

Signalling networks constructed by integrative analysis of multi-omics in peripheral blood mononuclear cells in systemic lupus erythematosus patients

Wencong Song

Jinan University

Fengping Zheng

Shenzhen People's Hospital, Jinan University

Shaoying Huang

Jinan University

Wanxia Cai

Jinan University Institute for Economic and Social Research

Haiyan Yu

Jinan University

Xiaoping Hong

Jinan University

Jingquan He

Jinan University

Weier Dai

University of Texas at Austin

Lianghong Yin

Jinan University

Donge Tang

Jinan University

Dongzhou Liu

Jinan University

Yong Dai (✉ daiyong22@aliyun.com)

Jinan University <https://orcid.org/0000-0002-6840-9158>

Research article

Keywords: Systemic lupus erythematosus, Multi-Omics, Integrative Analysis

Posted Date: April 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-60254/v1>

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

[Read Full License](#)

Abstract

Background: Systemic Lupus Erythematosus (SLE) is a complex chronic autoimmune disease with productive autoantibodies and multiple organ systems damage. Currently, effective diagnosis and treatment for SLE are still challenged and the mechanism investigation of SLE will be helpful for developing the efficiency of disease diagnosis and treatment. Yet, the mechanism of SLE has not been clearly revealed. Thereafter, it needs more studies on the mechanism of SLE. We aim at investigating and verification critical signaling pathways underneath the molecular mechanism of SLE disease and constructing a systematic signaling network by integrating data from different omics.

Methods: 20 pairs of experimental groups (the SLE patients and healthy people) were recruited in our study. Omics technologies and biomolecules included genomics (DNA), whole transcriptomics (miRNAs, lncRNAs, circRNAs and mRNAs), proteomics (proteins) were used for collecting high throughput data from peripheral blood mononuclear cells (PBMCs) of subjects in our study. Joint analysis method of Multi-omics was taken advantage of in our studies for investigating some disease-related regulatory signaling networks.

Results: Potential biomolecules that contain unique expression patterns and disease related signaling pathways in SLE were identified in our study. These pathways and biomolecules include complement and coagulation cascades (C1q, C6), Systemic lupus erythematosus (HIST1H2BA, C1QC), Oxidation (NFYB, WFS1, ATXN3L), apoptotic (PIK3R3, GADD45G, CTSF), JAK-STAT (TYK2, STAT1/2), TNF signaling, Cell proliferation, cellular senescence et al. Signaling pathways identified by different omics were connected with each other by sharing some common molecules and constructed the potential systematic regulation networks in SLE.

Conclusions: Multi-omics technologies and the integrative analysis method develop some novel potential biomarkers and systemic signaling networks, which promotes our understanding of SLE disease and extend information for improving studies of diagnosis and treatment.

Introduction

SLE is a worldwide disease with approximate 70/100,000 prevalence in China [1, 2]. There is no specific or efficient diagnosis and treatment for SLE, thus it needs more studies on the etiology of SLE disease.

SLE is the autoimmune disease with a large number of autoantibodies production, which results in multiple tissues and organs' damage. Early diagnosis and treatment improve patient prognosis [3]. Up till now, the diagnosis and treatment of SLE disease are largely relied on the autoantibodies laboratory tests and the immune-suppressive drug respectively. However, the existing diagnosis and treatment methods for SLE disease contain some drawbacks, for example, many autoantibodies are non-specific, which lead to activation of polyclonal B cell [4], while immune-suppressive drugs such as steroids might increase the susceptibility to dangerous infections. To improve the diagnosis and treatment of SLE disease, more studies are needed for the SLE molecular pathogenesis mechanism investigation. Notably, exploration of

SLE disease relative signaling pathways will be helpful for understanding disease's molecular mechanism, and investigating diagnostic biomarkers and the drug targets.

To unveil the molecular mechanism of SLE, great efforts have been made in searching for signalling pathways [5]. By utilising Omics technologies, many susceptibility genes associated with SLE pathogenesis were identified [6-9]. Previous studies have shown that genomic DNA methylation were associated with SLE disease, and the unique DNA methylation patterns of some adhesion molecules such as LFA-1, CD70 were identified in SLE patients [10, 11]. Besides, H4 acetylation level [12] and expression level of miRNAs such as miR-142, miR-25, miR-125A, miR-146 were correlated to the disease activity of SLE [13]. mRNA expression level of some genes like interferon (IFN), stat1 and prf1 were demonstrated as good indicators for the SLE disease, as well as the expression level of proteins like IFNA, STAT1[14], PRF1[15], K1 and K10, K2e, K5, K6, K14 and K16. Those genes mentioned above were involved in particular signaling pathways that have been demonstrated correlated with SLE, such as JAK-STAT (STAT1, STAT4)[14], superoxide dismutase (SOD) , NADPH [16], apoptosis (Fas, Bcl2, TNF and IFN) [17, 18] etcetera.

Omics technologies have developed our understanding on the molecular mechanism of SLE and provided an amount of information for improving the treatment and diagnosis of SLE. Nonetheless, due to the complex pathogenesis of SLE, studies by one or two omics methods are limited. Therefore, the exact molecular mechanism of SLE has not been well elucidated so far. The integrative analysis of data generated by multi-omics in recent years has provided new ways for biomedical research. The joint analysis study combines data of messenger RNA (mRNA) and long non-coding RNAs (lncRNA) in SLE [19]. Cell proliferation, mRNA processing and translation-related pathways were identified in induced pluripotent stem cells from SLE patients' urine by integrative analysis of mRNA, microRNA and protein [5]. The vital role of Coagulation and complement pathway in the pathogenesis of SLE were studied by integrated analysis of Transcriptomics, Proteomics and Metabolomics [20]. Nevertheless, to explore the mechanism of SLE disease, the number of studies on integrative analysis of multi-omics are far more enough.

In our study, peripheral blood mononuclear cells (PBMCs) were collected from both SLE patients and healthy subjects. With the help of integrated analysis method, unique methylation profiles of simultaneous genes and distinctive expression profiles of Long non-coding RNAs (lncRNA), microRNAs (miRNAs), circular RNAs (circRNAs), messenger RNAs (mRNAs) and proteins in SLE patients were identified. At the same time, the systematic signaling regulation networks of SLE were constructed in here.

Methods And Materials

Blood samples collection

Blood samples were collected from both 20 healthy controls (H_B) and 20 SLE patients (SL_B) in Shenzhen People's Hospital (Guangdong, China). Recruited criteria: SL_B patients diagnosed as having SLE disease while H_B without SLE or other immune diseases or treated with immunosuppressant. The study was approved by the ethics committee of Shenzhen People's Hospital, and the experiments were undertaken and carried out by the guidelines of the Declaration of Helsinki of 1995 (as revised in Edinburgh 2000). Besides, all participants signed informed consent forms.

PBMCs preparation

PBMCs from each donor were obtained by using the Ficoll density gradient separation method (Sigma, USA). Then PBMCs was stored at -80°C for further treatment.

