions on the corresponding genes, the location of region1384_rev suggests a tremendous possibility of a regulatory role upon CHD2, and thus regulating a large group of genes by chromatin modification. Moreover, we observed several rat ESTs and orthologous non-coding genes of mouse and human within this region, verifying the existence of this lncRNA. (Figure 5)

Fig. 5. Observations supporting the existence of lncRNA.

A dark grid indicates that evidence was found for the corresponding lncRNA, while a white grid indicates no such evidence was found. Rat ncGene stands for rat non-coding genes; Ortho-ncGene stands for orthologous non-coding genes; N-scan and SGP stands for N-scan and SGP gene prediction, respectively; EST stands for expressed sequence tags. The lncRNAs are sorted by the number of evidences found.



Evidence was also found in support of the existence of other differentially expressed lncRNAs and their potential functions on gene regulation and signal transduction. Four

lncRNAs (region0877_for, region1066_for, region0112_for, and region0138_for) were identified sequentially adjacent to zinc-finger proteins, which were generally found as a component of transcription factors; besides, region0867_for was identified neighboring a gene coding transcription factor, and region0283_rev was found near a gene coding a DNA binding protein; these observations indicates that lncRNA may regulate the expression of transcription factor genes. Moreover, two lncRNAs (region0314_for and region0007_for) were identified adjacent to genes coding G-protein regulation proteins (Rho GTPase activating protein 5 and Rho guanine nucleotide exchange factor), which implies that lncRNAs may also play roles in cell signaling. Region1374_rev was found next to a small-nuclear RNA that is involved in snRNA modifications, and region0430.1_for was next to an RNA motif binding protein; this indicates that lncRNAs may also be involved in RNA regulations.

The "K4-K36" domain only represents a transcription active region that may be a gene, it is unable to differentiate an intronic lncRNA from a novel exon. Therefore our strategy focused on intergenic ncRNAs only, we may have missed the lncRNAs located in intronic regions and untranslated regions. In addition, we required the expression of all candidate lncRNAs be higher on one of the two strands to eliminate noises and error in sequencing data. Because of this criterion, we may also have missed the lncRNAs that are transcribed on an antisense strand of a non-coding gene. Nevertheless, once we have new transcriptome data with deeper sequencing and longer reads, we will be able to identify more lncRNAs from intragenic regions.

In the future, we plan to conduct RT-PCR experiments to validate the transcriptional activity of the lncRNAs which are likely to be associated to gene regulations in alcohol preference. We also plan to sequence more brain regions of rats at different drinking levels, thus to characterize the tissue related functions of lncRNA and their association with different drinking scores.

IV. Methods

A. Eliminating protein-coding regions

The annotation library of rat genes and orthologous regions in human and mouse were downloaded from UCSC table browser. To remove non-coding genes from the library, the genes that cannot be mapped to UniProt accessions were eliminated. Then the putative lncRNA regions were superimposed to the protein-coding genes and classified into three categories according to the overlap length, 1) non-overlapping, 2) complete overlapping and 3) partial overlapping. The putative lncRNA regions not overlapping with protein-coding regions were retained for next step, while the complete overlapping regions were discarded. For partially overlapped regions, we truncated the overlap and shortened the lncRNA regions, if transcriptional activity were detected outside the overlap.

B. Determining transcriptional strand preference

To determine the transcriptional strand preference within the putative lncRNA regions, we firstly calculated two transcription intensities in RPKM for each lncRNA region in each sample, one is for forward strand and the other is for reverse strand. Then we conducted a Friedman test to compare the transcriptional intensities on different strands. A putative lncRNA region with p-value<0.01 was defined as significantly strand biased and the higher expressed strand were defined as the sense strand, while the insignificant regions were eliminated. Both the RPKM calculation and Friedman test were implemented with Partek® Genomics Suite® software, version 6.6 Copyright ©; 2016 Partek Inc., St. Louis, MO, USA.

