

## Research Article

# Integration of Genome-Wide DNA Methylation and Transcription Uncovered Aberrant Methylation-Regulated Genes and Pathways in the Peripheral Blood Mononuclear Cells of Systemic Sclerosis

Honglin Zhu,<sup>1,2</sup> Chengsong Zhu,<sup>2</sup> Wentao Mi,<sup>2</sup> Tao Chen,<sup>2</sup> Hongjun Zhao <sup>1,2</sup>,  
Xiaoxia Zuo,<sup>1</sup> Hui Luo <sup>1,2</sup> and Quan-Zhen Li <sup>1,2</sup>

<sup>1</sup>Department of Rheumatology & Immunology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008, China

<sup>2</sup>Department of Immunology and Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390, USA

Correspondence should be addressed to Quan-Zhen Li; [quan.li@utsouthwestern.edu](mailto:quan.li@utsouthwestern.edu)

Received 9 April 2018; Revised 16 June 2018; Accepted 25 July 2018; Published 2 September 2018

Academic Editor: Elena Schiopu

Copyright © 2018 Honglin Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objective.** Systemic sclerosis (SSc) is a systemic connective tissue disease of unknown etiology. Aberrant gene expression and epigenetic modifications in circulating immune cells have been implicated in the pathogenesis of SSc. This study is to delineate the interaction network between gene transcription and DNA methylation in PBMC of SSc patients and to identify methylation-regulated genes which are involved in the pathogenesis of SSc. **Methods.** Genome-wide mRNA transcription and global DNA methylation analysis were performed on PBMC from 18 SSc patients and 19 matched normal controls (NC) using Illumina BeadChips. Differentially expressed genes (DEGs) and differentially methylated positions (DMPs) were integrative analyzed to identify methylation-regulated genes and associated molecular pathways. **Results.** Transcriptome analysis distinguished 453 DEGs (269 up- and 184 downregulated) in SSc from NC. Global DNA methylation analysis identified 925 DMPs located on 618 genes. Integration of the two lists revealed only 20 DEGs which harbor inversely correlated DMPs, including 12 upregulated (ELANE, CTSG, LTBR, C3ARI, CSTA, SPII, ODF3B, SAMD4A, PLAUR, NFE2, ZYX, and CTSZ) and eight downregulated genes (RUNX3, PRF1, PRKCH, PAG1, RASSF5, FYN, CXCR6, and F2R). These potential methylation-regulated DEGs (MeDEGs) are enriched in the pathways related to immune cell migration, proliferation, activation, and inflammation activities. Using a machine learning algorithm, we identified six out of the 20 MeDEGs, including F2R, CXCR6, FYN, LTBR, CTSG, and ELANE, which distinguished SSc from NC with 100% accuracy. Four genes (F2R, FYN, PAG1, and PRKCH) differentially expressed in SSc with interstitial lung disease (ILD) compared to SSc without ILD. **Conclusion.** The identified MeDEGs may represent novel candidate factors which lead to the abnormal activation of immune regulatory pathways in the pathogenesis of SSc. They may also be used as diagnostic biomarkers for SSc and clinical complications.

## 1. Introduction

Systemic sclerosis (SSc) is a chronic autoimmune disease characterized by microvascular dysfunction, immune abnormalities, chronic inflammation, and interstitial and perivascular fibrosis in the skin and internal organs [1]. Pulmonary arterial hypertension (PAH) and interstitial lung disease (ILD) are the most common pulmonary complications in SSc,

and PAH is the leading cause of mortality in SSc patients [2]. SSc is a clinically heterogeneous disease with unknown etiopathogenesis and presents with distinct subphenotypes [3]. The exact cellular and molecular mechanism of SSc remains unclear.

Considerable evidences from genome-wide association studies (GWAS) support the inheritable nature of the disease by the identification of susceptibility genes that have been

TABLE 1: Characteristics of clinical subjects and healthy controls.

Clinical characteristics	Controls (n=19)	SSc (n=18)
Age (Mean $\pm$ SD)	39.8 $\pm$ 6.5	46.2 $\pm$ 12.4
Sex (M/F)	5/14	7/11
Disease duration (months)	NA	48.7 $\pm$ 37.3
Organ involvement		
Interstitial lung disease	NA	10
PAH	NA	1
Raynaud's phenomena	NA	16
Gastrointestinal involvement	NA	9
Digital ulcer	NA	5
Serological characteristics		
Anti-Scl-70	NA	12
Anti-RNP	NA	4
Anti-ANA	NA	16
Anti-dsDNA	NA	0
Anti-Jo-1	NA	0
Anti-Sm	NA	0
Anti-SSa	NA	4
Anti-SSb	NA	0
Anti-ANCA	NA	1
Medications		
Prednisone	NA	17
Cyclophosphamide	NA	7

attributed to immune regulation, inflammation, transcription, kinase activity, DNA cleavage, and repair in SSc [4–6]. In addition, the environmental influence via epigenetic mechanisms, particularly altered status of DNA methylation, contributes to the environment-host interaction in the development of SSc [7, 8]. DNA methylation profiling in SSc has been obtained by Illumina methylation arrays, which identifies genes with differential methylation. It has been reported that more than 60 methylation-regulated genes were associated with autoimmunity in SSc [5, 9, 10]. However, there were no studies that investigate the correlation between gene expression and DNA methylation at global genome level in SSc. The advances in genome-wide technologies have provided unprecedented opportunities to expand our view of the relationship among the genome, methylome, and transcriptome. The integration of epigenetic and genetic data promises to gain insight into the mechanisms affecting epigenetic alteration, gene expression, and disease susceptibility [11, 12].

SSc is a systemic disease and involves multiple organs, tissues, and cell types. Previous data were generated from fibroblast [13, 14], purified CD4<sup>+</sup> T cells [8, 14, 15], whole blood cells [16], and skin biopsy [17]. Peripheral blood mononuclear cells (PBMCs) represent a broad spectrum of cell types including T cells, B cells, NK cells, monocytes, and dendritic cells, which all played major roles in immunological events. Its ready accessibility, in particular in repeat sampling for serial comparison of profiling, has made PBMC an ideal target in obtaining a comprehensive picture of immune status.

In this study, we performed a comprehensive screening for global patterns of aberrant DNA methylation and gene expression in the PBMC of SSc and healthy control in order to identify methylation-regulated genes and the enriched pathways and to understand the functional consequence of DNA methylation aberrancy on the regulation of gene expression in SSc.

## 2. Materials and Methods

**2.1. SSc Patients and Controls.** 18 SSc patients (14 diffuse cutaneous and 4 limited cutaneous SSc) and 19 normal controls enrolled in the study in the Department of Rheumatology, Xiangya Hospital, Central South University, Changsha, China. All patients and controls are Han Chinese. The diagnosis of all patients meets the ACR classification criteria for SSc [18]. The clinical features of SSc patients included in this study are shown in Table 1. ILD was diagnosed with high-resolution computed tomography (HRCT) when ground-glass opacity or reticulation was detected in nondependent portions of lung or ground-glass opacity and reticulation was found in dependent portions of lung that persisted on prone imaging and the presence of honeycombing and traction bronchiectasis [19–21]. The gastrointestinal involvement was evaluated by GI symptoms and nutrition status, including reflux, bloating, distension, constipation, diarrhoea, anorectal incontinence, weight loss, and malnutrition [22]. The institutional review board at Xiangya Hospital, Central South University, approved this study. All study participants signed a written informed consent prior to participation.

**2.2. DNA and RNA Isolation.** Peripheral blood samples were obtained from SSc patients and normal controls. The PBMC were separated from heparinized blood by density gradient centrifugation over Ficoll-Hypaque plus (GE Healthcare, Piscataway, NJ, USA). Total RNA was isolated from PBMC by standard phenol-chloroform extraction using Trizol reagent (Invitrogen Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Genomic DNA was isolated from whole blood cells using genomic DNA extraction kits (Life Technologies, Gaithersburg, MD).

**2.3. mRNA Transcription Profiling.** The genome-wide mRNA transcriptome in PBMC from 18 SSc patients and 19 NC was analyzed using Illumina humanHT-12 v4.0 BeadChip arrays bearing 48,700 human gene transcripts following vendor's instruction. Briefly, 500 ng total RNA from each sample was used to generate biotin-labeled cRNA probe which was then hybridized onto genechip array. After staining with Cy3-labeled streptavidin (Amersham, Piscataway, NJ, USA), slide was scanned on an Illumina HiScan scanner and the image was analyzed using GenomeStudio v3 software (Illumina, Inc.) to generate bead-summarized data. Further analysis on the gene expression data was performed using lumi package in R [48]. The ReMOAT annotation of gene expression data was used to include only "perfect" and "good" annotated probes [49]. Exploratory quality-control analyses revealed no strong batch effects. After quality control and preprocessing of Illumina arrays, differential gene expression between disease and normal group was performed using limma package in R and multiple comparisons correction was performed in R [50]. For genes with multiple transcripts, we selected the transcript with the highest fold change between SSc and normal. The differentially expression genes were called using our criteria of BH- (Benjamini-Hochberg-) adjusted  $p < 0.05$  and absolute fold change  $> 1.5$ .

**2.4. Genome-Wide DNA Methylation Analysis.** DNA methylation status of 485,000 CpG sites across the entire genome was interrogated using the Illumina Human Methylation 450K BeadChip arrays (Illumina, San Diego, CA). This methylation array covers 99% of RefSeq genes, with an average of 17 CpG sites per gene across the promoter region, 50 sites in untranslated region (5'-UTR), first exon, gene body, and 3'-UTR. It covers 96% of CpG islands, non-CpG-islands methylation sites, and microRNA promoter regions. Genomic DNA (1 $\mu$ g) extracted from PBMC was bisulfite converted using EZ DNA Methylation kit (Zymo Research Corp, Orange, CA), and a cyclic denaturation step was used during the conversion reaction. Whole-genome amplification, fragmentation, resuspension, hybridization, washing, extending and staining, and scanning were performed according to the manufacturer's instructions. The arrays were scanned using Illumina HiScan scanner and generated the raw IDAT format files which were analyzed using the Chip Analysis Methylation Pipeline (ChAMP) implemented in R and available from Bioconductor [51]. Briefly, raw IDAT files were used as input data and probes were filtered by their raw intensity values using a detection  $p$ -value threshold of 0.01. Probes corresponding to the X and Y chromosomes

were removed from the dataset as both male and female samples were being analyzed. Locus-by-locus analyses were conducted using the nonparametric Wilcoxon rank-sum test, and multiple comparisons correction was performed in R [50]. A CpG site was considered statistically differentially methylated only if the BH-adjusted  $p$  value  $< 0.05$  between the tested groups and the absolute difference of the median  $\beta$ -value is greater than 0.12. For genes with multiple probes measuring DNA methylation, we selected the probe with the highest fold change value for DNA methylation.