Whole-genome bisulfite sequencing (WGBS)-DNA methylation

Genomic DNA library construction and sequencing

Total DNA was extracted by QIAamp Fast DNA Tissue Kit (Qiagen, Dusseldorf, Germany) and measured by spectrophotometer. The fragmented DNA samples by using sonication were subjected to bisulfite conversion. The Accel-NGS Methyl-Seq DNA Library Kit (Swift, MI, USA) was utilized for attaching adapters to single-stranded DNA fragments. Then, the pair-end $2\times 150\text{bp}$ sequencing was conducted by an Illumina HiSeq 4000 platform.

Bioinformatics analysis for WGBS data in PBMCs

Sequence quality verification was performed by the FastQC online analysis tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). WALT was used for mapping qualified reads to the reference genome[21], and samtools was used for further deduplication of reads[22]. The DNA methylation level was determined by using perl scripts and MethPipe[23]. The R package-MethylKit was utilized for calculation of differentially methylated regions (DMRs)[24].

Whole transcriptomics

Small-RNA library construction and sequencing

Total RNA was extracted from PBMCs by using TRIzol reagent (Invitrogen, CA, USA). TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, USA) was used for small RNA sequencing library construction, and then the library was sequenced by an Illumina HiSeq 2500 with 50 base single-end reads.

Differential expression analysis for RNAs

The raw datasets were processed by the program of and Cutadapt and ACGT101-miR (LC Sciences, Houston, TX, US)[25]. Qualified reads were mapped to the human reference genome (GRCh38) by Bowtie

2[26] and TopHat2[27]. Differential expression levels of mRNAs, miRNAs, circRNAs and lncRNAs were calculated [28] and Fisher's exact test according to $|\log_2(\text{fold change})| > 1$ and $p\text{-value} < 0.05$.

Bioinformatics analysis for mRNAs and lncRNAs

Transcripts Assembly: Firstly, qualified reads were sorted by Cutadapt[25], then verified by FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped to the genome of humans by the usage of Bowtie2[26] and Tophat2[27]. The mapped reads of each sample were assembled using StringTie[28]. Then, all PBMSs' transcriptomes were merged to reconstruct a comprehensive transcriptome. Finally, tools of StringTie[28] and Ballgown [29] were utilised for calculating transcripts' expression.

lncRNAs identification

First of all, the known mRNAs and transcripts shorter than 200 bp were castoff. Then CPC[30], CNCI [31] and Pfam[32] were used for prediction of transcripts.

Localization of lncRNAs on genome

With the help of program Circos[33], localization and abundance of lncRNAs in the genome could be present in the way of diagrams. The lncRNAs in diagram were subdivided into six categories (*I, J, O, U, X, K*) according to their class code generated by StringTie: (*I*) A transfrag falling entirely within a reference intron (Intronic); (*J*) Potentially novel isoform or fragment at least one splice junction is shared with a reference transcript; (*O*) Generic exonic overlap with a reference transcript; (*U*) Unknown, intergenic transcript (intergenic); (*X*) Exonic overlap with reference on the opposite strand (antisense); (*K*) The known, lncRNAs .

Prediction of miRNA targets

TargetScan[34] and Miranda[35] were tools for prediction of miRNA targets.

Bioinformatics analysis for circRNAs

Firstly, Cutadapt[25] was used for un-quality reads filtering, then sequence quality was verified by online analysis tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Bowtie2[26] and Tophat2[27] were used for humans genome mapping. Unmapped reads were mapped by using tophat-fusion[28]. CIRCEplorer [26] was used to assemble the mapped reads to circRNAs; Then, back splicing reads were identified in unmapped reads. The expression level of circRNAs from different samples or groups was calculated. By the R package edgeR[27], the comparisons with $p\text{-value} < 0.05$ were selected.

Bioinformatics analysis of miRNAs

RNA families (rRNA, tRNA, snRNA, snoRNA) and repeats were applied to filter non-miRNA sequences in the ACGT101-miR program (LC Sciences, Houston, Texas, USA). Subsequently, the remaining sequences

were mapped to human precursors in miRBase 22.0 by BLAST to identify both known miRNAs and novel 3p- and 5p- derived miRNAs.

The Prediction of Target Genes of miRNAs

Two computational target prediction algorithms (TargetScan 5.0 and Miranda 3.3a) were used for the prediction of genes targeted by these most abundant miRNAs.

Proteomics-iTRAQ

Samples were eluted in lysis buffer and further processed by the procedures in the previous report [20] and finally get the supernatant aliquot. Protein was digested with Trypsin Gold (Promega). Then peptides were dried and reconstituted in 0.5 M TEAB and processed according to the manufacturer's introduction. The peptides were subjected to nano-electrospray ionization. Intact peptides were detected in the Orbitrap. The details of the protocol were referred from a previous study [20].

Bioinformatics analysis for proteomics

Msquant search engine (version 2.3.02) and related database were utilized for protein identification [36]. Each confident protein contains at least one unique peptide. The median peptide ratio in Mascot was used for normalization of the quantitative protein ratios.

Functional enrichment analysis

The databases of Gene Ontology (GO) (<http://www.geneontology.org>) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg>) were utilized for predicting the underlying biological functions and pathways of the candidate target genes. DNA methylation, ceRNAs (lncRNAs, circRNAs), miRNAs, mRNAs and proteins related genes were put into the database for annotation, visualization and integrated discovery (DAVID) for GO and KEGG pathway analysis.

Statistical analysis

Before further analysing of data, all variables were go through normality and equal tests. Unpaired Student's *t* test was used for analysis the difference of normal distributed data while nonparametric tests were used for non-normal distributed data. Data were analysed by GraphPad Prism (Version 6, CA) and GraphPad Prism uses the abbreviation mean±SEM. $P < 0.05$ was considered as statistically different.

Signaling networks construction

TargetScan (v. 5.0) [37], miRanda (v. 3.3a) [38] were used for predicting potential regulation networks of miRNA-mRNA and lncRNA-miRNA-mRNA. The regulatory networks of methylation-mRNA, methylation-miRNA-mRNA and protein-mRNA were unveiled in PBMCs-SLE. The regulation networks in proteins-proteins was predicted by the STRING database (<http://string-db.org/>) and KEGG database. The

overlapping predictions between two datasets were considered as effective target pairs, and all signaling networks were constructed by Cytoscape (v. 3.5.0) [39].

qRT-PCR verified the identified genes in the PBMCs from SLE patients and healthy people

The technical verification of identified RNA sequencing data in the PBMCs of SL_B and H_B was under the help of the Quantitative real-time polymerase chain reaction (qRT-PCR). Total RNA was extracted from the PBMCs described above and subjected to first-strand cDNA synthesis using the TRUEScript 1st Stand cDNA SYNTHESIS Kit (Aidlab, Beijing, China). The specific primers used in the qRT-PCR are listed in supplement material (Table S10), and qRT-PCR was performed using an SYBR PrimeScript miRNA RT-PCR Kit (TianGen Biotech, Beijing, China). GAPDH and U6 were used as internal controls. Three independent biological replicates for each gene. The comparative $2^{-\Delta\Delta C_t}$ analyzed method was used in this study.

Results

Twenty both sex and age match SLE patients (SL_B) and healthy subjects (H_B) with no autoimmune disease were recruited. The SLE diagnostic criteria in this study was based on ACR SLE classification criteria and the characteristics of the enrolled people are present in supplement material (Table S11).

