

# Integrative multi-omics analysis revealed SNP-lncRNA-mRNA (SLM) networks in human peripheral blood mononuclear cells

Wei Xia<sup>1,2,3</sup> · Xiao-Wei Zhu<sup>1,2</sup> · Xin-Bo Mo<sup>1,2</sup> · Long-Fei Wu<sup>1,2</sup> · Jian Wu<sup>4</sup> · Yu-Fan Guo<sup>4</sup> · Ke-Qin Zeng<sup>4</sup> · Ming-Jun Wang<sup>4</sup> · Xiang Lin<sup>1,2</sup> · Ying-Hua Qiu<sup>1,2</sup> · Lan Wang<sup>1,2</sup> · Pei He<sup>1,2</sup> · Fang-Fei Xie<sup>1,2</sup> · Peng-Fei Bing<sup>1,2</sup> · Xin Lu<sup>1,2</sup> · Yao-Zhong Liu<sup>5</sup> · Neng-Jun Yi<sup>6</sup> · Fei-Yan Deng<sup>1,2</sup> · Shu-Feng Lei<sup>1,2</sup>

Received: 8 December 2016 / Accepted: 20 February 2017 / Published online: 28 February 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Long non-coding RNAs (lncRNAs) serve as important controller of cellular functions via regulating RNA transcription, degradation and translation. However, what are the regulation patterns of lncRNAs on downstream mRNA and how the upstream genetic variants regulate lncRNAs are largely unknown. We first performed a comprehensive expression quantitative trait locus (eQTL) analysis (MatrixeQTL package, R) using genome-wide lncRNA expression and SNP genotype data from human peripheral blood mononuclear cells (PBMCs) of 43 unrelated individuals. Subsequently, multi-omics integrative network analysis was applied to construct SNP-lncRNA-mRNA (SLM)

interaction networks. The causal inference test (CIT) was used to identify lncRNA-mediated (epi-) genetic regulation on mRNA expressions. Our eQTL analysis detected 707 pairs of *cis*-effect associations ( $p < 5.64E-06$ ) and 6657 trans-effect associations ( $p < 3.51E-08$ ), respectively. We also found that top significant *cis*-eSNPs were enriched around the lncRNA transcription start site regions, and that enrichment patterns of *cis*-eSNPs differs among different lncRNA sizes (small, medium and large). The constructed SLM interaction networks (1 primary networks and four small separate networks) showed various complex interaction patterns. Especially, the in-depth CIT detected 50 significant lncRNA-mediated SLM trios, and some hot-spots (e.g., SNPs: rs926370, rs7716167 and rs16880521; lncRNAs: HIT000061975 and ENST00000579057.1). This study represents the first effort of dissecting the SLM interaction patterns in PBMCs by multi-omics integrative network analysis and causal inference test for clearing the regulation chain. The results provide novel insights into the regulation patterns of lncRNA, and may facilitate investigations of PBMC-related immune physiological process and immunological diseases in the future.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-017-1771-1) contains supplementary material, which is available to authorized users.

✉ Shu-Feng Lei  
leisf@suda.edu.cn

- 1 Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University, Suzhou 215123, Jiangsu, People's Republic of China
- 2 Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Suzhou 215123, Jiangsu, People's Republic of China
- 3 Department of Statistical, The Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu, People's Republic of China
- 4 Department of Rheumatology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, People's Republic of China
- 5 Department of Biostatistics and Bioinformatics, Center of Genomics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
- 6 Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

## Abbreviations

|         |  |
|---------|--|
| 3'UTRs  | 3' Untranslated regions  |
| CIT     | Causal inference test  |
| DAVID   | The database for annotation, visualization, and integrated discovery |
| elncRNA | eQTL lncRNA  |
| eProbe  | eQTL probe   |
| eQTL    | Expression quantitative trait locus                                  |
| eSNP    | eQTL SNPs  |
| FDR     | False-discovery rate   |
| GO      | Gene ontology  |
| GWAS    | Genome-wide association studies                                      |

|            |   |
|------------|---|
| KEGG       | Kyoto encyclopedia of genes and genomes |
| LD         | Linkage disequilibrium                  |
| lncR-eQTLs | lncRNA eQTLs                            |
| lncRNAs    | Long non-coding RNAs                    |
| PBMCs      | Peripheral blood mononuclear cells      |
| PCC        | Pearson correlation coefficients        |
| RA         | Rheumatoid arthritis                    |
| SLM        | SNP-lncRNA-mRNA                         |
| SNPs       | Single nucleotide polymorphisms         |
| TES        | Transcription end site                  |
| TFBS       | Transcription factor binding sites      |
| TSS        | Transcription start site                |

## Introduction

Long non-coding RNA (lncRNA) is a class of non-protein-coding transcripts with more than 200 nucleotides in length, which has varied secondary structures and spatial features that allow them to bind DNA, RNA or protein partners (Guttman et al. 2009; Ponting et al. 2009). Previous studies have pointed out that lncRNAs may play vital roles in a wide range of biological and cellular processes, including development, localization, alternative splicing, chromatin remodeling, cell cycle control and cell apoptosis, migration and metabolism (Cabianca et al. 2012; Gupta et al. 2010; Kino et al. 2010; Klattenhoff et al. 2013; Meola et al. 2012; Tripathi et al. 2010; Willingham et al. 2005). Recently, evidences showed that lncRNAs can serve as important controller of cellular functions via regulating RNA transcription and epigenetic modulation, and enhancing RNA degradation (Mercer et al. 2009; Prensner and Chinnaiyan 2011; Wang and Chang 2011).

Over the past decade, genome-wide association studies (GWAS) have identified a large number of SNPs that are associated with complex traits and diseases (Altshuler et al. 2008; Lettre and Rioux 2008; McCarthy and Hirschhorn 2008). Surprisingly, only 7% of these identified variations are located in protein-coding regions (Kumar et al. 2012; Pennisi 2011), leaving a majority of them located at the non-coding intervals, including lncRNAs (Pennisi 2010), suggesting that these SNPs are not likely to alter the protein-coding sequence or protein structure, but most likely to involve other regulation models including regulating gene transcription. After large scale GWAS research, a big challenge is to dissect the functions of the identified SNPs that have shown associations with diseases, and to illustrate the functional mechanisms underlying the associations. Considering that mRNA, miRNA, lncRNA serve as intermediate phenotypes linking functional SNPs and disease phenotypes, exploring the relationship between SNPs and these intermediate phenotypes may facilitate understanding the identified associations. Expression quantitative trait locus

(eQTL) analysis (Schadt et al. 2003) is a powerful method to detect the effects of genetic variation on gene expression or human transcripts (Borel et al. 2011; Huan et al. 2015; Kumar et al. 2013; Veyrieras et al. 2008). So far, most of the previous eQTL studies have focused on mRNA expression and only in two levels. Studies on lncRNA expression regulation, including lncRNA eQTL (lncR-eQTL) analysis, are rare, and no study on multi-omics levels. It has been recognized that the expression of lncRNA is under complex regulation (e.g., genetic variations), and lncRNAs can serve as important intermediate controllers in regulating downstream molecular and cellular functions, including mRNA expression. Therefore, revealing the genome-wide SNP-lncRNA-mRNA (SLM) interaction patterns, in disease-relevant cells or tissues, would facilitate uncovering functional mechanisms underlying the associations, mediated by lncRNAs especially.

In this study, we conducted integrative analyses of multiple omics data to reveal genome-wide SLM interaction patterns in peripheral blood mononuclear cells (PBMCs), which consist of several important immunity cells (monocyte, T and B lymphocyte), and play a decisive role in the process of immune response. First, we performed a comprehensive genome-wide lncR-eQTL analysis, and described the distribution characteristics, functional annotation and enrichment for the regulatory SNPs (eSNPs). Second, to discover genome-wide SLM interaction patterns, we applied an integrative network approach that combined eQTL analysis with network analysis. Third, we performed causal inference test (CIT) (Millstein et al. 2009) to assess the regulation chain of SLM. To the best of our knowledge, this is the first effort of dissecting the SLM interaction patterns in PBMCs by multi-omics integrative network analysis and CIT. The results provide novel insights into the regulation patterns of lncRNA, and may facilitate investigations of PBMC-related immune physiological process and immunological diseases in the future.

