

# The Long Noncoding RNA Differential Expression in Peripheral Blood Leukocyte from Schizophrenia Patients by RNA Sequencing

Jie Wu (✉ [nokookon@outlook.com](mailto:nokookon@outlook.com))

Kunming Medical University <https://orcid.org/0000-0002-5806-2963>

Zijun Liu

Kunming Medical University Sixth Affiliated Hospital

Yunqiao Zhang

Kunming Medical University Sixth Affiliated Hospital

Zhaowei Teng

Kunming Medical University Sixth Affiliated Hospital

Xu You

Kunming Medical University Sixth Affiliated Hospital

Ye Ruan

Kunming Medical University Sixth Affiliated Hospital

Changqing Gao

Kunming Medical University Sixth Affiliated Hospital

Yong Zeng

Kunming Medical University Sixth Affiliated Hospital

Zhaowei Teng

Kunming Medical University Sixth Affiliated Hospital

---

## Research article

**Keywords:** Schizophrenia, Long non-coding RNA, RNA Sequencing, Bioinformatics, IL1RAP

**Posted Date:** December 1st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-113426/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The diagnosis of schizophrenia (SCZ) depends on the evaluation of clinical symptoms, and there is no objective biomarker. Surveys have found that long non-coding RNA (lncRNA) may be affected in the pathogenesis of SCZ. There are also different genes in the expression of peripheral blood (PBL) in SCZ patients.

**Methods:** We profiled transcriptome analysis of PBL in 50 patients with schizophrenia and 50 controls without psychiatric diagnoses, reconstructed PBL transcriptome information using RNA-seq, predicted lncRNA-mRNA interaction via "RNAplex", a hierarchical classification-Spielman correlation coefficient approach was used to analyze the correlation between lncRNA and protein-coding gene expression among samples, and used systematic bioinformatics methods (Go/Pathway) to perform lncRNA functional annotation and qPCR experimental verification. Predicting functional sites for sequences using the database PROSITE, NCBI, UCSC, JASPAR.

**Results:** We screened 94 lncRNA and 1179 mRNA differential expressions in PBL, of which 46 new lncRNAs were identified for the first time. Enrichment into lncRNA involves biological processes and signaling pathways related to the neutrophil activation involved in immune response. According to Spearman correlation coefficient analysis, 81 lncRNA and 410 mRNA have expression correlation ( $p < 0.01$  and  $|r| \geq 0.4$ ), QPCR in independent samples verified that the core node of the lncRNA-mRNA co-expression network IL1RAP-TCONS\_00138311 variable shear is indeed highly expressed in SCZ patients,  $2^{-\Delta\Delta Ct}$  is 0.56, the area under the ROC curve is 0.924. The top four ranked transcription factors were predicted to be HSF1, HSF2, HSF4, and FOXA1.

**Conclusions:** Combined with sequence function analysis, it showed that the transcription factors FOXA1, HSF1, HSF2, HSF4, etc. may mediate the activation of IL1B-induced NF- $\kappa$ B pathway and other inflammatory pathways through the regulation of IL1RAP alternative splicing transcripts TCONS\_00138311, thereby participating in the pathogenesis of SCZ. We propose that the frequency of Differential lncRNA in peripheral blood could be used as novel biomarker for distinguishing SCZ from health.

## Background

Schizophrenia (SCZ) is a serious mental disorder of uncertain etiology with a high incidence of morbidity, recurrence, and disability, with a lifetime prevalence of approximately 1% in the adult population [1], and a high prevalence of residual symptoms, with approximately 10-15% of patients requiring lifelong care [2]. SCZ is still diagnosed empirically by clinicians and there are no objective laboratory indicators. Elucidating the etiology and pathological mechanism of SCZ, looking for clinical diagnostic biomarkers and targeted drug therapy has valuable application and significance.

Advanced in Genome-Wide Association Studies (GWAS) indicate that the occurrence of SCZ is affected by a combination of genetic and environmental factors [3]. Recent years' studies have shown that long

non-coding RNA (lncRNA) may be involved in the pathogenesis of SCZ [4]. lncRNA conceived to play important roles in many life activities such as dose compensation effect, epigenetic regulation, cell cycle regulation and cell differentiation regulation, participate in many important regulatory processes such as transcription activation and transcription interference [5]. Pro-inflammatory cytokines such as IL-6, TNF- $\alpha$ , and IFN- $\gamma$  have been reported in the literature to be significantly elevated in SCZ, whereas lncRNAs such as TMEVPG1 enhance the expression of IFN- $\gamma$  and NRON regulates the transcription factor NFAT, thereby regulates IL-6 and IFN- $\gamma$  expression; conversely, these inflammatory cytokines can stimulate and alter lncRNA expression. lncRNA is mostly located in the nucleus and is highly expressed in the mammalian brain and genome [6], Because brain tissue and PBL gene expression have a common regulatory pathway [7], multiple cytokines secreted in brain tissue are also present in peripheral blood leukocytes, therefore, many lncRNAs that are highly expressed in brain tissue are also clearly expressed in PBL, detection of lncRNA expression levels has clear feasibility and reliability.

## Methods

### Research object

Blood samples used for sequencing were collected from the psychiatric inpatient or outpatient department of the Sixth Affiliated Hospital of Kunming Medical University, the Second People's Hospital of Honghe Prefecture and the Second People's Hospital of Yuxi City, test groups conform to the diagnosis of schizophrenia in the fourth edition of the American Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), furthermore, in line with the initial onset or did not take antipsychotic drugs 5 weeks before enrollment, Han people aged 15 to 60 years old, control group was recruited at the same time.

### RNA-Sequencing of PBL

Drawing 5~10ml of EDTA anticoagulant venous blood, performed centrifugation within 60 minutes. Extracting total RNA by TRIzol Reagent (Ambion<sup>®</sup>), using VAHTS Total RNA-seq (H/M/R) Library Prep Kit for Illumina<sup>®</sup> (NR603) kit for total RNA purification and library construction, the constructed sequencing library was used by Illumina HiSeq TM 4000(China, Illumina) to RNA sequencing (RNA-Seq). Illumina HiSeq TM4000 is currently the only expression profile chip that can achieve 30 times repeat of the probe, and the chip result and qPCR correlation coefficient  $R^2=0.97$  [8]. RNA-Seq compared with the other sequencing techniques, has a broader dynamic range, provides a better estimate of relative expression levels of any genomic region with higher technical reproducibility, and facilitates alternative splicing detection[9]. Along with these advantages, RNA-seq has been used to reconstruct the entire organism transcriptome[10]. Conducting novel lncRNA predictions for the selected new transcripts, predicting the coding capacity of new transcripts using CPC, CNCI and SwissProt, and the intersection of these transcripts without coding potential and protein annotation information was taken as a reliable prediction.

### Gene Co-Expression Network Construction

Differential analysis of lncRNA expression between groups using edgeR [11], selecting differential transcripts using FDR and log<sub>2</sub>FC, the filter conditions are FDR<0.05 and |log<sub>2</sub>FC|>1. On account of lncRNA expression is low compared to mRNA levels, a hierarchical classification-Spielman correlation coefficient approach was used to analyze the correlation between lncRNA and protein-coding gene expression among samples. Using cytoscape software to build co-expression network.

### **LncRNA action modes**

The action modes of lncRNAs involved in regulation is divided into 3 categories: antisense, cis and trans [12], we respectively predict target genes by the modes of action in which each lncRNA is likely to be involved in the following manner. Antisense action target gene prediction using the RNAplex to find short interactions between two long-stranded RNAs, to find the complementary binding between antisense lncRNAs and mRNAs; cis action target gene prediction annotating lncRNAs within 10kb upstream or downstream of a gene, to analysis of their potential to intersect with regions where cis-acting elements are located; trans action target gene prediction using Pearson's correlation coefficient method to analyze the correlation of lncRNA and protein-coding gene expression between samples, protein-coding genes with absolute correlation values greater than 0.999 were taken.

### **Targeted mRNA enrichment analysis (Go/Pathway)**

Using the DAVID database (V.6.8) and the R package "clusterProfiler" [13], Gene Ontology (GO) enrichment analysis of 410 differential mRNAs, to annotate function of lncRNAs [14]. Signaling pathways were also annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [15].