**2.5. Real-Time PCR Validation of the DEGs.** DEGs were validated using Taqman assays on a 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). The assay was performed by using a TaqMan RNA-to-CT 1-Step kit (Applied Biosystems) in a total volume of 20  $\mu$ l, which contained a final concentration of 900 nM sense and antisense primers (Supplementary Table 1), 250 nM Taqman gene probe, 1 $\times$  TaqMan RT Enzyme Mix, and 1 $\times$  TaqMan RT-PCR Mix. The cDNA amplification was monitored using 7900HT Fast Real-Time PCR system under the conditions of 48 $^{\circ}$ C for 15 min, 95 $^{\circ}$ C for 10 min, and 40 cycles of 95 $^{\circ}$ C for 15 s and 60 $^{\circ}$ C for 1 min. This assay was carried out in triplicate for each sample, including a no-template control. The relative quantity (RQ) of the gene expression in each sample was calculated by normalizing to housekeeping gene GAPDH.

**2.6. Integrative Analysis of Transcriptome and DNA Methylation.** Each methylation probe was mapped to the nearest transcript starting site. Transcription information of hg19 was fetched from UCSC Genome browser database and further processed using the Bioconductor Genomic Feature package. A probe was mapped to the nearest gene if the distance between the probe and the nearest gene's transcription starting site was less than 10 kilobases. We retained only the subset of probes associated with genes that were represented on the gene expression microarray. This resulted in the retention of 16,750 genes associated with the CpG probes in methylation and transcription probes in gene expression. Pearson correlation coefficients for each annotated gene were first calculated among all possible pairs of methylation probe sets and gene expression probe sets between SSc and normal control group. The methylation-expression probe set pair with the maximum absolute correlation coefficient was then chosen for each gene.

**2.7. Support Vector Machines.** Support vector machines (SVMs) are supervised learning models traditionally employed for classification analysis through constructing a model based on a separating plane that maximizes the margin between different classes [52]. A set of differential expression genes (DEGs) will be used to evaluate whether selected DEGs will correctly classify normal control from SSc patient samples. We begin by choosing radial kernel and tune the optimal model parameters (i.e., cost and gamma) to achieve the best diagonal performance on hold-on-one-out cross-validation test [53]. The SVM is trained using data from all but one of the sample. The sample not used in training is then assigned a class by the SVM. A single SVM experiment

consists of a series of hold-one-out experiments, each sample being held out and tested exactly once. The e1071 R package is used for implement the SVM analysis.

**2.8. Bioinformatics Analysis.** Differentially methylated and/or expressed genes were analyzed for Gene Ontology (GO) enrichments in comparison to all genes available on the Illumina Infinium Human Methylation 450K platform using the DAVID Functional Annotation Tool [54]. Genes for which expression levels change in concordance with DNA methylation changes were analyzed for gene network and biological processes enrichment using IPA (Ingenuity Pathway Analysis: <http://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>). Meanwhile, GAGE package in R [55] was used for detection of gene enrichment in KEGG pathway and GO analysis; the enriched pathway was visualized by Pathview package in R [56].

### 3. Results

**3.1. Identification of DEGs in PBMC of SSc.** We performed a genome-wide probe-based differential gene expression analysis using 47,312 probes after QC procedure to identify differentially expression probes including covariates for age, sex, and the first two principal components from expression values of all the probes, which is routinely included in GWAS analyses to control for population stratification [57]. By comparison of the whole-genome transcription profile of PBMC between patients with SSc and NCs, we identified 590 differentially expressed transcripts (Supplementary Table 2-1), representing 453 unique differentially expressed genes (DEGs) in SSc, which include 184 upregulated DEGs and 269 downregulated DEGs (Figure 1(a); Supplementary Table 2-2). Hierarchical clustering of the DEGs demonstrated that the patterns of expression profiles clearly distinguished NC from SSc patients with the exception that four SSc were misclassified in the NC group (Figure 1(b); Supplementary Table 2). Among the upregulated DEGs, 10 DEGs belong to the family of IFN-inducible genes, 6 DEGs are complement components, and 3 DEGs are  $\alpha$ -defensins (Table 2). Induction of interferon signature genes and complement activation were reported in a number of autoimmune diseases including SSc and its complications [58–61]. Human  $\alpha$ -defensins encode human neutrophil peptides (HNPs) and function as enhancer of phagocytosis by macrophages. Defensins were associated with mucosal immunity [62] and were implicated in the pathogenesis of ILD in SSc [63]. By examining the downregulated DEGs, we found 19 genes are associated with NK cells (Table 2), implying a compromised NK cell function. The DEG profiling identified in our study highlighted a crucial contribution of innate immunity in the pathogenesis of SSc.

Pathway analysis using Ingenuity Pathway Analysis software (IPA) indicated that the 453 DEGs are most significantly involved in the pathways of NK cell signaling, inflammasome, crosstalk between dendritic cells and NK cells, Tec kinase signaling, and NF- $\kappa$ B signaling (Figure 1(c)). Among the 184 overexpressed DEGs, 48 genes are associated with rheumatic diseases, 40 genes are involved in viral infection, and 34 genes

are related to inflammatory response (Supplementary Figure 1).

DEGs were also analyzed for GO enrichments using the DAVID Functional Annotation Tool [54]. GO analysis demonstrated that the DEGs were significantly enriched in “immune system process”, “immune response”, and “regulation of immune system process”. Additionally, GO for downregulated genes were enriched in “NK cell mediated immunity” (Supplementary Figure 2).

**3.2. Identification of Differentially Methylated Positions (DMPs) in SSc.** By performing a locus-by-locus differential DNA methylation analysis with age, sex, and the first two principal components as covariates, we identified a total of 925 differential methylated positions (DMPs) in SSc patients, among which 782 DMPs (85%) were hypomethylated and 143 DMPs (15%) were hypermethylated (Figure 2(a); Supplementary Table 3). The two-dimensional hierarchical clustering with DMGs resulted in separation of SSc from normal controls, with only minimal number of misclassified samples (one NC and two SSc) (Figure 2(b)).

Next, we investigated the genome distribution of the 925 DMPs in PBMC of SSc. Surprisingly, only 5% (45 of 925) of DMPs were located on CpG islands and the rest were found in non-CpG islands, mostly in open sea regions, followed by shore and shelf (Figure 2(c) and Supplementary Table 4). Based on the position of the DMPs, we found that about 30% DMPs were located in the promoter regions, including TSS1500, TSS200, 5'UTR, and the first exon, whereas the majority of DMPs were found within nonpromoter regions, mostly in gene body, followed by IGR and 3'UTR. Hypomethylated or hypermethylated DMPs demonstrated a very comparable genomic distribution proportion within CpG island/non-CpG island and promoter/nonpromoter regions (Figure 2(c) and Supplementary Table 3).

The 925 DMPs were mapped onto 618 unique genes and defined as differentially methylated genes (DMGs). Some of the DMGs have been previously reported in SSc and other autoimmune diseases [64]. Several interferon-inducible genes, including PAPP9, MX1, IFI44L and PLSCR1A, were among the top hypomethylated genes. PF4 and HLA-C were top hypermethylated genes. PF4 is a marker for activation of endothelial cell and coagulation that was implicated in vasculopathy of SSc [65–67]. HLA-C is coexpressed with certain killer cell immunoglobulin-like receptors (KIRs) in SSc and SLE [68] and may involve in the innate immunity.

Pathway analysis of the 618 DMGs indicated the DMGs were enriched in natural killer (NK) cell signaling, antigen presentation, D-myo-inositol 1, 3, 4-trisphosphate biosynthesis, and cytotoxic T cell mediated apoptosis (Figure 2(d)).

GO analysis revealed that the differentially hypomethylated genes were significantly enriched in the processes of regulation including alternative splicing, phosphoprotein, and polymorphism, which may result in modification of transcription and translation by DNA methylation aberrancy, whereas the differentially hypermethylated genes were considerably enriched in adaptive immunity and intracellular signal transduction (BH-adjusted  $p < 0.05$ ) (Supplementary Figure 3).

TABLE 2: Differentially expressed gene family in PBMC of SSc patients.

	<b>Gene family</b>	<b>Genes</b>
Down-regulated	NK cell associated gene family	CD2, CD96, CD160 KLRB1, KLRC1, KLRC2, KLRC3, KLRD1, KLRD1, KLRF1, KLRG1 KIR2DL3, KIR2DL5A, KIR2DS1, KIR2DS3, KIR2DS4, KIR3DL1 NCR3
		Natural cytotoxicity triggering receptor 3
		Natural killer cell triggering receptor Perforin 1
		Interferon, alpha-inducible protein 27
Up-regulated		IFI27
		Interferon, gamma-inducible protein 30
		IFI30
		Interferon induced protein 35
		IFI35
		Interferon induced transmembrane protein 3
		IFI1M3
		Janus kinase 1
		JAK1
		RNASE I, II, III
	S100A11	
	STAT4	
	Signal transducer and activator of transcription 4	
	Complement component	
	Defensin $\alpha$	
	CIQA, CIQB, CIQC, C2, C3ARI, C5ARI DEFA1, DEFA1B, DEFA3	

\* These CD molecules are also found in T cells.