Summarization of KEGG enrichment and GO analysis for mono-omics

We summarized the data of KEGG enrichment and GO analysis from single omics which including Genomics (DNA methylation), Transcriptomics (lncRNAs, circRNAs, miRNAs and mRNAs), Proteomics (proteins) (Fig. 1-3). We use the interactive Venn diagram (Fig. 1E) to display the overlap of KEGG pathways and GO terms in single omics, and the detail of the KEGG pathways and GO terms in mono-omics were showed in the table (Table 1).

Different single omics shared some common regulatory pathways which showed in the Venn diagram (Fig. 1E) and table 1, and these common pathways including Apoptotic process, Cell adhesion et al.

Table 1 KEGG pathways and GO terms in different single omics

KEGG pathways and GO terms in single omics (pathway and items' number)	Name of KEGG pathways and GO terms
miRNA (47)	Ubiquitin mediated proteolysis, RAS signaling, PI3K-Akt signaling, mTOR signaling, Apoptotic process, Cell differentiation, Oxidation process, Cell adhesion, GTPase signaling, regulation of transcription, transcription, ...
DNA methylation (38)	Apoptotic process, Oxidation process, GTPase signaling, Cell proliferation, Cell adhesion, Cell differentiation, T cell receptor signaling, RNA polymerase, Response to stimulus, ...
lncRNA (43)	GTPase signaling, Apoptotic process, Cell differentiation, Immune response, Cell proliferation, Oxidation process, Systemic lupus erythematosus, Cell adhesion, Signal transduction, Transcription, DNA-Templated, ...
circRNA (38)	Apoptotic process, GTPase signaling, Oxidation reduction process, Cell adhesion, Inflammatory response, Cellular response to hypoxia, Rheumatoid arthritis, Protein transport, Membrane organization, ...
mRNA (38)	Toll-like receptor signaling, TNF signaling, Systemic lupus erythematosus, Rheumatoid arthritis, Apoptotic process, Antigen processing and presentation, Immune response, Cell adhesion, Cell differentiation, Cell proliferation, GTPase signaling, Inflammatory response, ...
Protein (39)	Oxidation process, Immune response, Apoptotic process, Antigen processing and presentation, T cell receptor signaling, Dopaminergic synapse, Fructose and mannose metabolism, ...

Integrative analysis of multi-omics

Methylation-mRNA

GO analysis was conducted for functional analysis of methylation modified genes. GO terms were enriched by features of methylated genes (the differential mRNAs expression of methylated genes were selected by the sorted criterial: $|\log_2(\text{fold change})| \geq 1$ and $P\text{-value} < 0.05$). BP GO terms contain the Oxidation-reduction process et al. CC GO terms include Mitochondrion and Mitochondrial inner membrane et al. MF GO terms contain ATP binding (Fig. 4B). KEGG showed that the differential expression of Methylation modified genes were involved in the pathways of JAK-STAT et al. (Fig. 4A).

miRNA-mRNA

GO analysis was conducted for the functional analysis of miRNA targeted genes. GO terms were enriched by miRNA targeted genes (the differential mRNAs expression of which were put in the sorted criterial:

$|\log_2(\text{fold change})| \geq 1$ and $P\text{-value} < 0.05$). BP GO terms contain apoptotic process et al. CC GO terms include mitochondrion et al. MF GO terms contain ATP binding et al. (Fig. 4B). KEGG enrichment showed that the differential expression of miRNA targeted genes were involved in signaling pathways including Apoptosis et al (Fig. 4A).

Methylation-miRNA-mRNA

GO analysis was conducted for the functional annotation of methylated genes targeted by miRNA. GO terms of biological processes (BP), molecular functions (MF) and cellular components (CC) were selected, which were enriched by the miRNA targeted methylated genes (Sorted criteria: $|\log_2(\text{fold change})| \geq 1$ and $P\text{-value} < 0.05$). In the enriched GO terms, BP contains positive regulation of GTPase activity, MF contains ATP binding et al.(Fig. 4B).

KEGG showed that the differential expressed genes involving in the integrative data of methylation-miRNA-mRNA was involved in pathways of Cellular Senescence, Antigen Processing and Presentation and TNF signaling et al.(Fig. 4A).

mRNA-Protein

GO terms were enriched for differential expressed proteins in the integrative data of protein-mRNA (Sorted criteria was the expression levels of proteins in the joint analysis data fulfil the following requirements: (fold change ≥ 1.2 or fold change ≤ 0.83) and $P\text{-value} < 0.05$). In our study, GO terms including 25 BP, 10 MF and 15 CC. In the GO analysis, BP GO terms contain Apoptotic process, Cell proliferation et al (Fig. 4B).

KEGG enrichment analysis showed that the differential expressed genes in the integrative analysis of protein-mRNA were involved in signaling pathways including Systemic Lupus Erythematosus, Complement and Coagulation Cascades et al (Fig. 4A).

We also summarized the KEGG and GO data from figure 1, and calculate the overlap of KEGG pathways or GO terms in the joint analysis of multi-omics, and found that different multi-omics shared some common KEGG pathways or GO terms (Table 2). Those common KEGG pathways and GO terms include apoptosis, oxidation, inflammation process et al.

Table 2 KEGG and GO analysis summarization in integrative analysis of multi-omics

Integrative analysis of multi-omics (number of KEGG pathways and GO terms)	Name of overlap KEGG pathways and GO terms
DNA methy-mRNA (41)	JAK-STAT pathway, Oxidation process, Cell differentiation, Cell adhesion, GTPase signaling, Cell proliferation, Proteolysis...
miRNA-mRNA (32)	Apoptotic process, GTPase signaling, Immune response, Cell differentiation, Natural killer cell mediated cytotoxicity, Antigen processing and presentation, Signaling transduction...
DNA methy-mRNA-miRNA (43)	TNF signaling, GTPase signaling, Cell adhesion, Cellular senescence, Antigen processing and presentation, Cell cycle...
mRNA-protein (41)	Apoptotic process, Cell proliferation, GTPase signaling, Systemic lupus erythematosus, Complement and coagulation cascades, Natural killer cell mediated cytotoxicity, ...

Further analysis of the common KEGG pathways and GO terms in both single omics and multi-omics

Basing on the summarization of mono-omics and multi-omics described above, there were some common KEGG pathways and GO terms identified in our study (Table 1 and Table 2). Besides, part of these common pathways and GO terms in our study which containing significant differential expressed molecules (mRNA or protein) were selected and displayed (Fig. 5A-C). Data also unveiled many critical molecules that acting as bridges connecting different biological processes and signaling pathways, such as hsa-miR543_R+1 connects the Th1 cell differentiation pathway with Th17 cell differentiation pathway (Fig. 5B), mPI3K3 connects the cellular senescence signaling and TNF signaling (Fig. 5B).

The network was made up of 6 genes (mRNA), 3 pathways, 9 miRNAs and 62 lncRNAs. Yellow circles represent target pathways, red triangles symbolize genes and green arrow stand for miRNAs, blue circles indicate lncRNAs. The grey lines stand for connection between different pathways, genes, miRNAs and lncRNAs.

