



Identification of potential biomarkers for systemic lupus erythematosus by integrated analysis of gene expression and methylation data

Wenjing Zhang^{1,2} · Guixin Liang¹ · Huifeng Zhou¹ · Xuedan Zeng¹ · Zhiwen Zhang¹ · Xia Xu³ · Kuan Lai¹

Received: 11 May 2022 / Revised: 7 December 2022 / Accepted: 18 December 2022 / Published online: 3 January 2023
© The Author(s), under exclusive licence to International League of Associations for Rheumatology (ILAR) 2023

Abstract

Introduction Systemic lupus erythematosus (SLE) is a heterogeneous and chronic autoimmune disease. Aberrant DNA methylation occurs during various processes of SLE development regulating the mRNA expression of interrelated genes. This study aims to screen potential DNA methylation markers for SLE.

Methods Gene expression and methylation datasets were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between SLE patients and healthy controls were screened using the limma R package, and differentially methylated positions (DMPs) and regions (DMRs) were identified using dmpfinder and bumpHunter (minfi). Additionally, the DNA methylation markers to distinguish SLE patients from healthy controls were explored through receiver operating characteristic (ROC) curves and logistic regression analyses. Finally, we validated the results of the bioinformatic analysis by pyrosequencing.

Results In total, 91 DEGs, 90,092 DMPs, 15 DMRs, and 13 DMR-associated genes were identified. Through the integrative analysis of DEG- and DMR-associated genes, we identified five type I interferon (IFN)-related genes as key epigenetic-driven genes in SLE. GO enrichment analysis showed that the five SLE-associated epigenetic-driven genes were mainly enriched in the type I IFN signaling pathway involved in immune response and defense response to virus. Moreover, we identified two SLE-specific DNA methylation markers, three SLE without lupus nephritis (SLE-LN⁻)-specific DNA methylation markers, and two SLE with lupus nephritis (SLE-LN⁺)-specific DNA methylation markers by stepwise logistic regression.

Conclusions Overall, our study demonstrates potential DNA methylation markers of SLE, SLE-LN⁻, and SLE-LN⁺, which may help the diagnosis, boost the development of new epigenetic therapy, and contribute to individualized treatment.

Key Points

- This study identified five type I IFN-related genes as key epigenetic-driven genes in SLE, which support the importance of the type I IFN pathway in the pathogenesis of SLE
- We identified novel DNA methylation biomarkers in SLE, SLE-LN⁻, and SLE-LN⁺ by a comprehensive analysis of bioinformatics methods and executed experimental validation, and binary logistic regression analysis showed that they have excellent potential
- These results may provide new insights into the biological mechanisms of SLE, and identify reliable biomarkers for SLE, SLE-LN⁻, and SLE-LN⁺, which may contribute to diagnosis and individualized treatment

Keywords DNA methylation · Epigenetic · Gene expression · Systemic lupus erythematosus · Type I interferon

Wenjing Zhang and Guixin Liang have contributed equally to this work and share the first authorship.

✉ Kuan Lai
lai_kuan@163.com

¹ Department of Dermatology, Nanfang Hospital, Southern Medical University, Baiyun District, No. 1838, North Guangzhou Avenue, Guangzhou 510515, China

² Department of Dermatology, The First Affiliated Hospital of Jinan University, Guangzhou 510632, China

³ Guangzhou Institute of Dermatology, Guangzhou 510030, China

Abbreviations

SLE	Systemic lupus erythematosus
PBMCs	Peripheral blood mononuclear cells
GEO	Gene Expression Omnibus
DEGs	Differentially expressed genes
DMPs	Differentially methylated positions
DMRs	Differentially methylated regions
ROC	Receiver operating characteristic
IFN	Interferon
LN	Lupus nephritis
FDR	False discovery rate
GO	Gene ontology
PPI	Protein–protein interaction network
AUC	Area under the curve

Introduction

Systemic lupus erythematosus (SLE) is a heterogeneous and chronic autoimmune disease characterized by abnormal production of autoantibodies, which results in multiple organ damage. Although the pathogenesis is not known, environmental factors (such as virus infection and exposure to chemicals and radiation), lack of vitamin D, genetic and epigenetic factors, and sex hormones participate in the development of SLE [1, 2]. Epigenetic modifications, primarily including DNA methylation, histone modifications, and non-coding RNAs, generate hereditary alterations in gene expression without changing the DNA coding sequence [3]. The importance of epigenetic dysfunction in complex human diseases has been widely demonstrated in autoimmune diseases [4, 5]. DNA methylation, primarily on CpG dinucleotides, is one of the most studied epigenetic modifications. The best-known function of DNA methylation is to regulate nearby gene expression [6]. Altered DNA methylation in SLE results in immune dysregulation [7]. Aberrant DNA methylation participates in various processes of SLE development by regulating the mRNA expression of interrelated genes. Epigenetic alterations are reversible, and some drugs targeting DNA methylation have already been approved by the United States Food and Drug Administration (FDA), while several other drugs are currently being tested [8]. Epigenetic drugs, in combination with other therapies, may be a novel treatment option for SLE in the future. Thus, we performed an integrated analysis of DNA methylation alterations and mRNA expression differences in the peripheral blood mononuclear cells (PBMCs) of patients with SLE and healthy controls in order to identify novel DNA methylation markers that could serve as potential therapeutic targets for SLE.

Materials and methods

Microarray data

DNA expression and methylation microarray data (GSE81622 and GSE82218) of Chinese SLE patients and corresponding healthy controls were acquired from the GEO database (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov/geo/>) in the National Center for Biotechnology Information (NCBI). The microarray expression data collected were based on an Illumina HumanHT-12 V4.0 expression beadchip (GPL10558). The methylation microarray data collected were based on an Illumina Methylation 450 K array platform (GPL13534), which covered roughly 450,000 CpGs that randomly separate in different gene regions, including TSS1500, TSS200, 5'UTR, 1stExon, body, and 3'UTR. TSS1500 and TSS200 are regions from 201 to 1500 bases and 1 to 200 bases of the upstream of transcriptional start site (TSS), respectively. The “5'UTR (5' untranslated region)” is considered to be the region between the TSS and the first initiation codon. The “1stExon (the first exon)” is one of the most extensively studied translated regions that is generally influenced by methylation status. “Body” stands for the sequence from the first initiation codon to the stop codon of a gene. The “3'UTR (3' untranslated region)” is the area between the stop codon and poly-A tail. TSS1500, TSS200, 5'UTR, and 1st exon were referred to as promoter regions.

Data process

Raw gene expression profiles data were pre-processed by R and Bioconductor packages. After background correction, logarithm transformation and normalization were conducted. The DNA methylation microarrays were normalized and batch-corrected through *minfi*, *impute*, and *wateRmelon* R packages. The mean β value of <0.1 and sex chromosome probes were filtered out.

Differential expression analysis

Differentially expressed genes (DEGs) were screened by using the *limma* package in R (version 4.0.3) [9]. The Benjamini and Hochberg false discovery rate (FDR) [10] was employed to adjust the raw *P*-values for multiple testing. DEGs were screened with *P*-values using a FDR of <0.01 and $\log_2\text{FC} > 1$ as the cutoff criteria. A heatmap based on the significant DEGs was drawn using the R package *pheatmap*.