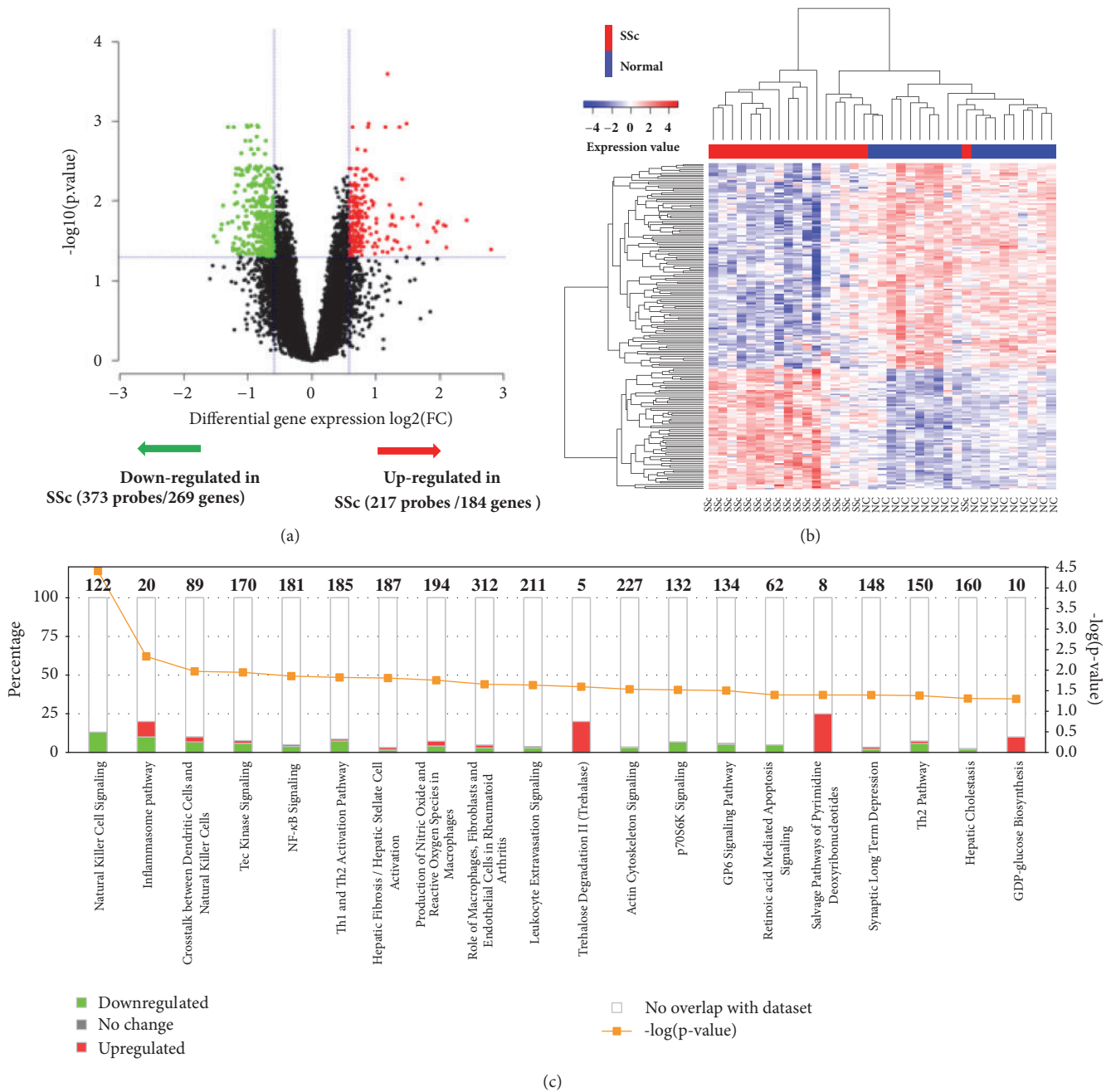
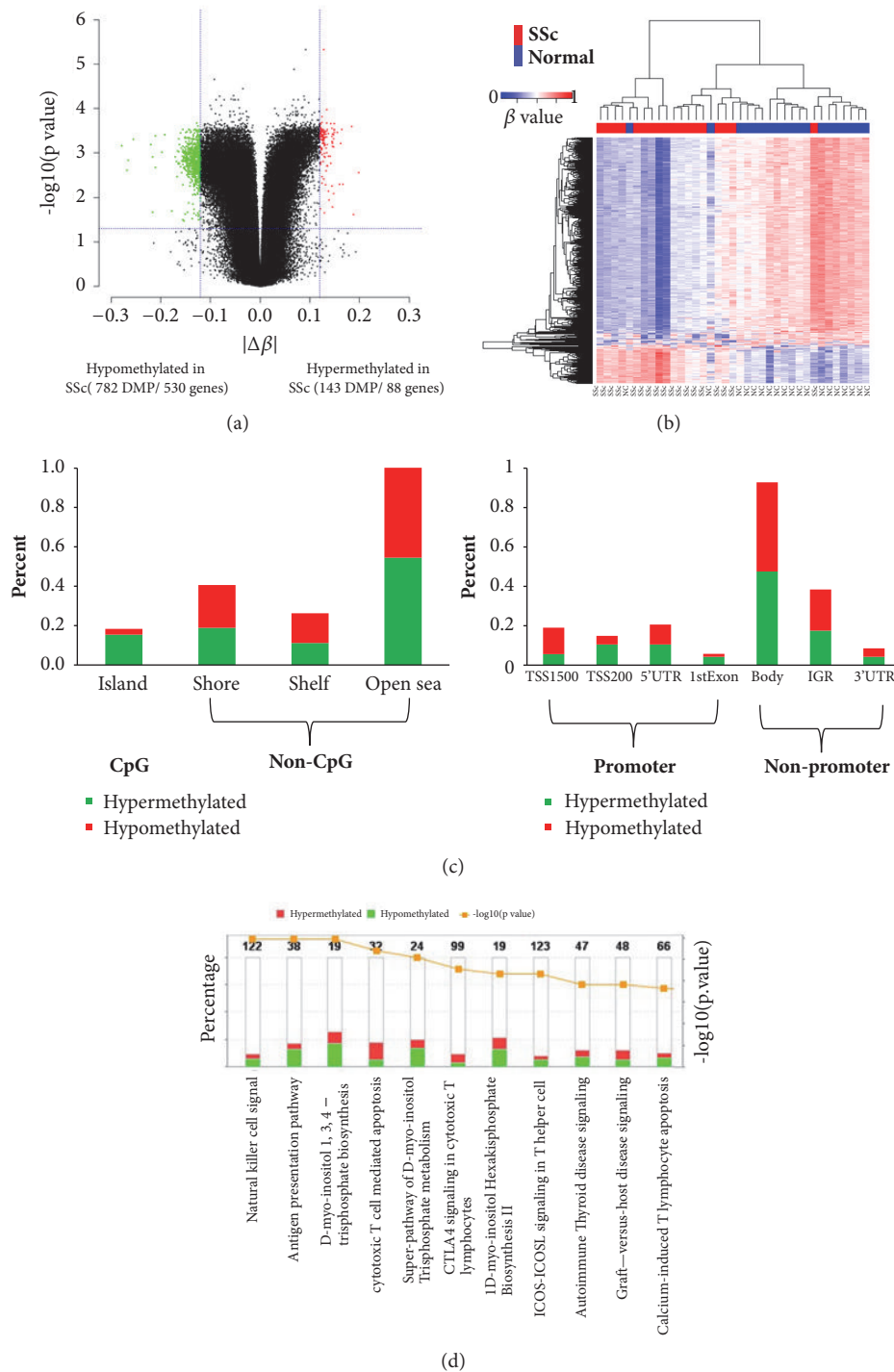


FIGURE 1: Identification of differentially expressed genes between systemic sclerosis and normal. (a) Volcano plot of the differential gene expression analysis. X-axis: fold change difference (log<sub>2</sub> scale); y-axis: BH-adjusted p values for each probe (-log<sub>10</sub> scale). The vertical dotted lines represent absolute cutoff value of 1.5-fold change. The horizontal dotted line represents the significant cutoff of p < 0.05 (see Supplementary Table 2). (b) Two-dimensional hierarchical clustering of differential gene expression probes. Probes are shown in rows; samples are shown in columns. (c) Canonical pathway analysis of the differentially expressed genes in PBMC from SSc patients as compared with normal controls. Statistically significant pathways were shown. The 1<sup>st</sup> y-axis indicates percentage of differential expression genes involved in the pathway. The 2<sup>nd</sup> y-axis shows BH-adjusted p-value of enrichment analysis.

3.3. Identification of DEGs Associated with Altered DNA Methylation. To assess the functional relevance between the changes on gene expression and associated DNA methylation in PBMC of SSc patients, we integrated the gene expression profiles with DNA methylation profiles by Entrez gene ID. By integration of the 453 DEGs and 618 DMGs, we only identified 25 common genes which showed significant

changes on their gene expression level (p < 0.05, |FC| > 1.5) and also contained at least one significantly altered DNA methylation site (p < 0.05, |Δβ| > 0.12) on the gene (Figure 3(a)). Among the 25 methylation associated genes, 20 genes demonstrated an inverse correlation on the changes between gene expression and DNA methylation (Figures 3(b)–3(d); Table 3), including 12 upregulated genes



**FIGURE 2: Identification of DNA methylation differences between systemic sclerosis and normal.** (a) Volcano plot of the differential DNA methylation analysis. X-axis: mean  $\beta$ -value difference; y-axis: BH-adjusted  $p$  values for each probe ( $-\log_{10}$  scale). The vertical dotted lines represent 12% change in  $\beta$ -values. The horizontal dotted line represents the significant cutoff of  $p < 0.05$ . Seven genes, ANKFY1, FAM49B, GNG7, INPP5A, LAX, PITPNM2, and VOPPI, showed both significant hypermethylation and hypomethylation (see text and Supplementary Table 2). (b) Two-dimensional hierarchical clustering was performed using the differential Infinium DNA methylation probes across all samples ( $n = 37$ ). Probes are in rows; samples are in columns. (c) Proportions of probes from genes with associated CpG islands, shelf, shore, and open sea (left) and proportions of probes from genes with associated probe locations, categorized as gene body, intergenic, 3'UTR, 5'UTR, 1<sup>st</sup> exon, TSS1500, and TSS200 (right). (d) Canonical pathway analysis of the differentially expressed genes in PBMC from SSc patients as compared with normal controls. Statistically significant pathways were shown. The 1<sup>st</sup> y-axis indicates percentage of differential expression genes involved in the pathway. The 2<sup>nd</sup> y-axis shows BH-adjusted  $p$ -value of enrichment analysis.