Technical validation of RNA sequencing data

The technical validation of sequencing data was conducted by RT-PCR. There was a panel of 12 genes and 3 ceRNAs were selected from the top differential expressed ones. Our technical validation data (Fig. 6A) showed that some genes including IFN α , ENST00000506362, ENST00000274764 and ENST00000578377 were up-regulated while others were down-regulated in the SLE patients, such as TYK2(NM_003331.4), ENST00000493202, ENST0000052956, ENST00000263413 and ENST00000262741. The selected genes mentioned above were consistent to our sequencing data (Table S1, S4-8).

Three of our selected panel genes including ENST00000494726, STAT1 (NM_139266.2) and ENST00000375769 contain similar expression trends in both RT-PCR and sequencing data, but they were not significant in RT-PCR results. Besides, the non-coding RNAs were also validated by RT-PCR, and raw data was showed in Table S9.

Discussion

In this study, hundreds of SLE unique expressed molecules and several SLE disease-related signaling pathways were identified. These pathways including T-cell receptor (42), Oxidation-reduction process, regulation of Apoptotic process, Complement and Coagulation Cascades, JAK-STAT etcetera (Table 1-2).

It was found the Complement and Coagulation Cascades played important role in the development of SLE disease [20]. we found that mRNA expression of some Complement and Coagulation pathway-related genes including C6, C1QC were significantly elevated in our study, which was consistent to the study mentioned above by Liang, et al. JAK-STAT might have an essential role in pathogenesis of SLE disease by regulating IFN regulatory factor-related genes[40], and the role of JAK-STAT cascades in SLE was also verified in our previous study on 6-methyladenine (6mA). In our current study, we found JAK-STAT signaling related genes such as PRLR, SOCS6, GHR contain unique profile in both methylation and mRNA expression in SLE patients (Table S1). Our proteomic data also demonstrated that several core genes in JAK-STAT signaling, including STAT1 and STAT3 were elevated substantially. Has-miR-1249-5p was found to regulate the JAK-STAT signaling, and miRNAs (miR-2110, miR-7110) regulate mRNAs including PIK3R3 and PDGFB (Table S3). miRNAs and mRNAs, such as miR-432-5p and STAT1 were reported by previous studies that contain unique expression in SLE disease [41, 42] (Table S2). The Oxidation process plays important role in the SLE development[16], which was also supported by our data. Similarly, regulation networks like Apoptotic process, Autoimmune, Innate immune process, Cell Proliferation process and JAK-STAT signaling were discussed in SLE studies previously[10, 17, 40, 43, 44], and the function of all these signaling networks mentioned above were also verified by our study (Fig. 1-5). The data of lncRNAs, miRNAs and proteins were combined and constructed a systematic regulation network of SLE disease in our study (Table S2). The SLE disease allied signaling network unveiled by our study provides a systematic insight for understanding the potential molecular mechanism underlies the SLE disease.

This is the first study in SLE that combining 3 omics methods and 6 molecules at the same time. Specifically, the omics technologies and the molecules used in our study included genomics (DNA), whole transcriptomics (ceRNAs (lncRNAs, circRNAs), miRNAs, mRNAs) and proteomics (proteins). By utilizing the joint analysis method of multi-omics, bunch of potential biomarkers and regulation signaling networks were identified. This study helps us further understand the SLE disease and expands our view on systematic signaling regulation networks for the molecular mechanism of SLE.

The limitation of single omics

Recently, searching for new biomarkers has been considered as emerging event on SLE research [45]. Single omics studies focusing on a single molecule are limited as they neither fully reveal the complex pathogenesis nor display the complex interaction of molecules at different omics' levels in SLE disease.

The barriers of integrative analysis of multi-omics

The integrative analysis of multi-omics in recent years has provided new method for biomedical research. The complex immune responses in SLE disease are usually result from different bio-molecules interaction and cross-talk of different signaling networks. Therefore, integrating omics data of different omics platforms becomes an effective tool for SLE disease study. Nevertheless, the integrative analysis of omics is not flawless and it has some limitations too.

Experimental and Technical Variation

Different generations of platforms (for example, LC-MS/MS) were used for detecting proteomic in various ranges of chemicals identity and abundance. Such technical heterogeneity results in challenging in the joint analysis of omics studies. Technical variation between our study and the previous reports may contribute to the data variation. In the future, ways for tackling those problems may include using standard control samples and appropriate statistical models (eg, mixed-effect model) to adjust for batch effect and harmonize these measurement variations.

Analytical challenges

We got some valuable data from the joint analysis of the multi-omics, for example, some SLE relevant biological markers and regulated signalling pathways were dug out in our study. The various types of omics data we got here were fundamentally different: some data are discrete and static, while others are opposite. The differential omics data sets contribute to the challenge of integrated analysis of multi-omics. Firstly, the statistical framework for the joint analysis of multiple disparate data sets should be robust and reproducible. Secondly, combination of multi-dimensional methods, such as Bayesian models [46], neural networks[47] is needed for simultaneously analyzation of several data sets.

Accuracy and validation

High confidence is critical for the clinical application, thus independent technologies including established clinical tests, such as enzymatic and single-assay tests are needed for our data validation. Besides, the amount of blood samples from both SLE patients and health controls are not enough, and more samples are needed for the verification of the markers and pathways unveiled by our study.

Interpretation

Different kind of omics' technologies and the joint analysis of omics produce large amount of data. However, we face difficulty when we try to do the date interpretation especially for some rare and novel molecular events, cause those events are often beyond the ones that can be functionally validated. Therefore, more studies will be needed in the future for unveiling the function of unknowing molecules.

Conclusion

By integrative analysis of multi-omics, we unveiled some common regulation signaling pathways in SLE disease. Signaling pathways identified in this study included Systemic Lupus Erythematosus,

Complement and Coagulation Cascades, TNF signaling, JAK-STAT, Oxidation (GTPase activity, mitochondrion), Apoptosis et al. (Fig. 6B). Besides, some critical biological molecules were also unveiled in our study, such as hsa-miR543_R+, mPI3K3, PRF1, mTTLL7, RASGRP4 and so forth (Fig. 6A). These molecules mentioned above shared by different signaling pathways and function as the linkage among pathways. With the help of the shared molecules between signaling pathways, different signaling pathways correlate with each other and construct complex regulation networks crosstalk. These networks crosstalk might help us better understand the pathogenesis of SLE disease systematically and promote drug targets and diagnostic biomarkers investigation in SLE disease.

Abbreviations

SLE: Systemic Lupus Erythematosus; PBMCs: peripheral blood mononuclear cells; IFN: interferon; SOD: superoxide dismutase; lncRNA: long non-coding RNAs; mRNA: messenger RNA; circRNAs: circular RNAs; miRNAs: microRNAs; WGBS: Whole genome bisulfite sequencing; DMRs: differentially methylated regions; IAM: iodoacetamide; TEAB: tetraethyl-ammonium bromide; HPLC: high performance liquid chromatography; qRT-PCR: Quantitative real-time polymerase chain reaction; BP: biological processes; MF: molecular functions (MF); CC:cellular components; GO: gene ontology; KEGG: Kyoto encyclopaedia of genes and genomes. TNF: Tumor Necrosis Factor.