C. Refining exon structures

To reduce computational time, we eliminated all the RNA-seq reads outside the putative lncRNA regions and on the antisense strand. Then SAMtools was used to pileup all the reads. A list of expressed chunks was generated based on the pileup file. If the distance between two chunks was less than 25 basepairs, they were merged into one chunk. If the number of reads covered by a chunk was less than 8, this chunk was discarded. The derived chunks were defined as putative exons and mapped to the lncRNA regions. A refflat file was generated for each lncRNA regions, annotating their coordinates and putative exon structures.

D. Detecting protein-coding potential

The sequences of the putative exons were extracted and aligned against SwissProt with BlastX. In the cases that several alignments were generated for one exon sequence, we only retained the one with the highest alignment score. Then we calculated the average $\log_{10}$(E-value) of all exons for each lncRNA to evaluate their protein-coding potential. LncRNA regions with an average $\log_{10}$(E-value) < -3 were discarded; the rest regions were defined as non-coding potential regions.

E. Deriving significantly correlated lncRNA-gene pairs with generalized linear model

It is reasonable to assume that number of sequence tags identified in each lncRNA region follows a Poisson distribution; we therefore used a generalized linear model to infer the relationship between the expression intensity of lncRNAs and protein-coding genes.

$$\begin{aligned} Y_{ik} &\sim Poisson\left(\mu_{ijk}\right) \\ \log\left(\mu_{ijk}\right) &= N_{ik} + \alpha + \beta_1 s_k + \beta_2 g_{jk} + \beta_3 b_k \end{aligned} \quad (1)$$

In this model, i, j, and k are the indices of lncRNA, gene and animal. Y_{ik} is the number of RNA-seq read counts in the region of lncRNA i in animal k; μ_{ijk} is the expected value of Y_{ik}; N_{ik} is a constant value that serves as a normalization factor to balance sample and lncRNA specific variation. Here, $N_{ij}=\log(K_i)+\log(M_k)$, where K_i is the length of exon model of lncRNA i and M_k is the total number of mappable reads for sample k. s_k is the strain of animal k (P or NP); g_{ik} is the transcription intensity (RPKM) of gene i in animal k; b_k is the batch effect of the experiment (Run1 or Run2). The

significance of β_2 was used to evaluate the correlation between lncRNA i and gene j.

To simplify the model and derive the most trustworthy results, we assumed that there is no interaction between strain and gene transcription intensity. To rule out the cases that the interaction may exist, we used another generalized linear model to identify these cases.

$$\begin{aligned} &Y_{ik} \sim Poisson(\mu_{ijk}) \\ &\log(\mu_{ijk}) = N_{ik} + \alpha + \beta_1 s_k + \beta_2 g_{jk} + \beta_3 b_k + \beta_4 s_k g_{jk} \end{aligned} \quad (2)$$

If the coefficient of strain-gene interaction term β_4 was significant ($p \leqslant 0.05$), then the corresponding lncRNA-gene pair was discarded.

To increase the reproducibility of the results, we required that all lncRNAs and genes should have transcriptional intensities more than 0.5 RPKM in at least 8 samples. All lncRNA-gene pairs involving ineligible lncRNA or genes were discarded.

REFERENCES

[1] A. H. Mokdad, J. S. Marks, D. F. Stroup, and J. L. Gerberding, "Actual causes of death in the United States, 2000," JAMA, vol. 291, pp. 1238-45, Mar 10 2004.

[2] P. H. Blackman, "Actual causes of death in the United States," JAMA, vol. 271, pp. 659-60; author reply 660-1, Mar 2 1994.

[3] H. J. Edenberg, J. Reynolds, D. L. Koller, H. Begleiter, K. K. Bucholz, P. M. Conneally, et al., "A family-based analysis of whether the functional promoter alleles of the serotonin transporter gene HTT affect the risk for alcohol dependence," Alcohol Clin Exp Res, vol. 22, pp. 1080-5, Aug 1998.

[4] D. M. Dick, H. J. Edenberg, X. Xuei, A. Goate, S. Kuperman, M. Schuckit, et al., "Association of GABRG3 with alcohol dependence," Alcohol Clin Exp Res, vol. 28, pp. 4-9, Jan 2004.