TABLE 3  
(a) 12 upregulated MeDEGs identified in SSc

HUGO <sup>a</sup>	HUGO gene name <sup>a</sup> : function <sup>b</sup>	Function in SSc or related pathology and reference
C3AR1	Complement C3a Receptor 1: Receptor for the chemotactic and inflammatory peptide anaphylatoxin C3a. This receptor stimulates chemotaxis, granule enzyme release and superoxide anion production.	Inflammatory response. [23]
CSTA	cystatin A: cysteine proteinase inhibitor, encoding Cystatin A, underlies exfoliative ichthyosis	Protease inhibitor in cell-cell adhesion in the basal and lower suprabasal layers. [24]
CTSZ	Cathepsin Z, lysosomal cysteine proteinase and member of the peptidase C1 family: susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis	Angiogenesis [25, 26]
CTSG	Cathepsin G, granule serine protease	Development of anti-cathepsin G antibodies in sera, function unknown. Angiogenesis [27] destruction of the lung matrix and acute inflammation propagation [28].
ELANE	Elastase, Neutrophil Expressed: mutations implicated in severe Congenital and cyclic neutropenia	Destruction of extracellular matrix [29, 30]
LTBR	Lymphotoxin Beta Receptor, a member of the tumor necrosis factor receptor superfamily	development of natural killer cells
NFE2	Nuclear Factor, Erythroid 2: Component of the NF-E2 complex essential for regulating the beta-globin control region	Transcription factor
ODF3B	Outer Dense Fiber Of Sperm Tails 3B	Involved in the pathogenic CD4 T cell in MS
PLAUR	Plasminogen Activator, Urokinase Receptor	Coagulation/fibrolytic system [31, 32] Angiogenesis [33, 34]
SAMDC4A	Sterile Alpha Motif Domain Containing 4A: posttranscriptional regulator	No report on SSc or related disease
SPI1	Spi-1 Proto-Oncogene: transcriptional regulator	Transcription factor
ZYX	Zyxin: actin regulator in vascular smooth muscle	Coagulation [35] integrity of vasculature [36]

(b) 8 downregulated MeDEGs identified in SSc

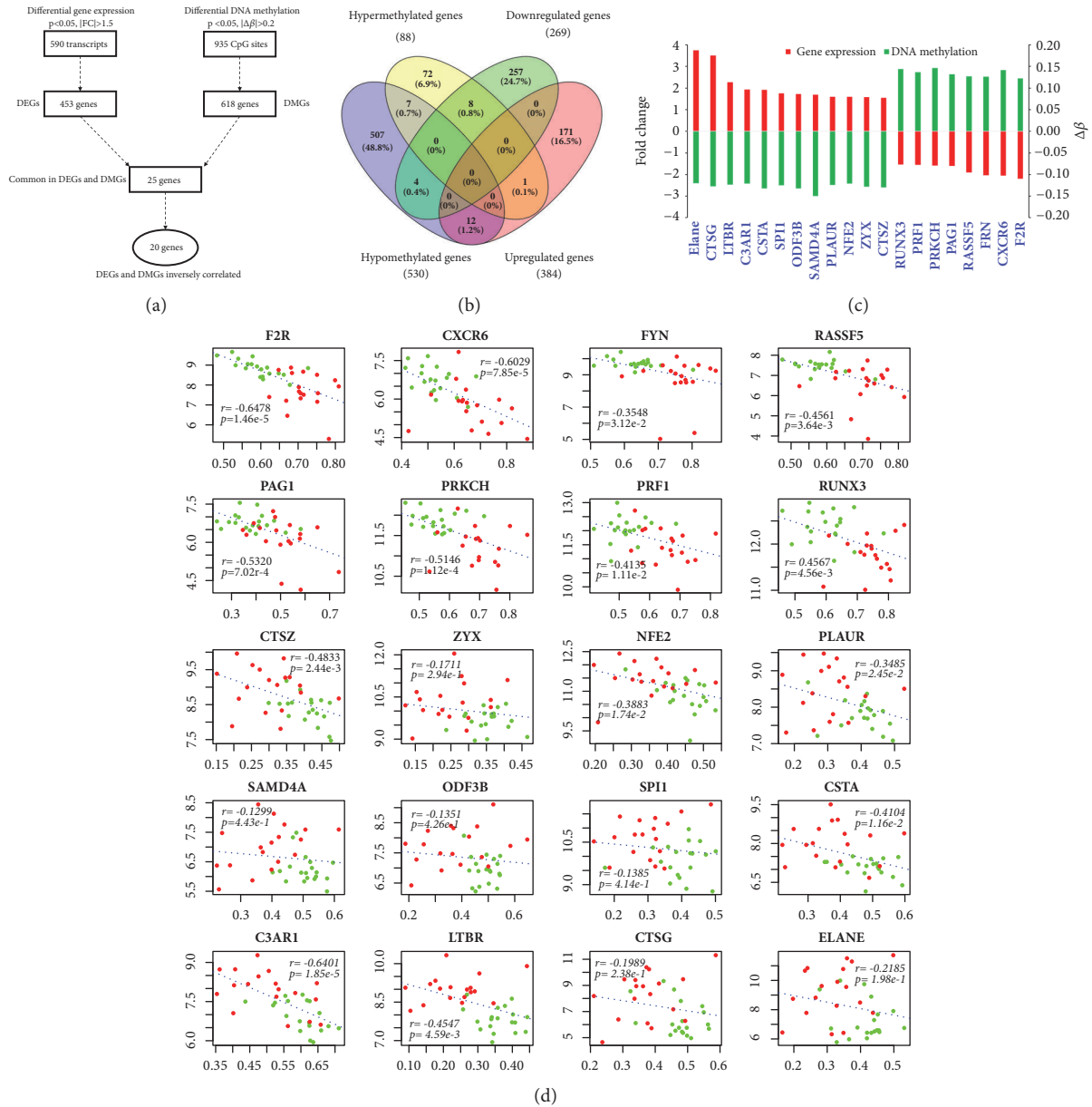
HUGO <sup>a</sup>	HUGO gene name <sup>a</sup> : function <sup>b</sup>	Function in SSc or related pathology and reference
CXCR6	C-X-C Motif Chemokine Receptor 6, receptor for the C-X-C chemokine CXCL16	Angiogenesis [37, 38]
F2R	Coagulation Factor II Thrombin Receptor	Coagulation/fibrolytic system, fibrosis [39–41]
FYN	Src Family Tyrosine Kinase	Fibrosis [40, 42]
PAG1	Phosphoprotein Membrane Anchor With Glycosphingolipid Microdomains 1	Proliferation of lymphocytes
PRF1	Perforin 1, pore forming protein 1, crucial effector of T and NK cell-mediated cytotoxicity, highly homologous to complement component C9	cytotoxic activity [43–45] vascular damage [46]
PRKCH	Protein Kinase C Eta, calcium-insensitive, but activated by diacylglycerol (DAG) and phosphatidylserine	Associated in stroke, but no report in SSc or related diseases
RASSF5	Ras Association Domain Family Member 5, Potential tumor suppressor.	No report on SSc or related disease
RUNX3	Runt Related Transcription Factor 3	Transcription factor, induction and suppressive function of Foxp3+ inducible T reg [47]

<http://www.genecards.org>.

<sup>a</sup>Human Genome Organization nomenclature.

<sup>b</sup>Gene function from GeneCards website.





**FIGURE 3: The genes that shows the most significant changes in DNA methylation and gene expression.** (a) Flowchart of identification of genes showing coordinately changed DNA methylation and gene expression. (b) Venn diagram of differential DNA methylation and differential gene expression analysis. (c) The top-rank relevant differentially methylated sites with different gene expression in SSc compared with controls. The 1<sup>st</sup> y-axis represents fold change of gene expression between SSc and normal control. The 2<sup>nd</sup> y-axis represents  $\Delta\beta$  of DNA methylation between SSc and normal control. (d) Correlation plots of DNA methylation versus gene expression in SScs and normal for the most significant changes in DNA methylation and gene expression.

that were hypomethylated, and eight downregulated genes that were hypermethylated (Figures 3(c)-3(d)). The negative correlations between gene expression and DNA methylation indicate that these genes are possibly methylation-regulated DEGs (MeDEGs) in the PBMC of SSc.

To further define the regulatory effect of the DNA methylation sites on gene expression, we investigated the location of each DMP in the genomic regions of the 20 genes. We found the DMPs were located in the promoter regions of 12 genes (60%), including hypomethylated DMPs on seven

upregulated genes (CSTA, CTSG, CTSZ, ELANE, LTBR, NFE2, and ODF3B) and hypermethylated DMPs on the five downregulated genes (CXCR6, FYN, PAG1, PRF1, and RUNX3). For the rest eight MeDEGs, the DMPs were found in the gene body (Supplementary Table 5). There is evidence that gene bodies are also involved in the regulation of gene expression [69, 70]. The genomic distribution of the DMPs suggested the DEGs may be methylation-regulated and yet to be confirmed. However, our data revealed only 4.4% of DEGs (20 of 453) harbored significant altered methylation