## Materials and methods

### Sample

A total of 43 female subjects from Chinese Han ethnicity were recruited. Peripheral blood (15 ml) was collected by phlebotomy and stored in sodium citrate supplemented vacuum tubes. PBMCs were isolated by density gradient centrifugation using Lymphoprep (Sigma, life science, USA) within 4 h after phlebotomy. The isolated PBMCs were treated with TRIzol reagent (Invitrogen, Carlsbad, California, USA) to avoid RNA degradation, and then stored at  $-80^{\circ}\text{C}$ . The study was approved by the ethical committee

of Soochow University. The written informed consent was obtained from all the subjects.

### Genome-wide genotyping and expression profiling

The experiment was performed in the laboratory of CapitalBio Corporation (Beijing, China). In the discovery stage, DNA was isolated from PBMCs using phenol–chloroform extraction and ultrapurification method. Affymetrix Genome-Wide Human SNP Array 6.0 chips were employed for SNP genotyping by following the protocol recommended by the manufacturer. Systematic quality controls using the PLINK software and R statistics were applied on the raw genotyping data to filter out unqualified samples and SNPs. In particular, SNPs with a minor allele frequency less than 5%, or a call-rate less than 95% were excluded, finally resulting in 551,745 SNPs used in further analysis.

Total RNA was extracted using TRIzol reagent according to the instructions recommended by the manufacturer and then quantified by using NanoDrop ND-1000 (Thermo Scientific, Wilmington, Delaware, USA) spectrophotometer. RNA integrity was determined with 1% formaldehyde denaturing gel electrophoresis. Genome-wide lncRNA and mRNA expression was profiled using lncRNA + mRNA Human Gene Expression Microarray V4.0 (CapitalBio Corp, Beijing, China) according to the manufacturer's instructions. The data was extracted by Agilent Feature Extraction (V10.7) and was summarized, normalized and controlled for quality using GeneSpring GX program (V12.0). Then, log<sub>2</sub> transformation was applied to the data using the Adjust Data function of Multiexperiment Viewer (MeV) software. Finally, the data visualization was performed with Java Treeview software. Subsequently, after further filtering out probes with detection rate less than 80% and/or incomplete annotation information, a total of 17,190 unique lncRNA probes and 21,323 unique mRNA probes were used for further analysis.

### lncRNA-eQTL analysis

When detecting *cis*-lncR-eQTL, SNPs were confined within 1 Mb distant from the transcription start site (TSS) or transcription end site (TES) of lncRNAs. For each lncRNA-SNP pair, a multivariate linear regression analysis was conducted, after adjusting for disease states using DAS28 (van Riel and Fransen 2005) and age, to detect association between SNPs and lncRNA expression using R package *MatrIXeQTL* (Huang and Cai 2013; Qi et al. 2014; Shabalin 2012). Benjamini-Hochberg false-discovery rate (FDR) was used to correct for multiple testing. Under the FDR level of 5%, the significance threshold for association was 5.64E−06 and 3.51E−08 for *cis*- and

*trans*-lncRNA-eQTLs, respectively. For significant probe-SNP pairs, we defined the corresponding SNP, probe, and lncRNA as eSNP, eProbe, and elncRNA, respectively. The contribution to elncRNA expression variability (beta-value) by each eSNP was calculated in the linear regression analysis.

### lncRNA structure and classification

lncRNA Structure annotation was obtained from the ENSEMBL, UCSC, RNAdb, UCR, Human lncRNA Catalog, RefSeq lncRNA table, and so on, under human genome reference version 37 (hg19). For each lncRNA, the chromosome physical locations of its TSS and TES were obtained from the fields “start” and “end” of the database annotation table, respectively.

To comprehensively display the distribution of *cis*-eSNPs in lncRNA transcripts, we divided the *cis*-acting region into discrete bins as follows. First, since there is dramatic variation in gene size, three categories based on transcript length were generated: small lncRNAs (0–10 kb), medium lncRNAs (10–50 kb) and large lncRNAs (>50 kb). The *cis*-acting region is defined as spanning from TSS to TES. Transcribed regions were split into fixed numbers of bins: each small gene was split into 10 bins of equal size, medium genes into 25 bins and large genes into 15 bins (Veyrieras et al. 2008). The regions outside the transcript were also assigned to bins based on their distance to the TSS (for the upstream region) or TES (downstream). The bins outside the transcript were 1 kb as a unit for small and medium genes and 10 kb as a unit for large genes.

The lncRNAs were classified into five major types: intergenic, intronic, bidirectional (or divergent), antisense, and unknown, based on the referred categorization of lncRNAs in the GENCODE gene annotation (Derrien et al. 2012).

### Annotation and enrichment analysis of eSNPs

To determine whether the significant lncRNA-eQTL SNPs were enriched within certain functional domains on the genome, we annotated the eSNPs using SNPInfo (Xu and Taylor 2009), which integrate multiple data resources [such as the ENCODE Project (Kellis et al. 2014), Ensembl (Hubbard et al. 2002), PolyPhen (Adzhubei et al. 2010), SNPs3D (Yue et al. 2006) and several miRNA databases]. PolyPhen is a software tool which predicts possible impact of amino acid substitution on the structure and function of human proteins using straightforward physical and evolutionary comparison. SNPs3D is a website which assigns molecular functional effects of non-synonymous SNPs based on structure and sequence analysis. In particular, Svm structure means SNPs can affect protein function primarily by decreasing protein stability; Svm profile means

SNP are disease association mutations. We used a random fraction of the total SNP data, which is 10-fold sized than the eSNP dataset, to generate a null distribution. Finally, we evaluated the enrichment for each functional category of eSNPs through comparing these two datasets using a Fisher's exact test.

### Network construction and module identification

#### *Construction of the SNP-lncRNA-mRNA (SLM) interaction network*

Combining the strength of eQTL analysis with network analysis, we build an integrative network to understand the complex relationship among SNP genotype, lncRNA expression, and mRNA gene expression. The integrative network is composed of various types of relationships among SNPs, lncRNAs and mRNAs, including SNP-lncRNAeQTL association pairs, correlation networks of elncRNAs and mRNAs expression, and SNP-mRNA eQTL association pairs. First, we selected SNP-lncRNA pairs for *cis*-eQTL effects at a FDR level of 0.05. Second, to detect elncRNAs and mRNAs co-expression, we calculated the Pearson correlation coefficients (PCC) of each elncRNA paired with mRNA. The  $PCC < -0.9$  or  $>0.9$  was considered as statistically significant co-expression. Third, eQTL analysis was performed for eSNPs and mRNAs significantly coexpressed with elncRNAs ( $PCC < -0.9$  or  $>0.9$ ). Either *cis*- or *trans*-eQTLs with  $FDR < 0.05$  would be selected as SNP-mRNA association pairs. The network was constructed using open source bioinformatics software Cytoscape3.2.1 (Institute of Systems Biology in Seattle).

#### *Linkage disequilibrium (LD) analysis*

For eSNPs close to each other in physical locations along the chromosomes, we conducted LD analysis in HapView using the data of 1000 Genome Project (Abecasis et al. 2012).

#### *Annotation and enrichment analysis of emRNA and module identification*

The database for annotation, visualization, and integrated discovery (DAVID) (Huang et al. 2009) is an online analysis resource, which provides a comprehensive set of functional annotation tools for researchers to identify the enrichment characteristics of the identified genes. Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted for emRNAs with DAVID online analysis. Then, the modules representing new sub-networks were

constructed by combining the emRNAs with co-expressed lncRNAs and associated SNP.

#### *Causal inference analysis of lncRNA-mediated genetic risks for mRNAs*

The SNP-lncRNA-mRNA regulation relationships were assessed using the causal inference test (CIT) (Millstein et al. 2009) to test the regulation chain and to select the lncRNA-mediated trios. Briefly, the CIT has statistical tests for four conditions, all of which must be met for the lncRNA-mediated causal classification: (1) SNP and mRNA are associated, (2) SNP is associated with lncRNA after adjusting for mRNA, (3) lncRNA is associated with mRNA after adjusting for SNP, and (4) SNP is independent of mRNA after adjusting for lncRNA (Liu et al. 2013). The CIT *p* value was defined as the maximum of the component test *p* values, and a multivariate linear regression was used in the four component tests.