### **Quantitative Polymerase Chain Reaction (qPCR) Validate of LncRNA**

In this study, the selected core lncRNAs were subjected to qPCR for verification, the filtering conditions are FDR<0.05, |log<sub>2</sub>FC|>1, FPKM>0.5 and GO/ Pathway enrichment analysis of functional lncRNAs. β-Actin was used as an internal reference gene, to compare controls group (CK) and test group (T) by the  $2^{-\Delta\Delta Ct}$  value. The lncRNA expression rates between the CK and T groups were compared by the independent samples t-test. ROC curve statistical test for whether lncRNA can be used as a possible diagnostic biomarker for SCZ. SPSS 22.0 software was used to analyze and process the data, significant differences are indicated by p<0.05.

### **Functional Analysis of qPCR-validated LncRNA Sequences**

Acquisition of QPCR-validated transcript sequences, predicting functional sites for sequences using the PROSITE database [16], If there is no transcription, translation function or is not itself a transcription factor, and then use NCBI blast, predicting whether lncRNAs are transcripts of target genes and their homologues [17], conduct a conservative analysis via UCSC database [18], finding transcription factors that bind to lncRNAs. Finally, the JASPAR database was used to predict transcription factors and binding sites [19], selecting transcription factors with the same direction of transcription.

# Results

## Clinical Characteristics of Sequencing Samples

Statistical differences were analyzed between the sex and age of the 50 CK and 50 T groups collected. Analysis using the independent samples t-test yields: the mean age (mean  $\pm$  standard deviation) of the T and CK groups was  $35.56 \pm 10.04$  and  $32.52 \pm 8.04$ , respectively, and no significant difference ( $p > 0.05$ ), there was no significant difference in gender composition ( $p > 0.05$ ).

## RNA-seq Quality Assessment

Total RNA of the sequencing samples was analyzed for quality by an Agilent 2100 Bioanalyzer. The OD260/OD280 values are all between 1.8-2.0 (Fig. 1a), the two-sample independent repeated correlation is shown in Fig (Fig. 1b), suggesting that the quality of the total RNA isolated and purified in this experiment meets the requirements of sequencing. The euclidean distance algorithm is used to calculate the expression distance of each sample gene, and the Pearson algorithm is used to calculate the distance between the samples, then established cluster map according to the distance (Fig. 1c). 5106 new lncRNA transcripts obtained by CPC, CNCL and SwissProt prediction (Fig. 1d). Principal component analysis of sequencing samples showed that PC1 principal component was 79.4%, PC2 was 9.3%, and PC1+PC2 could distinguish 88.7% of the overall variance (Fig. 1e).

## Gene Co-Expression Network Construction

Selecting for differential expression of lncRNA and mRNA with  $FDR < 0.05$  and  $|\log_2FC| > 1$ . 94 differentially expressed lncRNAs were obtained (Fig. 2a,c), 70 were up-regulated and 24 were down-regulated, 46 new lncRNAs were identified for the first time, of which XLOC\_079942, LINC02362, ACTA2-AS1, AL162431.2, CLSTN2-AS1 were the top five differential fold change lncRNAs in up-regulated expression ( $|\log_2FC| > 5$ ), and AL390763.2, XLOC\_078139, XLOC\_043880, AC008892.1, AL034397.3 are the top five lncRNAs with differential fold change in down-regulated expression ( $|\log_2FC| > 4$ ). 1179 differentially expressed mRNAs were obtained (Fig. 2b,d), 702 were up-regulated and 477 were down-regulated, of which CNOT2, RNF19B, MECP2, TNFAIP3, ALDO are the top five mRNAs with differential fold change in upregulated expression ( $|\log_2FC| > 9$ ), and AHR, ARMC8, HBP1, ERVFC1-1, COTL1 are the top five mRNAs with differential fold change in downregulated expression ( $|\log_2FC| > 7$ ). Heat map of 94 lncRNA and 1179 mRNA expression, exhibiting distinct group information after clustering between groups, The up-regulated expression genes could clearly distinguish from the down-regulated expression group, and the differential genes could significantly represent the schizophrenia group and thus distinguish from the normal group, suggesting that the sequenced differential genes may serve as potential biomarkers of schizophrenia.

According to Spearman correlation coefficient analysis, 81 lncRNA and 410 mRNA have expression correlation ( $p < 0.01$  and  $|r| \geq 0.4$ ), 1152 mRNAs showed positive correlation with lncRNA, 300 mRNAs showed negative correlation with lncRNA (Fig. 3), the obtained data used cytoscape software to construct the lncRNA-mRNA co-expression network, and independent individuals were excluded. Exhibiting some

core lncRNAs in the co-expression network, such as PDCD1, MEG3, LINC00599, PSMA3-AS1, CLSTN2-AS1, AC008892.1, XLOC078139, etc, targeting multiple mRNAs. The core node IL1RAP was also found to be linked to several genes, suggesting that it could be a candidate gene for late functional validation.

### **LncRNA Action Modes**

Target gene prediction of the mode of action in which lncRNAs might be involved, yielding one lncRNA with an antisense interaction with mRNA, 22 lncRNAs present cis action, the trans-acting lncRNA is the most, up to 94 (Fig. 4). Among them, XLOC\_079942, XLOC\_081142, LINC00958, LINC00599, LINC01052, CLSTN2-AS1, AL034397.3, LINC01287 are more trans-associated with mRNA, as the core node, IL1RAP is associated with multiple lncRNA. The results show that lncRNAs mainly affect mRNA expression in a co-expression manner and thus participate in gene expression regulation and transcription.

### **Results of GO/Pathway Enrichment Analysis**

GO/Pathway enrichment analysis of 410 target mRNAs, filtered at  $P < 0.05$ . Enriched to 271 Biological Process (BP) terms, among them, the first four items that are significantly related to neutrophil activation, neutrophil degranulation, neutrophil activation involved in immune response, neutrophil mediated immunity, and leukocyte migration are all related to immune response (Fig. 5a,b); 23 Cellular Component (CC) terms, including specific granule lumen, secret granule lumen, cytoplasmic vesicle lumen, etc (Fig. 5c,d); 16 Molecular Function (MF) terms, including protein binding, transcription factor binding, transcription regulatory region DNA binding etc. (Fig. 5e,f). The first four of these BPs share the genes CAPN1, PLA2G2B, OLR1, CTSD, ABCA13, LCN2, MS4A3, and TCN1 are all involved in the pathogenesis of neuropsychiatric disorders.

The KEGG enrichment results also prove that the main enrichment pathways for targeted mRNAs are concentrated in the immune response, such as C-type lectin receptor signaling pathway, NOD-like receptor signaling pathway, TNF signaling pathway, NF- $\kappa$ B signaling pathway and so on (Fig. 6a-d).

### **QPCR Validation Results**

Mean Ct values of  $\beta$ -Actin were detected in peripheral blood samples from CK and T groups using  $\beta$ -Actin as an experimental internal reference, at 15.44-15.24, the CV% was 3.38-3.05. We performed qPCR experiments on the lncRNAs selected by the prenatal bioinformatics analysis to meet the conditions of  $FDR < 0.05$  and  $|\log_2FC| > 1$ , and the differential expression of  $FPKM > 0.5$  between groups. The experimental results were as follows: IL1RAP-TCONS\_00138311 was down-regulated in the T group relative to the CK group, which was 0.56 times ( $P < 0.05$ ) (Fig. 7a, Table 1); CCR3-TCONS\_00134168 was up-regulated in the T group relative to the CK group, and was 4.68 times that in the CK group ( $P < 0.05$ ) (Fig. 7b, Table 1); Thus the sequencing results are believed to be accurate and credible, and the sequencing results can be further analyzed and functionally verified.

Taking IL1RAP-TCONS\_00138311 and CCR3-TCONS\_00134168 as test variables, and CK group and T group as state variables, the ROC curve was constructed. The results showed that the expression of

IL1RAP-TCONS\_00138311 and CCR3-TCONS\_00134168 in peripheral blood leukocytes of SCZ patients had certain effects on the predicted value ( $P < 0.05$ ). Using IL1RAP-TCONS\_00138311 as a diagnostic indicator,  $AUC = 0.924$ , 95% of the confidence interval is 0.873-0.974 and the Jorden index reaches its maximum value 0.74 when  $FPKM = 15.48$  in PBL, at which point the sensitivity was 80.0% and the specificity was 94.0% (Fig. 7c, Table 1). Principal component analysis using IL1RAP-TCONS\_00138311 sequencing expression value shows that PCA1 plus PCA2 components can explain 24.3% of the overall variance (Fig. 7d).