[5] H. J. Edenberg, W. N. Strother, J. N. McClintick, H. Tian, M. Stephens, R. E. Jerome, et al., "Gene expression in the hippocampus of inbred alcohol-preferring and -nonpreferring rats," Genes Brain Behav, vol. 4, pp. 20-30, Feb 2005.

[6] D. M. Dick, A. Agrawal, M. A. Schuckit, L. Bierut, A. Hinrichs, L. Fox, et al., "Marital status, alcohol dependence, and GABRA2: evidence for gene-environment correlation and interaction," J Stud Alcohol, vol. 67, pp. 185-94, Mar 2006.

[7] H. J. Edenberg, X. Xuei, H. J. Chen, H. Tian, L. F. Wetherill, D. M. Dick, et al., "Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis," Hum Mol Genet, vol. 15, pp. 1539-49, May 1 2006.

[8] T. Liang, M. W. Kimpel, J. N. McClintick, A. R. Skillman, K. McCall, H. J. Edenberg, et al., "Candidate genes for alcohol preference identified by expression profiling in alcohol-preferring and -nonpreferring reciprocal congenic rats," Genome Biol, vol. 11, p. R11, 2010.

[9] H. J. Edenberg, "Common and rare variants in alcohol dependence," Biol Psychiatry, vol. 70, pp. 498-9, Sep 15 2011.

[10] J. Song, D. L. Koller, T. Foroud, K. Carr, J. Zhao, J. Rice, et al., "Association of GABA(A) receptors and alcohol dependence and the effects of genetic imprinting," Am J Med Genet B Neuropsychiatr Genet, vol. 117B, pp. 39-45, Feb 2003.

[11] H. J. Edenberg, D. M. Dick, X. Xuei, H. Tian, L. Almasy, L. O. Bauer, et al., "Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations," Am J Hum Genet, vol. 74, pp. 705-14, Apr 2004.

[12] J. C. Wang, A. L. Hinrichs, H. Stock, J. Budde, R. Allen, S. Bertelsen, et al., "Evidence of common and specific genetic effects: association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome," Hum Mol Genet, vol. 13, pp. 1903-11, Sep 1 2004.

[13] D. M. Dick, J. C. Wang, J. Plunkett, F. Aliev, A. Hinrichs, S. Bertelsen, et al., "Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1," Alcohol Clin Exp Res, vol. 31, pp. 1645-53, Oct 2007.

[14] H. J. Edenberg, X. Xuei, L. F. Wetherill, L. Bierut, K. Bucholz, D. M. Dick, et al., "Association of NFKB1, which encodes a subunit of the transcription factor NF-kappaB, with alcohol dependence," Hum Mol Genet, vol. 17, pp. 963-70, Apr 1 2008.

[15] D. M. Dick, F. Aliev, J. C. Wang, S. Saccone, A. Hinrichs, S. Bertelsen, et al., "A Systematic single nucleotide polymorphism screen to fine-map alcohol dependence genes on chromosome 7 identifies association with a novel susceptibility gene ACN9," Biol Psychiatry, vol. 63, pp. 1047-53, Jun 1 2008.

[16] H. J. Edenberg, J. Wang, H. Tian, S. Pochareddy, X. Xuei, L. Wetherill, et al., "A regulatory variation in OPRK1, the gene encoding the kappa-opioid receptor, is associated with alcohol dependence," Hum Mol Genet, vol. 17, pp. 1783-9, Jun 15 2008.

[17] J. C. Wang, R. Grucza, C. Cruchaga, A. L. Hinrichs, S. Bertelsen, J. P. Budde, et al., "Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence," Mol Psychiatry, vol. 14, pp. 501-10, May 2009.

[18] A. C. Chen, Y. Tang, M. Rangaswamy, J. C. Wang, L. Almasy, T. Foroud, et al., "Association of single nucleotide polymorphisms in a glutamate receptor gene (GRM8) with theta power of event-related oscillations and alcohol dependence," Am J Med Genet B Neuropsychiatr Genet, vol. 150B, pp. 359-68, Apr 5 2009.