## Results

### Genome-wide identification of lncRNA-eQTLs

We systematically examined how SNPs regulate lncRNA expression through a genome-wide eQTL analysis with 551,745 SNPs and 17,190 lncRNA transcripts in PBMC from 43 unrelated individuals. Of a total of 9,484,496,550 tests, 6,206,063 were for *cis*-pairs (local) and 9,478,290,487 were for *trans*-pairs (distal) (Fig. 1). The QQ plot not only showed the high statistical power of the MatrixeQTL, but also indicated that biological functional effects of *cis*-eQTL were more significant than *trans*-eQTL. We identified 71,407 *cis*-SNP-probe combinations and millions of *trans*-SNP-probe combinations under a significant threshold of 0.01. After controlling the FDR at 0.05 for *cis*- and *trans*-eQTLs, respectively, we identified 707 significant *cis*-SNP-probe associations, corresponding to 498 unique *cis*-eSNPs and 187 unique *cis*-elncRNAs, as well as 6657 significant *trans*-SNP-probe associations corresponding to 1117 unique *trans*-eSNPs and 658 unique *trans*-elncRNAs.

### Distribution characteristics of eSNPs

For all *cis*-eQTLs with *p* value  $< 0.01$ , the *p* values were plotted against the distance between each eSNP and the TSS of its corresponding lncRNA transcript (Fig. 2). In general, top significant eSNPs were enriched around the TSS regions, and such a pattern is independent of the type of lncRNAs (Intergenic, intronic, divergent, antisense and unknown) (Fig. S1).

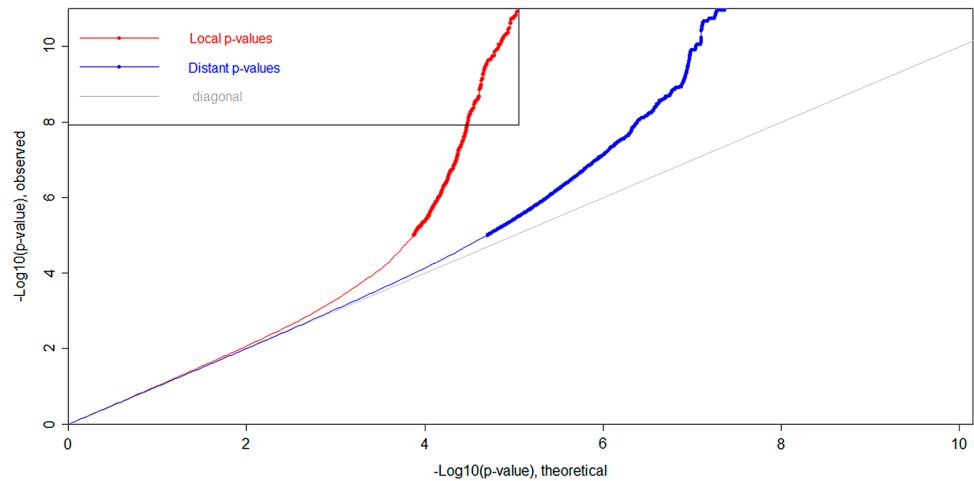
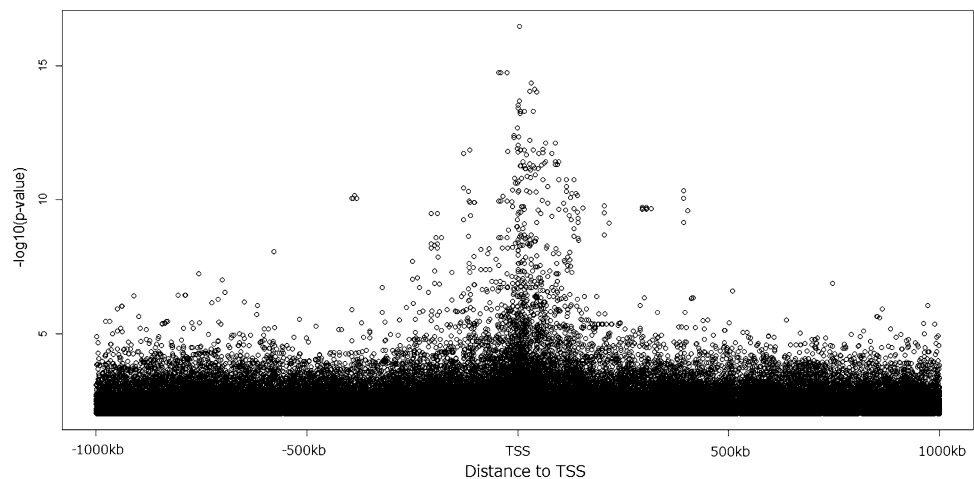
**Fig. 1** Q-Q plots of local (*cis*-) and distal (*trans*-) eQTLs**Fig. 2** The distribution of eQTL association significance ( $-\log_{10} p$  value) against the physical distance (kb) from lincRNA transcript start site for all eSNPs with  $p$  value less than 0.01

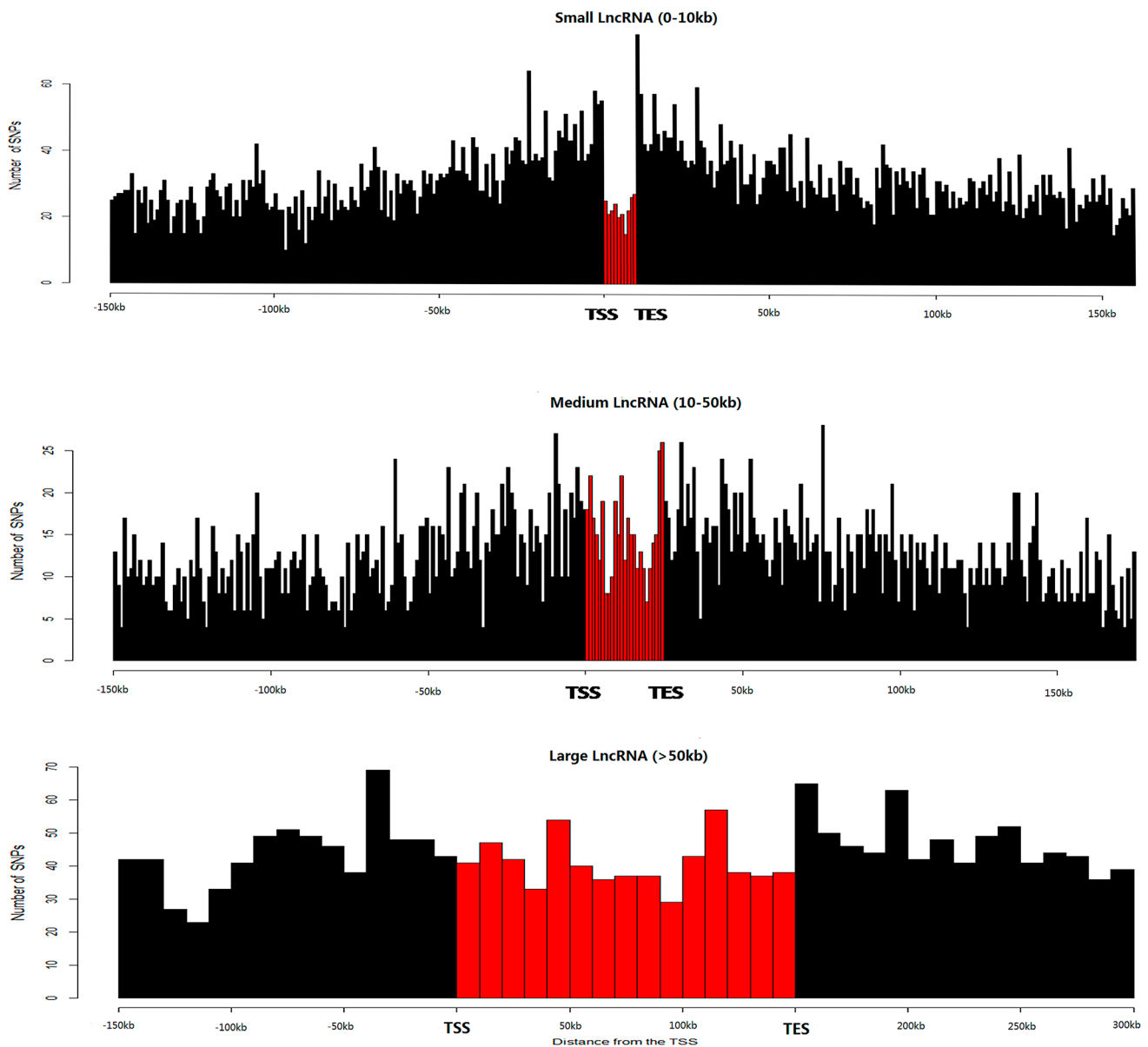
Figure 3 shows the histograms of *cis*-eSNPs locations ( $p < 0.01$ ), as a function of lincRNA sizes. For lincRNAs of small size, the eSNPs were mainly located on the upstream adjacent to the TSS and the downstream near the TES regions. In contrast, there were fewer eSNPs in the body region of the transcripts. For lincRNAs of medium size, the eSNPs were enriched around the TSS and TES regions and formed a peak in the body region. For lincRNAs of large size, the distribution of eSNPs appeared to have no obvious peaks. Generally, when taking account various sizes of lincRNAs together, eSNPs are most densely located around the TSS and TES.