**Table 1:** QPCR and ROC Analysis Results

LncRNA	IL1RAP-TCONS_00138311	CCR3-TCONS_00134168
Sequence	F:AGGGGAAGGGAATCAACAAATAG	F:TCAAGACTTCGTGGCTTAAACAATA
5' to 3')	R:GGGGCGTGGCATGTAACC	R:GGAACTCCATACCTGAAAGACCCTA
CK	7.13±0.69	14.59±0.88
T	8.58±1.06	12.93±1.43
$2^{-\Delta\Delta Ct^*}$	0.56	4.68
AUC	0.9236	0.6486
P-value*	0.009***	0.016**

\* $\Delta\Delta Ct = \text{Test group } \Delta Ct - \text{Control group } \Delta Ct$

\* $2^{-\Delta\Delta Ct} = \text{Expression ratio}$ ,  $2^{-\Delta\Delta Ct} > 1 = \text{Up-regulated}$ ,  $2^{-\Delta\Delta Ct} < 1 = \text{Down-regulated}$ .

### IL1RAP-TCONS\_00138311 Sequence Functional Analysis

The selected transcripts were predicted by CPC, CNCI, and SwissProt databases, and none had coding potential or protein annotation information. The acquires IL1RAP-TCONS\_00138311 transcript sequence uses the NCBI blast function, enter Query Length 1311, bits to 1023 length, and matches to Sequence ID: NG\_029105.2 (43715-44737, Score=1890, Expect=0.0, Identities=100%, Gaps= 0%), identified as a new IL1RAP transcript. Enter the upstream 2000bp and downstream 100bp sequences as the gene promoter region, and conduct a conservative analysis through the UCSC database to find transcription factors that can bind to lncRNA (Fig. 8a). Finally, using the JASPAR database, the score is set to 500, relative profile score threshold set 90%, select the transcription factor in the same transcription direction as IL1RAP-TCONS\_00138311, the top four ranked transcription factors were predicted to be HSF1, HSF2, HSF4, and FOXA1, and the binding sites are as follows (Fig. 8b).

## Discussion

The results show that lncRNAs mainly affect mRNA expression in a co-expression manner and thus participate in gene expression regulation and transcription. IL1RAP as a core node was also found to be linked to several genes. The variable clipping TCONS\_00138311 of IL1RAP was Differential expression verified by qPCR experimental, and a higher ROC area 0.924 can specifically distinguish between schizophrenic patients and healthy people.

GO/Pathway analysis enriched to 271 Biological Process (BP) terms, among them, the first four of these BPs share the genes CAPN1, PLAU, OLR1, CTSD, ABCA13, LCN2, MS4A3, and TCN1 are all involved in the pathogenesis of neuropsychiatric disorders. OLR1, CTSD, and MS4A3 are involved in the occurrence and development of neurodegenerative diseases. Interaction between OLR1 and LDL component L5 inhibits NGF-induced activation of the PI3k/Akt cascade and suppresses cell viability and neuronal differentiation induced by nerve growth factor (NGF) [20]. CTSD is expressed at high levels in almost all tissues of the brain and is essential for maintaining a steady state of lysosome-dependent proteins in the brain [21], and defective or reduced activity of CTSD expression can lead to pathological protein aggregation in Parkinson disease (PD), and Alzheimer disease (AD) [22]. The MS4A3 gene encodes a member of the transmembrane 4A gene family that exhibits a unique expression pattern in hematopoietic cells and non-lymphoid tissues, which may play a role in signal transduction and may function as a subunit associated with receptor complexes, and polymeric SNP association analysis of the gene showed a correlation between the gene and AD genes and cognitive decline [23]. CAPN1, PLAU, and TCN1 are involved in neuroprotection. CAPN1 calpain is a calcium-dependent protease, which is related to synaptic plasticity and neuroprotection in the mammalian brain [24] and regulates Alzheimer's disease-related protein signaling pathways[25]. PLAU is a urokinase-type plasminogen activator ligand that promotes neuroprotection with the receptor PLAU or its plasma membrane-binding partner platelet-derived growth factor receptor beta (PDGFR $\beta$ ) [26]. The TCN1 gene encodes a member of the vitamin b12-binding protein family, which is a major component of secondary granules in neutrophils, and vitamin B12 deficiency disrupts neurodevelopment during pregnancy and throughout childhood [27], adequate vitamin B12 is also necessary for adult neurocognitive function [28], elevated TCN1 expression is significantly associated with poorer memory [29]. ABCA13 and LCN2 were significantly associated with schizophrenia. The ABCA13 gene encodes a transmembrane transport protein, a C-terminal motif involved in lipid transport in lipid metabolism[30], and dysregulation of lipid metabolism is strongly associated with psychiatric disorders, with ABCA13 increasing susceptibility to schizophrenia, bipolar disorder, and major depression [31]. LCN2 plays an important role in mediating inflammatory activity by binding to chemotactic peptides, leukotrienes and platelet-activating factors [32], significantly increased in schizophrenia [33]. Others, such as TNIP1 and IL1B regulate NF-k $\beta$  activation and play a role in autoimmunity and tissue homeostasis; MGST1, CAMP, TNFRSF14, and FOS are involved in cell chemotaxis, immune mediator induction, and inflammatory response regulation.

The KEGG enrichment results also prove that the main enrichment pathways for targeted mRNAs are concentrated in the immune response, such as C-type lectin receptor signaling pathway, NOD-like receptor signaling pathway, TNF signaling pathway, NF- $\kappa$ B signaling pathway and so on (Fig. 6a-d). The nuclear factor-  $\kappa$ B (NF- $\kappa$ B) complex provides direct transcriptional control of interferon-induced

transmembrane protein (IFITM) and cytokines [34,35,36] that are overexpressed in schizophrenia role[37]. In the prefrontal cortex (PFC) of schizophrenic subjects, the mRNA levels of two NF- $\kappa$ B family members NF- $\kappa$ B1 and NF- $\kappa$ B2 increased, while the NF- $\kappa$ B site-binding protein Schnurri, which inhibits the function of NF- $\kappa$ B-2 mRNA levels are low [38]. These findings suggest that higher NF- $\kappa$ B transcriptional activity may play an important role in elevating IFITM and cytokine mRNA levels in schizophrenic PFC.

## Conclusions

In this RNA-seq, we used more samples for whole transcriptome sequencing and screened the differentially expressed lncRNAs, circRNAs, microRNAs and smallRNAs in the peripheral blood of SCZ patients. Among them, the variable clipping TCONS\_00138311 of IL1RAP caught our attention. IL1RAP is an essential protein for the IL1 receptor [39], regulation of IL1B-induced activation of the NF- $\kappa$ B pathway and other inflammatory pathways, its close relationship with SCZ. IL1B inflammatory pathway plays an important role in the development of SCZ [40], there are currently four main types of IL1RAP transcripts and protein isomers, among them mL1RAP and sL1RAP are the main types, they are different in molecular structure, tissue distribution and biological function [41,42], mL1RAP has intact extracellular, transmembrane and intracellular segments, causing IL1B downstream signaling and activating transcription and translation of IL1B effector genes [43] while sL1RAP only the extracellular segment can bind to IL1B, as the absence of an intracellular segment does not activate the IL1B pathway. Hence, we intend to investigate the regulation of IL1RAP variable splicing by the transcription factors FOXA1, HSF1, HSF2, HSF4, etc., in order to clarify whether IL1RAP aberrant variable splicing is involved in the pathogenesis of SCZ, or whether IL1RAP-TCONS\_00138311 new-transcripts induce the activation of NF- $\kappa$ B pathway and other inflammatory pathways through IL1B, and thus participate in the development of SCZ (Fig. 8c).

## Abbreviations

SCZ: Schizophrenia; PBL: peripheral blood ; GWAS: Genome-Wide Association Studies; lncRNA: Long non-coding RNA; IL1RAP: Interleukin 1 receptor accessory protein; HSF1: Heat shock transcription factor 1; HSF2: Heat shock transcription factor 2; HSF4: Heat shock transcription factor 4; FOXA1: Forkhead box A1.