Declarations

Ethics approval and consent to participate

The study was approved by the ethics committee of Shenzhen People's Hospital (the committee's reference number: LL-KY-2019589), and the experiments were undertaken and carried out by the guidelines of the Declaration of Helsinki of 1995 (as revised in Edinburgh 2000). Besides, all participants signed informed consent forms.

Consent for publication

Not applicable

Availability of data and supporting materials section

The raw data for small RNA sequencing and ribosomal RNA-depleted sequencing in our study are available from [Gene Expression Omnibus: GSE146410] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Other data including proteomics generated or analysed during this study are included in this submitted article [and its supplementary information files].

Authors' contributions

Wencong Song and Yong Dai conceived and designed the experiments. Shaoying Huang, Fengping Zhen, Donge Tang acquired the data. Haiyan Yu, Jingquan He, Wan-Xia Cai, Xiaoping Hong, Lianghong Yin, Dongzhou Liu, Weier Dai analysed data. Wencong Song and Yong Dai draft this manuscript. All authors revised and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Science Foundation for Young Scientists of China (Grant No32000641), the Key Research and Development Program of Guangdong Province (No. 2019B020229001), the Science and Technology Plan of Shenzhen (No. JCYJ20200109144218597 and JCYJ20190807145815129), Sanming Project of Medicine in Shenzhen (No. SYJY201704 and No. SYJY201705), Shenzhen Key Medical Discipline Construction Fund (No. SZXK011).

Acknowledgements

Not applicable

References

1. Osio-Salido E, Manapat-Reyes H. Epidemiology of systemic lupus erythematosus in Asia. *Lupus*. 2010;19(12):1365-1373.
2. Rojas-Villarraga A, Torres-Gonzalez J-V, Ruiz-Sternberg A-M. Safety of hormonal replacement therapy and oral contraceptives in systemic lupus erythematosus: a systematic review and meta-analysis. *PloS one*. 2014;9(8):e104303.
3. Doria A, Zen M, Canova M, Bettio S, Bassi N, Nalotto L *et al*. SLE diagnosis and treatment: when early is early. *Autoimmunity reviews*. 2010;10(1):55-60.
4. Egner W. The use of laboratory tests in the diagnosis of SLE. *Journal of clinical pathology*. 2000;53(6):424-432.
5. Tang D, Chen Y, He H, Huang J, Chen W, Peng W *et al*. Integrated analysis of mRNA, microRNA and protein in systemic lupus erythematosus-specific induced pluripotent stem cells from urine. *BMC genomics*. 2016;17(1):488.
6. Barilla-LaBarca M-L, Atkinson JP. Rheumatic syndromes associated with complement deficiency. *Current opinion in rheumatology*. 2003;15(1):55-60.
7. Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL *et al*. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nature genetics*. 2008;40(2):204.

8. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X *et al.* A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature genetics*. 2009;41(11):1228.
9. Fernando MM, Freudenberg J, Lee A, Morris DL, Boteva L, Rhodes B *et al.* Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. *Annals of the rheumatic diseases*. 2012;71(5):777-784.
10. Richardson B. DNA methylation and autoimmune disease. *Clinical Immunology*. 2003;109(1):72-79.
11. Lee W-W, Yang Z-Z, Li G, Weyand CM, Goronzy JJ. Unchecked CD70 expression on T cells lowers threshold for T cell activation in rheumatoid arthritis. *The Journal of Immunology*. 2007;179(4):2609-2615.
12. Zhang Z, Song L, Maurer K, Petri MA, Sullivan KE. Global H4 acetylation analysis by ChIP-chip in systemic lupus erythematosus monocytes. *Genes and immunity*. 2010;11(2):124.
13. Ding S, Liang Y, Zhao M, Liang G, Long H, Zhao S *et al.* Decreased microRNA-142-3p/5p expression causes CD4+ T cell activation and B cell hyperstimulation in systemic lupus erythematosus. *Arthritis & Rheumatism*. 2012;64(9):2953-2963.
14. Fang TJ, Lin YZ, Liu CC, Lin CH, Li RN, Wu CC *et al.* Methylation and gene expression of histone deacetylases 6 in systemic lupus erythematosus. *International journal of rheumatic diseases*. 2016;19(10):968-973.
15. Kozłowska A, Hrycaj P, Łacki JK, Jagodziński PP. Perforin level in CD4+ T cells from patients with systemic lupus erythematosus. *Rheumatology international*. 2010;30(12):1627-1633.
16. Shah D, Mahajan N, Sah S, Nath SK, Paudyal B. Oxidative stress and its biomarkers in systemic lupus erythematosus. *Journal of biomedical science*. 2014;21(1):23.
17. Munoz LE, van Bavel C, Franz S, Berden J, Herrmann M, Van Der Vlag J. Apoptosis in the pathogenesis of systemic lupus erythematosus. *Lupus*. 2008;17(5):371-375.
18. Salmon M, Gordon C. The role of apoptosis in systemic lupus erythematosus. *Rheumatology*. 1999;38(12):1177-1183.
19. Luo Q, Li X, Xu C, Zeng L, Ye J, Guo Y *et al.* Integrative analysis of long non-coding RNAs and messenger RNA expression profiles in systemic lupus erythematosus. *Molecular medicine reports*. 2018;17(3):3489-3496.
20. Liang Y, Xie S-B, Wu C-H, Hu Y, Zhang Q, Li S *et al.* Coagulation cascade and complement system in systemic lupus erythematosus. *Oncotarget*. 2018;9(19):14862.
21. Chen H, Smith AD, Chen T. WALT: fast and accurate read mapping for bisulfite sequencing. *Bioinformatics*. 2016;32(22):3507-3509.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
23. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PloS one*.

- 2013;8(12):e81148.
24. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology*. 2012;13(10):R87.
 25. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal*. 2011;17(1).
 26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-359.
 27. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013;14(4):R36.
 28. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*. 2015;33(3):290.
 29. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology*. 2015;33(3):243.
 30. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research*. 2007;35(Web Server issue):W345-349.
 31. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research*. 2013;41(17):e166-e166.
 32. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al.* The Pfam protein families database. *Nucleic acids research*. 2011;40(D1):D290-D301.
 33. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* Circos: an information aesthetic for comparative genomics. *Genome research*. 2009;19(9):1639-1645.
 34. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4.
 35. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome biology*. 2010;11(8):R90.
 36. Naryzhny SN, Maynskova MA, Zgoda VG, Ronzhina NL, Kleyst OA, Vakhrushev IV *et al.* Virtual-experimental 2DE approach in chromosome-centric human proteome project. *Journal of proteome research*. 2015;15(2):525-530.
 37. Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*. 2007;27(1):91-105.
 38. Manyam G, Ivan C, Calin GA, Coombes KR. targetHub: a programmable interface for miRNA–gene interactions. *Bioinformatics*. 2013;29(20):2657-2658.
 39. Liu J, Li H, Zheng B, Sun L, Yuan Y, Xing C. Competitive Endogenous RNA (ceRNA) Regulation Network of lncRNA–miRNA–mRNA in Colorectal Carcinogenesis. *Digestive diseases and sciences*.