Differential methylation positions and regions analysis

The methylation levels based on β values were produced for CpG probes ranging from 0 (the CpG site is unmethylated) to 1 (the CpG site is fully methylated) and the differentially

methylated positions (DMPs) associated with each variable in our study using the function `dmpFinder` in `minfi` package [11]. This function tests each genomic position for the association between methylation and a phenotype of interest by testing the linear regression for continuous phenotypes and *F*-test for categorical phenotypes. CpG sites with *P*-values after a FDR of <0.05 were defined as DMPs. Furthermore, differentially methylated regions (DMRs) were predicted using the `Bumphunter` function in `minfi` package. The cutoff value, which is a user-defined numeric value that determines the upper and lower bounds of the genomic profiles that will be used as candidate regions, was set to 0.2, and the number of permutations was set to 250 times. Regions with family-wise error (FWER) <0.5 were identified as significant SLE-associated DMRs. Then, the DMRs were annotated using a web server called `wANNOVAR` (<http://wannovar.wglab.org/>) [12] to prioritize candidate genes in SLE. The intersection analyses were performed through `VennDiagram` in R package.

Functional enrichment analysis and protein–protein interaction (PPI) network construction

Further gene ontology (GO) analysis was conducted with `g:Profiler` [13] using the interaction of screened genes. A STRING online database was used to plot the protein–protein interaction network (PPI).

Clinical sample collection

PBMCs were collected from six patients with SLE [three SLE with lupus nephritis (SLE-LN⁺) and three SLE without lupus nephritis (SLE-LN⁻)] and three age- and gender-matched healthy controls. All patients signed informed consent. This study was approved by the Ethics Review Board at Nanfang Hospital, Southern Medical University (Guangzhou, Guangdong, China).

Pyrosequencing

Genomic DNA was extracted from the PBMCs of six SLE patients and three controls using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The concentration and purity of the DNA samples were determined using a spectrophotometer (Nano-Drop 2000, Thermo Scientific, Waltham, MA, USA). The bisulfite conversion of a total of 500 ng of purified DNA from each sample was performed with the QiagenEpiTect Bisulfite Kit (Qiagen, Cat. #59,104, Hilden, Germany) according to manufacturer's instructions. The bisulfite-converted DNA was amplified with Kapa Biosystems Hs Polymerase with `Dntps250u KK5516` (Kapa, Boston, MA, USA) with the following reaction setup: 34.8 μ L PCR-grade

water, 10 μ L 5 \times KAPA buffer, 1 μ L dNTPs (10 mM/each), 1 μ L forward primer (50 pM/ μ L), 1 μ L reverse primer (50 pM/ μ L), 2 μ L template DNA (25 ng), and 0.2 μ L Taq polymerase (5 U/ μ L) in a total reaction volume of 50 μ L. Each reaction was run on the following thermal cycle conditions: one cycle of an initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, and extension at 72 °C for 1 min, with a final extension at 72 °C for 7 min, and hold at 4 °C. The amplicons were then subjected to pyrosequencing with PyroMark Q96/48 ID (Qiagen, Hilden, Germany). Primers for pyrosequencing assays were designed by PyroMark Assay Design software (version 2.0 Qiagen). All primers used are presented in Supplementary data1.

Statistical analysis

Correlations were determined using the Pearson or Spearman test. Logistic regression was used to calculate probabilities and corresponding 95% confidence intervals (CIs). Receiver operating characteristic (ROC) curves were constructed, and the area under the curve (AUC) was used to evaluate the value of DMPs as biomarkers for SLE. A stepwise logistic regression model was used to select methylation diagnostic markers. A binary logistic regression model was selected to fit the ROC curve. All statistical analyses were conducted using SPSS 23.0 (SPSS Inc., Chicago, IL, USA). Visualizations were performed using GraphPad Prism 7 (GraphPad Software Inc., San Diego, CA, USA).

Results

Identification of DEGs in PBMCs isolated from patients with SLE

The gene expression profile data of 30 patients with SLE (including 15 SLE-LN⁻ and 15 SLE-LN⁺) and 25 healthy controls were obtained from the GEO database. After standardization and gene annotation, a total of 91 DEGs were obtained between SLE and healthy samples, of which we identified 48 upregulated mRNAs and 43 downregulated mRNAs. The volcano map is shown in Fig. 1A, and the expression heatmap of the DEGs is shown in Fig. 1B.

Identification of DMPs in PBMCs isolated from patients with SLE

The distribution of the DNA methylation levels in each sample before and after normalization is shown in Fig. 2A and B. The methylation levels of all of the target positions in the

array demonstrated the known bimodal distribution in SLE and healthy controls (Fig. 2C). Following the post-quality control steps, methylation data for 365,793 CpGs were analyzed using DMPfinder to identify the DMPs associated with SLE. In total, statistically significant differences in 90,092 CpG positions were found in the GSE82218 dataset. All DMPs from each autosomal chromosome are shown in the Manhattan chart in Fig. 2E, indicating that all of the DMPs were evenly distributed throughout the chromosomal complement. In order to specifically evaluate the methylation patterns and preferential genome location, we then examined the distribution of all of the significant DMPs with a q value $< 1 \times 10^{-5}$ ($n = 1184$) in various functional genomic regions. As shown in Fig. 2D, in assessing the six categories according to the Illumina annotations, we found that 44.49% of significant DMPs occurred in the promoter regions and 49.85% of significant DMPs were found in the gene body.

Identification of DMRs in PBMCs isolated from patients with SLE

Bumphunter analysis showed significant differences in DNA methylation in 15 regions (Supplementary data 2). Overall, we annotated the DMRs using wANNOVAR

and obtained 13 DMR-associated genes, including *PARP9*, *IFI44L*, *MX1*, *CMPK2*, *PDE7A*, *IFI44*, *IFIT1*, *FANCC*, *IFIT3*, *NLRC5*, *PLSCR1*, *USP18*, and *ODF3B*. GO analysis was performed with these genes. As shown in Fig. 3A, 24 biological processes were recognized with an adjusted P -value < 0.05 . These biological processes were mainly enriched in response to type I interferon signaling pathway, immune response, and defense response to a virus. The PPI network of the 13 genes was established using the STRING database (Fig. 3B).

Identification of epigenetic-driven genes in SLE

We then assessed the intersection of DEGs and DMR-associated genes and identified five genes (*IFI44*, *IFI44L*, *IFIT1*, *MX1*, and *USP18*) as key epigenetic-driven genes in SLE (Fig. 4A), identified 16 significant DMPs which mapped to the five genes based on a nominal P -value cutoff of $< 1 \times 10^{-5}$, including cg07107453 and cg01079652 (*IFI44*), cg05696877, cg17980508, cg00458211, cg00855901, cg13304609, cg03607951 and cg06872964 (*IFI44L*), cg05552874 (*IFIT1*), cg16785077, cg08924203, cg22862003, cg21549285, and cg26312951 (*MX1*), and