## Declarations

### Ethics approval and consent to participate

IRB approval was given by both the Sixth Affiliated Hospital of Kunming Medical University, the Second People's Hospital of Honghe Prefecture and the Second People's Hospital of Yuxi City. Voluntary written informed consent was given by all participants (informed consent has obtained from a parent or guardian for participants under 16 years old)

## Consent for publication

Not applicable.

## Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This study was supported by the National Natural Science Foundation of China (81760253, 81960254) and the grant of Applied Basic Research Foundation of Yunnan Province (CN) 2018FE001-(009). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

YZ conceived and supervised the project. JW designed and implemented the methods, and conducted the experiments. ZJL, YQZ, ZWT, XY, YR, CQG contributed to data acquisition. JW, ZJL wrote the manuscript. All authors approved the manuscript.

## Acknowledgments

We thank Wenyu Zhang and Ziqiao Feng (Kunming Medical University Sixth Affiliated Hospital) for their assistance in recruiting and sequencing subjects for participation in the study.

## References

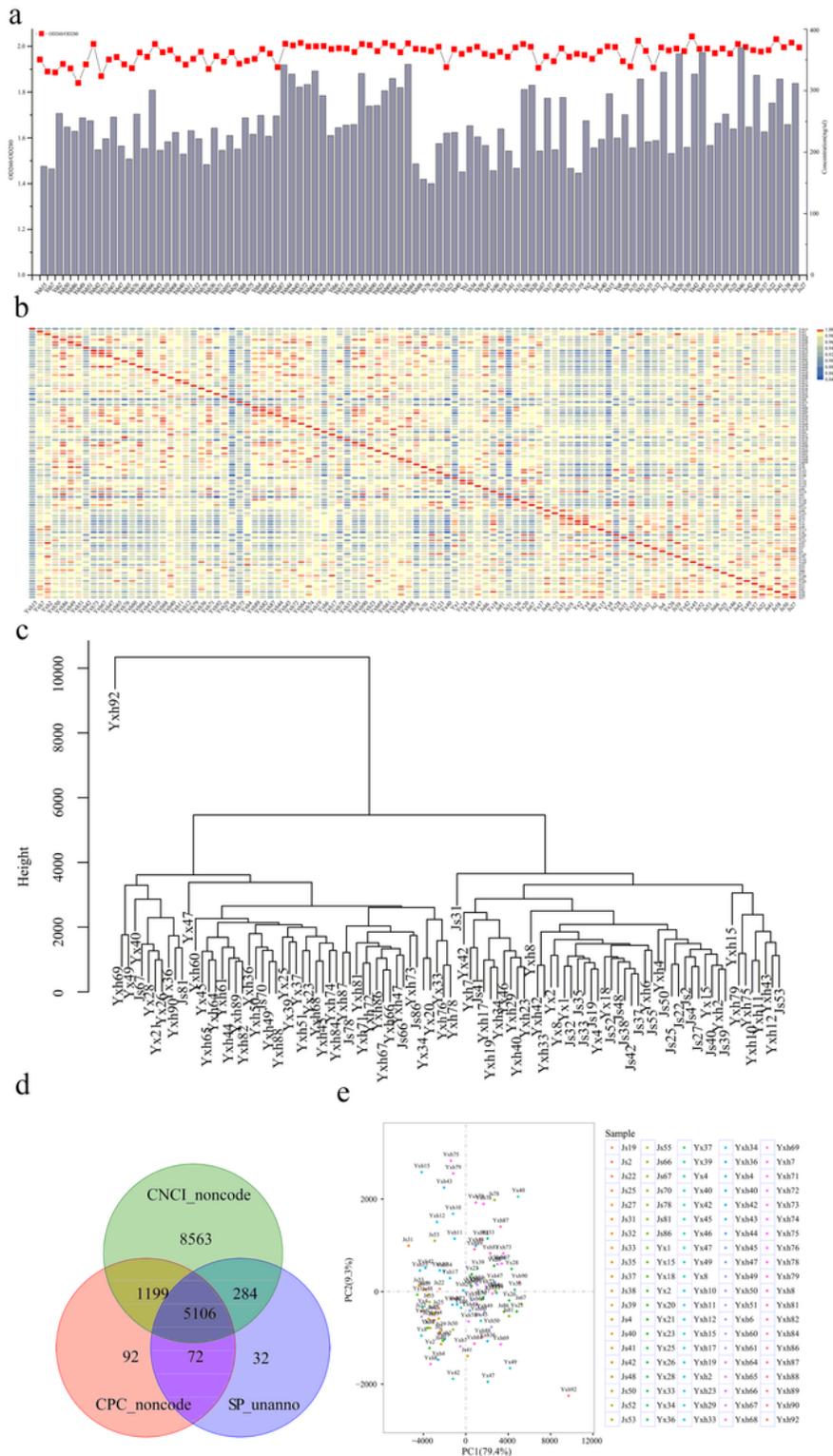
1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *Plos Medicine*, 2005;2(5):413-33.
2. Shaloo G, Gina I, Kevin J, Kristel VI. Productivity loss and resource utilization, and associated indirect and direct costs in individuals providing care for adults with schizophrenia in the EU5. *Clinicoeconomics & Outcomes Research Ceor*, 2015;7(11):593-602.
3. Tim BB, Stephan R, Silviualin B, Sang HL, Naomi RW, Pablo VG, et al. Genome-wide association study reveals greater polygenic loading for schizophrenia in cases with a family history of illness. *Am J Med Genet B Neuropsychiatr Genet*, 2016;171(2):276-89.
4. Shengdong C, Xinyang S, Wei N, Lingming K, Mingjun H, Wanshuai L, et al. Aberrant expression of long non-cding RNAs in schizophrenia patients. *Medical ScienceMonitor International Medical*

- Journal of Experimental & Clinical Research, 2016;22(9):3340-51.
5. Jie Sun, Meng Zh, Zhitao M, Dapeng H, Zhenzhen W, Chuanxing L, et al. Systematic analysis of genomic organization and structure of long non-coding RNAs in the human genome. *Febs Letters*, 2013;587(7):976-82.
  6. Jinneng C. The functional role of long non-coding RNAs and epigenetics. *Biological Procedures Online*, 2014;16(1):11.
  7. Christopher HW, Akul S, Josué PS, Stephen JG, Ming TT. The utility of gene expression in blood cells for diagnosing neuropsychiatric disorders. *International Review of Neurobiology*, 2011;101:41-63.
  8. Sol AJ, Jong LP, Jonghwan K, Jeong HK, Yong SK, Jin CK, et al. Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing. *Genomics Inform*. 2019;17(3):32-8.
  9. John CM, Christopher EM, Shrikant MM, Matthew S, Yoav G. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 2008, 18(9):1509-17.
  10. Tnicolai S, Kapila G, George AMC, Torsten O. Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. *Trends Parasitol* 2011;27(10):434-41.
  11. Robinson MD, Mccarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Biogeosciences*, 2010;26:139-40.
  12. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, 2011;43(6):904-14.
  13. Guangchuang Y, Ligen W, Yanyan H, Qingyu H. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics A Journal of Integrative Biology*, 2012;16(5):284-87.
  14. Gene OC, Mulder N. The Gene Ontology (GO) project in 2006. *Nucleic Acids Research*, 2005;34:322-46.
  15. Minoru K, Susumu G. KEGG: Kyoto Encyclopedia of Genes and Genomes[J]. *Nucleic Acids Res*. 2000;28(1):27-30.
  16. Christian JA, Lorenzo C, Edouard de C, Petra SL, Virginie B, Amos B, et al. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res*. 2010;38:161-66.
  17. Nuala AL, Mathew WW, J Rodney B, Stacy C, Diana H, Rich MV, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):733-45.
  18. Laurence RM, Ann SZ, Angie SH, Donna K, Robert MK, Matthew W, et al. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*, 2013;41:64-9.
  19. Oriol F, Jaime A CM, Aziz K, Robin van der L, Xi Z, Phillip A R, et al. JASPAR 2020: Update of the Open-Access Database of Transcription Factor Binding Profiles. *Nucleic Acids Res*, 2020;48(1):87-92.
  20. Jizyuh W, Chioulia L, Chingtien L, Chenyen L. Electronegative Low-Density Lipoprotein L5 Impairs Viability and NGF-Induced Neuronal Differentiation of PC12 Cells via LOX-1. *International Journal of Molecular Sciences*, 2017;18(8):1744-89.