2019;64(7):1868-1877.

40. Kawasaki M, Fujishiro M, Yamaguchi A, Nozawa K, Kaneko H, Takasaki Y *et al*. Possible role of the JAK/STAT pathways in the regulation of T cell-interferon related genes in systemic lupus erythematosus. *Lupus*. 2011;20(12):1231-1239.
41. Rai G, Rai R, Saeidian AH, Rai M. Microarray to deep sequencing: transcriptome and miRNA profiling to elucidate molecular pathways in systemic lupus erythematosus. *Immunologic research*. 2016;64(1):14-24.
42. Dong J, Wang Q, Zhou C, Ma X, Zhang Y. Activation of the STAT1 signalling pathway in lupus nephritis in MRL/lpr mice. *Lupus*. 2007;16(2):101-109.
43. Bradshaw S, Zheng WJ, Tsoi LC, Gilkeson G, Zhang XK. A role for Fli-1 in B cell proliferation: implications for SLE pathogenesis. *Clinical Immunology*. 2008;129(1):19-30.
44. de la Fuente H, Richaud-Patin Y, Jakez-Ocampo J, González-Amaro R, Llorente L. Innate immune mechanisms in the pathogenesis of systemic lupus erythematosus (SLE). *Immunology letters*. 2001;77(3):175-180.
45. Liu C-C, Ahearn JM. The search for lupus biomarkers. *Best practice & research clinical rheumatology*. 2009;23(4):507-523.
46. Fridley BL, Lund S, Jenkins GD, Wang L. An Bayesian Integrative Genomic Model for Pathway Analysis of Complex Traits. *Genetic epidemiology*. 2012;36(4):352-359.
47. Holzinger ER, Dudek SM, Frase AT, Pendergrass SA, Ritchie MD. ATHENA: the analysis tool for heritable and environmental network associations. *Bioinformatics*. 2013;30(5):698-705.

Figures

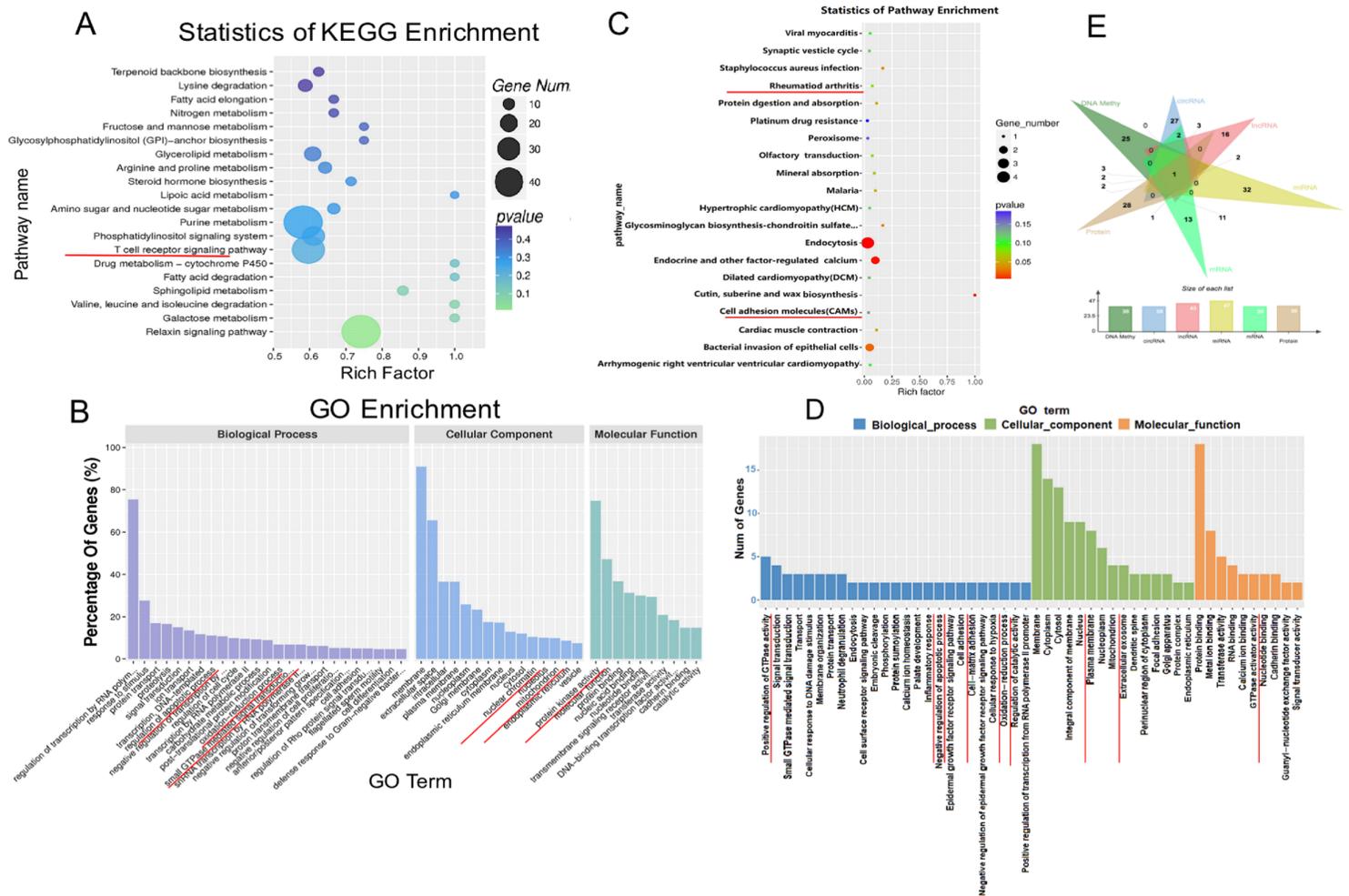


Figure 1

KEGG enrichment and GO analysis for both Chromosome methylation and circRNAs. For KEGG enrichment, the horizontal axis indicates the Rich factor, which represents the enrichment level of methylation targeted genes (A) and circRNA hosting genes (C). The vertical axis displays KEGG pathways that modified by methylation or regulated by circRNAs.

the number of transcripts in GO terms (B) and the percentage of differential expressed proteins in specific GO terms (D). The GO terms were sorted by the threshold of fold changes of specific target genes ($|\log_2$ fold change) ≥ 1 and P-value < 0.05 in mRNA and (fold change ≥ 1.2 or fold change ≤ 0.833 and P-value < 0.05) in protein. The items of interested were marked in Red line underneath.