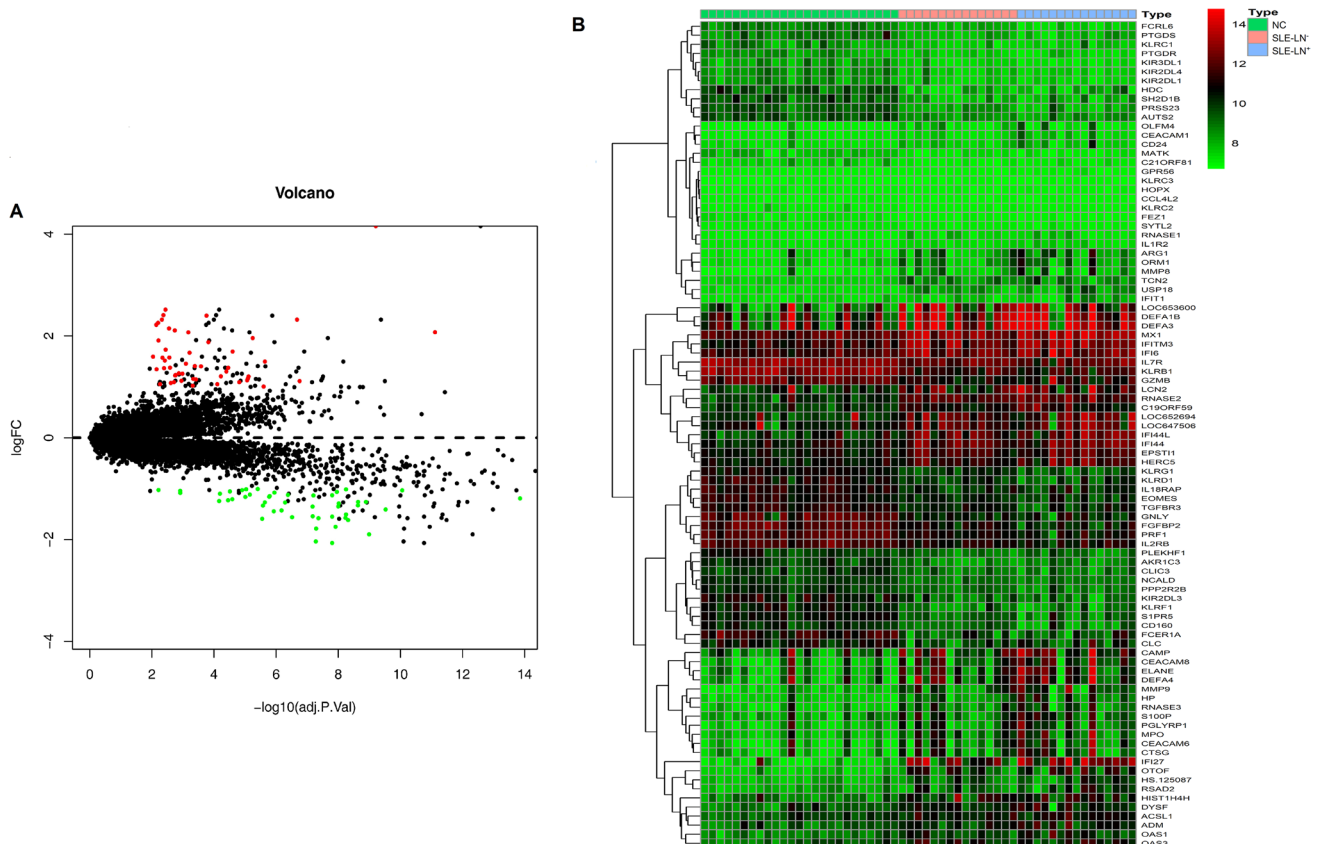


Fig. 1 Identification of DEGs between SLE and healthy controls samples. **A** A volcano plot of DEGs. The red points indicate genes with high expression in SLE, and the green points indicate genes with

low expression according to the threshold of the FDR. **B** An expression heatmap of DEGs. NC, normal controls; SLE-LN⁻, SLE without lupus nephritis; SLE-LN⁺, SLE with lupus nephritis

cg14293575 (*USP18*). GO analysis was performed with the five SLE-associated epigenetic-driven genes (*IFI44*, *IFI44L*, *IFIT1*, *MX1*, and *USP18*). As shown in Fig. 4B, 10 biological processes were recognized with an adjusted *P*-value < 0.05. They were mainly enriched in type I interferon signaling pathway, immune response, and defense response to the virus. The PPI network of the SLE-associated epigenetic-driven genes was established using the STRING database (Fig. 4C). Clustering analysis based on the identified genes and DMPs found on the identified genes discriminated most SLE patient samples from the controls (Figs. 4D, E).

DNA hypomethylation related to the overexpression of SLE-associated epigenetic-driven genes

In order to explore the effect of gene methylation on expression, we analyzed the correlation of the five SLE-associated epigenetic-driven gene expressions and the methylation levels of the significant DMPs which mapped to the five genes in 25 healthy samples and 30 SLE samples by Pearson or Spearman correlation analysis. The results showed that cg07107453 and cg01079652 hypomethylation was related to the overexpression of *IFI44*. Hypomethylation of cg05696877, cg17980508, cg00458211, cg00855901, cg13304609, cg03607951, and cg06872964 was related to the overexpression of *IFI44L*. cg05552874 hypomethylation was related to the overexpression of *IFIT1*. Hypomethylation of cg16785077, cg08924203, cg22862003, cg21549285, and cg26312951 was related to the overexpression of *MX1*, and cg14293575 hypomethylation was related to the overexpression of *USP18* (Supplementary data 3).

Assessment of diagnostic efficacy on screened methylation sites

In order to further explore the diagnostic value of each screened methylation site, the AUC was calculated based on the DNA methylation data of 30 patients with SLE (including 15 SLE-LN⁻ and 15 SLE-LN⁺) and 25 healthy controls (Supplementary data 4). The AUC on the 16 DMPs were up to 0.9, which meant that these DNA methylation alterations could have significant characteristics to distinguish SLE, SLE-LN⁻, and SLE-LN⁺ from control subjects.

Subsequently, a discrimination model using the DMP panel was constructed to predict the probability of diagnosis for SLE, SLE-LN⁻, and SLE-LN⁺. The stepwise logistic regression model showed that cg05696877 (*IFI44L*) and cg03607951 (*IFI44L*) could be used as a panel of biomarkers to distinguish SLE from healthy controls. Binary logistic regression analysis found that the ROC curves of the two methylation markers revealed an AUC value of 0.999 (95% CI: 0.994–1.000, sensitivity 96.67%, specificity 100%) (Fig. 5A). cg21549285 (*MX1*), cg05696877 (*IFI44L*), and cg07107453 (*IFI44*) were

able to be used as a panel of biomarkers to distinguish SLE-LN⁻ from healthy controls. The ROC curves of the three methylation markers revealed an AUC value of 1.000 (95% CI: 1.000–1.000, sensitivity 100%, specificity 100%) (Fig. 5B). cg22862003 (*MX1*) and cg05696877 (*IFI44L*) were also able to be used as a panel of biomarkers to distinguish SLE-LN⁺ from healthy controls. The ROC curves of the two methylation markers revealed an AUC value of 0.997 (95% CI: 0.989–1.000, sensitivity 100%, specificity 96%) (Fig. 5C). The AUC on these DMP panels were up to 0.99, which meant that these panels were significant characteristics that could distinguish SLE, SLE-LN⁻, and SLE-LN⁺ from control subjects.

The DNA methylation levels of the DMPs mentioned above were further confirmed by pyrosequencing (Fig. 6). Compared with healthy controls, the methylation levels of cg05696877 (*IFI44L*) and cg03607951 (*IFI44L*) decreased in SLE, while the levels of cg21549285 (*MX1*), cg05696877 (*IFI44L*), and cg07107453 (*IFI44*) decreased in SLE-LN⁻, and the levels of cg22862003 (*MX1*) and cg05696877 (*IFI44L*) decreased in SLE-LN⁺. The results showed that the methylation level of key methylation markers may be closely associated with the diagnosis of SLE, SLE-LN⁻, and SLE-LN⁺.

