

## ORIGINAL ARTICLE

# Multi-omics integrative analysis identified SNP-methylation-mRNA: Interaction in peripheral blood mononuclear cells

Yi-Hua Lu<sup>1,2,3</sup>  | Bing-Hua Wang<sup>1,3</sup> | Fei Jiang<sup>1,3</sup> | Xing-Bo Mo<sup>1,3</sup> | Long-Fei Wu<sup>1,3</sup> | Pei He<sup>1,3</sup> | Xin Lu<sup>1,3</sup> | Fei-Yan Deng<sup>1,3</sup>  | Shu-Feng Lei<sup>1,3</sup>

<sup>1</sup>Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Suzhou, P. R. China

<sup>2</sup>Department of Epidemiology and Health Statistics, School of Public Health, Nantong University, Nantong, P. R. China

<sup>3</sup>Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Suzhou, P. R. China

## Correspondence

Shu-Feng Lei and Fei-Yan Deng, Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Suzhou, Jiangsu, P. R. China.

Emails: leisf@suda.edu.cn; fdeng@suda.edu.cn

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 81401343, 81541068 and 81872681; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20150346; Startup Fund from Soochow University, Grant/Award Number: Q413900112 and Q413900712

## Abstract

Genetic variants have potential influence on DNA methylation and thereby regulate mRNA expression. This study aimed to comprehensively reveal the relationships among SNP, methylation and mRNA, and identify methylation-mediated regulation patterns in human peripheral blood mononuclear cells (PBMCs). Based on in-house multi-omics datasets from 43 Chinese Han female subjects, genome-wide association trios were constructed by simultaneously testing the following three association pairs: SNP-methylation, methylation-mRNA and SNP-mRNA. Causal inference test (CIT) was used to identify methylation-mediated genetic effects on mRNA. A total of 64,184 significant cis-methylation quantitative trait loci (meQTLs) were identified (FDR < 0.05). Among the 745 constructed trios, 464 trios formed SNP-methylation-mRNA regulation chains (CIT). Network analysis (Cytoscape 3.3.0) constructed multiple complex regulation networks among SNP, methylation and mRNA (eg a total of 43 SNPs simultaneously connected to cg22517527 and further to PRMT2, DIP2A and YBEY). The regulation chains were supported by the evidence from 4DGenome database, relevant to immune or inflammatory related diseases/traits, and overlapped with previous eQTLs from dbGaP and GTEx. The results provide new insights into the regulation patterns among SNP, DNA methylation and mRNA expression, especially for the methylation-mediated effects, and also increase our understanding of functional mechanisms underlying the established associations.

## KEYWORDS

Causal Inference Test, DNA methylation, integrative analysis, MeQTLs, multi-omics

## 1 | INTRODUCTION

Previous genome-wide association studies (GWASs) have identified a large number of SNPs associated with complex traits and diseases.<sup>1-3</sup> A big challenge after GWAS is to explain the functions of the identified SNPs, and to illustrate the mechanisms underlying the associations. Notably, only 7% of the identified variations

are located in protein-coding regions,<sup>4,5</sup> and the majority of diseases-associated variations are unlikely to affect the protein functions by changing the amino acid sequence. DNA methylation plays an important role in regulating expression of target gene. DNA methylation at promoter is dynamically linked to gene activity, and could directly influence the patterns of gene expression and cellular differentiation.<sup>6</sup> Aberrant DNA methylation in

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

cancer was associated with abnormally regulated expression of normal cellular genes.<sup>7</sup> Recently, disrupted DNA methylation patterns were established as a contributor to metabolic syndrome,<sup>8,9</sup> schizophrenia<sup>10,11</sup> and inflammatory or autoimmune disorders.<sup>12,13</sup> The feature and distribution of DNA methylation have been studied in a variety of tissues/cells, and these genome-wide maps of DNA methylation have revealed interesting features and provided important insights into its potential functions in genome regulation.<sup>14,15</sup> However, the functional mechanism underlying the variation in DNA methylation itself is still largely unknown. Previous studies have revealed that DNA methylation at specific loci can be influenced by sequence variations.<sup>16-19</sup> So far, how these genetic variations exert their effects was largely unknown, for example, it is unknown whether the SNPs exert their effects on DNA methylation and ultimately affects the gene expression.

Genome-wide expression quantitative trait locus (eQTL) analysis is a well-known method to explore genetic effect of SNP on gene expression. This approach has extensively been used to investigate the associations between SNP and other target phenotype (eg methylation level<sup>11,16-19</sup>) or between two phenotypes (eg DNA methylation and gene expression<sup>15</sup>). Methylation quantitative trait locus (meQTL) analysis is a kind of improved eQTL method that has been used to investigate the associations between SNPs and the methylation levels.<sup>16-19</sup> Another study has assessed the associations between DNA methylation and gene expression, that is, expression quantitative trait methylation (eQTM).<sup>15</sup> However, such QTLs studies were largely limited to testing one single pair of association only. Because of lack of multiple-omics data from the same set of samples, the complex triangle relationship between SNP, DNA methylation and mRNA expression was undefined yet.

This study conducted multi-omics integrative analyses as well as causal inference test (CIT)<sup>20</sup> to reveal the complex connections among SNP, DNA methylation and mRNA expression, and identified DNA methylation-mediated regulation effects in peripheral blood mononuclear cells (PBMCs), a commonly used target cell in immunology studies.

## 2 | MATERIALS AND METHODS

### 2.1 | Subjects and PBMC isolation

The study was approved by the Institutional Research Ethic Board at Soochow University. Table S1 showed the basic characteristics of 43 study subjects. The human subjects included 43 unrelated Chinese Han adult females from Suzhou city of China, which were recruited originally for identifying risk molecules of rheumatoid arthritis. Subjects were excluded from serious diseases involving vital organs (brain, liver, kidney, heart or lung). All subjects signed informed consent forms before entering this project. A total of 15 ml peripheral blood was collected and stored in sodium-citrate-supplemented vacuum tubes. PBMCs were isolated using density gradient centrifugation using Lymphoprep (Sigma, life science, USA). PBMCs were divided into two equal parts, one for DNA extraction, and the other

for RNA extraction after treatment of Trizol reagent (Invitrogen, Carlsbad, CA) to avoid RNA degradation.