After running MatrxQTL program, the beta-value for each eQTL pair represented the effect size for the influence of SNP variation on lincRNA expressions. Figure 4 shows the distribution of beta-value of *cis*-eQTLs and *trans*-eQTLs at FDR = 0.05 level. We found that the beta-values of *cis*-effect were mainly concentrated at  $\pm 0.8$  and formed two peaks. In contrast, the *trans*-effect beta-values were mainly concentrated at  $\pm 2$ . The

data suggest that the *trans*-eQTLs tend to have a greater impact on lincRNA expressions than *cis*-eQTLs.

#### Functional annotation and enrichment analysis for *cis*- and *trans*-eSNPs

We also examined the functional properties of the eSNPs, including *cis*- and *trans*- eSNPs, respectively. We found that *cis*-eSNPs were significantly enriched for transcription factor binding sites (TFBS), splicing sites, microRNA binding sites, PolyPhen-, SNPs3D (svm profile)- and SNPs3D (svm structure)- positive regions (Fisher's exact  $p < 0.05$ ; Fig. 5a). In contrast to the diverse functions for *cis*-eSNPs, *trans*-eSNPs were solely enriched for SNPs3D (svm profile) and SNPs3D (svm structure) with a mediocre fold enrichment (Fig. 5b), suggesting that *trans*-eQTLs tend to have more spatial structure regulation than the sequence regulation. In addition, to find if the 498 *cis*-eSNPs are overlapped with disease-related GWAS SNPs, we also searched are two bioinformatics tools (GWAS integrator and phenotype-genotype



**Fig. 3** Distribution pattern of the significant *cis*-eSNPs ( $p < 0.01$ ) across the different lncRNA sizes (small, medium, and large). The *x*-axis refers to a typical region divided into a series of bins, as described in the “Materials and methods” section. The *y*-axis represents the number of eSNPs in bin(s). eSNPs inside lncRNAs are

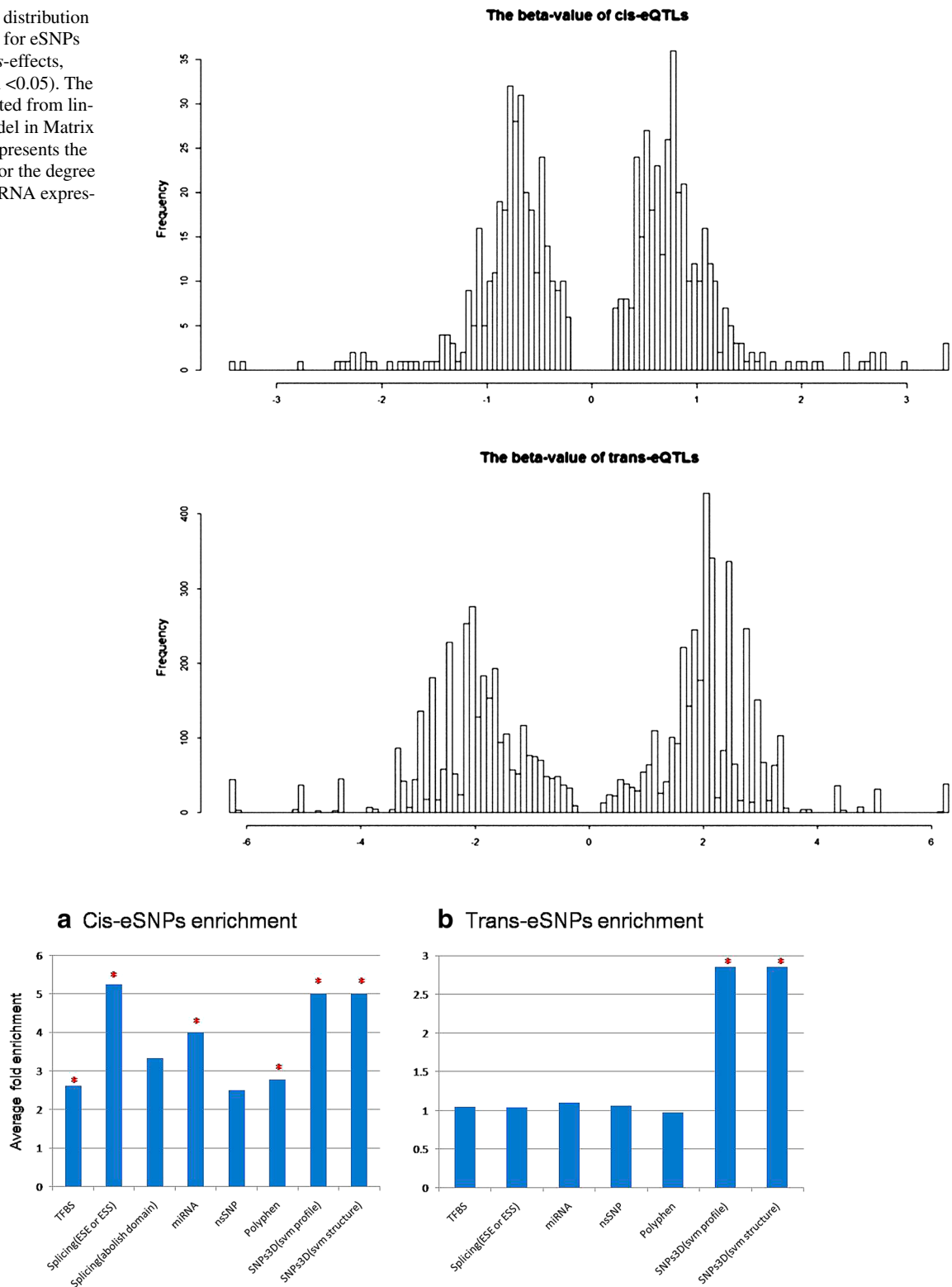
assigned to bins based on their physical location within the lncRNA. SNPs outside lncRNAs are assigned to bins based on their physical distance from the TSS (transcript start site). The distance from TSS to TES indicates the average transcript length for three types of lncRNAs

integrator) that provide robust lookup and analytic functionalities for published GWAS and meta-analysis studies. The results showed that 4 *cis*-eSNPs were associated with disease and traits (rs10263935 and rs801193 associated with aortic root size; rs11081443 with heart failure, and rs1533948 closely related to the secretion of lipoprotein).