Discussion

SLE is a systemic autoimmune disease characterized by the immune complex accumulation in blood vessels and connective tissue, with a complicated etiology [14]. Type I IFNs are one of the main pathogenic factors of SLE and have been shown to be related with other autoimmune diseases, such as Sjögren's syndrome, systemic sclerosis, rheumatoid arthritis, myositis, and virus-infected diseases [15–17]. SLE patients with elevated type I IFN levels are prone to nephritis and other serious complications [18, 19]. Type I IFNs act on renal resident cells and are involved in the entire process of renal injury [20]. Gene expression profile data demonstrated that the expression of type I IFN-regulated gene was elevated in the PBMCs of SLE patients and was termed a type I IFN signature [18], which was positively related to the clinical and laboratory indicators of disease activities [19]. In this study, five type I IFN-related genes were identified, including *IFI44*, *IFI44L*, *IFIT1*, *MX1*, and *USP18*, as key epigenetic-driven genes in SLE through the integrative analysis of DEG- and DMR-associated genes. GO enrichment analysis showed that these genes were mainly enriched in the type I IFN signaling pathway involved in immune response and defense response to a virus, which was consistent with previous studies [21–23].

IFI44 is a gene regulated by type I IFNs. In SLE patients, the expression of *IFI44* in PBMCs is related to IFN- α level in plasma [24]. *IFI44L* is a paralog gene of *IFI44*, considered to be an indicator of IFN signature, upregulated in the

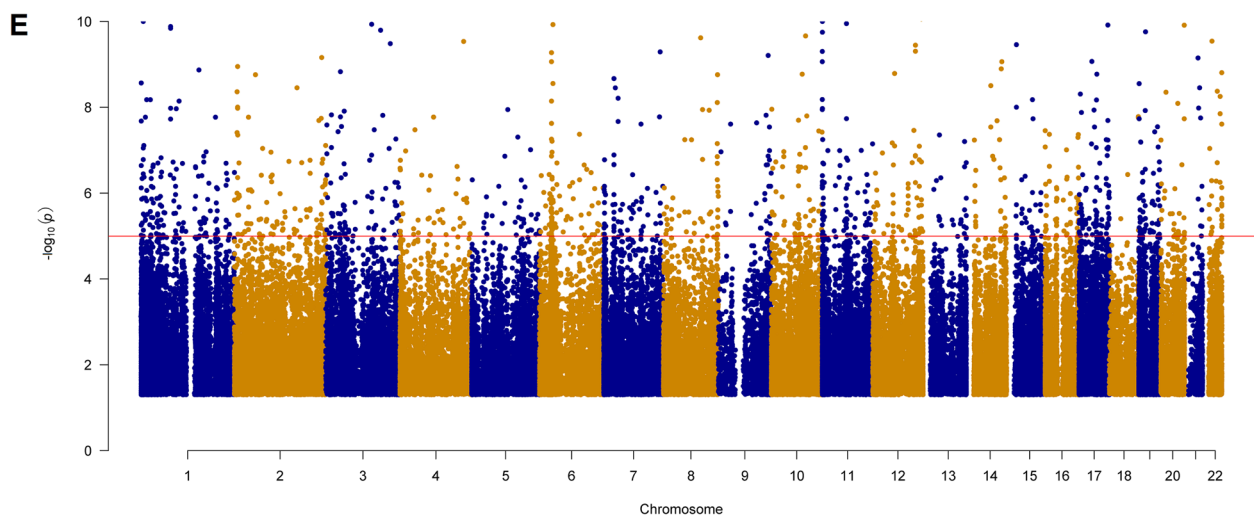
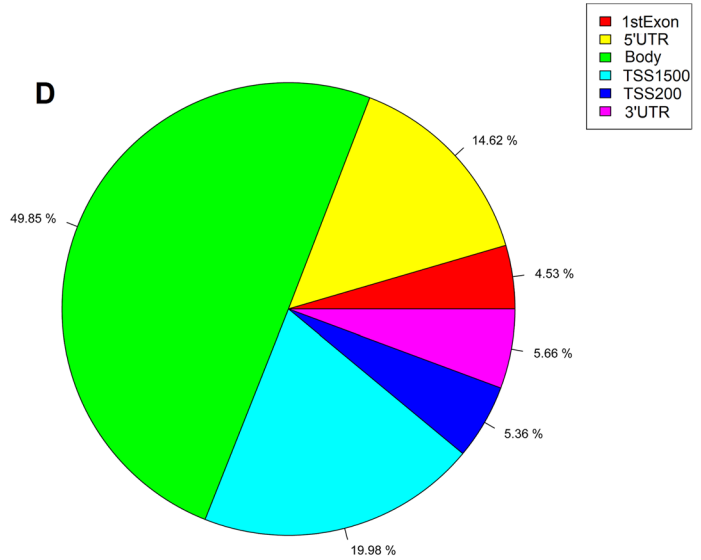
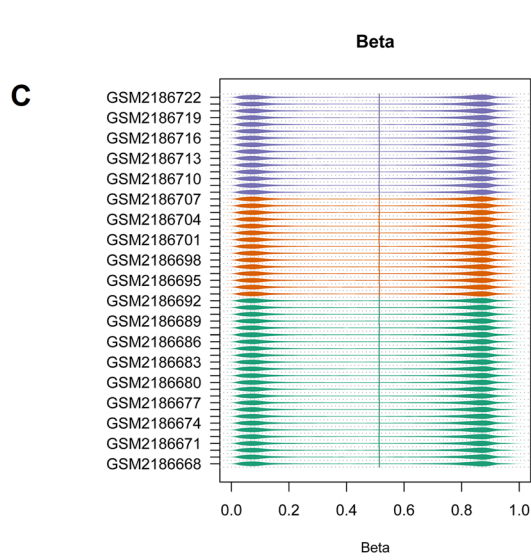
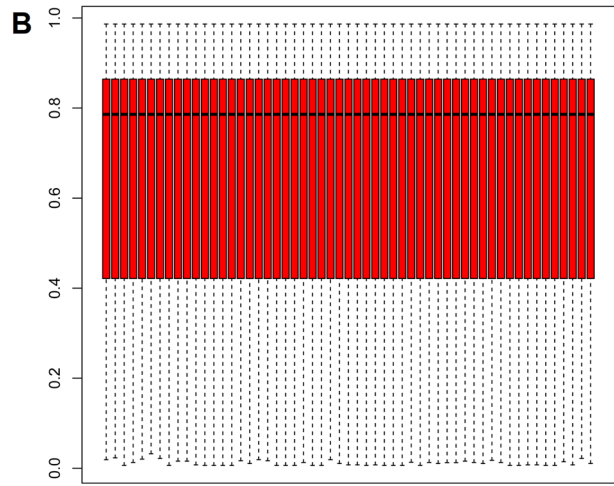
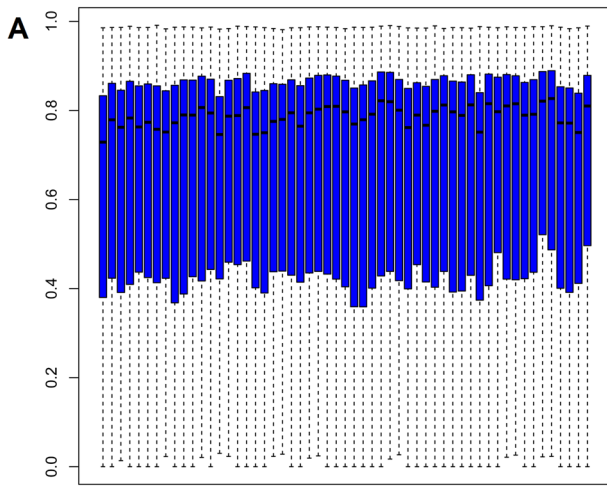