### 2.2 | Genome-wide SNP genotyping, DNA methylation profiling and transcriptome profiling

DNA was extracted from the isolated PBMCs using phenol-chloroform extraction and ultrapurification method.<sup>21</sup> The quality of extracted DNA was first tested by 0.8% agarose gel electrophoresis to check the integrity (usually > 10KB main band). The OD260/280 of 1.7-1.9 by NanoDrop ND-1000 (Thermo Scientific, Wilmington, Delaware) spectrophotometer was the QC cutoff of the DNA purification. Affymetrix Genome-Wide Human SNP Array 6.0 chips were used for SNP genotyping by following the protocol recommended by the manufacturer. The experiments were performed in the laboratory of CapitalBio Corporation (Beijing, China). We used the contrast QC greater than 0.4 for quality control. A total of 909,622 SNPs in each subject were genotyped. After excluded the SNPs with a minor allele frequency less than 5%, or a call-rate less than 95%, 551,745 SNPs were finally used in further analysis. All analyses are based on human reference genome 37 version assembly annotations.

DNA methylation profiling was performed with Illumina 450K Infinium Methylation BeadChip according to the manufacturer's instructions in the laboratory of CapitalBio Corporation (Beijing, China). DNA methylation data quality control consist of sample QC (subjects with more than 5% probes with a detection  $P > 0.05$  were removed) and probe QC (probes with a detection  $P > 0.05$  more than 5% subject were excluded). DNA methylation data normalization contains background adjustment, colour-bias adjustment, quantile normalization and beta mixture quantile (BMIQ) method normalization. The background adjustment was performed in GenomeStudio. Then, colour-bias adjustment and quantile normalization were performed in the R package lumi, followed by the BMIQ normalization to eliminate the bias between probe types. The methylation level was measured as beta ( $\beta$ ) =  $M/(M + U)$ , in which M was the methylated signals and U was the unmethylated signals. The  $\beta$  value ranges continuously from 0 (unmethylated) to 1 (fully methylated). After normalization, 416 285 methylated sites were left for further data processing.

Total RNA was extracted from PBMCs according to the instructions recommended by the manufacturer. The OD260/280 of  $\geq 1.8$  was the QC cutoff of the RNA purification (NanoDrop ND-1000 spectrophotometer). RNA integrity was determined with 1% formaldehyde denaturing gel electrophoresis and Agilent 2100 Bioanalyzer. 28S/18S rRNA ratio of  $\geq 1.5$  and RIN  $> 7$  were used as the eligible criteria. The transcriptome-wide mRNA expression was profiled using lncRNA&mRNA Human Gene Expression Microarray V4.0 (CapitalBio Corp, Beijing, China). The data were extracted by Agilent Feature Extraction (V10.7). The data summary, normalization and quality control were performed with GeneSpring GX program (V12.0). The log<sub>2</sub> transformation was applied using the Adjust Data function. The probes with less than 80% of detection rate and/or incomplete annotation information were filtered (Multiexperiment Viewer (MeV) software (<http://www.tm4.org>)). Subsequently, a total of 17,566 unique mRNA probes were used for further analysis.

**TABLE 1** Summary of associations in meQTLs, eQTLs and eQTLs

	meQTLs	eQTLs	eQTLs
Test	SNP & methylation	Methylation & expression	SNP & expression
Window size	1 MB	1 MB	1 MB
FDR	5%	5%	5%
Number of tests	144 470 159	10 944 256	5 880 162
Maximum <i>P</i> -value	2.22E-05	1.28E-05	4.39E-06
Cis-effect pairs	64 184	2 795	525
SNP	40 896	–	471
Methylation	16 033	2 090	–
mRNA	–	837	140

eQTLs, Expression quantitative trait loci, the association between SNP and gene expression; eQTLs, Expression quantitative trait methylation, the associations between DNA methylation and gene expression; FDR, Benjamini-Hochberg false-discovery rate; meQTLs, Methylation quantitative trait loci, the association between SNPs and methylation level.

### 2.3 | Quantitative trait locus analyses for three pairs (SNP & methylation, SNP & mRNA and methylation & mRNA)

All the three quantitative trait locus (QTL) analyses were performed with the MatrixEQTL package modelled in R software (freely available at <http://cran.r-project.org/>).<sup>15,22-25</sup> Here, we defined the QTL analyses for three association pairs (SNP & methylation, SNP & mRNA and methylation & mRNA) as meQTL, eQTL and eQTM, respectively. Multivariate linear regression analysis was conducted for each QTL analysis after adjusting for age and disease states of rheumatoid arthritis. Compared to trans-effects, cis-effects were much larger and be more stable.<sup>26-30</sup> So, in order to enhance the reliability of the results, this study focused on the cis-effect. SNPs located within 1 megabase (Mb) on either side of methylation sites was supposed to exert local effect (cis-meQTLs). For the cis-eQTL analysis, SNPs were also confined within 1 Mb distant from the transcription start site (TSS) or transcription end site (TES) of mRNAs. For the cis-eQTM analysis, methylation sites were confined within 1 Mb distant from TSS or TES of mRNAs. Benjamini-Hochberg false-discovery rate (FDR) was used to correct for multiple testing.

### 2.4 | CIT for DNA methylation-mediated genetic effect on mRNA expression

To identify DNA methylation-mediated effect on mRNA expression, we first constructed the associated trios according to the analysis results of the above three pairs. The associated trios were generated according to the physical positions of SNPs and methylation sites in genes with known official names. Herein, CIT<sup>31,32</sup> was applied to identify methylation-mediated association between SNPs and mRNAs. Briefly, the causal inference simultaneously requires the following four criteria: SNP and mRNA expression is associated; SNP is associated with methylation level after adjusting for mRNA; Methylation is associated with mRNA after adjusting for SNP; and SNP is independent of mRNA expression after adjusting for

methylation level.<sup>31,32</sup> The covariates, age and disease status, were adjusted in the above four tests. The maximum of the test *P*-values was reported as the CIT *P*-Value. CIT was performed in SAS 9.2 software (SAS Institute Inc, Cary, North Carolina).