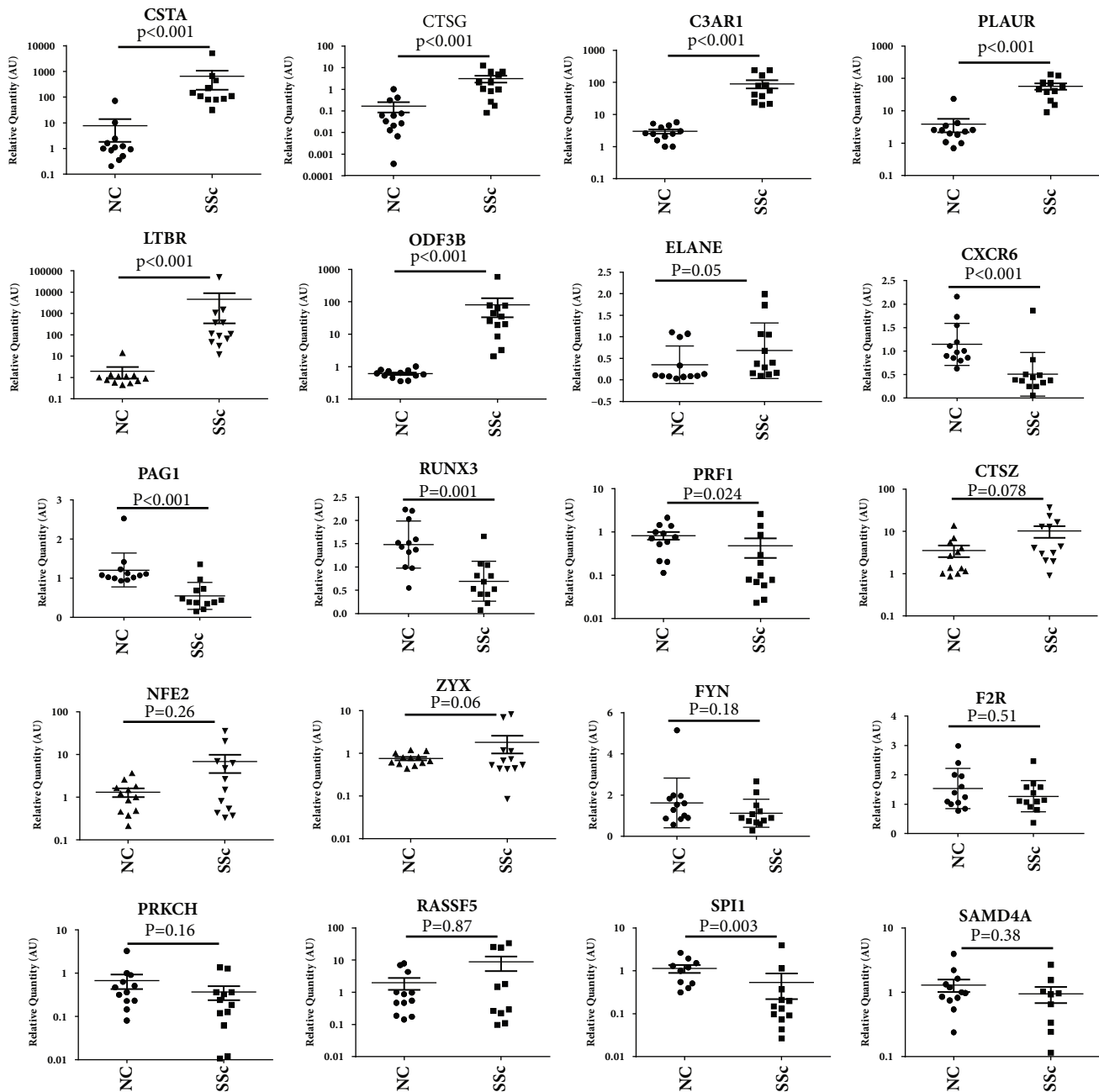


FIGURE 4: QPCR validation of gene expression: 20 MeDEGs were validated by QPCR in a separate cohort of SSc (n=12) and NC (n=12) samples. 11 MeDEGs (7 upregulated and 4 downregulated) were confirmed to be differentially expressed between SSc and NC ( $p<0.05$ ) by QPCR and 6 MeDEGs (3 upregulated and 3 downregulated) showed the difference in expression but did not reach statistical significance. Inconsistency in the detection of the expression for three MeDEGs by microarray or QPCR was also noted.

sites, implying that DNA methylation may not play a major role on the gene expression profile in the PBMC of SSc.

In addition, five genes showed positive correlation between DNA methylation and gene expression changes. One gene, HLA-C, was hypermethylated and upregulated, and four genes, MBP, DNKTR, GNG2 and SPTBN1, were hypomethylated and downregulated in SSc (Figure 3(b); Supplementary Table 5).

3.4. Validation of the Expression of MeDEGs by QPCR. The expression of the 20 MeDEGs was further validated by quantitative RT-PCR (QPCR) in a separate cohort of SSc (n=12) and NC (n=12). 11 MeDEGs were confirmed to be differentially expressed in SSc ( $p<0.05$ ), including seven upregulated genes (CSTA, CTSG, C3AR1, PLAUR, LTBR, ODF3B, and ELANE) and four downregulated genes (CXCR6, PAG1, RUNX3, and PRF1) (Figure 4). Six MeDEGs showed an

increasing trend (CTS2, NFE2, and ZXY) or decreasing trend (FYN, F2R, and PRKCH), but the change was not statistically significant, possibly because of the small sample size. The overall consistency on the transcriptional level between the two assays by microarray and QPCR was observed in 17 of 20 MeDEGs. However, inconsistency was also noted in three genes (RASSF5, SPI1, and SMAD4A). Further confirmation is undergoing with an expanded sample collection.

Some MeDEGs have been reported as susceptibility genes in SSc, such as hypomethylated RUNX3 [13], which was in consistence with our result. Most of them were not associated with SSc and their role in SSc remains elusive. However, ample evidence indicated that the MeDEGs are involved in the characteristic pathology of SSc, including extracellular matrix remodeling (CSTA, CTS2, CTSG, and ELANE), angiogenesis (CTS2, CTSG, and CXCR6), coagulation/fibrotic system (PLAUR and F2R), inflammatory response (C3AR1), and transcription factors associated with DEGs (NEF2, SPI1, and RUNX3) (Table 3).

**3.5. Association of DNA Methylation and Gene Expression Profiling with Clinical Phenotype in SSc.** Identification of 20 MeDEGs in SSc has enabled us to differentiate the disease status and/or stage of SSc. We performed an exploratory two-dimensional (2D) hierarchical clustering with the 20 MeDEGs across the 37 samples with the 20 MeDEGs. As expected, profiling with either DEGs (Figure 5(a)) or DMGs (Figure 5(b)) resulted in separate sample clusters, SSc and NCs, with a few exceptions. On the other hand, the mis-clustered samples may be associated with the pathogenic transition of disease status. In order to precisely differentiate SSc from NC, we applied SVM algorithm on the 20 MeDEGs. By fitting various combinations of the 20 MeDEGs, we further identified a set of six DEGs, i.e., F2R, CXCR6, FYN, LTBR, CTSG, and ELANE, that completely separate the SSc and NC population. All the 19 NCs are predicted as healthy controls, and the 18 SSc patients are all classified as SSc (Figure 5(c)). The six MeDEGs may be potentially used as diagnostic markers.

Diffuse and limited are two subsets of SSc, which are classified mainly by clinical features. To answer the question whether the MeDEGs may differentiate dsSSc from lcSSc, we assessed the expression for the 20 MeDEGs in the subsets. However, we did not find the difference in gene expression or methylation between dsSSc and lcSSc (data not shown). As we only have four lcSSc and 14 dsSSc, the assessment needs to be validated in an expanded population.

Interstitial lung disease (ILD) frequently complicates SSc and can be a debilitating disorder with poor prognosis. We then attempted to characterize a profiling that distinguishes ILD from non-ILD SSc patients. Initial analysis with locus-by-locus DMP or transcript-by-transcript DEG was not distinguishable between ILD and non-ILD SSc patients (data not shown). We then reexamined the level of mRNA expression and DNA methylation of the 20 MeDEGs in NC, ILD, and non-ILD SSc patients. The level of mRNA expression and DNA methylation was evaluated across three groups. The expression for four underexpressed MeDEGs, F2R, FYN, PAG1, and PRKCH, was significantly reduced in SSc with ILD

as compared to SSc without ILD ( $p < 0.05$ ) (Figure 5(d)). The corresponding methylation level for the four genes tended to be increasing, but the difference was not significant. The finding can be interpreted as a correlation between down-regulated expression for F2R, FYN, PAG1, and PRKCH and the development of ILD in SSc. Further, these four MeDEGs may be used as diagnostic or prognostic markers for SSc with pulmonary involvement.

Serum Scl-70 antibodies are the hallmark of systemic sclerosis and are found in up to 60% of patients with this connective tissue disease. Scl-70 antibodies are more common in patients with extensive cutaneous involvement and interstitial pulmonary fibrosis and are considered as a poor prognostic sign. Likewise, characterization of a gene profile in patients with positive Scl-70 antibody with the 20 MeDEGs was attempted but failed to identify a distinguishable gene set (data not shown).

**3.6. Gene Networks Affected by 20 MeDEGs.** Finally, we explored the function connection and molecular interaction networks associated with the 20 MeDEGs using pathway analysis tools. A total of 4 interaction networks were identified (Table 4). The highest scoring (score = 41) molecular interaction network is related to cellular movement and immune cell trafficking which involved 15 of the 20 MeDEGs (Figure 6(a)). The MeDEGs are most significantly enriched in the molecular pathways associated with cell-to-cell signaling and interaction ( $p = 1.26 \times 10^{-11}$ ), cellular movement ( $p = 1.56 \times 10^{-11}$ ), hematological system development and function ( $p = 1.56 \times 10^{-11}$ ), immune cell trafficking ( $p = 1.56 \times 10^{-11}$ ), and cellular growth and proliferation ( $p = 4.6 \times 10^{-10}$ ) (Figure 6(b)). Most of the MeDEGs are connected with cell migration, trafficking and chemotaxis (11 genes, Figure 6(c)), endothelial adhesion and immune cell activation (10 genes, Figure 6(d)), proliferation of immune cells (14 genes, Figure 6(e)), and inflammatory and immune response (8 genes, Figure 6(f)). The results indicated that the MeDEGs play a pivotal role in the cascades of inflammatory response by orchestrating such a sequential process of inflammation: initiation, activation, and amplification of immune response.