Fig. 2 Identification of differentially methylated position (DMPs) between SLE and healthy controls samples. **A, B** Distribution of DNA methylation levels in each sample before and after normalization. **C** The methylation levels of all target sites in the array demonstrated the known bimodal distribution in SLE and healthy controls. **D** This figure shows the intragenic DMP distribution. Approximately 50% of the DMPs are located in the gene body and 50% are located in the promoter regions. **E** A Manhattan plot of all DMPs. The red line shows the significant DMPs with a q value $< 1 \times 10^{-5}$

peripheral blood and synovial tissue of patients with SLE [25, 26]. *IFIT1* is induced by type I IFN treatment or viral infection and is an effector molecule that restricts viral translation [27]. *IFIT1* may interact with Rho/Rac guanine nucleotide exchange factor and participate in SLE immune response [28]. *MX1* is a key downstream gene of type I IFN and plays a role in mediating the IFN-induced antiviral response against a large variety of viruses [29]. The expression of *MX1* in PBMCs has been used as a sensitive biomarker for the efficacy of LN therapy [30]. Liu et al. showed that *USP18* is necessary for Th17 differentiation and autoimmune response. *USP18* inhibits ubiquitination of the TAK1–TAB complex, thereby restricting IL-2 production and promoting IL-17 production, indicating that *USP18* may be targeted to treat autoimmune diseases [31]. Furthermore, it was found that the expression of *USP18* in peripheral blood B cells was significantly upregulated in SLE patients compared with healthy controls [32]. Our study found that the increased expression of *IFI44*, *IFI44L*, *IFIT1*, *MX1*, and *USP18*, in the PBMCs of SLE patients was associated with one or more hypomethylated CpG sites in the gene regulatory region. This indicated that these methylation sites may be involved in the pathogenesis of SLE by regulating the expression of these genes.

The available diagnostic markers for SLE, including anti-nuclear antibodies (ANAs), anti-double-stranded DNA (dsDNA) antibodies, and anti-Smith (Anti-SM) antibodies, have significant limitations. ANA exhibit a high sensitivity (100%), but a relatively low specificity (65%) [33]. Anti-dsDNA and anti-SM antibodies are highly specific in identifying most of the patients with suspected SLE (both are 99%); however, the sensitivity is low (30.2% vs 25.9%, respectively) [34]. Therefore, it is still important to explore novel diagnostic markers for SLE. Lupus nephritis (LN) is one of the most serious complications of SLE and has a high morbidity and mortality affecting approximately 40% of patients with lupus. Approximately 10% of patients with LN will develop end-stage renal disease [35]. Thus, early diagnosis is beneficial to improve the prognosis of LN. Currently, diagnosis and monitoring of patients with LN mainly rely on renal biopsy [36], which could cause damage to the health of the patient. Numerous studies have shown that DNA methylation positions can be used as efficient and non-invasive biomarkers for cancer and autoimmune diseases [25,

37–39]. Coit et al. [40] detected four hypomethylation CpG sites in *IFI44L* (cg05696877, cg06872964, cg13304609, and cg03607951) and one in *IFI44* (cg01079652) in the neutrophils from women patients with SLE. Xie et al. [41] have found five hypomethylation CpG sites in *IFI44L* (cg06872964, cg00458211, cg13304609, cg03607951, and cg05696877) and one in *IFI44* (cg07107453) in the PBMCs from patients with SLE by integrating the chip data. In this study, we identified two CpG sites in *IFI44L* (cg05696877 and cg03607951) and verified the hypomethylation status in PBMCs of SLE patients; it suggested that the methylation levels of cg05696877 and cg03607951 in the PBMCs may be useful biomarkers for SLE. A good diagnostic biomarker should be highly sensitive and specific for a disease. Zhao et al. [25] showed by examining the methylation status of two CpG sites within the *IFI44L* promoter, Chr1: 79 085 222 and Chr1: 79 085 250 (cg06872964), that the hypomethylation of the *IFI44L* promoter region has excellent AUC (0.968 and 0.982), sensitivity (91.5% and 93.6%), and specificity (95.3% and 96.4%) for diagnosing SLE. The methylation levels of these two CpG sites were significantly lower in SLE patients with renal damage compared to those without renal damage. A study by Karimifar et al. [42] showed that the DNA methylation level of the *IFI44L* promoter could distinguish patients with SLE from healthy individuals with a AUC of 0.639, sensitivity of 91.8%, and specificity of 38% but was not statistically significant between SLE patients with renal involvement and without renal involvement. The lack of significance may be due to the small sample size. In the present study, a combination of cg05696877 (*IFI44L*) and cg03607951 (*IFI44L*) from the logistic regression model demonstrated high AUC (0.999), sensitivity (96.67%), and specificity (100%). Compared to the two studies above, we found higher AUC, sensitivity, and specificity. In addition, we identified that cg21549285 (*MX1*), cg05696877 (*IFI44L*), and cg07107453 (*IFI44*) were able to be used as a panel of biomarkers to distinguish SLE-LN⁻ from healthy controls with high AUC (1.000), sensitivity (100%), and specificity (100%). cg22862003 (*MX1*) and cg05696877 (*IFI44L*) were able to be used as a panel of biomarkers to distinguish SLE-LN⁺ from healthy controls with high AUC (0.997), sensitivity (100%), and specificity (96%), which have never been reported before. Consistent with the conclusion made by Zhu et al. [43] who published the original datasets, we have found five key epigenetic-driven genes in SLE which enriched in type I IFN signaling pathway, identified reliable DNA methylation biomarkers from these genes for SLE, SLE-LN⁻ and SLE-LN⁺. Our study had some limitations. First, information on confounding factors such as disease severity, autoantibodies titers, and HLA genotyping was not available and hence may have contributed to the bias of the outcomes. Second, the analytical data were derived from public databases, with a limited number of cases. Sample data for validation