### 2.5 | Construction of SNP-methylation-mRNA interaction network

Based on CITs, the fulfilled association trios (SNP-methylation-mRNA) were selected to construct gene regulatory network. All network data were visualized using open source bioinformatics software Cytoscape 3.3.0 (Institute of Systems Biology in Seattle).<sup>33</sup>

### 2.6 | Linkage disequilibrium analysis

Since multiple nearby SNPs from SNP-methylation-mRNA regulatory chains were simultaneously connected to single methylation site, we conducted linkage disequilibrium (LD) analysis for those nearby SNPs in HaploView 4.2 using the data of 1000 Genomes Project<sup>34</sup> (2504 volunteer donors from various ethnic populations). We also obtained the LD measure  $r^2$ , which represents the statistical correlation between two SNPs of interest and is frequently used in LD mapping because of its statistical advantages and strong theoretical basis for population genetics. When  $r^2 = 0$ , it shows that the two loci are completely independent, and  $r^2 = 1$  means SNPs at the two loci sharing the same frequency.

### 2.7 | The identified SNP-methylation-mRNA chains overlapped with previous results

To find whether the identified associations in significant SNP-methylation-mRNA chains were reported by previous studies, we searched the databases of Phenotype-Genotype Integrator (PheGenI) ([www.ncbi.nlm.nih.gov/gap/PheGenI/](http://www.ncbi.nlm.nih.gov/gap/PheGenI/)), National Human Genome Research Institute (NHGRI) Catalog of published GWAS (GWAS Catalog), GTEx Portal (<https://www.gtexportal.org/>), and

the 4DGenome, a general repository for chromatin interaction data (<https://4dgenome.research.chop.edu/>).

### 3 | RESULTS

#### 3.1 | Cis-meQTL identification and distribution characteristics

A total of 551 745 genotyped SNPs and 416 285 methylation levels were tested. Among the 144 470 159 tested cis-meQTL pairs, 64 184 pairs were significant (FDR < 0.05), which corresponded to 40 896 unique SNPs and 16 033 unique methylation sites (Table 1 and Table S2). As shown in Figure 1, the plots significantly deviated from the reference line for both cis-meQTLs and trans-meQTLs, but more rapidly for cis-meQTLs, suggesting that the significant regulation effects of SNPs on methylation were relatively stronger for cis-meQTLs than for trans-meQTLs in general.

Specifically, for the methylation sites in CpG regions, we identified 42 131 (65.64%: 42 131/64 184) significant cis-meQTL pairs. These corresponding methylation sites were distributed at CpG island (N = 15 738, 24.52%), CpG shore (N\_shore and S\_shore, 0-2 Kb from CpG island) (N = 19 005, 29.61%) and CpG shelf (N\_shelf and S\_shelf, 2-4 Kb from CpG island) (N = 7 388, 11.51%) (Figure 2A). The shores and shelves were annotated according to their chromosome orientation from the p- to q-arms as in N- and S-shores, respectively. As presented in Figure 2B, according to annotation of gene region feature category by UCSC (<http://www.genome.ucsc.edu>), the corresponding SNPs were distributed in gene body (39.76%), TSS 1500 (-1500 to -200 bp to TSS) region (17.79%), TSS

200 (-200 bp to TSS) region (8.72%), 5'-UTR (8.40%), as well as 1st exons (3.30%).

The distribution of the physical distance from SNPs to their associated methylation sites for all the significant cis-meQTLs (FDR < 0.05) was presented in Figure 2C. An obvious peak in frequency around the methylation sites was observed, suggesting that the SNPs from the significant meQTLs were enriched around methylation sites. Similarly, the most effective SNPs were enriched around the methylation sites, indicating that the closer to methylation sites, the greater effects of the SNPs on methylations in general (Figure 2D).

#### 3.2 | Identification of the SNP-methylation-mRNA regulation chain

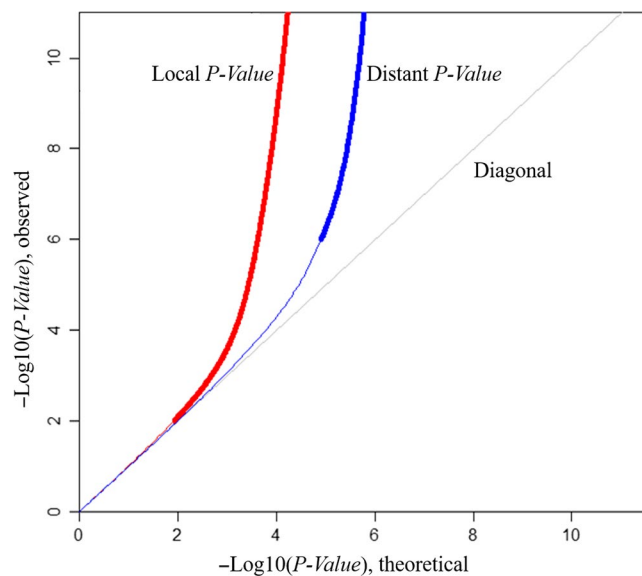
We further tested the associations of the following two pairs, methylation-mRNA by eQTLs and SNP-mRNA by eQTLs. As shown in Table 1, we identified 2,795 local significant methylation-mRNA associations (cis-eQTLs) (FDR < 0.05), and 525 significant local effect SNP-mRNA pairs (cis-eQTLs) (FDR < 0.05). According to the physical positions of SNPs and methylation sites in the corresponding genes with official names, the significant pairs from the above three association tests were matched and the overlapped pairs were selected to construct SNP-methylation-mRNA association trios. Consequently, 745 trios were generated, which corresponded to 272 unique SNPs, 65 unique methylations and 47 unique target mRNAs. In-depth CIT analysis showed that 464 trios fulfilled significant causal inference. Specifically, 191 SNPs were associated with mRNA expressions of 37 genes, which were dependently through 56 methylation sites. Notably, 21 SNPs were simultaneously associated with the same methylation site (cg22517527) and the same gene expression (PRMT2). The statistical results for the 464 methylation-mediated genetic effects on mRNA expressions were detailed in Table S3, including effect directions and P-values under all the tested conditions.