### Network construction for eSNP, elncRNA, and emRNA

Our integrative approach that combines eQTL analysis with the network analysis generated an integrative network that incorporates the comprehensive relationship among SNP variations, lncRNA and mRNA expressions. First, using the strength of association for *cis*-lncRNA-eQTL

**Fig. 4** Frequency distribution against beta-value for eSNPs with *cis*- and *trans*-effects, respectively (FDR <0.05). The beta value, generated from linear regression model in Matrix eQTL package, represents the eSNP effect size for the degree of variation in lncRNA expression



**Fig. 5** Annotation of *cis*- and *trans*- eSNPs: enrichment in functional elements. The annotation was based on online tools SNPInfo that rely on the ENCODE Project. **a** *Cis*-eQTL SNPs are enriched for functional elements and database of Polyphen and SNPs3D. **b** *Trans*-eQTL SNPs are enriched in Polyphen and SNPs3D databases. PolyPhen (polymorphism phenotyping): an automatic tool for prediction

of possible impact of an amino acid substitution on the structure and function of a human protein. SNPs3D: a web resource and database that provides and integrates as much information as possible on disease/gene relationships at the molecular level. Svm structure: SNPs can affect protein function primarily by decreasing protein stability. Svm profile: SNP are disease association mutations

at FDR <0.05, we structured the connection between 498 eSNPs and 187 lncRNAs; Second, based on the correlation analysis (applying the PCC < -0.9 or >0.9 as a significance threshold) between the 187 lncRNAs and 21,323 mRNA probes, we constructed a noncoding-coding gene co-expression relationship including 16 lncRNAs and 54 mRNAs. Last, after eQTL association analysis between 498 eSNPs and the 54 selected mRNAs, we identified 615 eSNP-mRNA pairs (*cis*- and *trans*-) consisted of 93 eSNPs and 54 mRNAs. Finally the complex combined networks from the three levels were presented in Fig. 6a. The whole network consisted of a primary network and several small separate parts, which showed the complex regulation among the SNP, lncRNA and mRNA. For example, rs2154339 connected to several lncRNA and mRNA (e.g., TCONS\_0011260, ESPL1, STEAP2, and PPOX). In addition, these lncRNA and mRNAs also connected to other SNP, lncRNA and mRNA. We also found an interesting module which consisted of one lncRNA (i.e., TCONS\_00005196), one mRNA (i.e., XLOC\_002198) and 13 SNPs (Fig. 6c, d). These SNPs are located close to each other in chromosome 2. As expected, 11 of the 13 SNPs were in strong linkage disequilibrium with each other through HapView analysis (Fig. 6c).

#### Annotation and enrichment analysis of emRNA and module identification

For the above 54 (epi-) genetically regulated emRNAs, the results of GO and KEGG pathway annotations are shown in Table S1, and their module networks are shown in Fig. 6b. Module 1 contained three genes representing four GO terms including “sodium ion binding”, “alkali metal ion binding”, “sodium ion transport” and “monovalent inorganic cation transport”. Module 2 contained four genes representing two GO terms including “metal ion transport” and “cation transport”. Modules 3–5 contained three, five and five genes, representing “electron carrier activity”, “cell fraction” and “organelle membrane”, respectively. These identified GO terms showed that these (epi-) genetically regulated mRNA-coding genes execute fundamental molecular functions in PBMC.

#### Identification of the regulation chain of SLM

Since the above constructed complex networks for SNP, lncRNA, and mRNA (Fig. 6a) were based on the correlation of each pair, the regulation relationships within the networks cannot be confidently inferred. Therefore, we performed an in-depth CIT analysis to identify the regulation chain of SLM, i.e., the lncRNA-mediated genetic effect on mRNA expression. We identified 86 SLM trios, which have significant association or correlations for any within-trio

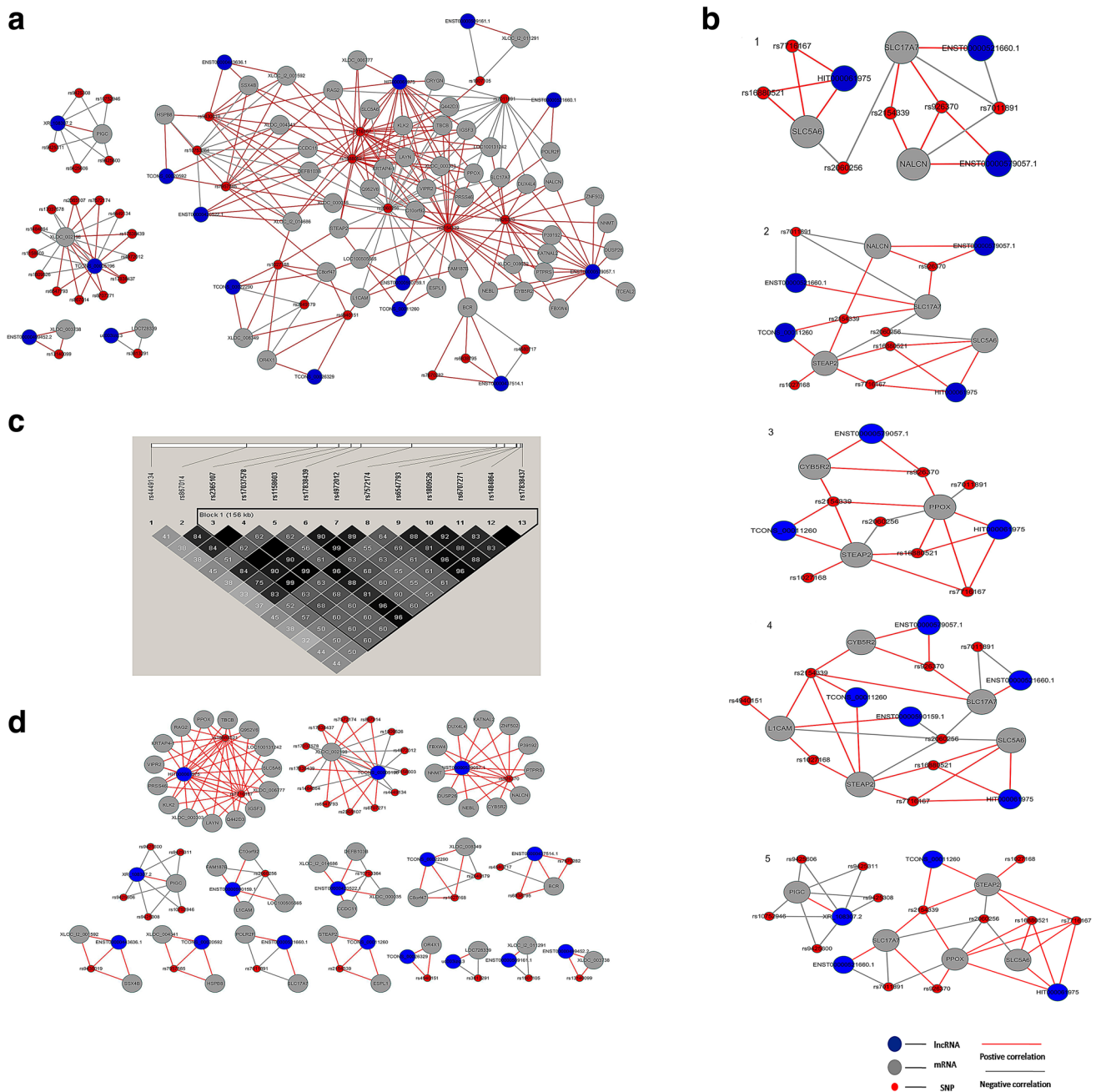
pairs (Fig. 6d). The CIT filtering steps were performed for the 86 trios. Finally, we discovered 50 significant lncRNA-mediated trios (Table 1), as well as some hotspots participating in multiple regulation chains. For example, eSNPs rs926370, rs7716167 and rs16880521 were involved in 11, 10 and 10 regulation chains, respectively. Besides, two lncRNAs (HIT000061975 and ENST00000579057.1) were involved in 20 and 11 chains, respectively. In total, 36 unique genes were regulated by 9 unique lncRNA and 13 eSNPs. As upstream regulation factors for mRNA expression, the number of identified nodes from the lncRNA (9) and SNP (13) levels are relatively less than that for mRNA (36).