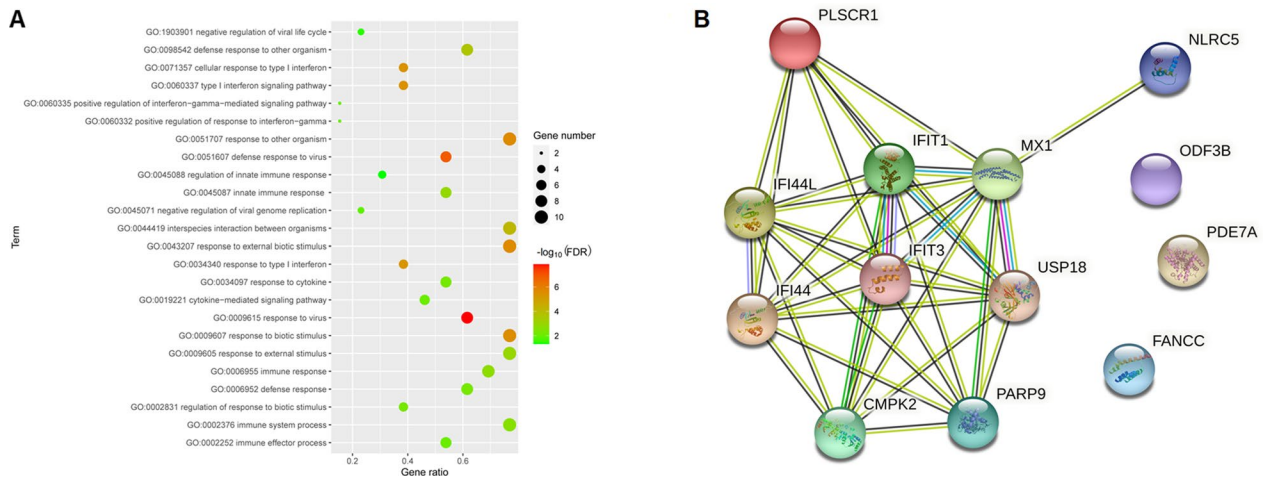


Fig. 3 Functional analysis of the 13 DMR-associated genes in patients with SLE. **A** Enriched GO terms of the 13 DMR-associated genes in the “biological process” category. **B** PPI network of the 13 DMR-associated genes in patients with SLE

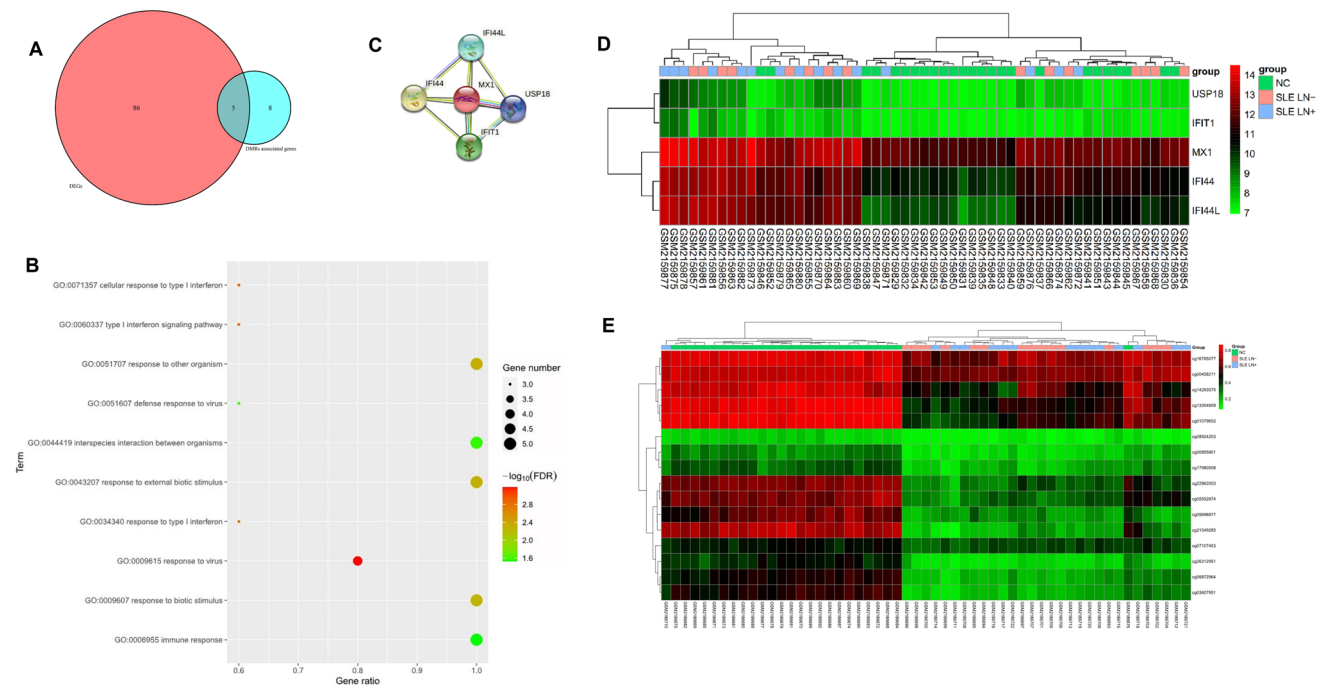


Fig. 4 Epigenetic-driven genes in SLE. **A** A Venn diagram showing the number of key epigenetic-driven genes in SLE. **B** Enriched GO terms of key epigenetic-driven genes in the “biological process” category. **C** PPI network of the key epigenetic-driven genes in SLE. **D** A heatmap of the key epigenetic-driven genes among all PBMC samples. Each column represents a sample, each row represents the gene expression level of all of the samples, and the sample clustering

tree appears at the top. **E** A heatmap of the DMPs on key epigenetic-driven genes among all PBMC samples. Each column represents a sample, each row represents the methylation level of all of the samples involved on one CpG site, and the sample clustering tree appears at the top. NC, normal controls; SLE-LN⁻, SLE without lupus nephritis; SLE-LN⁺, SLE with lupus nephritis

experiments was also limited, so it would be necessary to validate the current data in a larger clinical cohorts. Third, the gene expression and DNA methylation profiling were

performed in PBMCs, while different cell subsets in peripheral blood may have different transcription and methylation profiles during the development of the diseases.

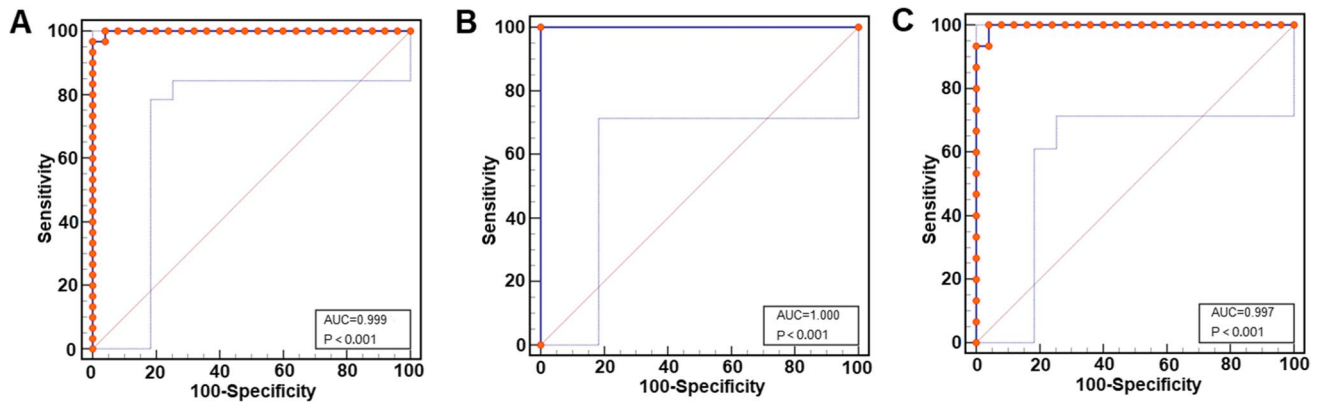


Fig. 5 ROC curves of the DNA methylation levels for methylation markers. **A** The ROC curve of a two-DMP panel for patients with SLE, the calculation was based on the DNA methylation data of 30 patients with SLE and 25 healthy controls. **B** The ROC curve of a three-DMP panel for patients with SLE without lupus nephritis, the

calculation was based on the DNA methylation data of 15 SLE-LN⁻ and 25 healthy controls. **C** The ROC curve of a two-DMP panel for patients with SLE with lupus nephritis, the calculation was based on the DNA methylation data of 15 SLE-LN⁺ and 25 healthy controls