#### 3.3 | Construction of SNP-methylation-mRNA interaction network

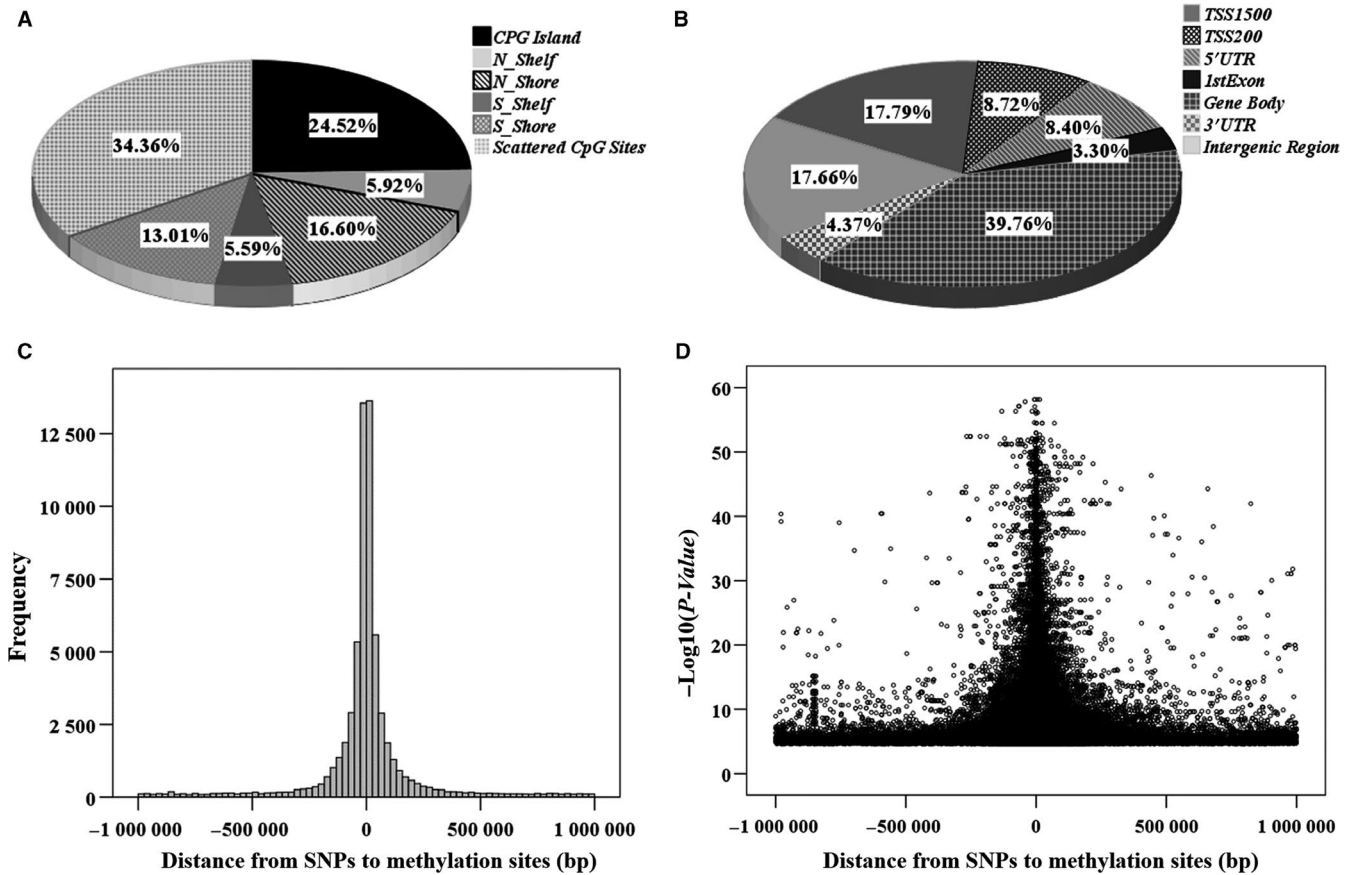
Figure 3 demonstrated the epi-genetic regulation patterns among 171 SNPs, 56 methylation sites and 37 target genes from 464 causal trios. We found a primary network and several small separate networks, which showed their complex regulation patterns. Notably, it was common that multiple SNPs are connected to a few limited methylation sites or mRNAs. For example, 43 SNPs simultaneously influenced methylation at cg22517527 and further regulated mRNA expression of PRMT2 (protein arginine methyltransferase 2), DIP2A (disco-interacting protein 2 homolog A) and YBEY (ybeY metallo-peptidase). Besides, 20 SNPs regulated YBEY expression through methylation at the cg20399509, and 30 SNPs regulated YBEY expression by methylation at the cg12516959.