21. Tokiko S, Yoshinori H, Ryo Y, Jun Y, Jian Z, Kiyoharu U, et al. Proteolytic degradation of glutamate decarboxylase mediates disinhibition of hippocampal CA3 pyramidal cells in cathepsin D-deficient mice. *Journal of Neurochemistry*, 2005;94(3):680-90.
22. Eija S, Sanna P, Petter S, Alekski H, Matti H, Jan M, et al. Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain*, 2006;129(6):1438-45.
23. Jasmine N, Gregory T, Daniel SE, Jennifer SY, Kristine Y. Gene-based aggregate SNP associations between candidate AD genes and cognitive decline. *Age*, 2016;38(2):41-51.
24. Soomaayeh H, Kahung L, Yan L, Yubin W, Michael RF, Xiaoning B, et al. Impaired cerebellar plasticity and eye-blink conditioning in calpain-1 knock-out mice. *Neurobiology of Learning & Memory*, 2020;170:106995.
25. Wenyue S, Qian Z, Yubin W, Athar C, Qingshun QL, Sujay D, et al. Deletion of the Capn1 Gene Results in Alterations in Signaling Pathways Related to Alzheimer's Disease, Protein Quality Control and Synaptic Plasticity in Mouse Brain. *Front Genet*, 2020;334(11):1-11.
26. Smith HW, Marshall CJ. Regulation of cell signalling by uPAR[J]. *Nature Reviews Molecular Cell Biology*, 2010;11(1):23-36.
27. Dror DK , Allen LH. Effect of vitamin B12 deficiency on neurodevelopment in infants: current knowledge and possible mechanisms. *Nutrition Reviews*, 2008,66(5):250-5.
28. Eileen M, Alastair M, David A, Ross C, Kerrie S, David W, et al. Cognitive impairment and vitamin B12: a review. *Int Psychogeriatr*. 2012;24(4):541-56.
29. Ibrahim AA, Torill U, Ole AA, Hansrichard B, Vidar MS, Timothy H, et al. Expression of TCN1 in Blood Is Negatively Associated With Verbal Declarative Memory Performance. *Sci Rep*. 2018;8(1):12654-65.
30. Michael LF, Keiichiro O, Glenn FS, Jennifer JM, Susan AB, Mason WF. ATP-binding Cassette Transporter A1 Contains a Novel C-terminal VFNFA Motif That Is Required for Its Cholesterol Efflux and ApoA-I Binding Activities. *Journal of Biological Chemistry*, 2004;279(46):48477-85.
31. Helen MK, Benjamin SP, Alan M, Mary PM, Dinesh CS, Allan FM, et al. A Cytogenetic Abnormality and Rare Coding Variants Identify ABCA13 as a Candidate Gene in Schizophrenia, Bipolar Disorder, and Depression. *Am J Hum Genet*. 2009;85(6):833-46.
32. Nils EM, Mads H, Kaj AJ, Jesper MH, Claus B, Bo FR, et al. Plasma Neutrophil Gelatinase Associated Lipocalin (NGAL) Is Associated With Kidney Function in Uraemic Patients Before and After Kidney Transplantation. *BMC Nephrology* 2012;13:2-8.
33. Li W, Yupeng D, Wei W, Xuyan F, Qi X. Elevation of plasma neutrophil gelatinase-associated lipocalin (NGAL) levels in schizophrenia patients. *Journal of Affective Disorders*, 2018;226:307-12.
34. Laura T, Anil KT, Jason B, Marta M, Guido F. The Nuclear Factor Kappa B Signaling Pathway: Integrating Metabolism With Inflammation. *Trends Cell Biol*. 2012;22(11):557-66.
35. Maria TD, Jorge M. The atypical PKCs in inflammation: NF- $\kappa$ B and beyond. *immunological reviews*, 2012;246(1):154-67.

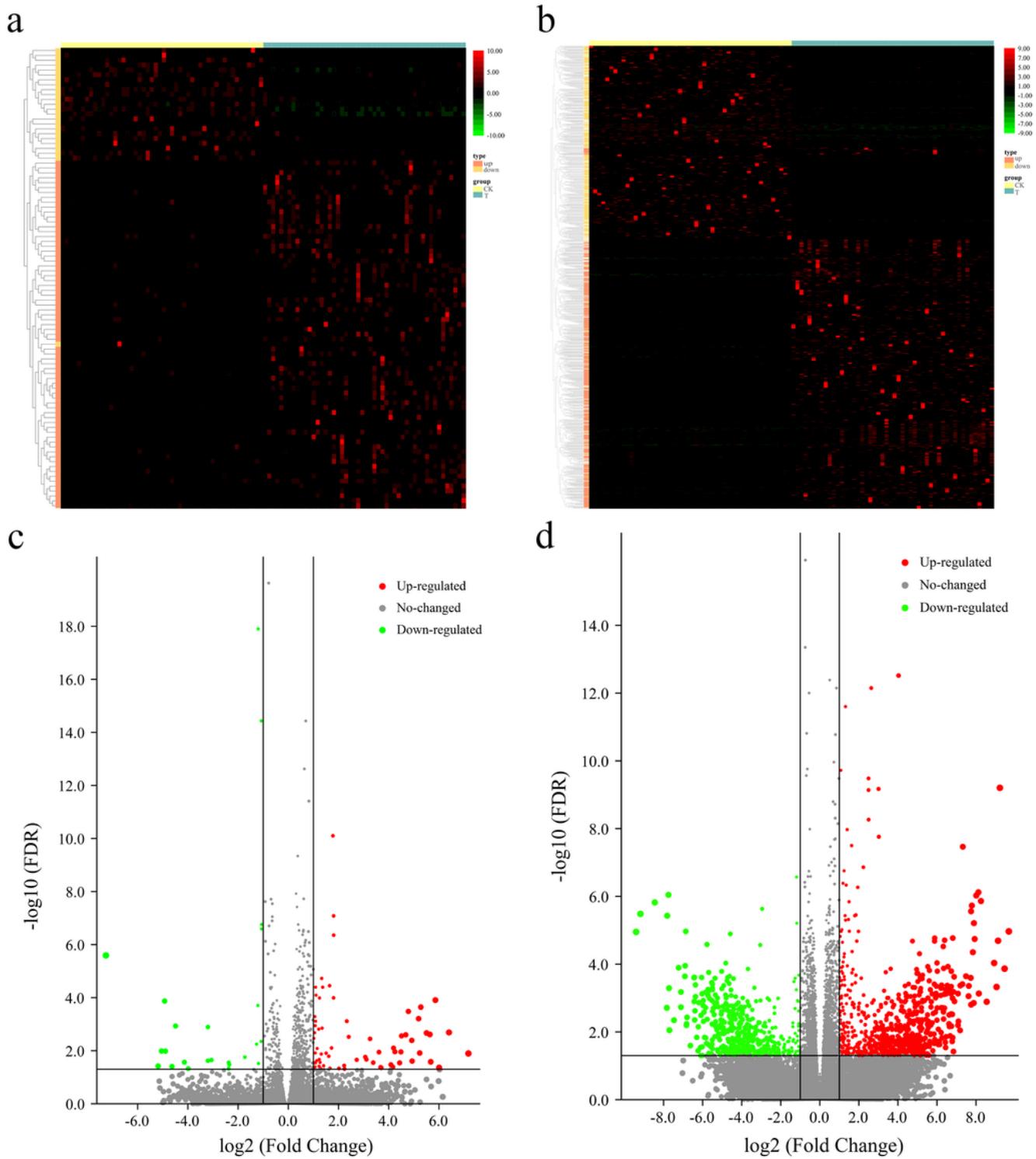
36. Keizo T, Katsunori K, Hideo H, Koji O, Hirotaka S, Satoko H, et al. Deficiency of Schnurri-2, an MHC Enhancer Binding Protein, Induces Mild Chronic Inflammation in the Brain and Confers Molecular, Neuronal, and Behavioral Phenotypes Related to Schizophrenia. *Neuropsychopharmacology*. 2013;38(8):1409-25.
37. David WV, Anjani C, Jessica RE, Kaitlyn MR, Annie EM, David AL. Molecular mechanisms and timing of cortical immune activation in schizophrenia. *Am J Psychiatry*. 2015;172(11):1112-21.
38. Keizo T, Katsunori K, Hideo H, Koji O, Hirotaka S, Satoko H, et al. Deficiency of Schnurri-2, an MHC Enhancer Binding Protein, Induces Mild Chronic Inflammation in the Brain and Confers Molecular, Neuronal, and Behavioral Phenotypes Related to Schizophrenia. *Neuropsychopharmacology*. 2013;38(8):1409-25.
39. Layé S, Liège S, Li KS, Moze E, Neveu PJ. Physiological Significance of the Interleukin 1 Receptor Accessory Protein. *Neuroimmunomodulation*, 2001;9(4):225-30.
40. Mostaid MdS, Dimitrakopoulos S, Cassandra W, Vanessa C, Cynthia SW, Everall IP, et al. An Interleukin-1 beta (IL1B) haplotype linked with psychosis transition is associated with IL1B gene expression and brain structure. *Schizophr Res*,2019;204:201-5.
41. Dirk ES, Roberta H, Della F, Heather M, Hongbo C, Ann MF, et al. The Soluble Form of IL-1 Receptor Accessory Protein Enhances the Ability of Soluble Type II IL-1 Receptor to Inhibit IL-1 Action. *Immunity*. 2003;18(1):87-96.
42. Liselotte EJ, Alexander SW. Expression of Alternatively Spliced interleukin-1 Receptor Accessory Protein mRNAs Is Differentially Regulated During Inflammation and Apoptosis. *Cell Signal*. 2003;15(8):793-802.
43. Wesche H, Korherr C, Kracht M, Falk W, Resch K, Martin MU. The Interleukin-1 Receptor Accessory Protein (IL-1RAcP) Is Essential for IL-1-induced Activation of Interleukin-1 Receptor-associated Kinase (IRAK) and Stress-activated Protein Kinases (SAP Kinases). *Journal of Biological Chemistry*. 1997;272(12):7727-31.