#### Discussions

By using multi-omics integrative strategy, this study comprehensively investigated the (epi-) genetic architecture underlying the variation of lncRNA expression, and illustrated SLM interaction/regulation patterns in human PBMCs. Our lncRNA eQTL analyses showed that the variation of lncRNA expression is under various extents of genetic regulation, and the *cis*-eSNPs tend to cluster around the TSS and be significantly enriched in various regulation elements effecting on transcription. Combining eQTL analysis with network analyses, the integrative approach generated a complex network incorporating comprehensive interactions (associations and correlations) among SNP variations, lncRNA, and mRNA expressions. In-depth CIT analysis dissected lncRNA-mediated (epi-) genetic regulation modules on mRNA expression. These results provide novel insights into molecular machinery in PBMC, and may shed light on PBMC-relevant human physiology and pathology.

Current accumulating evidence has shown that lncRNA appears to have similar regulation model with mRNA at the transcriptional stage. Similar with mRNA eQTLs reported by previous studies (Dimas et al. 2009; Murphy et al. 2010; Wen et al. 2015), this study discovered that *cis*-eSNPs for lncRNAs tends to cluster at the TSS, and be significantly enriched in various regulation elements that have effects on the transcription. Recent data from the ENCODE project suggests that non-coding RNAs were more often regulated by various transcription factors than others (Gerstein et al. 2012). The polymorphisms in DNA regulation elements may alter the binding efficiency of DNA with the transcriptional factors, resulting in variation in lncRNA expression. Similar with *cis*-eSNPs for mRNA, *cis*-eSNPs for lncRNA were found significantly enriched at transcription factor binding sites, splicing sites, microRNA binding sites, etc. The *cis*-eSNPs distribution profile for medium lncRNAs was also similar to mRNA (Veyrieras





**Fig. 6** Comprehensive network and regulation modules identification. *Red nodes* represent eQTL SNPs. *Blue nodes* represent lncRNA. *Gray nodes* represent mRNA. *Red edges* represent positive regulation between two nodes. *Gray edges* represent negative regulation between two nodes. **a** The integrative network. This network incorporates the comprehensive relationship between SNP variations, lncRNA, and mRNA expressions. **b** Module identified from network using online tools DAVID. *Module 1* represented four GO terms including “sodium ion binding”, “alkali metal ion binding”, “sodium ion transport” and “monovalent inorganic cation transport”.

*Module 2* represented two GO terms including “metal ion transport” and “cation transport”. *Module 3–5* represented “electron carrier activity”, “cell fraction” and “organelle membrane”, respectively. **c** Linkage disequilibrium (LD) analysis for a cluster of 13 SNPs regulating expression a specific gene. As we expected, strong linkage disequilibrium effects were emerged among 11 SNPs from the 13 SNPs through HapView analysis. *Each number in cell* represents  $R^2$  between neighboring SNPs. **d** CIT-generated SLMs: networks of lncRNA-mediated genetic regulation on mRNA expression (color figure online)

et al. 2008). Meanwhile, there were several differences of *cis*-eSNPs between lncRNA and mRNA. For example, we can find that *cis*-eSNPs of mRNA were in the enrichment

of splicing (abolish domain) and nsSNP (nonsynonymous SNP) that are two important regulations to ensure the diversity of proteins coded by mRNA (Westra et al. 2013), but

**Table 1** Significant lncRNA-mediated genetic effects on mRNAs expression identified by causal inference test

| SNP ID     | LncRNA ID         | Gene symbol    | S-G effect | S-L effect | L-G effect | S1 <i>p</i> value | S2 <i>p</i> value | S3 <i>p</i> value | CIT <i>p</i> value |
|------------|-------------------|----------------|------------|------------|------------|-------------------|-------------------|-------------------|--------------------|
| rs2649179  | TCONS_00022290    | C8orf47        | N          | N          | P          | 2.00E-04          | 4.63E-04          | 1.15E-12          | 4.63E-04           |
| rs926370   | ENST00000579057.1 | NALCN          | P          | P          | P          | 5.94E-04          | 4.56E-04          | 3.60E-15          | 5.94E-04           |
| rs926370   | ENST00000579057.1 | ZNF502         | P          | P          | P          | 5.40E-04          | 6.91E-04          | 1.47E-14          | 6.91E-04           |
| rs10753364 | ENST00000420522.1 | CCDC11         | N          | N          | P          | 3.17E-04          | 7.39E-04          | 2.40E-13          | 7.39E-04           |
| rs926370   | ENST00000579057.1 | DUSP26         | P          | P          | P          | 7.71E-04          | 6.12E-04          | 4.59E-13          | 7.71E-04           |
| rs2060256  | ENST00000590159.1 | LOC100505565   | N          | N          | P          | 2.25E-04          | 1.06E-03          | 8.04E-10          | 1.06E-03           |
| rs926370   | ENST00000579057.1 | NNMT           | P          | P          | P          | 1.06E-03          | 2.82E-04          | 1.38E-13          | 1.06E-03           |
| rs16880521 | HIT000061975      | KRTAP4-1       | P          | P          | P          | 5.20E-04          | 1.14E-03          | 5.34E-13          | 1.14E-03           |
| rs7716167  | HIT000061975      | KRTAP4-1       | P          | P          | P          | 5.20E-04          | 1.14E-03          | 5.34E-13          | 1.14E-03           |
| rs926370   | ENST00000579057.1 | NEBL           | P          | P          | P          | 4.68E-04          | 1.47E-03          | 6.18E-13          | 1.47E-03           |
| rs10753364 | ENST00000420522.1 | DEFB103B       | N          | N          | P          | 2.70E-04          | 1.58E-03          | 2.28E-11          | 1.58E-03           |
| rs926370   | ENST00000579057.1 | PTPRS          | P          | P          | P          | 4.27E-04          | 1.86E-03          | 1.38E-12          | 1.86E-03           |
| rs16880521 | HIT000061975      | LOC100131242   | P          | P          | P          | 4.99E-04          | 1.86E-03          | 3.20E-11          | 1.86E-03           |
| rs7716167  | HIT000061975      | LOC100131242   | P          | P          | P          | 4.99E-04          | 1.86E-03          | 3.20E-11          | 1.86E-03           |
| rs7957585  | TCONS_00020592    | XLOC_004341    | P          | P          | P          | 6.80E-06          | 1.90E-03          | 3.83E-13          | 1.90E-03           |
| rs9430019  | ENST00000443636.1 | XLOC_12_001592 | P          | P          | P          | 2.62E-05          | 2.03E-03          | 4.12E-15          | 2.03E-03           |
| rs926370   | ENST00000579057.1 | FBXW4          | P          | P          | P          | 3.88E-04          | 2.07E-03          | 9.31E-13          | 2.07E-03           |
| rs6838795  | ENST00000437514.1 | BCR            | P          | P          | P          | 7.95E-05          | 2.10E-03          | 8.10E-13          | 2.10E-03           |
| rs2649179  | TCONS_00022290    | XLOC_008349    | N          | N          | P          | 6.07E-05          | 2.58E-03          | 4.02E-13          | 2.58E-03           |
| rs926370   | ENST00000579057.1 | CYB5R2         | P          | P          | P          | 3.00E-04          | 2.68E-03          | 2.17E-13          | 2.68E-03           |
| rs926370   | ENST00000579057.1 | P39192         | P          | P          | P          | 3.24E-04          | 2.80E-03          | 1.26E-12          | 2.80E-03           |
| rs16880521 | HIT000061975      | Q952V6         | P          | P          | P          | 3.53E-04          | 3.24E-03          | 1.05E-10          | 3.24E-03           |
| rs7716167  | HIT000061975      | Q952V6         | P          | P          | P          | 3.53E-04          | 3.24E-03          | 1.05E-10          | 3.24E-03           |
| rs4580717  | ENST00000437514.1 | BCR            | N          | N          | P          | 3.22E-05          | 3.55E-03          | 4.68E-13          | 3.55E-03           |
| rs7670282  | ENST00000437514.1 | BCR            | P          | P          | P          | 3.22E-05          | 3.55E-03          | 4.68E-13          | 3.55E-03           |
| rs16880521 | HIT000061975      | IGSF3          | P          | P          | P          | 2.98E-04          | 3.63E-03          | 1.74E-11          | 3.63E-03           |
| rs7716167  | HIT000061975      | IGSF3          | P          | P          | P          | 2.98E-04          | 3.63E-03          | 1.74E-11          | 3.63E-03           |
| rs16880521 | HIT000061975      | PPOX           | P          | P          | P          | 2.93E-04          | 4.09E-03          | 9.20E-11          | 4.09E-03           |
| rs7716167  | HIT000061975      | PPOX           | P          | P          | P          | 2.93E-04          | 4.09E-03          | 9.20E-11          | 4.09E-03           |
| rs926370   | ENST00000579057.1 | KATNAL2        | P          | P          | P          | 2.67E-04          | 4.14E-03          | 5.26E-12          | 4.14E-03           |
| rs7011891  | ENST00000521660.1 | POLR2F         | N          | N          | P          | 1.58E-04          | 5.82E-03          | 1.26E-11          | 5.82E-03           |
| rs10753364 | ENST00000420522.1 | XLOC_12_014686 | N          | N          | P          | 8.06E-05          | 6.25E-03          | 3.10E-13          | 6.25E-03           |
| rs7011891  | ENST00000521660.1 | SLC17A7        | N          | N          | P          | 1.44E-04          | 6.93E-03          | 4.75E-11          | 6.93E-03           |
| rs926370   | ENST00000579057.1 | DUX4L4         | P          | P          | P          | 1.74E-04          | 7.11E-03          | 2.66E-12          | 7.11E-03           |
| rs16880521 | HIT000061975      | RAG2           | P          | P          | P          | 1.47E-04          | 8.12E-03          | 1.47E-12          | 8.12E-03           |
| rs7716167  | HIT000061975      | RAG2           | P          | P          | P          | 1.47E-04          | 8.12E-03          | 1.47E-12          | 8.12E-03           |
| rs16880521 | HIT000061975      | KLK2           | P          | P          | P          | 1.27E-04          | 1.02E-02          | 4.75E-12          | 1.02E-02           |
| rs7716167  | HIT000061975      | KLK2           | P          | P          | P          | 1.27E-04          | 1.02E-02          | 4.75E-12          | 1.02E-02           |
| rs1027168  | TCONS_00022290    | XLOC_008349    | P          | P          | P          | 9.26E-05          | 1.03E-02          | 3.50E-13          | 1.03E-02           |
| rs10753364 | ENST00000420522.1 | XLOC_000035    | N          | N          | P          | 4.13E-05          | 1.27E-02          | 4.31E-15          | 1.27E-02           |
| rs16880521 | HIT000061975      | VIPR2          | P          | P          | P          | 1.03E-04          | 1.41E-02          | 3.61E-10          | 1.41E-02           |
| rs7716167  | HIT000061975      | VIPR2          | P          | P          | P          | 1.03E-04          | 1.41E-02          | 3.61E-10          | 1.41E-02           |
| rs2060256  | ENST00000590159.1 | L1CAM          | N          | N          | P          | 2.06E-05          | 1.42E-02          | 6.41E-09          | 1.42E-02           |
| rs16880521 | HIT000061975      | LAYN           | P          | P          | P          | 7.52E-05          | 1.95E-02          | 1.39E-10          | 1.95E-02           |
| rs7716167  | HIT000061975      | LAYN           | P          | P          | P          | 7.52E-05          | 1.95E-02          | 1.39E-10          | 1.95E-02           |
| rs2060256  | ENST00000590159.1 | FAM187B        | N          | N          | P          | 7.64E-06          | 3.27E-02          | 5.64E-09          | 3.27E-02           |
| rs16880521 | HIT000061975      | PRSS46         | P          | P          | P          | 4.05E-05          | 3.60E-02          | 9.02E-11          | 3.60E-02           |
| rs7716167  | HIT000061975      | PRSS46         | P          | P          | P          | 4.05E-05          | 3.60E-02          | 9.02E-11          | 3.60E-02           |