#### 3.4 | LD analysis

LD analyses for the SNPs physically located closely in the same chromosome were performed by using the HaploView and the genotype



**FIGURE 1** Quantile-quantile plots of the associations from meQTL analyses. Local P-value: P-value from cis-meQTLs, in which the SNPs located within 1 megabase (Mb) on either side of methylation sites; Distant P-value: P-value from trans-meQTLs, in which the SNPs located outside 1 megabase (Mb) on either side of methylation sites



**FIGURE 2** The distribution of cis-meQTL associated methylation sites. (A) The distribution according to their positions in the UCSC CpG island region. (B) The distribution according to gene region feature category (UCSC). (C) The frequency distribution Note: Distance indicates the physical distance from SNP to their associated methylation site. (D) The distribution according to association significance ( $-\log_{10}(P\text{-value})$ ) against the physical distance from SNP to their associated methylation site

data of 1000 Genomes Project. As we expected, strong LD structure was detected for most of the analysed SNPs. As shown in Figure 4, we have found 7 LD blocks (3 in Chr21, 2 in Chr1, 1 in Chr11 and 1 in Chr6). For example, an SNP group (10 SNPs) in the Block 1 of Chr21 formed an extremely strong LD block (the first sub-network in Figure 3), which simultaneously influenced five methylation sites (cg00612595, cg02776659, cg09417038, cg12516959 and cg20399509) and further regulated mRNA expression of YBEY. In the Block 1 of Chr1, 21 SNPs also formed a strong LD block, which correspond to the 21 SNPs-cg04982190-NCF2 regulation chains (the second sub-network in Figure 3). Besides, another group (16 SNPs) had strong LD in chromosome 1, which were simultaneously connected to cg05043910 and further regulated mRNA expression of TSGA10 (the third sub-network in Figure 3).

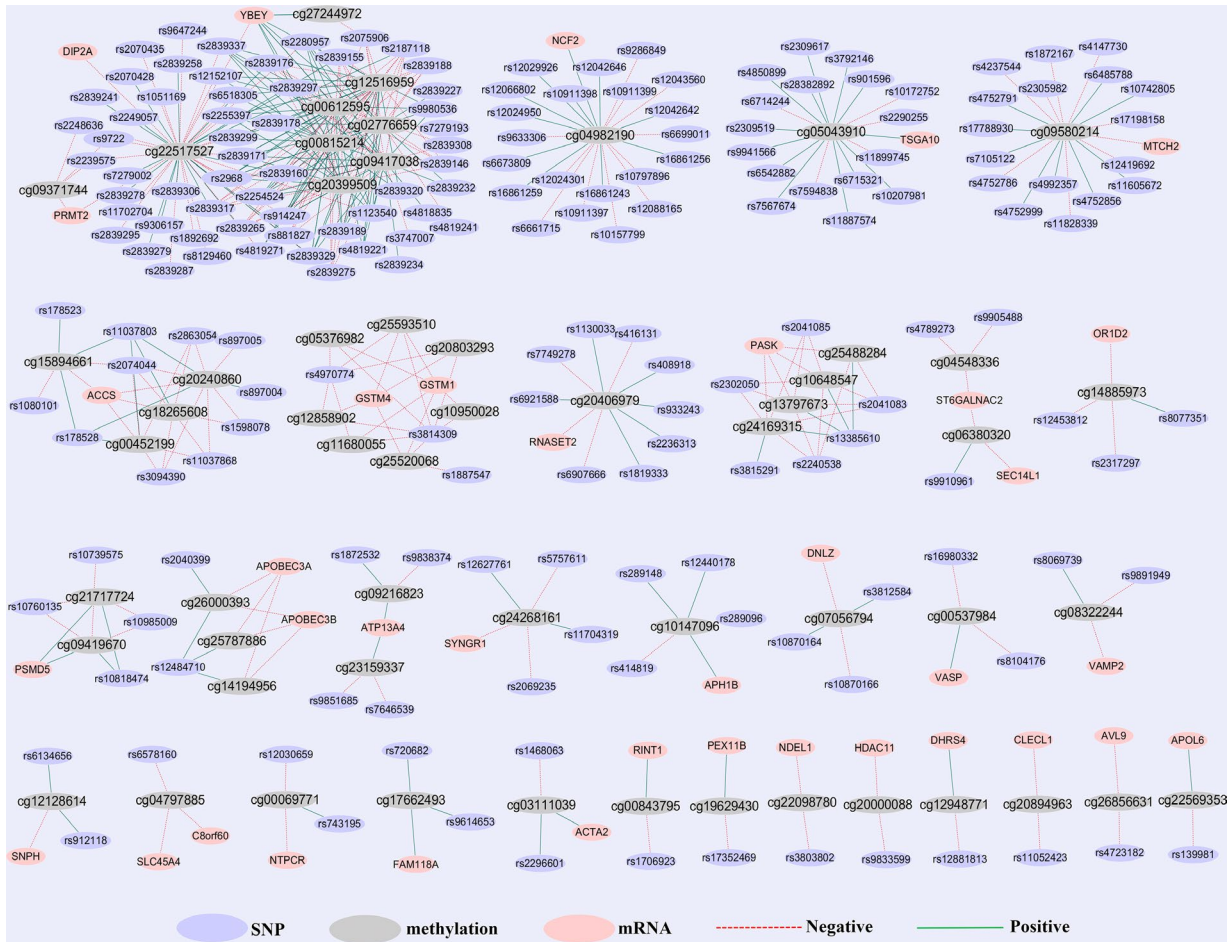
### 3.5 | Supported evidence from 4DGenome database for the identified SNP-methylation-mRNA chains

To find additional evidence to support the identified SNP-methylation-mRNA chains, we searched the 4DGenome database based on their physical positions (hg19). The information of the

SNP-Methylation-mRNA chains, their located interactions, detection methods and the detected cells/tissues were listed in Table S4. We found 89 significant interactions that correspond to 38 unique SNP-Methylation-mRNA chains in 23 types of cells/tissues including CD4 T cell, an important immunity related cell. Especially for the rs8069739-cg08322244-VAMP2, about 33 interactions were identified in a variety of cells, for example, in the CD4 T cells, rs8069739 and VAMP2 were located in Interactor B and cg08322244 was located in Interactor A.

### 3.6 | The SNP-methylation-mRNA chains relevant to immune or inflammatory related diseases/traits

Among the 191 SNPs and 37 genes involved in the identified causal trios, 19 SNPs and 7 genes were reported to be significantly associated with human diseases/traits (Table S5), as archived in the NHGRI database and the database of Genotypes and Phenotypes (dbGaP). The associated diseases/traits included rheumatoid arthritis, multiple sclerosis, systemic lupus erythematosus, inflammatory bowel disease and crohn disease, which are immune or inflammatory related diseases/traits.