## 4. Discussion

By comparison of the whole-genome transcriptome and DNA methylation in the PBMC of SSc and NC, we identified 453 DEGs and 618 DMGs which were significantly altered in SSc. However, integration of the DEGs and DMGs only derived 20 genes in which the gene expression is inversely correlated with the DNA methylation and is potential methylation-regulated DEGs (MeDEGs). To our knowledge, this is the first study to comprehensively investigate the effect of methylation on the gene transcription as the etiology of SSc at global genome level. The advantage of the integrative analysis lies in the simultaneous examination of the status or level change for both transcriptome and methylome which elucidated an immediate, real-time relationship between gene expression and DNA methylation. With the integrative analysis strategy,

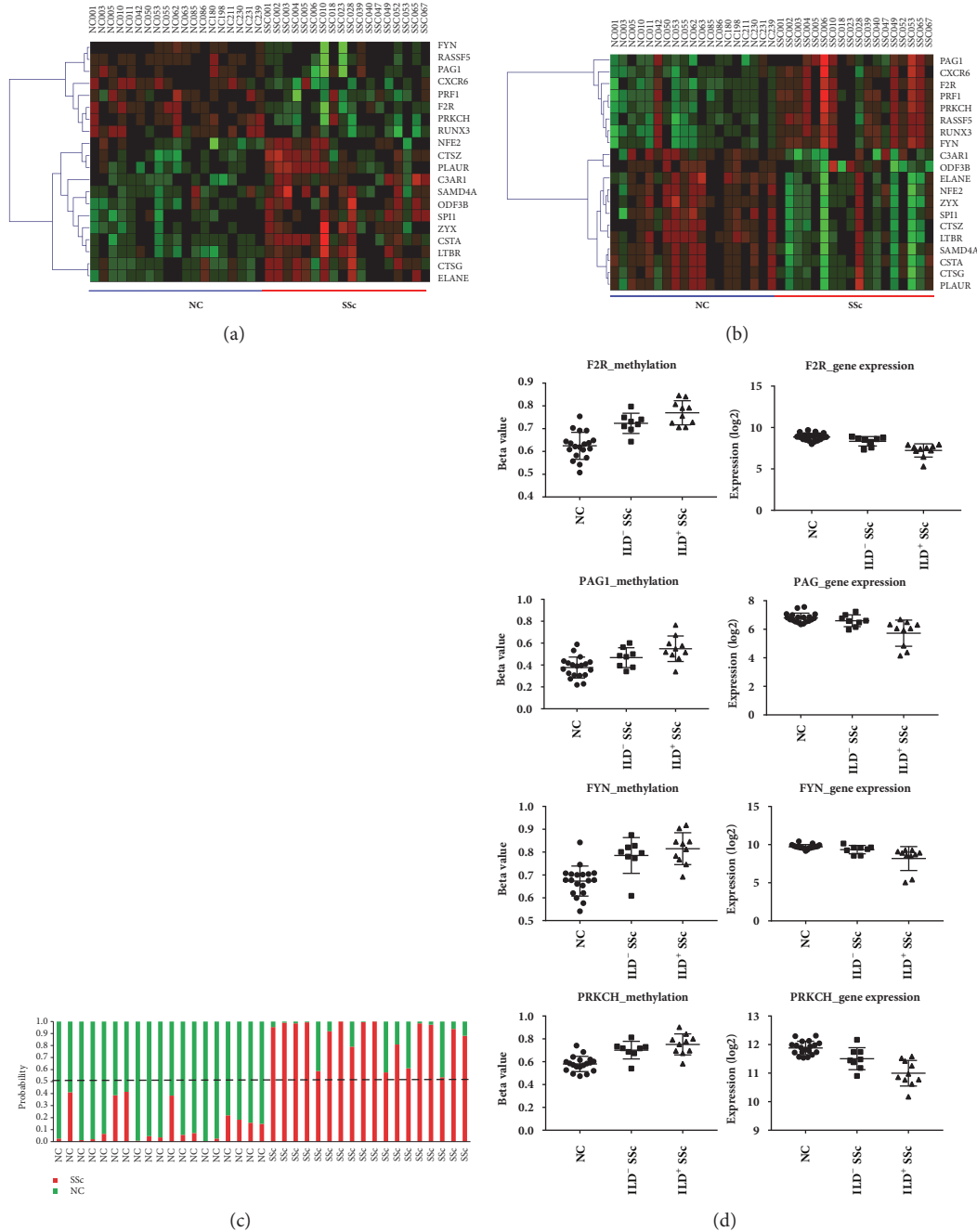


FIGURE 5: Association of clinical characteristics of SSc and 20 DEGs. (a) Heatmaps of gene expression and (b) DNA methylation show the most significant 20 correlated genes for SSc versus Normal. (c) Support vector machines analysis. The y-axis shows the group proportion inferred by SVM and sample ID is on the x-axis. (d) Array-based DNA methylation and mRNA expression among normal controls, the SSc with ILD, and SSc without ILD group,  $p$ -value  $< 0.05$  between SSc with and without ILD groups, are showed.

we have previously identified a large number of methylation-regulated genes in the PBMC of SLE patients [71]. In this study, we applied a more stringent criterion for identification of DEGs ( $p < 0.05$ ,  $|FC| > 1.5$ ) and DMGs ( $p < 0.05$ ,  $|\Delta\beta| > 0.12$ ) in the PBMC of SSc patients, and only 25 genes were overlapped between the two lists, in which 20 DEGs were inversely correlated with aberrant DNA methylation. The difference in methylation frequency in SLE and SSc may reflect a different underlying mechanism in disease spectrum.

We then investigated the 20 MeDEGs for their potential role in the pathogenesis of SSc. Pathway analysis revealed that these MeDEGs are significantly enriched in the molecular pathways related with cell-to-cell signaling and interaction, cell migration, immune cell trafficking, hematological system development and function, and cellular growth and proliferation. The dysregulation of these genes may promote the migration of immune cells (macrophage, leukocytes, and neutrophils) to the pathogenic tissues, activation, and

TABLE 4: Molecular interaction networks involved by the 20 MeDEGs.

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	<u>C3AR1</u> , Collagen Alpha1, Collagen(α), Cr3, <u>CTSG</u> , <u>CTS2</u> , Ecm, <u>ELANE</u> , ERK1/2, <u>F2R</u> , Fcεr1, Fibrinogen, <u>FYN</u> , IFN Beta, Ige, IL12 (complex), Immunoglobulin, Integrin, <u>LTBR</u> , Mapk, Nfat (family), P38 MAPK, <u>PAG1</u> , Par, <u>PLAUR</u> , <u>PRF1</u> , <u>PRKCH</u> , Pro-inflammatory Cytokine, Ras, <u>RASSF5</u> , <u>RUNX3</u> , <u>SAMD4A</u> , SPI1, <u>SRC</u> , Vegf	41	15	Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking
2	<u>ADCYAP1R1</u> , <u>ADGR13</u> , Akt, AZU1, caspase, <u>CCL3L1</u> , CD3, CD226, <u>CSDC2</u> , <u>CXCR6</u> , ERK, F2RL3, Gpct, GPR45, GPR63, HISTONE, IGK, IL8r, Interferon alpha, Jnk, KDM5D, LIME1, <u>NFE2</u> , <u>NFE2L3</u> , NFKB (complex), P2RY10, PI3K (complex), Pkc(s), <u>PRKCH</u> , RNA polymerase II, SIRT1, TCR, tubulin, UQCRI1, <u>ZYY</u>	8	4	Cell-To-Cell Signaling and Interaction, Cell-mediated Immune Response, Cellular Development
3	ESR1, <u>ODF3B</u> , PKD1	3	1	Connective Tissue Development and Function, Organ Morphology, Skeletal and Muscular System Development and Function
4	CASP1, CEP57, CNOT7, <u>CSTA</u> , CTSB, CTSH, CTSL, CUL2, CUL4B, DSP, FOXR1, GATA2, GLI2, HERC2, HOXA10, Hsp90, IVL, JUND, KRT1, LOR, MAP2K7, MAP3K1, METTL23, MYOC, PI3, PTN, RAD21, RASSF9, RASSF10, TCF3, TCTN3, UCHL5, USP53, ZDHHC17, ZIC1	2	1	Post-Translational Modification, Dermatological Diseases and Conditions, Hereditary Disorder

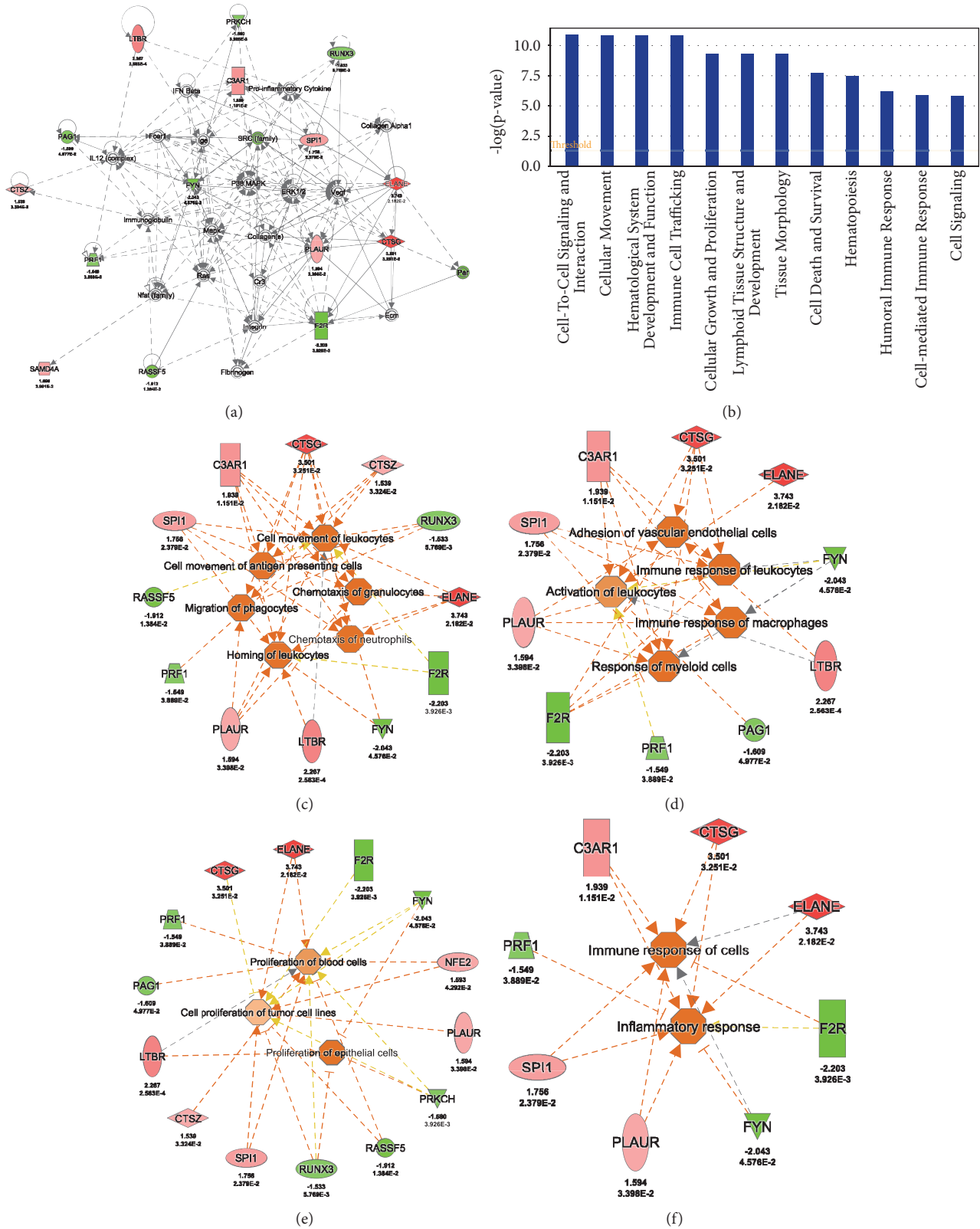


FIGURE 6: The diseases and functions analysis by integrated pathway analysis using 20 MeDEGs. (a) The top interaction networks involved by 15 MeDEGs. (b) Canonical pathway analysis of 20 MeDEGs in PBMC patients as compared with normal controls. Statistically significant pathways were shown. (c) Most genes (12/20) are involved with cell movements of leukocytes, cell movement of antigen presenting cells, migration of phagocytes, chemotaxis of granulocytes, homing of leukocytes, and chemotaxis of neutrophils. (d) The network from 10 DEGs is involved with adhesion of vascular endothelial cells, immune response of leukocytes, activation of leukocytes, immune response of macrophages, and response of myeloid cells. (e) Most genes (12/20) are involved with proliferation of blood cells, cell proliferation of tumor cell lines, and proliferation of epithelial cells. (f) The network from 8 genes is involved with immune response of cells and inflammatory response.