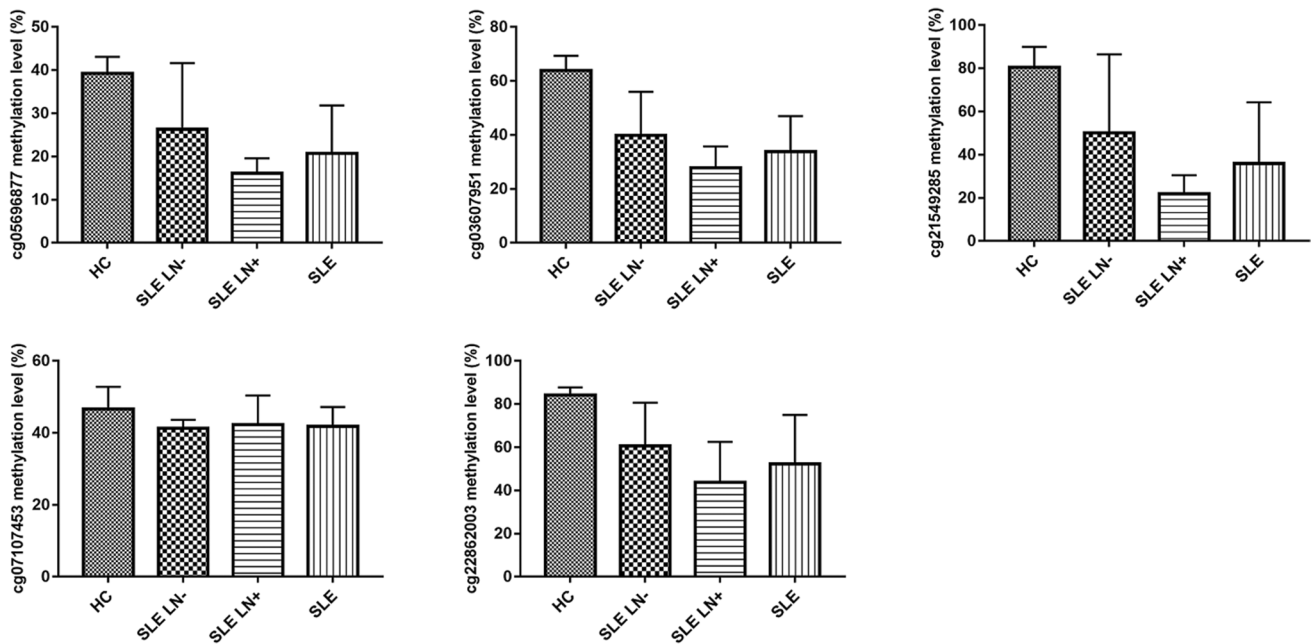


Fig. 6 Pyrosequencing validation. Methylation levels of cg05696877, cg03607951, cg21549285, cg07107453, and cg22862003 were measured in 6 SLE [3 SLE without lupus nephritis (SLE-LN⁻) and 3 SLE with lupus nephritis (SLE-LN⁺)] patients and 3 healthy controls (HC)

Conclusions

In summary, our study identified five type I IFN-related genes as key epigenetic-driven genes in SLE, which support the importance of the type I IFN pathway in SLE pathogenesis. Moreover, we identified novel DNA methylation biomarkers for SLE, SLE-LN⁻, and SLE-LN⁺ by a comprehensive analysis of bioinformatics methods and executed experimental validation, and binary logistic regression analysis showed

they have excellent potential. These results may provide new insights into the biological mechanisms of SLE and identify reliable biomarkers for SLE, SLE-LN⁻, and SLE-LN⁺, which may contribute to individualized treatment.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10067-022-06495-3>.

Acknowledgements We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Author contribution WJZ, ZWZ, HFZ, and KL participated in the conception and design, KL and XX provided the administrative support, HFZ, XDZ, GXL, and ZWZ participated in the provision of study materials or patients, WJZ participated in the collection and assembly of data, and data analysis and interpretation. All authors participated in manuscript writing and final approval of manuscript.

Funding This work was supported by the National Natural Science Foundation of China [grant number 82073440] and the Medical Science and Technology Research Fund project of Guangdong Province [grant number A2019180].

Data availability The data that support the findings of this study are openly available in GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

Declarations

Ethics approval and consent to participate This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Nanfang Hospital, Southern Medical University (Guangzhou, Guangdong, China).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publication All authors consent to the publication of this work.

Competing interests The authors declare no competing interests.

References

- Gatto M, Zen M, Ghirardello A, Bettio S, Bassi N, Iaccarino L et al (2013) Emerging and critical issues in the pathogenesis of lupus. *Autoimmun Rev* 12(4):523–536. <https://doi.org/10.1016/j.autrev.2012.09.003>
- AlSaleem A, AlE'Ed A, AlSaghier A, Al-Mayouf SM (2015) Vitamin D status in children with systemic lupus erythematosus and its association with clinical and laboratory parameters. *Clin Rheumatol* 34(1):81–84. <https://doi.org/10.1007/s10067-014-2811-z>
- Jiang J, Zhao M, Chang C, Wu H, Lu Q (2020) Type I interferons in the pathogenesis and treatment of autoimmune diseases. *Clin Rev Allergy Immunol* 59(2):248–272. <https://doi.org/10.1007/s12016-020-08798-2>
- Jeffries MA, Sawalha AH (2015) Autoimmune disease in the epigenetic era: how has epigenetics changed our understanding of disease and how can we expect the field to evolve? *Expert Rev Clin Immunol* 11(1):45–58. <https://doi.org/10.1586/1744666X.2015.994507>
- Gupta B, Hawkins RD (2015) Epigenomics of autoimmune diseases. *Immunol Cell Biol* 93(3):271–276. <https://doi.org/10.1038/icb.2015.18>
- Mo XB, Zhang YH, Lei SF (2021) Integrative analysis identifies potential causal methylation-mRNA regulation chains for rheumatoid arthritis. *Mol Immunol* 131:89–96. <https://doi.org/10.1016/j.molimm.2020.12.021>
- Hedrich CM, Mäbert K, Rauen T, Tsokos GC (2017) DNA methylation in systemic lupus erythematosus. *Epigenomics-UK* 9(4):505–525. <https://doi.org/10.2217/epi-2016-0096>
- Jones PA, Issa JP, Baylin S (2016) Targeting the cancer epigenome for therapy. *Nat Rev Genet* 17(10):630–641. <https://doi.org/10.1038/nrg.2016.93>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47. <https://doi.org/10.1093/nar/gkv007>
- Glickman ME, Rao SR, Schultz MR (2014) False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol* 67(8):850–857. <https://doi.org/10.1016/j.jclinepi.2014.03.012>
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD et al (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10):1363–1369. <https://doi.org/10.1093/bioinformatics/btu049>
- Yang H, Wang K (2015) Genomic variant annotation and prioritization with ANNOVAR and WANNOVAR. *Nat Protoc* 10(10):1556–1566. <https://doi.org/10.1038/nprot.2015.105>
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H et al (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47(W1):W191–W198. <https://doi.org/10.1093/nar/gkz369>
- Rahman A, Isenberg DA (2008) Systemic lupus erythematosus. *N Engl J Med* 358(9):929–939. <https://doi.org/10.1056/NEJMra071297>
- Zhang B, Liu L, Zhou T, Shi X, Wu H, Xiang Z et al (2020) A simple and highly efficient method of IFI44L methylation detection for the diagnosis of systemic lupus erythematosus. *Clin Immunol* 221:108612. <https://doi.org/10.1016/j.clim.2020.108612>
- Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C et al (2011) Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. *Ann Rheum Dis* 70(11):2029–2036. <https://doi.org/10.1136/ard.2011.150326>
- Mavragani CP, Crow MK (2010) Activation of the type I interferon pathway in primary Sjogren's syndrome. *J Autoimmun* 35(3):225–231. <https://doi.org/10.1016/j.jaut.2010.06.012>
- Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ et al (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* 100(5):2610–2615. <https://doi.org/10.1073/pnas.0337679100>
- Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J et al (2003) Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 197(6):711–723. <https://doi.org/10.1084/jem.20021553>
- Ding X, Ren Y, He X (2021) IFN-I mediates lupus nephritis from the beginning to renal fibrosis. *Front Immunol* 12:676082. <https://doi.org/10.3389/fimmu.2021.676082>
- Kong J, Li L, Zhimin L, Yan J, Ji D, Chen Y et al (2019) Potential protein biomarkers for systemic lupus erythematosus determined by bioinformatics analysis. *Comput Biol Chem* 83:107135. <https://doi.org/10.1016/j.compbiolchem.2019.107135>
- Rönnblom L, Alm GV, Eloranta ML (2011) The type I interferon system in the development of lupus. *Semin Immunol* 23(2):113–121. <https://doi.org/10.1016/j.smim.2011.01.009>
- Sim TM, Ong SJ, Mak A, Tay SH (2022) Type I interferons in systemic lupus erythematosus: a journey from bench to bedside. *Int J Mol Sci* 23(5). <https://doi.org/10.3390/ijms23052505>
- Hua J, Kirou K, Lee C, Crow MK (2006) Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies. *Arthritis Rheum* 54(6):1906–1916. <https://doi.org/10.1002/art.21890>
- Zhao M, Zhou Y, Zhu B, Wan M, Jiang T, Tan Q et al (2016) IFI44L promoter methylation as a blood biomarker for systemic lupus erythematosus. *Ann Rheum Dis* 75(11):1998–2006. <https://doi.org/10.1136/annrheumdis-2015-208410>
- Nzeusseu TA, Galant C, Theate I, Maudoux AL, Lories RJ, Houssiau FA et al (2007) Identification of distinct gene expression profiles in the synovium of patients with systemic lupus