**FIGURE 3** The constructed networks based on the significant SNP-methylation-mRNA regulatory chains. Cytoscape 3.3.0 was used to establish the networks. The SNP, DNA methylation and mRNA from significant CIT trios were imported. Purple nodes represent SNPs, grey nodes represent DNA methylation and pink nodes represent mRNA. Red dot edges represent negative regulation between two nodes. Green solid lines represent positive regulation between two nodes

### 3.7 | The identified SNP-methylation-mRNA chains overlapped with previous eQTLs from dbGaP and GTEx

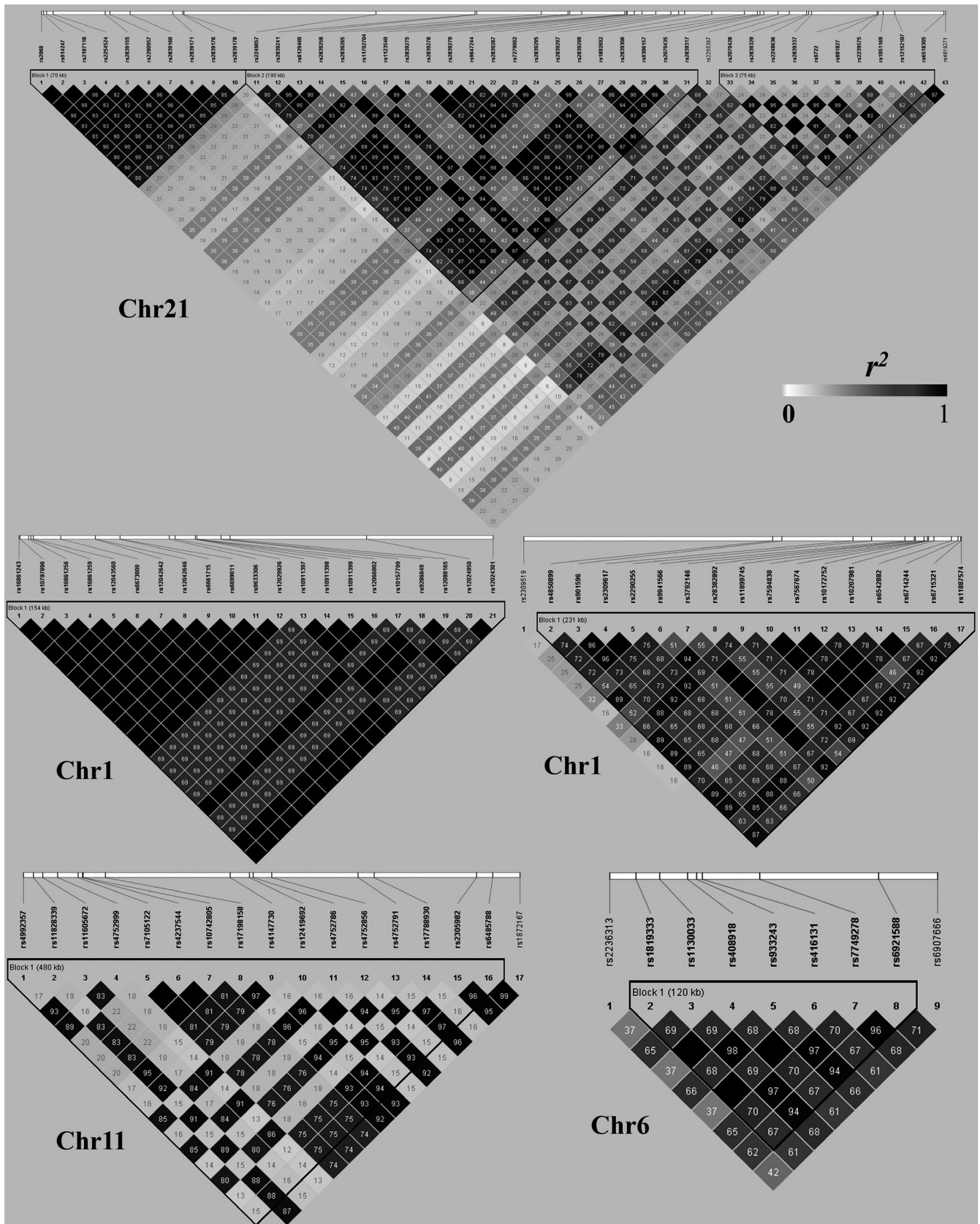
The identified regulation chains also probably provide functional explanations for previous eQTL results. After searched the dbGaP according to SNP ID, 16 eQTLs had been reported in lymphoblastoid and liver, which correspond to 13 unique SNPs and 6 genes. Meanwhile, the GTEx Portal was also retrieved for more supportive information, and it was found 112 eQTLs results containing 74 unique SNPs and 10 genes (Table S6).

## 4 | DISCUSSION

DNA methylation is known as an important epigenetic regulatory factor in mediating the correlations between genetic variants and mRNAs.<sup>35,36</sup> EQTL analysis is a typical and powerful statistical method in explaining the functional link between SNP and disease/trait.<sup>27,37-39</sup> Several studies have supported that cis-acting QTLs

have large effect sizes that can be detected in a relatively small sample less than 100 subjects.<sup>16,40-42</sup> A large number of cis-meQTL associations identified in our study have suggested that DNA methylation levels were under strong genetic influence. Previous meQTL studies also found abundant local effects in other tissues.<sup>22,43-45</sup> Furthermore, an obvious peak observed around the methylation sites suggested that the physical distance seemed to have large effect on their associations, that is, the more close to methylation site, the greater effect for the associations. Our findings suggested that methylation sites were typically associated with SNPs in close proximity. The significant SNPs in our cis-meQTL analysis were mostly located in gene regions, including gene body and the promoter. It is because that genetic influence on the human methylome involves heterogeneous processes and is predominantly dependent on local sequence context at the meQTL sites.<sup>45,46</sup>