proliferation of immune cells and may eventually lead to local pathogenesis. Some of the gene may also induce adhesion and proliferation of vascular endothelial cells which can cause abnormal blood vessel growth [72]. Among the upregulated MeDEGs, ELANE, CTSG, and CTSZ are three proteases involved in remodeling of extracellular matrix. ELANE which encodes neutrophil elastase, together with matrix metalloproteinase 12 (MMP12), has been reported in the development cystic fibrosis in lungs [73]. CTSG and CTSZ are cathepsin proteases produced by macrophages. Cathepsin G leads to destruction of the lung matrix and continued propagation of acute inflammation [28]. Cathepsin Z promotes tumor invasion by interacting with integrins and the extracellular matrix [74]. The role of these genes in SSc has yet to be defined, but we speculated that overexpression of ELANE, CTSZ, and CTSG may involve dysregulation of extracellular matrix and subsequent fibrosis. Fibrosis in SSc is preceded by vascular injury, which in turn leads to tissue hypoxia and capillary rarefaction. Vascular injury potentially activates coagulation cascade and production of thrombin, which may trigger myofibroblast differentiation.

Urokinase-type Plasminogen Activator (PLAUR) is another upregulated MeDEG which is associated with vascular injury in SSc. Iwamoto et al. detected overproduction of PLAUR in vascular smooth muscle cells in SSc [75]. PLAUR induced cell proliferation and suppressed apoptosis of human pulmonary artery smooth muscle cells, resulting in hyperplasia of vascular smooth muscle cells (VSMCs) which is characteristic of proliferative vasculopathy in SSc. The pathogenic role of PLAUR has been supported by detection of elevated level of soluble Urokinase-type Plasminogen Activator receptor in both diffuse and limited cutaneous SSc [76].

ZYX is a novel MeDEG which is upregulated in SSc. A recent publication revealed that Zyxin, a scaffold protein encoded by ZYX, was required in TGF- $\beta$  signaling pathway in response to hypoxia [77]. Nevertheless, the upregulated Zyxin level contributed to cell-matrix adhesion through TGF- $\beta$  signaling [78]. TGF- $\beta$  is known as a pleiotropic cytokine and causes fibroblast activation [79], and ZYX may work in concert with TGF- $\beta$  during the profibrotic process.

Several genes were downregulated and hypermethylated in PBMC of SSc, including F2R, CXCR6, PAG1, PRF1, and RUNX3. Pathway analysis showed that these genes may participate in the activation and proliferation of immune cells. F2R is a gene encoding coagulation factor II (thrombin) receptor, PAR1, which involved in the regulation of thrombotic response. Imbalance of coagulation system followed by autoimmunity has been implicated in the development of fibrosis in lungs [39–41]. PAR1 activation also led to lung vascular leakage. PAR1 immunoreactivity was detected in SSc skin [80]. A recent study showed that thrombin inhibition with dabigatran was protective, thus significantly inhibiting protease-activated receptor-1 (PAR1) activation and the development of pulmonary fibrosis [39]. CXCR6 is the chemokine receptor for angiogenic CXCL16. The increased expression CXCL16 and its receptor in dermal endothelial cells of SSc suggesting CXCL16-CXCR6 may play a role on angiogenesis in SSc skin [37]. PRF1 is a member of granzyme

B-Perforin family which is released from NK cells as cytotoxic factor. PRF1 has been reported to be hypermethylated in blood cells of SSc [81] which is in consistent with our result. RUNX3 is a transcription factors participating in the maintenance of genomic stability. Hypomethylated RUNX3 has been reported in dermal fibroblasts of diffuse and limited systemic sclerosis [18].

Other than the aforementioned MeDEGs, most of the DEGs were not associated with their change in DNA methylation and vice versa. As an example, a group of interferon (IFN) signature genes were identified to be significantly upregulated in the DEG profiling, including IFI27, IFI30, IFI35, IFITM3, STAT4, etc. However, no altered methylation sites on these genes were identified. On the other hand, we noticed another set of IFN-inducible genes, PAPP9, MX1, IFI44L, and PLSCR1, which were significantly hypomethylated but no difference on the gene expression level in SSc compared with NC, suggesting the overexpression of IFN-inducible gene was not directly regulated by DNA methylation in PBMC of SSc.

Another group of DEGs with no associated DNA methylation change is NK cell related genes. We found the decreased expression of dozens of genes related to NK cells which implied the impaired innate immunity. Previous studies in peripheral blood were reported with discrepant results. In consistent with our data, defective NK activity was supported by decreased percentage or absolute numbers of NK cells [82, 83], whereas other studies found either normal NK cell counts [84] or increased frequency/number of NK cells in dcSSc with being normal in lcSSc [85]. Our data indicated numbers of genes for NK receptors, KLRs, KIRs, PRF1, and NKTR, may contribute to the reduced NK cell activity. Characteristic KIR genotype, KIR2DS2+/KIR2DL2-, has been associated with SSc [86]. A recent publication suggested the interplay between activating or inhibitory KIR genes with HLA ligands as a critical index of SSc predisposition [87].

In addition to genes showing a reversed correlation between expression and DNA methylation, we also identified four genes, MBP, GNG2, SPTBN1, and NKTR, that were concomitantly downregulated and hypomethylated and one gene, HLA-C, which was concomitantly upregulated and hypermethylated. Apparently regulation of these genes may be governed by other machinery. Prior studies showed that a polymorphic microRNA binding site in the 3'UTR of HLA-C [88]. In fact 24 miRNAs were found differentially expressed in SSc skin [89]. A transcriptional factor Oct1 consensus binding site [90] may account for the differential allele-specific expression level of HLA-C. Transcription of MBP is regulated by heterogeneous nuclear ribonucleoprotein (hnRNP) [91] or alternative splicing [92], indicating distinct regulatory mechanism on gene expression.

The main limitation for this study is that the sample size is relatively small which limited our ability to fully analyze subgroup comparisons within SSc, such as ILD versus non-ILD and dcSSc versus lcSSc. The variation in patient age and disease duration may also impact the interpretation of the results. Future studies should assess methylated gene profile in a large SSc population, which will be selected with more cautiousness on the uniformity of disease onset and duration.

The identification of differential expressed gene families in the present study, IFN-related genes and NK cells associated genes, warrants a follow-up investigation in isolated T cells and NK cells.

In conclusion, integration of genome-wide mRNA transcription and DNA methylation profiles distinguished 20 methylation-regulated DEGs. The low overlap frequency between DEGs and DMGs suggested that, for most of the genes in PBMC of SSc, the DNA methylation and gene transcription are two independent molecular events. The identification of MeDEGs in the peripheral circulation of patients suggested a systemic immune response in SSc. Gene network analysis revealed that the 20 MeDEGs comprehensively participate the cascade of immune cell activation, migration, proliferation, and inflammatory response and eventually lead to local pathogenesis. These data indicate that abnormal DNA methylation on susceptibility genes may attribute to the sustained activation of PBMC in peripheral circulation. The gene panel of six MeDEGs by SVM demonstrated high accuracy (100% sensitivity and 100% specificity) in diagnosing SSc. Analysis revealed four under expressed MeDEGs may contribute to the development of ILD. Identification of novel candidate susceptibility genes for SSc may be beneficial in developing biomarkers for diagnostic and prognostic purposes. This study delineated an overall picture regarding epigenetic regulation of autoimmunity in SSc. Our observation laid the groundwork for further diagnostic and mechanistic studies of SSc that could lead to in-depth understanding of the disease.

### Data Availability

The microarray data for gene expression and methylation has been submitted to the GEO database with accession number GSE117931, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117931>.

### Conflicts of Interest

The authors declare that they have no conflicts of interest relating to the conduct of this study or the publication of this manuscript.

### Authors' Contributions

Honglin Zhu, Chengsong Zhu, and Wentao Mi contribute equally.

### Acknowledgments

The authors thank the participants and the staff in Division of Rheumatology, Xiangya Hospital, for making this study possible. They thank Yun Lian in Microarray Core, UT Southwestern Medical Center, for microarray data submission to GEO database. This work was supported by grants from National Natural Science Foundation of China (81671621, 81270852, 81373206, and 81401357) and Scientific Research Fund of Xiangya Hospital (2013Q10).

## Supplementary Materials

*Supplementary 1.* **Supplementary Figure 1:** the diseases and functions analysis by integrated pathway analysis using 184 upregulated genes and 269 downregulated genes. **Supplementary Figure 2:** GO analysis of the downregulated genes (left) and upregulated genes (right) in PBMC from SSc patients as compared to normal controls. Statistically significant GO terms are shown. **Supplementary Figure 3:** GO analysis of the hypomethylated genes (left) and hypermethylated genes (right) in PBMC from SSc patients as compared to normal controls. Statistically significant GO terms are shown.

*Supplementary 2.* **Supplementary Table 1:** the list of primer sequence used in QPCR validation of the 20 MeDEGs.

*Supplementary 3.* **Supplementary Table 2:** the complete list of differentially expressed genes (DEGs) identified in PBMC of SSc patients.