## Figures



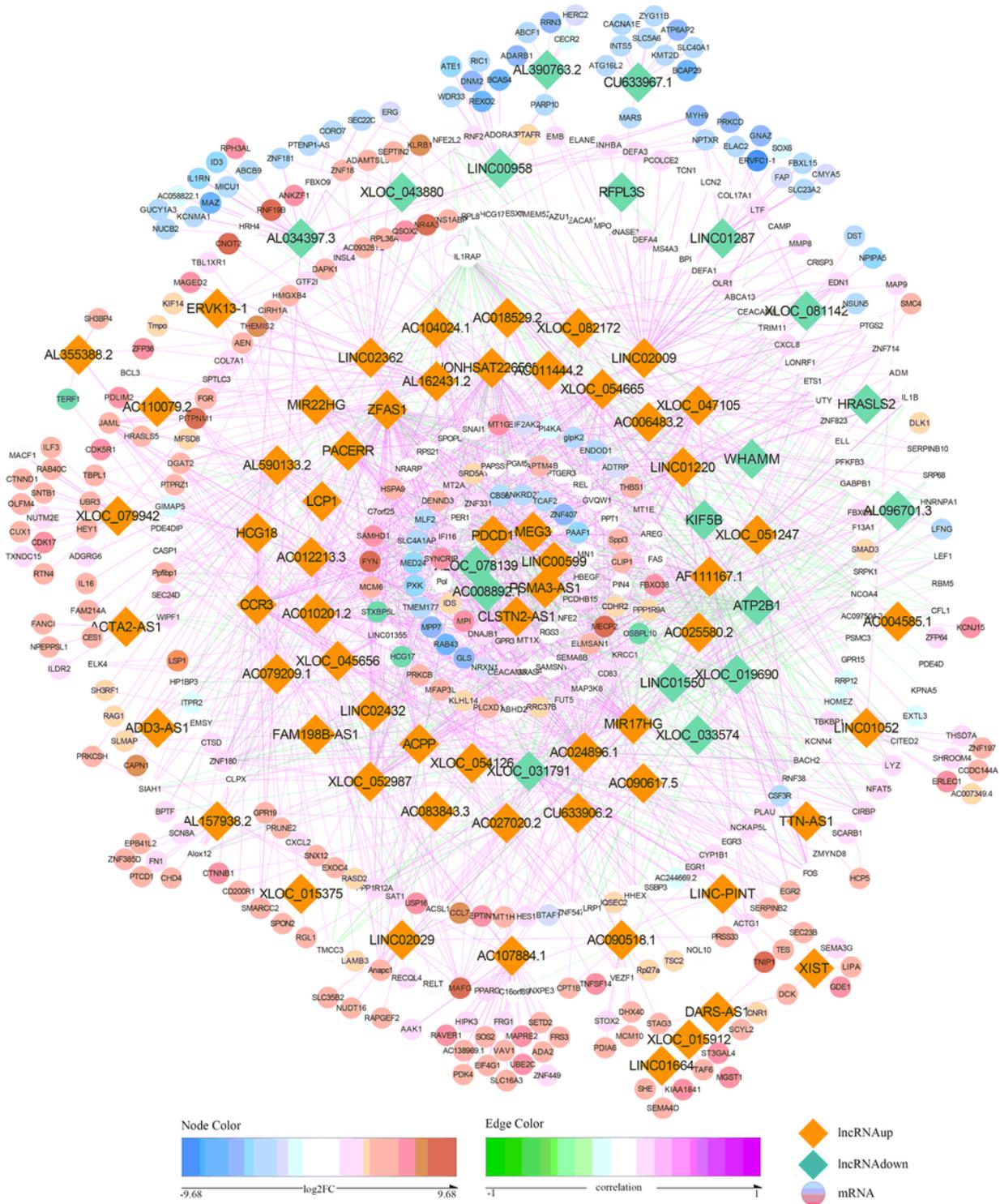
**Figure 1**

(a) Total RNA quality map of sequencing samples. (b) Correlation scatter plot of sequencing samples. (c) Sample clustering map. (d) Venn diagram of the new transcript of LncRNA, the intersection was predicted by CPC, CNCI and SwissProt (5106 new LncRNA). (e) PCA plots of sequencing samples, PC1 principal component can explain 79.4% of the overall variance of the expression of all transcripts, PC1 and PC2 can explain 88.7% of the overall variance.