This study represents the first efforts of conducting integrative multiple omics analyses by multiple QTL tests and in-depth CIT to more comprehensively reveal methylation-mediated regulation effects. CIT provides a highly interpretable quantitative measure for a trio of variables when association between two implies causation



**FIGURE 4** Linkage disequilibrium (LD) analysis. HaploView 4.2 was used to analyse the linkage disequilibrium. The SNPs used in LD analysis are from the significant SNP-methylation-mRNA regulatory chains. The shades of colour represent  $r^2$ , deeper colour represent the higher value of  $r^2$ . Each number in cell represents  $r^2$  between neighbouring SNPs. The black cells without numbers means  $r^2 = 1$

and the third is a potential mediator.<sup>32</sup> This method is theoretically and computationally accessible in disentangling molecular relationships,<sup>47</sup> and was proposed as a novel statistical framework in which existing notions of causal mediation were formalized into a hypothesis test.

The identified 464 regulation chain of SNP-methylation-mRNA suggest that DNA methylation can emerge both as marker and determinant.<sup>15</sup> For instance, we found 43 SNPs further regulate gene expression of PRMT2, DIP2A and YBEY by regulating methylation levels of cg22517527. Two SNPs (rs4970774 and rs3814309) negatively regulated the GSTM1 and GSTM4 by negatively controlling seven methylated sites collectively. The SNP rs12484710 positively regulated cg26000393, cg25787886 and cg14194956 so that to influence the expression of APOBEC3A and APOBEC3B. LD analysis was conducted in SNPs physically located closely by using the HaploView and 1000 Genomes Project. Strong LD structure was identified for most of the analysed SNPs. Nevertheless, the SNPs closest to the significant methylation site is likely to be most effective in regulation, but when multiple SNPs are in strong LD in the detected region, it is challenging to discriminate which SNP within the same LD region is truly causative. Moreover, whether and how the SNPs influence the methylation structure and then regulate gene expression is unclear yet. These regulatory networks can provide a basis for future functional studies. Herein, we propose that some DNA sequence variants, through changing the methylation levels, hence influence the gene expressions.

To find more evidence to support the identified SNP-methylation-mRNA regulatory chains, we searched the 4DGenome. 4DGenome is a public database that archives and disseminates chromatin interaction data. It covers all major experimental technologies including chromosome conformation capture (3C), chromosome conformation capture-on-chip (4C), chromosome conformation capture carbon copy, ChIA-PET, Hi-C, Capture-C and a computational prediction of IM-PET for detecting chromatin interactions. The interaction paired sequence tags archived in this database implies the corresponding pair of DNA regions is in close physical proximity and probably functionally interacts. We found 89 results including 38 unique SNP-methylation-mRNA chains overlapped with the 4DGenome database. The results from 4DGenome database provide additional understanding of chromatin architecture how SNP, methylation and mRNA are tightly related.

The identified regulation chains probably provide functional explanations for the associations of SNPs and diseases. We searched for the Phenotype-Genotype Integrator and GWAS Catalog. PBMCs consist of several important immunity cells, which play a decisive role in the process of immune or inflammatory response. It was found some SNPs and genes in SNP-methylation-mRNA chains were related to the immune or inflammatory disease/trait by searching for the database. For instance, RNASET2 regulated by rs1819333 is associated with inflammatory bowel diseases and Crohn disease, which further illustrated that significant association of SNPs and genes in SNP-methylation-mRNA chains, may play a decisive role in the process of immune response in PBMCs. These overlapped results

suggested that the methylation is a probable functional mechanism in connecting SNP with the susceptibility to diseases. We further validated the associations between SNPs and mRNAs in SNP-methylation-mRNA chains in dbGAP and GTEx Portal, and found 16 overlap results in dbGAP and 112 overlap results in GTEx Portal.

This study had several potential limitations. First, the relatively small sample size may offer limited power in detecting minor- or modest-effect meQTLs/eQTLs/eQTLs. Second, the inferred interaction patterns were based on multi-omics data, further cellular and molecular experiments will be helpful to validate the findings. Third, it was probably inappropriate to extend the PBMC expression regulatory pattern to other cells or tissues because of the high tissue- or cell-specificity as mentioned above. Fourth, the subjects concerning only women may limit the extensions of the results in male or mixed sample (both male and female) because gender-specific genetic architecture is common in humans. Last, the cross-reactive probes and polymorphic CpGs in Illumina 450K Infinium Methylation BeadChip probably have confusing impacts on methylation readouts.<sup>48,49</sup>

In summary, our study comprehensively investigated the (epi-) genetic architecture underlying the variation of methylation expression, and illustrated SNP-methylation-mRNA regulation pattern by in-depth CIT analysis in human PBMCs by using multi-omics integrative strategy. The results provide new insights into the regulation patterns among SNP, DNA methylation and mRNA expression, especially for the methylation-mediated effects, and also increase our understanding of functional mechanisms underlying the established associations. The results would further facilitate the investigations of PBMC-related immune physiological process and immunological diseases in the future.

## DATA SHARING STATEMENT

The microarray data for methylation have been submitted to the GEO database with accession number GSE111942.

## ACKNOWLEDGEMENTS

The study was supported by the National Natural Science Foundation of China (81872681, 81401343 and 81541068), the Natural Science Foundation of Jiangsu Province (BK20150346) and the Startup Fund from Soochow University (Q413900112, Q413900712).

## CONFLICT OF INTEREST

The authors declared that we have no conflict of interest to this work.

## AUTHOR CONTRIBUTION

Study conception and design: Shu-Feng Lei, Fei-Yan Deng and Yi-Hua Lu; Acquisition of data: Long-Fei Wu, Fei Jiang, Pei He and Xin Lu; Analysis and interpretation of data: Yi-Hua Lu, Bing-Hua Wang,



and Xing-Bo Mo; Drafting of manuscript: Yi-Hua Lu; Critical revision: Fei-Yan Deng and Shu-Feng Lei. All authors have participated in revising the manuscript critically and gave their final approval of the version to be submitted.