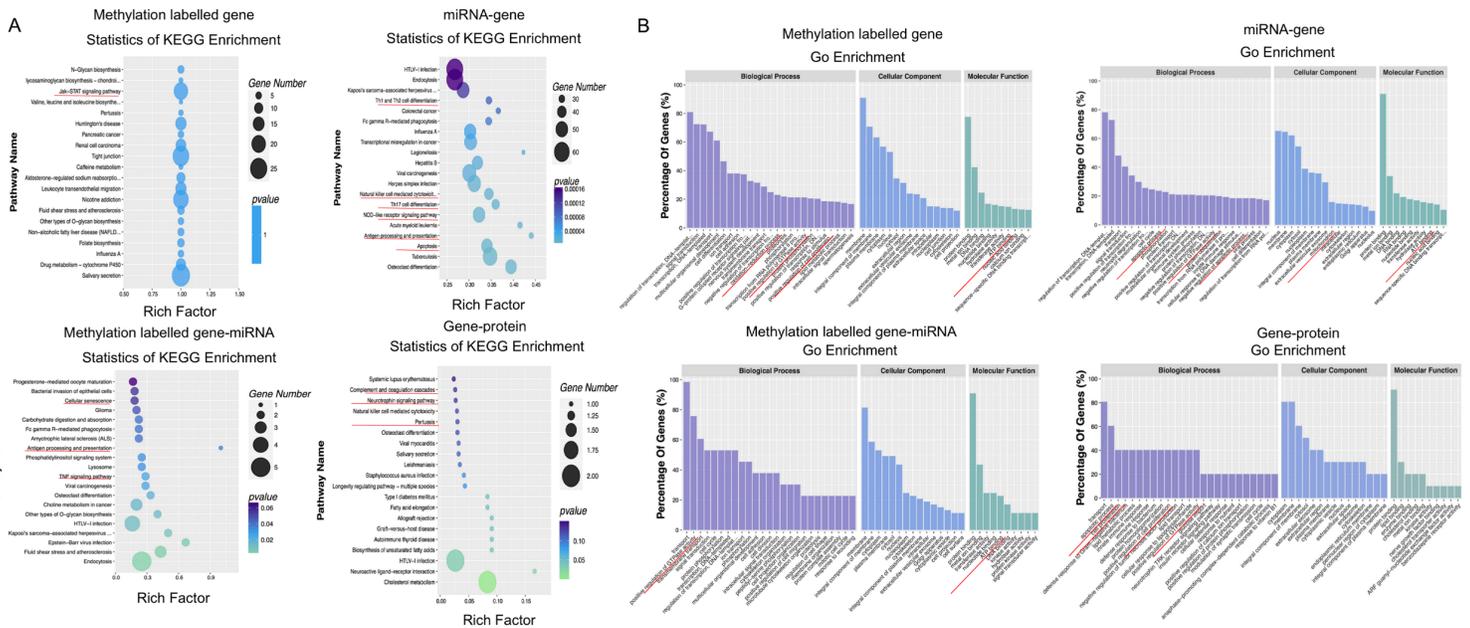


Figure 4

KEGG enrichment and GO analysis of the integrative omics' data. (A-B) The KEGG enrichment and GO analysis: For methylation-mRNA, miRNA-mRNA and methylation-miRNA-mRNA, genes involved in specific GO terms and KEGG pathways were sorted by the filter threshold $|\log_2$ fold change ≥ 2 and P-value < 0.05 . For protein-mRNA, data were sorted by the filter threshold (for mRNA : $|\log_2$ fold change ≥ 2 and P-value < 0.05 ; for protein: FC >1.2 or FC <0.8333 and P-value < 0.05 , FC means fold change). (A) the horizontal axis indicates the Rich factor, which represents the enrichment level of the relative molecule, while the vertical axis displays the regulatory pathways involved in SLE disease. (B) The horizontal axis indicates the names of the GO terms of CC, BP, MF, respectively. The vertical axis displays the number of related genes involved in particular GO terms. The items of interested were marked in Red line underneath.

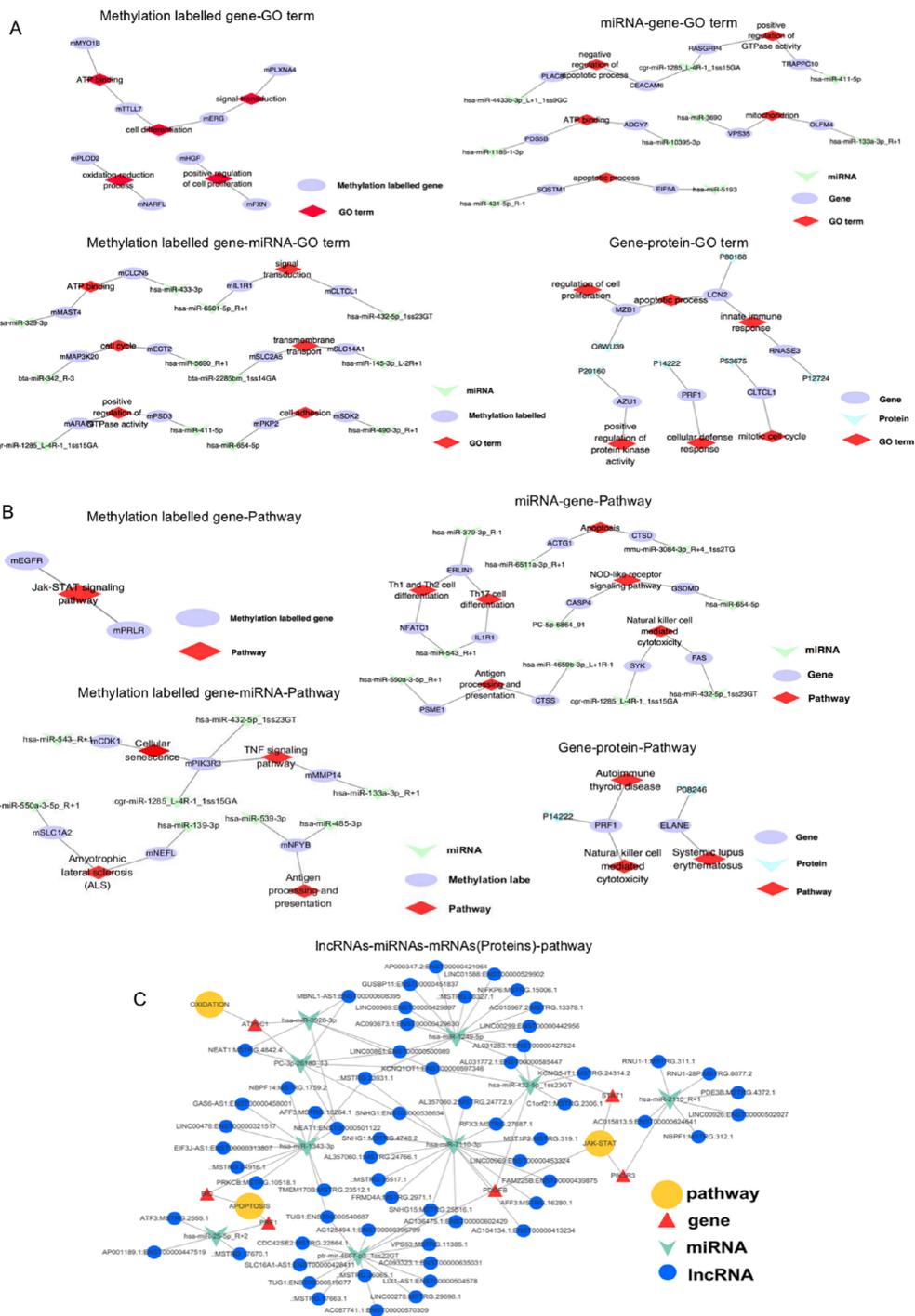
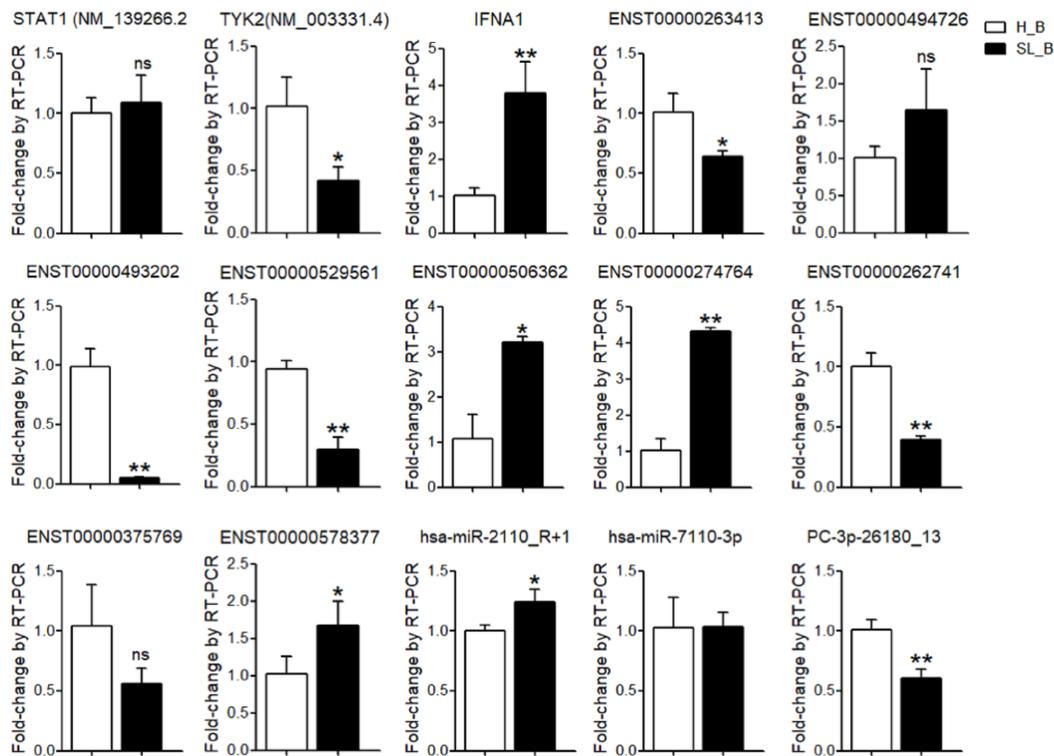