[19] A. C. Chen, N. Manz, Y. Tang, M. Rangaswamy, L. Almasy, S. Kuperman, et al., "Single-nucleotide polymorphisms in corticotropin releasing hormone receptor 1 gene (CRHR1) are associated with quantitative trait of event-related potential and alcohol dependence," Alcohol Clin Exp Res, vol. 34, pp. 988-96, Jun 2010.

[20] J. Derringer, R. F. Krueger, N. Manz, B. Porjesz, L. Almasy, E. Bookman, et al., "Nonreplication of an association of SGIP1 SNPs with alcohol dependence and resting theta EEG power," Psychiatr Genet, vol. 21, pp. 265-6, Oct 2011.

[21] V. Nieratschker, A. Batra, and A. J. Fallgatter, "Genetics and epigenetics of alcohol dependence," J Mol Psychiatry, vol. 1, p. 11, 2013.

[22] R. Zhang, Q. Miao, C. Wang, R. Zhao, W. Li, C. N. Haile, et al., "Genome-wide DNA methylation analysis in alcohol dependence," Addict Biol, vol. 18, pp. 392-403, Mar 2013.

[23] S. P. Farris, A. Z. Pietrzykowski, M. F. Miles, M. A. O'Brien, P. P. Sanna, S. Zakhari, et al., "Applying the new genomics to alcohol dependence," Alcohol, vol. 49, pp. 825-36, Dec 2015.

[24] E. P. Consortium, "An integrated encyclopedia of DNA elements in the human genome," Nature, vol. 489, pp. 57-74, Sep 6 2012.

[25] X. C. Quek, D. W. Thomson, J. L. Maag, N. Bartonicek, B. Signal, M. B. Clark, et al., "lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs," Nucleic Acids Res, vol. 43, pp. D168-73, Jan 2015.

[26] J. A. Goodrich and J. F. Kugel, "Non-coding-RNA regulators of RNA polymerase II transcription," Nat Rev Mol Cell Biol, vol. 7, pp. 612-6, Aug 2006.

[27] I. Martianov, A. Ramadass, A. Serra Barros, N. Chow, and A. Akoulitchev, "Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript," Nature, vol. 445, pp. 666-70, Feb 8 2007.

[28] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long non-coding RNAs: insights into functions," Nature Reviews Genetics, vol. 10, pp. 155-159, Mar 2009.

[29] A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, et al., "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes," Genome Res, vol. 15, pp. 1034-50, Aug 2005.

[30] J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, and J. T. Lee, "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome," Science, vol. 322, pp. 750-6, Oct 31 2008.

[31] Y. Ogawa, B. K. Sun, and J. T. Lee, "Intersection of the RNA interference and X-inactivation pathways," Science, vol. 320, pp. 1336-41, Jun 6 2008.

[32] X. Wang, S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, et al., "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription," Nature, vol. 454, pp. 126-30, Jul 3 2008.

[33] Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, and K. W. Kinzler, "The antisense transcriptomes of human cells," Science, vol. 322, pp. 1855-7, Dec 19 2008.

[34] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, et al., "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals," Nature, vol. 458, pp. 223-7, Mar 12 2009.

[35] P. A. Fujita, B. Rhead, A. S. Zweig, A. S. Hinrichs, D. Karolchik, M. S. Cline, et al., "The UCSC Genome Browser database: update 2011," Nucleic Acids Res, vol. 39, pp. D876-82, Jan 2011.

[36] L. Lumeng, T. D. Hawkins, and T. K. Li, "NEW STRAINS OF RATS WITH ALCOHOL PREFERENCE AND NONPREFERENCE," in Alcohol and Aldehyde Metabolizing Systems, ed: Academic Press, 1977, pp. 537-544.

[37] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," J Mol Biol, vol. 215, pp. 403-10, Oct 5 1990.

[38] W. Gish and D. J. States, "Identification of protein coding regions by database similarity search," Nat Genet, vol. 3, pp. 266-72, Mar 1993.

[39] D. Rodriguez, G. Bretones, V. Quesada, N. Villamor, J. R. Arango, A. Lopez-Guillermo, et al., "Mutations in CHD2 cause defective association with active chromatin in chronic lymphocytic leukemia," Blood, vol. 126, pp. 195-202, Jul 9 2015.