- erythematosus. *Arthritis Rheum* 56(5):1579–1588. <https://doi.org/10.1002/art.22578>
27. Diamond MS (2014) IFIT1: a dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine Growth Factor Rev* 25(5):543–550. <https://doi.org/10.1016/j.cytogfr.2014.05.002>
 28. Ye S, Pang H, Gu YY, Hua J, Chen XG, Bao CD et al (2003) Protein interaction for an interferon-inducible systemic lupus associated gene, IFIT1. *Rheumatology (Oxford)* 42(10):1155–1163. <https://doi.org/10.1093/rheumatology/keg315>
 29. Haller O, Kochs G (2011) Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. *J Interferon Cytokine Res* 31(1):79–87. <https://doi.org/10.1089/jir.2010.0076>
 30. Feng X, Wu H, Grossman JM, Hanvivadhanakul P, FitzGerald JD, Park GS et al (2006) Association of increased interferon-inducible gene expression with disease activity and lupus nephritis in patients with systemic lupus erythematosus. *Arthritis Rheum* 54(9):2951–2962. <https://doi.org/10.1002/art.22044>
 31. Liu X, Li H, Zhong B, Blonska M, Gorjestani S, Yan M et al (2013) USP18 inhibits NF- κ B and NFAT activation during Th17 differentiation by deubiquitinating the TAK1-TAB1 complex. *J Exp Med* 210(8):1575–1590. <https://doi.org/10.1084/jem.20122327>
 32. Fan H, Zhao G, Ren D, Liu F, Dong G, Hou Y (2017) Gender differences of B cell signature related to estrogen-induced IFI44L/BAFF in systemic lupus erythematosus. *Immunol Lett* 181:71–78. <https://doi.org/10.1016/j.imlet.2016.12.002>
 33. Solomon DH, Kavanaugh AJ, Schur PH (2002) Evidence-based guidelines for the use of immunologic tests: antinuclear antibody testing. *Arthritis Rheum* 47(4):434–444. <https://doi.org/10.1002/art.10561>
 34. Flechsig A, Rose T, Barkhudarova F, Strauss R, Klotsche J, Dähnrich C et al (2017) What is the clinical significance of anti-Sm antibodies in systemic lupus erythematosus? A comparison with anti-dsDNA antibodies and C3. *Clin Exp Rheumatol* 35(4):598–606
 35. Morales E, Galindo M, Trujillo H, Praga M (2021) Update on lupus nephritis: looking for a new vision. *Nephron* 145(1):1–13. <https://doi.org/10.1159/000511268>
 36. Davidson A, Aranow C, Mackay M (2019) Lupus nephritis: challenges and progress. *Curr Opin Rheumatol* 31(6):682–688. <https://doi.org/10.1097/BOR.0000000000000642>
 37. Xu W, Xu M, Wang L, Zhou W, Xiang R, Shi Y et al (2019) Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. *Signal Transduct Target Ther* 4:55. <https://doi.org/10.1038/s41392-019-0081-6>
 38. Zhao X, Zhang L, Wang J, Zhang M, Song Z, Ni B et al (2021) Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. *J Transl Med* 19(1):35. <https://doi.org/10.1186/s12967-020-02698-x>
 39. Song G, Zhu X, Xuan Z, Zhao L, Dong H, Chen J et al (2021) Hypermethylation of GNA14 and its tumor-suppressive role in hepatitis B virus-related hepatocellular carcinoma. *Theranostics* 11(5):2318–2333. <https://doi.org/10.7150/thno.48739>
 40. Coit P, Yalavarthi S, Ogenovskii M, Zhao W, Hasni S, Wren JD et al (2015) Epigenome profiling reveals significant DNA demethylation of interferon signature genes in lupus neutrophils. *J Autoimmun* 58:59–66. <https://doi.org/10.1016/j.jaut.2015.01.004>
 41. Xie S, Zeng Q, Ouyang S, Liang Y, Xiao C (2021) Bioinformatics analysis of epigenetic and SNP-related molecular markers in systemic lupus erythematosus. *Am J Transl Res* 13(6):6312–6329
 42. Karimifar M, Pakzad B, Karimzadeh H, Mousavi M, Kazemi M, Salehi A et al (2019) Interferon-induced protein 44-like gene promoter is differentially methylated in peripheral blood mononuclear cells of systemic lupus erythematosus patients. *J Res Med Sci* 24:99. https://doi.org/10.4103/jrms.JRMS_83_19
 43. Zhu H, Mi W, Luo H, Chen T, Liu S, Raman I et al (2016) Whole-genome transcription and DNA methylation analysis of peripheral blood mononuclear cells identified aberrant gene regulation pathways in systemic lupus erythematosus. *Arthritis Res Ther* 18:162. <https://doi.org/10.1186/s13075-016-1050-x>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.