Figure 5

Signaling networks constructed by the represented data in GO and KEGG enrichment of the integrative analysis of multi-omics. (A) GO analysis of “Methylation labeled gene-GO terms”, “miRNA-gene-GO terms”, “Methylation labeled gene-miRNA-GO terms” and “Gene-protein-GO terms”. (B) KEGG enrichment analysis of “Methylation labeled gene-Pathway”, “miRNA-gene-Pathway”, “Methylation labeled gene-miRNA-Pathway” and “Gene-protein-Pathway”. Data was selected from the nodes of signaling network in

integrative analysis of multi-omics (Fig. 4A). The selected network nodes, including miRNAs, genes or proteins were the most significant differential expressed molecules in the KEGG enrichment of multi-omics integrative analysis. (C) Construction of lncRNAs-miRNAs-mRNAs (Proteins)-pathway systematic signaling regulation network. The network was made up of 6 genes (mRNA), 3 pathways, 9 miRNAs and 62 lncRNAs. Yellow circles represent target pathways, red triangles symbolize genes and green arrow stand for miRNAs, blue circles indicate lncRNAs. The grey lines stand for connection between different pathways, genes, miRNAs and lncRNAs.

A



B

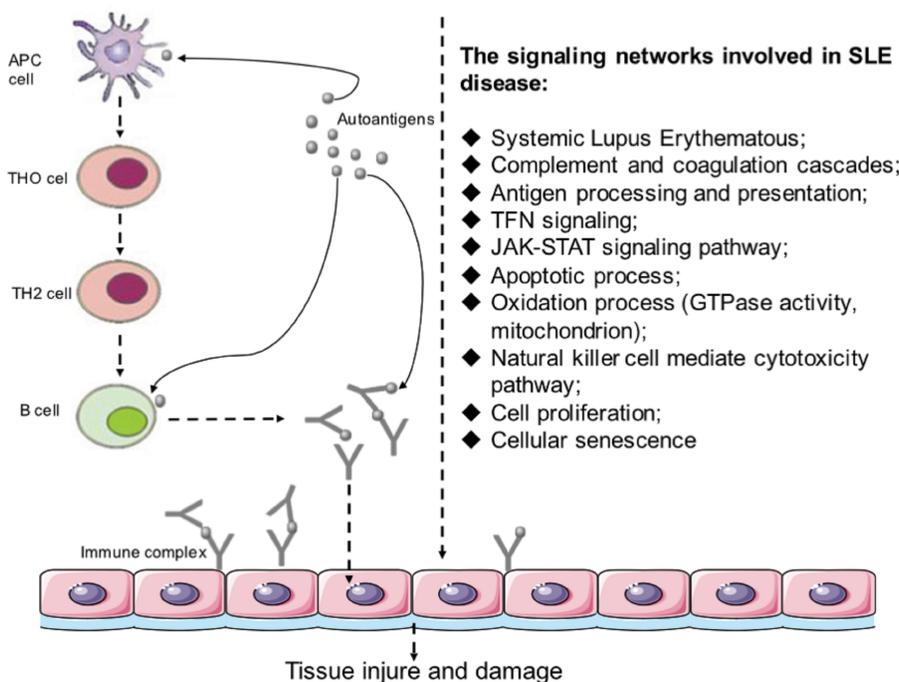


Figure 6

Technical validation of sequencing data and summarization of molecular regulation networks. (A) Twelve genes and 3 non-coding RNAs were analyzed by qRT-PCR, and their relative expression levels were normalized to housekeeping genes such as GAPDH, U6. The data are represented in the way of means \pm SEM. * $p < 0.05$ and ** $p < 0.01$ mean significant difference between SLE and healthy group, while ns means no significance with $p \geq 0.05$. (B) The whole picture of molecular regulation networks in SLE. We summarize a putative molecular regulation network of SLE by utilising integrated analysis of multi-omics.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1JAKSTAT.xlsx](#)
- [TableS2Network.xlsx](#)
- [TableS3pathwaymRNAmiRNAIncRNA.xlsx](#)
- [TableS4apoptosis.xlsx](#)
- [tableS5IntegrativeTablemethylationmRNA.xlsx](#)
- [tableS6circRNAmiRNAmRNA.xlsx](#)
- [tableS7circRNAupmiRNAdownmRNAup.xlsx](#)
- [tableS8SLBVSHBmRNAupload.xlsx](#)
- [tableS9DifferentiallyexpressionmiRNAs.xlsx](#)
- [TableS1011primersandclinicaldata.docx](#)