**Figure 2**

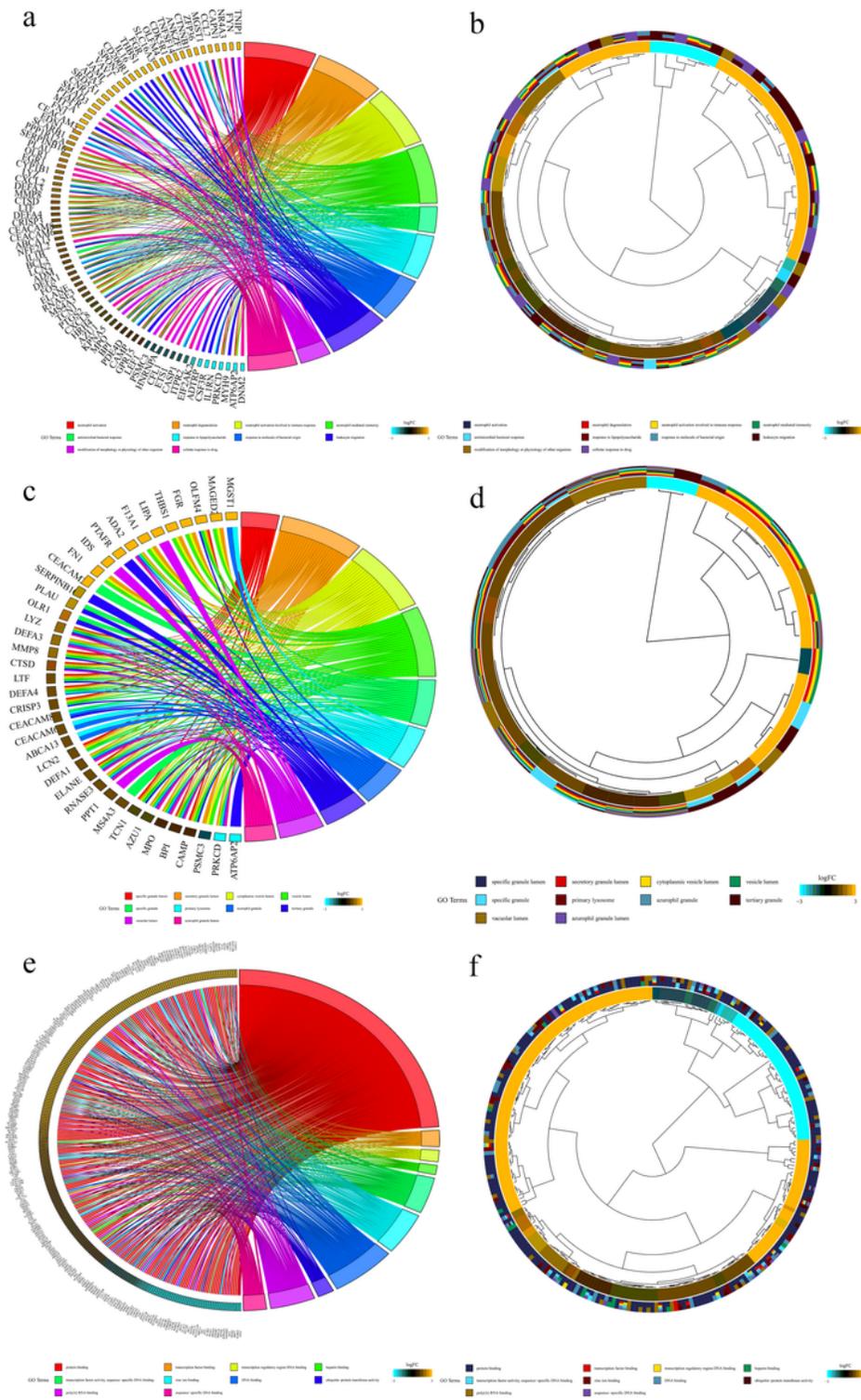
(a) Heatmap of expression profiles for the 94 lncRNAs. (b) Heatmap of expression profiles for 1179 mRNAs. (c) Volcano plot of the FDR values as a function of weighted fold-change for lncRNAs. (d) Volcano plot of the FDR values as a function of weighted fold-change for mRNAs.



**Figure 3**

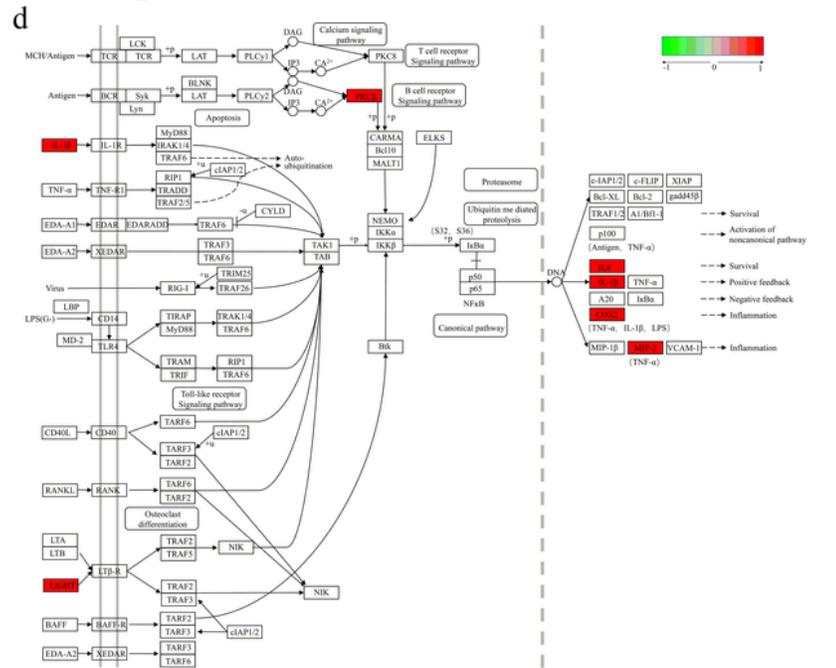
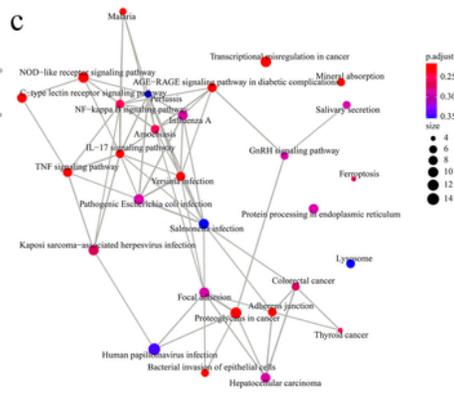
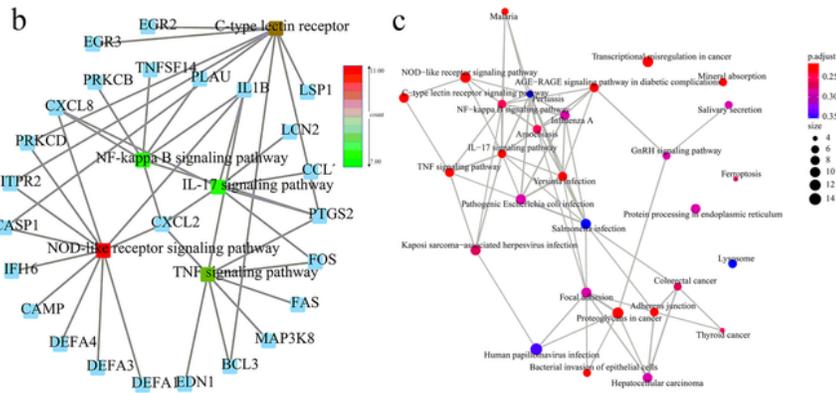
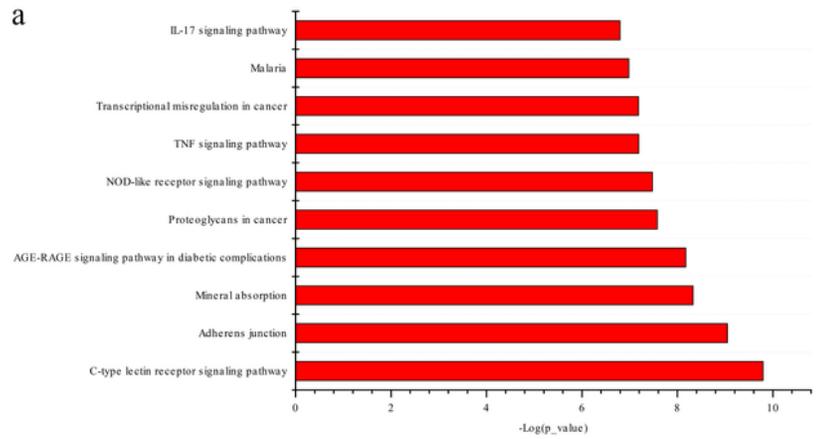
Predicted lncRNA and mRNA co-expression network in schizophrenia. The co-expression network was established between the 81 significantly expressed lncRNAs and 410 significantly differentially expressed mRNAs that had Spearman correlation coefficients ( $p < 0.01$  and  $|r| \geq 0.4$ ). Red through blue color indicates high to low expression level. Purple through green color indicates high to low correlation level.