**Table 1** continued

| SNP ID    | LncRNA ID      | Gene symbol | S-G effect | S-L effect | L-G effect | S1 <i>p</i> value | S2 <i>p</i> value | S3 <i>p</i> value | CIT <i>p</i> value |
|-----------|----------------|-------------|------------|------------|------------|-------------------|-------------------|-------------------|--------------------|
| rs1027168 | TCONS_00022290 | C8orf47     | P          | P          | P          | 2.96E−05          | 3.97E−02          | 1.69E−11          | 3.97E−02           |
| rs7957585 | TCONS_00020592 | HSPB8       | P          | P          | P          | 3.86E−07          | 4.07E−02          | 2.70E−10          | 4.07E−02           |

This table presented 50 trios of lncRNA-mediated genetic effect on mRNA expression from causal inference test (CIT). S-L, L-G, or S-G effect means the effect of SNP on lncRNA, lncRNA on mRNA, and SNP on mRNA, respectively. *P* positive regulation, *N* negative regulation. S1 *p* value was the association between SNP and mRNA, S2 *p* value was the association between SNP and lncRNA after adjusting for mRNA, S3 *p* value was the association between lncRNA and mRNA after adjusting for SNP. CIP *p* value was the maximum *p* value of S1, S2 and S3

the *cis*-eSNPs of lncRNA are not enrichment in the two functional elements. In addition, previous study has shown that lncRNA and 3′ untranslated regions (3′UTRs) in protein-coding RNAs are significantly similar in structural features and sequence composition, which produce the miRNAs can combine target lncRNA in anywhere in body of lncRNA, rather than mRNA which only can be combined in the 3′ untranslated regions of mRNA (Niazi and Valadkhan 2012).

Recent studies have revealed that abnormal expression of lncRNAs is correlated with various biological process and human diseases (Huarte and Rinn 2010; Spizzo et al. 2012). LncRNA has then been considered to be one of the most important intermediate phenotype on regulating mRNA expression. However, as a relatively novel kind of transcripts, the genetic architecture underlying lncRNA variation and the regulation relationship between lncRNA and mRNA are barely known. With the development of biotechnology, more and more researchers have paid attentions to multidimensional data, and explored complex regulation relationships between molecules in cells or tissues. This study showed that multi-omics integrative analysis is powerful and effective strategy in dissecting complex interactions. With the use of lncR-eQTL results alone, it is very difficult to identify causal variant, causal gene, or underlying mechanism by which a SNP or lncRNA affects a phenotype. By integrating multi-level molecular data, we obtained evidences showing that eSNP and elncRNA have strong associations with gene expression levels. Integrative network analyses together with CIT analyses have uncovered the lncRNA-mediated genetic influences on gene expressions, i.e., a SNP regulates lncRNA expression, which in turn alters mRNA expression. Based on the results, attentions were drawn to several hotspots with interesting correlations. For example, eSNP rs926370 probably regulates lncRNA NST00000579057.1 expression, subsequently altering mRNA expressions of multiple genes, including FBXW4, CYB5R2, NALCN, etc. The wide variety of biological regulation patterns of lncRNA, such as modular scaffolds assembling diverse combinations of regulatory proteins (Schonrock et al. 2012), make it possible to up-regulate or down-regulate the target gene expression. In addition,

we found that lncRNA-mediated genes in PBMC are mainly involved in common physiological processes (e.g., “Sodium ion binding”, “Sodium ion transport”).

This study has several potential limitations. First, the inferred interaction patterns were based on multi-omics data, further cellular and molecular experiments will be helpful to validate the findings. Second, the results were derived from PBMC, thus the findings may not be applicable to other cells or tissues.

In summary, we conducted a genome-wide SNP association study for PBMC lncRNA expression, identified significant eSNPs and characterized their distribution patterns. Follow-up multi-omics integrative analyses identified complex SLM interaction/regulation networks in PBMC. The findings provide novel insights into the molecular machinery for PBMC gene expression, and shed lights on PBMC-relevant human physiology and pathology.

**Acknowledgements** The study was supported by Natural Science Foundation of China (81473046, 31271336, 81373010, 81502868, 31401079, 81401343, 81541068), the Natural Science Research Project of Jiangsu Provincial Higher Education (16KJA330001), the Natural Science Foundation of Jiangsu Province (BK20130300, BK20150346), the Project funded by China Postdoctoral Science Foundation (2014M551649), the Startup Fund from Soochow University (Q413900112, Q413900712), and a Project of the Priority Academic Program Development of Jiangsu Higher Education Institutions.

#### Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest.

#### References

- Abecasis GR et al (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65. doi:10.1038/nature11632
- Adzhubei IA et al (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249. doi:10.1038/nmeth0410-248
- Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. *Science* 322:881–888. doi:10.1126/science.1156409
- Borel C et al (2011) Identification of *cis*- and *trans*-regulatory variation modulating microRNA expression levels in human fibroblasts. *Genome Res* 21:68–73. doi:10.1101/gr.109371.110

- Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D (2012) A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149:819–831. doi:[10.1016/j.cell.2012.03.035](https://doi.org/10.1016/j.cell.2012.03.035)
- Derrien T et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789. doi:[10.1101/gr.132159.111](https://doi.org/10.1101/gr.132159.111)
- Dimas AS et al (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325:1246–1250. doi:[10.1126/science.1174148](https://doi.org/10.1126/science.1174148)
- Gerstein MB et al (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489:91–100. doi:[10.1038/nature11245](https://doi.org/10.1038/nature11245)
- Gupta RA et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071–1076. doi:[10.1038/nature08975](https://doi.org/10.1038/nature08975)
- Guttman M et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227. doi:[10.1038/nature07672](https://doi.org/10.1038/nature07672)
- Huan T et al (2015) Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun* 6:6601. doi:[10.1038/ncomms7601](https://doi.org/10.1038/ncomms7601)
- Huang T, Cai YD (2013) An information-theoretic machine learning approach to expression QTL analysis. *PLoS One* 8:e67899. doi:[10.1371/journal.pone.0067899](https://doi.org/10.1371/journal.pone.0067899)
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57. doi:[10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211)
- Huarte M, Rinn JL (2010) Large non-coding RNAs: missing links in cancer? *Hum Mol Genet* 19:R152–R161. doi:[10.1093/hmg/ddq353](https://doi.org/10.1093/hmg/ddq353)
- Hubbard T et al (2002) The Ensembl genome database project. *Nucleic Acids Res* 30:38–41
- Kellis M et al (2014) Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA* 111:6131–6138. doi:[10.1073/pnas.1318948111](https://doi.org/10.1073/pnas.1318948111)
- Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3:ra8. doi:[10.1126/scisignal.2000568](https://doi.org/10.1126/scisignal.2000568)
- Klattenhoff CA et al (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152:570–583. doi:[10.1016/j.cell.2013.01.003](https://doi.org/10.1016/j.cell.2013.01.003)
- Kumar V, Wijmenga C, Withoff S (2012) From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Semin Immunopathol* 34:567–580. doi:[10.1007/s00281-012-0312-1](https://doi.org/10.1007/s00281-012-0312-1)
- Kumar V et al (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 9:e1003201. doi:[10.1371/journal.pgen.1003201](https://doi.org/10.1371/journal.pgen.1003201)
- Lettre G, Rioux JD (2008) Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet* 17:R116–R121. doi:[10.1093/hmg/ddn246](https://doi.org/10.1093/hmg/ddn246)
- Liu Y et al (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 31:142–147. doi:[10.1038/nbt.2487](https://doi.org/10.1038/nbt.2487)
- McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 17:R156–R165. doi:[10.1093/hmg/ddn289](https://doi.org/10.1093/hmg/ddn289)
- Meola N, Pizzo M, Alfano G, Surace EM, Banfi S (2012) The long noncoding RNA Vax2os1 controls the cell cycle progression of photoreceptor progenitors in the mouse retina. *RNA* 18:111–123. doi:[10.1261/rna.029454.111](https://doi.org/10.1261/rna.029454.111)
- Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155–159. doi:[10.1038/nrg2521](https://doi.org/10.1038/nrg2521)
- Millstein J, Zhang B, Zhu J, Schadt EE (2009) Disentangling molecular relationships with a causal inference test. *BMC Genet* 10:23. doi:[10.1186/1471-2156-10-23](https://doi.org/10.1186/1471-2156-10-23)
- Murphy A et al (2010) Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum Mol Genet* 19:4745–4757. doi:[10.1093/hmg/ddq392](https://doi.org/10.1093/hmg/ddq392)
- Niazi F, Valadkhan S (2012) Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* 18:825–843. doi:[10.1261/rna.029520.111](https://doi.org/10.1261/rna.029520.111)
- Pennisi E (2010) Genomics. 1000 Genomes Project gives new map of genetic diversity. *Science* 330:574–575. doi:[10.1126/science.330.6004.574](https://doi.org/10.1126/science.330.6004.574)
- Pennisi E (2011) The biology of genomes. Disease risk links to gene regulation. *Science* 332:1031. doi:[10.1126/science.332.6033.1031](https://doi.org/10.1126/science.332.6033.1031)
- Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629–641. doi:[10.1016/j.cell.2009.02.006](https://doi.org/10.1016/j.cell.2009.02.006)
- Prensner JR, Chinnaiyan AM (2011) The emergence of lncRNAs in cancer biology. *Cancer Discov* 1:391–407. doi:[10.1158/2159-8290.CD-11-0209](https://doi.org/10.1158/2159-8290.CD-11-0209)
- Qi J, Asl HF, Bjorkegren J, Michoel T (2014) kruX: matrix-based non-parametric eQTL discovery. *BMC Bioinform* 15:11. doi:[10.1186/1471-2105-15-11](https://doi.org/10.1186/1471-2105-15-11)
- Schadt EE et al (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302. doi:[10.1038/nature01434](https://doi.org/10.1038/nature01434)
- Schonrock N, Harvey RP, Mattick JS (2012) Long noncoding RNAs in cardiac development and pathophysiology. *Circ Res* 111:1349–1362. doi:[10.1161/CIRCRESAHA.112.268953](https://doi.org/10.1161/CIRCRESAHA.112.268953)
- Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28:1353–1358. doi:[10.1093/bioinformatics/bts163](https://doi.org/10.1093/bioinformatics/bts163)
- Spizzo R, Almeida MI, Colombatti A, Calin GA (2012) Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* 31:4577–4587. doi:[10.1038/onc.2011.621](https://doi.org/10.1038/onc.2011.621)
- Tripathi V et al (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39:925–938. doi:[10.1016/j.molcel.2010.08.011](https://doi.org/10.1016/j.molcel.2010.08.011)
- van Riel PL, Fransen J (2005) DAS28: a useful instrument to monitor infliximab treatment in patients with rheumatoid arthritis. *Arthritis Res Ther* 7:189–190. doi:[10.1186/ar1820](https://doi.org/10.1186/ar1820)
- Veyrieras JB, Kudaravalli S, Kim SY, Dermizakis ET, Gilad Y, Stephens M, Pritchard JK (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 4:e1000214. doi:[10.1371/journal.pgen.1000214](https://doi.org/10.1371/journal.pgen.1000214)
- Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. *Mol Cell* 43:904–914. doi:[10.1016/j.molcel.2011.08.018](https://doi.org/10.1016/j.molcel.2011.08.018)
- Wen X, Luca F, Pique-Regi R (2015) Cross-population joint analysis of eQTLs: fine mapping and functional annotation. *PLoS Genet* 11:e1005176. doi:[10.1371/journal.pgen.1005176](https://doi.org/10.1371/journal.pgen.1005176)
- Westra HJ et al (2013) Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat Genet* 45:1238–1243. doi:[10.1038/ng.2756](https://doi.org/10.1038/ng.2756)
- Willingham AT et al (2005) A strategy for probing the function of non-coding RNAs finds a repressor of NFAT. *Science* 309:1570–1573. doi:[10.1126/science.1115901](https://doi.org/10.1126/science.1115901)
- Xu Z, Taylor JA (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 37:W600–W605. doi:[10.1093/nar/gkp290](https://doi.org/10.1093/nar/gkp290)
- Yue P, Melamud E, Moulton J (2006) SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinform* 7:166. doi:[10.1186/1471-2105-7-166](https://doi.org/10.1186/1471-2105-7-166)