## ORCID

Yi-Hua Lu  <https://orcid.org/0000-0002-7186-5373>

Fei-Yan Deng  <https://orcid.org/0000-0001-7553-5358>

## REFERENCE

- McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet.* 2008;17(R2):R156-R165.
- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* 2008;322(5903):881-888.
- Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet.* 2008;17(R2):R116-R121.
- Pennisi E. The biology of genomes. Disease risk links to gene regulation. *Science.* 2011;332(6033):1031.
- Kumar V, Wijemga C, Withoff S. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Semin Immunopathol.* 2012;34(4):567-580.
- Ambrosi C, Manzo M, Baubec T. Dynamics and context-dependent roles of DNA methylation. *J Mol Biol.* 2017;429(10):1459-1475.
- Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. *Biochem Biophys Res Commun.* 1983;111(1):47-54.
- de Toro-Martin J, et al. Methylation quantitative trait loci within the TOMM20 gene are associated with metabolic syndrome-related lipid alterations in severely obese subjects. *Diabetol Metab Syndr.* 2016;8:55.
- Guenard F, et al. Genetic regulation of differentially methylated genes in visceral adipose tissue of severely obese men discordant for the metabolic syndrome. *Transl Res.* 2017;184:p. 1-11 e2.
- Lin D, Chen J, Ehrlich S, et al. Cross-Tissue Exploration of Genetic and Epigenetic Effects on Brain Gray Matter in Schizophrenia. *Schizophr Bull.* 2017;44(2):443-452.
- Hoffmann A, Ziller M, Spengler D. The Future is The Past: Methylation QTLs in Schizophrenia. *Genes (Basel).* 2016;7(12):E104.
- Rushton MD, Reynard LN, Young DA, et al. Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Hum Mol Genet.* 2015;24(25):7432-7444.
- Hedrich CM, Mäbert K, Rauen T, Tsokos GC. DNA methylation in systemic lupus erythematosus. *Epigenomics.* 2017;9(4):505-525.
- Murrell A, Heeson S, Cooper WN, et al. An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum Mol Genet.* 2004;13(2):247-255.
- Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife.* 2013;2:e00523.
- Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 2011;12(1):R10.
- Li Y, Zhu J, Tian G, et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* 2010;8(11):e1000533.
- Gamazon ER, Badner JA, Cheng L, et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry.* 2013;18(3):340-346.
- Grundberg E, Meduri E, Sandling J, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet.* 2013;93(5):876-890.
- Chen LS, Emmert-Streib F, Storey JD. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol.* 2007;8(10):R219.
- Di Pietro F, Ortenzi F, Tilio M, Concetti F, Napolioni V. Genomic DNA extraction from whole blood stored from 15- to 30-years at -20 degrees C by rapid phenol-chloroform protocol: a useful tool for genetic epidemiology studies. *Mol Cell Probes.* 2011;25(1):44-48.
- Gutierrez-Arcelus M, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet.* 2015;11(1):e1004958.
- Huang T, Cai YD. An information-theoretic machine learning approach to expression QTL analysis. *PLoS ONE.* 2013;8(6):e67899.
- Qi J, Asl HF, Björkegren J, Michael T. kruX: matrix-based non-parametric eQTL discovery. *BMC Bioinformatics.* 2014;15:11.
- Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28(10):1353-1358.
- Montgomery SB, Sammeth M, Gutierrez-Arcelus M, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* 2010;464(7289):773-777.
- Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature.* 2010;464(7289):768-772.
- Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature.* 2005;437(7063):1365-1369.
- Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. *Nat Genet.* 2007;39(10):1217-1224.
- Stranger BE, Forrest MS, Clark AG, et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 2005;1(6):e78.
- Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol.* 2013;31(2):142-147.
- Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet.* 2009;10:23.
- Adami S, Palacios S, Rizzoli R, Levine AB, Sutradhar S, Chines AA. The efficacy and safety of bazedoxifene in postmenopausal women by baseline kidney function status. *Climacteric.* 2014;17(3):273-284.
- Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13(7):484-492.
- Schubeler D. Molecular biology. Epigenetic islands in a genetic ocean. *Science.* 2012;338(6108):756-757.
- Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015;16(4):197-212.
- Heinig M, Petretto E, Wallace C, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature.* 2010;467(7314):460-464.
- Smith R, Sheppard K, DiPetrillo K, Churchill G. Quantitative trait locus analysis using J/qlt. *Methods Mol Biol.* 2009;573:175-188.
- Kilpinen H, Waszak SM, Gschwind AR, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science.* 2013;342(6159):744-747.
- del Rosario R-H, Poschmann J, Rouam SL, et al. Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat Methods.* 2015;12(5):458-464.

42. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660.
43. Kerkel K, Spadola A, Yuan E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008;40(7):904-908.
44. Shi J, Marconett CN, Duan J et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat Commun*. 2014;5:3365.
45. McClay JL, Shabalin AA, Dozmorov MG, et al. High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol*. 2015;16:291.
46. Harris AR, Nagy-Szakal D, Pedersen N, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis*. 2012;18(12):2334-2341.
47. Nakkuntod J, Sukkapan P, Avihingsanon Y, Mutirangura A, Hirankarn N. DNA methylation of human endogenous retrovirus in systemic lupus erythematosus. *J Hum Genet*. 2013;58(5):241-249.
48. Chen Y-a, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-209.
49. Daca-Roszak P, Pfeifer A, Žebracka-Gala J, et al. Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: implications for comparative population studies. *BMC Genom*. 2015;16:1003.

**How to cite this article:** Lu Y-H, Wang B-H, Jiang F, et al. Multi-omics integrative analysis identified SNP-methylation-mRNA: Interaction in peripheral blood mononuclear cells. *J Cell Mol Med*. 2019;23:4601-4610. <https://doi.org/10.1111/jcmm.14315>