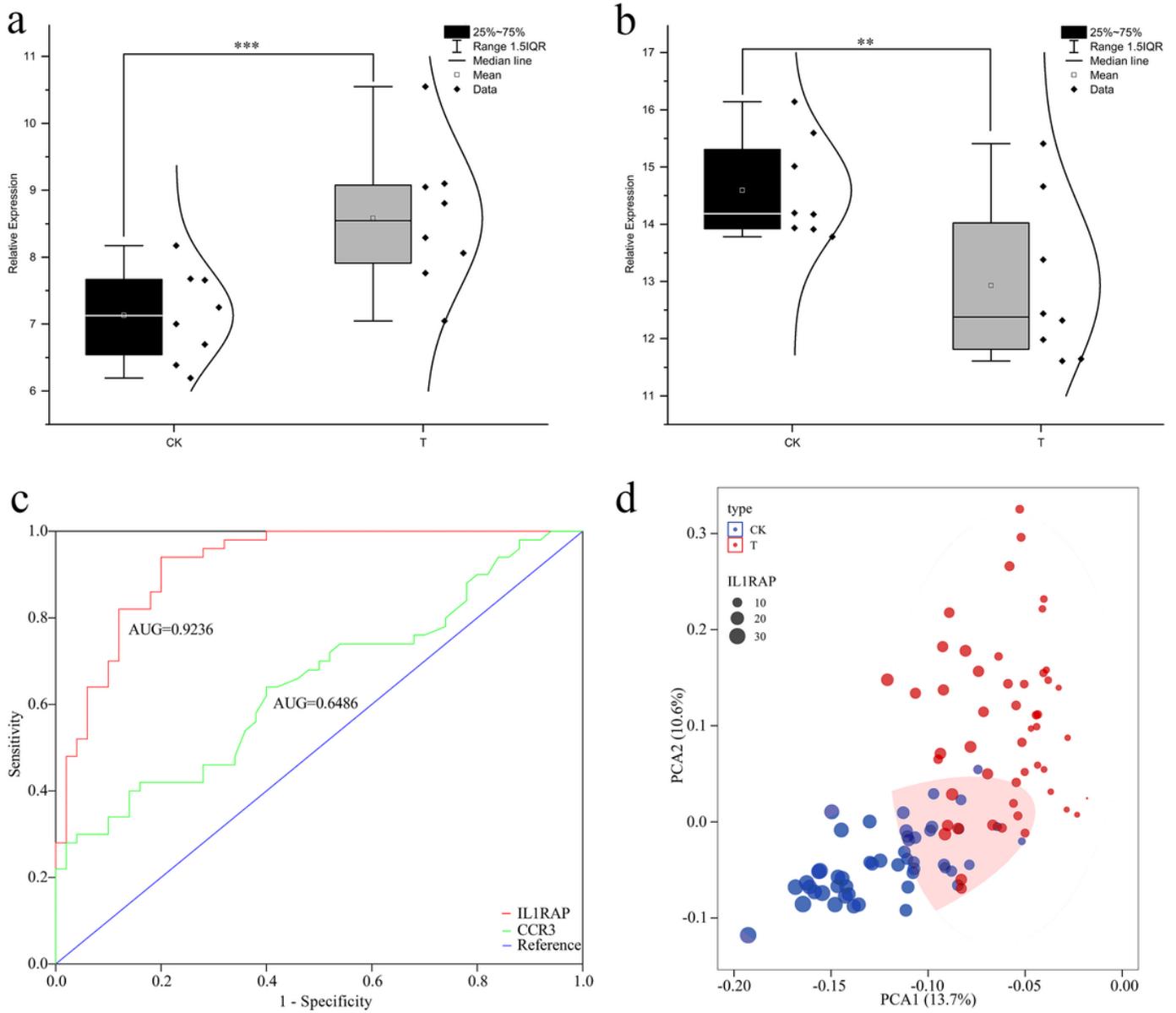
**Figure 5**

TOP 10 GO analysis of the significantly correlated mRNAs targeted by lncRNAs. (a) Chord plot displays of the relationship between gene and BP\_term. (b) Hierarchical clustering of the gene expression profiles in each BP\_term. (c) Chord plot displays of the relationship between gene and CC\_term. (d) Hierarchical clustering of the gene expression profiles in each CC\_term. (e) Chord plot displays of the relationship between gene and MF\_term. (f) Hierarchical clustering of the gene expression profiles in each MF\_term.



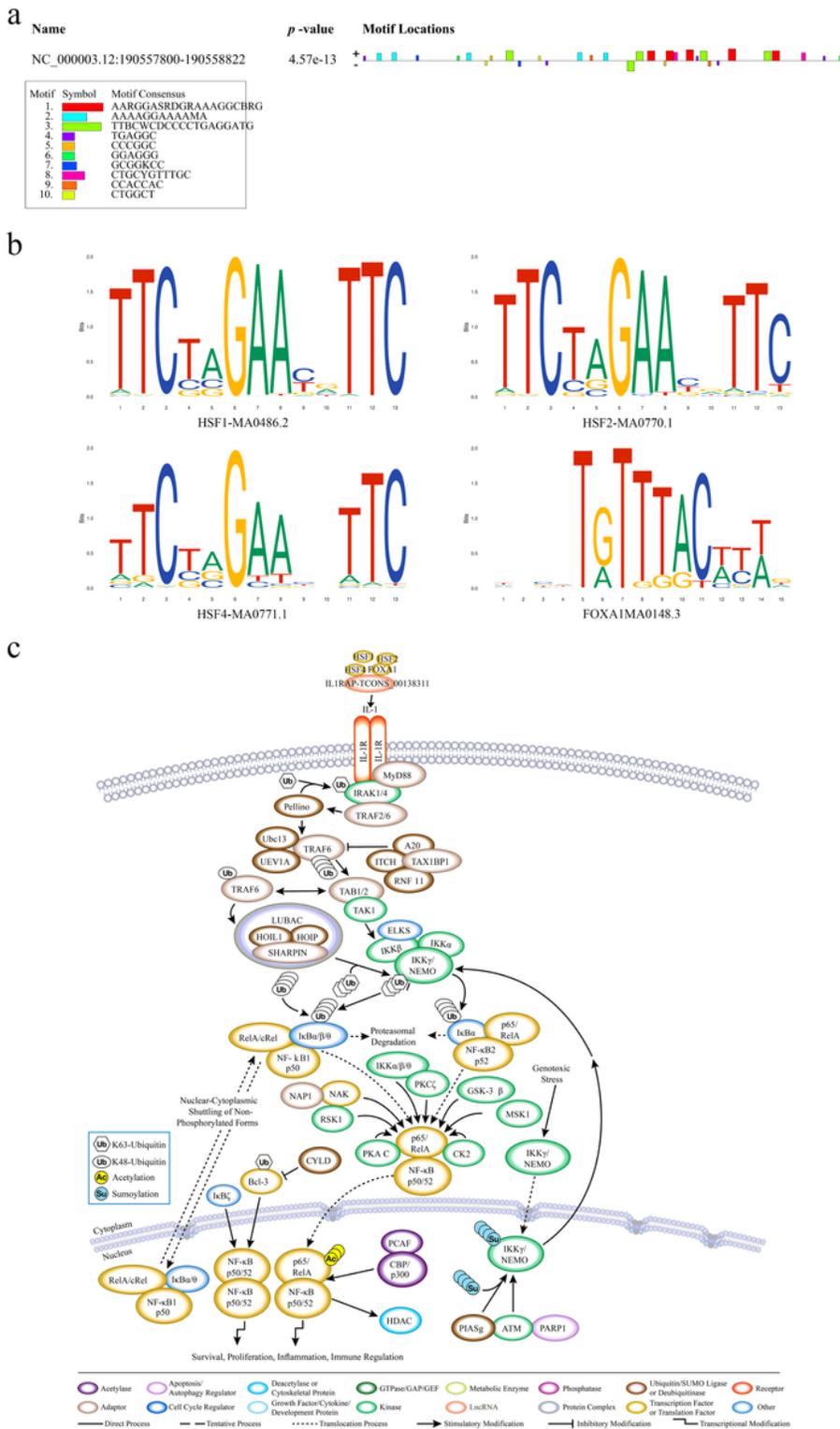
**Figure 6**

(a) TOP 10 KEGG analysis of the significantly correlated mRNAs targeted by lncRNAs. (b) Enrichment gene and top 5 Immune pathway co-expression network TOP 5 Immune pathway. (c) Co-expression network of KEGG pathway with common genes. (d) Enrichment gene of NF- $\kappa$ B signaling pathway.



**Figure 7**

(a) IL1RAP-TCONS\_00138311 expression level  $\Delta Ct(x \pm S)$ . (b) CCR3-TCONS\_00134168 expression level  $\Delta Ct(x \pm S)$ . (c) ROC curves predicted by lncRNA on SCZ. (d) IL1RAP-TCONS\_00138311 labeled in sample principal component analysis



**Figure 8**

(a) Predicted motif\_locations of IL1RAP-TCONS\_00138311. (b) The 4 most correlated transcription factors' sequence logo. The height of the base represents the information content of the base (the information content represents the size of the frequency of occurrence), and the information content of the base stack represents the size of the conservatism of the site. (c) IL1RAP-TCONS\_00138311/IL1B/NF- $\kappa$ B action mode diagram.