*Supplementary 4.* **Supplementary Table 3:** the complete list of differentially methylated positions (DMPs) identified in PBMC of SSc patients.

*Supplementary 5.* **Supplementary Table 4:** genomic distribution of differentially methylated positions.

*Supplementary 6.* **Supplementary Table 5:** the complete list of 20 MeDEGs in PBMC of SSc patients.

## References

- [1] R. Lafyatis, "Connective tissue disease: SSc-fibrosis takes flight with Wingless inhibition," *Nature Reviews Rheumatology*, vol. 8, no. 8, pp. 441-442, 2012.
- [2] T. A. Winstone, D. Assayag, P. G. Wilcox et al., "Predictors of mortality and progression in scleroderma-associated interstitial lung disease: A systematic review," *CHEST*, vol. 146, no. 2, pp. 422-436, 2014.
- [3] Y. Y. Ho, D. Lagares, A. M. Tager, and M. Kapoor, "Fibrosis—a lethal component of systemic sclerosis," *Nature Reviews Rheumatology*, vol. 10, no. 7, pp. 390-402, 2014.
- [4] P. Chairta, P. Nicolaou, and K. Christodoulou, "Genomic and genetic studies of systemic sclerosis: A systematic review," *Human Immunology*, vol. 78, no. 2, pp. 153-165, 2017.
- [5] P.-S. Tsou and A. H. Sawalha, "Unfolding the pathogenesis of scleroderma through genomics and epigenomics," *Journal of Autoimmunity*, vol. 83, pp. 73-94, 2017.
- [6] J. Jin, Y.-C. Chou, M. Lima, D. Zhou, and X. Zhou, "Systemic sclerosis is a complex disease associated mainly with immune regulatory and inflammatory genes," *The Open Rheumatology Journal*, vol. 8, no. 1, pp. 29-42, 2014.
- [7] W. Lei, Y. Luo, K. Yan et al., "Abnormal DNA methylation in CD4<sup>+</sup> T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis," *Scandinavian Journal of Rheumatology*, vol. 38, no. 5, pp. 369-374, 2009.
- [8] X. Lian, R. Xiao, X. Hu et al., "DNA demethylation of CD40L in CD4<sup>+</sup> T cells from women with systemic sclerosis: A possible explanation for female susceptibility," *Arthritis & Rheumatology*, vol. 64, no. 7, pp. 2338-2345, 2012.



- [9] E. Rosato, C. Letizia, M. Proietti et al., "Plasma adrenomedullin and endothelin-1 levels are reduced and Raynaud's phenomenon improved by daily tadalafil administration in male patients with Systemic Sclerosis," *Journal of Biological Regulators and Homeostatic Agents*, vol. 23, no. 1, pp. 23–29, 2009.
- [10] J. C. A. Broen, T. R. D. J. Radstake, and M. Rossato, "The role of genetics and epigenetics in the pathogenesis of systemic sclerosis," *Nature Reviews Rheumatology*, vol. 10, no. 11, pp. 671–681, 2014.
- [11] H. T. Bjornsson, M. D. Fallin, and A. P. Feinberg, "An integrated epigenetic and genetic approach to common human disease," *Trends in Genetics*, vol. 20, no. 8, pp. 350–358, 2004.
- [12] R. Jaenisch and A. Bird, "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals," *Nature Genetics*, vol. 33, pp. 245–254, 2003.
- [13] N. Altorok, P.-S. Tsou, P. Coit, D. Khanna, and A. H. Sawalha, "Genome-wide DNA methylation analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals common and subset-specific DNA methylation aberrancies," *Annals of the Rheumatic Diseases*, vol. 74, no. 8, pp. 1612–1620, 2015.
- [14] Y. Wang, P.-S. Fan, and B. Kahaleh, "Association between enhanced type I collagen expression and epigenetic repression of the FLII gene in scleroderma fibroblasts," *Arthritis & Rheumatology*, vol. 54, no. 7, pp. 2271–2279, 2006.
- [15] Y. Wang, Y. Shu, Y. Xiao et al., "Hypomethylation and over-expression of ITGAL (CD11a) in CD4+ T cells in systemic sclerosis," *Clinical Epigenetics*, vol. 6, no. 1, 2014.
- [16] P. Matatiele, M. Tikly, G. Tarr, and M. Gulumian, "DNA methylation similarities in genes of Black South Africans with Systemic lupus erythematosus and Systemic sclerosis," *Journal of Biomedical Science*, vol. 22, no. 1, article no. 34, 2015.
- [17] Y. Wang and B. Kahaleh, "Epigenetic repression of bone morphogenetic protein receptor II expression in scleroderma," *Journal of Cellular and Molecular Medicine*, vol. 17, no. 10, pp. 1291–1299, 2013.
- [18] F. Van Den Hoogen, D. Khanna, J. Fransen et al., "OP0033 Classification Criteria for Systemic Sclerosis: Preliminary Results," *Annals of the Rheumatic Diseases*, vol. 72, no. Suppl 3, pp. A59.2–A59, 2014.
- [19] E. A. Kazerooni, F. J. Martinez, A. Flint et al., "Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring," *American Journal of Roentgenology*, vol. 169, no. 4, pp. 977–983, 1997.
- [20] J. B. Orens, E. A. Kazerooni, F. J. Martinez et al., "The sensitivity of high-resolution CT in detecting idiopathic pulmonary fibrosis proved by open lung biopsy: A prospective study," *CHEST*, vol. 108, no. 1, pp. 109–115, 1995.
- [21] American Thoracic Society (ATS) and European Respiratory Society (ERS), "Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement," *American Journal of Respiratory and Critical Care Medicine*, vol. 161, no. 2, pp. 646–664, 2000.
- [22] C. P. Denton and D. Khanna, "Systemic sclerosis," *The Lancet*, vol. 390, no. 10103, pp. 1685–1699, 2017.
- [23] C. Kemper and J. P. Atkinson, "T-cell regulation: With complements from innate immunity," *Nature Reviews Immunology*, vol. 7, no. 1, pp. 9–18, 2007.
- [24] D. C. Blaydon, D. Nitoiu, K.-M. Eckl et al., "Mutations in CSTA, encoding cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion," *American Journal of Human Genetics*, vol. 89, no. 4, pp. 564–571, 2011.
- [25] M.-L. Kang, J.-E. Kim, and G.-I. Im, "Vascular endothelial growth factor-transfected adipose-derived stromal cells enhance bone regeneration and neovascularization from bone marrow stromal cells," *Journal of Tissue Engineering and Regenerative Medicine*, vol. 11, no. 12, pp. 3337–3348, 2017.
- [26] A. Buhler, S. Berger, F. Bengsch et al., "Cathepsin proteases promote angiogenic sprouting and laser-induced choroidal neovascularisation in mice," *Experimental Eye Research*, vol. 115, pp. 73–78, 2013.
- [27] T. J. Wilson, K. C. Nannuru, M. Futakuchi, and R. K. Singh, "Cathepsin G-mediated enhanced TGF-beta signaling promotes angiogenesis via upregulation of VEGF and MCP-1," *Cancer Letters*, vol. 288, no. 2, pp. 162–169, 2010.
- [28] I. Craciun, A. M. Fenner, and R. J. Kerns, "N-Arylacyl O-sulfonated aminoglycosides as novel inhibitors of human neutrophil elastase, cathepsin G and proteinase 3," *Glycobiology*, vol. 26, no. 7, pp. 701–709, 2016.
- [29] J. H. Kristensen, M. A. Karsdal, J. M. B. Sand et al., "Serological assessment of neutrophil elastase activity on elastin during lung ECM remodeling," *BMC Pulmonary Medicine*, vol. 15, no. 1, 2015.
- [30] T. Hara, F. Ogawa, K. Yanaba et al., "Elevated serum concentrations of polymorphonuclear neutrophilic leukocyte elastase in systemic sclerosis: association with pulmonary fibrosis," *The Journal of Rheumatology*, vol. 36, no. 1, pp. 99–105, 2009.
- [31] J. M. Mahoney, J. Taroni, V. Martyanov et al., "Systems Level Analysis of Systemic Sclerosis Shows a Network of Immune and Profibrotic Pathways Connected with Genetic Polymorphisms," *PLoS Computational Biology*, vol. 11, no. 1, 2015.
- [32] M. Manetti, Y. Allanore, and L. Revillo, "A genetic variation located in the promoter region of the UPAR (CD87) gene is associated with the vascular complications of systemic sclerosis," *Arthritis & Rheumatology*, vol. 63, pp. 247–256, 2011.
- [33] S. D'Alessio, G. Fibbi, and M. Cinelli, "Matrix metalloproteinase 12-dependent cleavage of urokinase receptor in systemic sclerosis microvascular endothelial cells results in impaired angiogenesis," *Arthritis & Rheumatology*, vol. 50, pp. 3275–3285, 2004.
- [34] S. Serrati, M. Cinelli, F. Margheri et al., "Systemic sclerosis fibroblast inhibit in vitro angiogenesis by MMP-12-dependent cleavage of the endothelial cell urokinase receptor," *The Journal of Pathology*, vol. 210, no. 2, pp. 240–248, 2006.
- [35] X. Han, P. Li, Z. Yang et al., "Zyxin regulates endothelial von Willebrand factor secretion by reorganizing actin filaments around exocytic granules," *Nature Communications*, vol. 8, 2017.
- [36] R. J. Saphirstein, Y. Z. Gao, Q. Q. Lin, and K. G. Morgan, "Cortical actin regulation modulates vascular contractility and compliance in veins," *The Journal of Physiology*, vol. 593, no. 17, pp. 3929–3941, 2015.
- [37] B. J. Rabquer, P.-S. Tsou, Y. Hou et al., "Dysregulated expression of MIG/CXCL9, IP-10/CXCL10 and CXCL16 and their receptors in systemic sclerosis," *Arthritis Research & Therapy*, vol. 13, no. 1, article R18, 2011.
- [38] T. Isozaki, A. S. Arbab, C. S. Haas et al., "Evidence that CXCL16 is a potent mediator of angiogenesis and is involved in endothelial progenitor cell chemotaxis: studies in mice with K/BxN serum-induced arthritis," *Arthritis & Rheumatology*, vol. 65, no. 7, pp. 1736–1746, 2013.

- [39] B. S. Shea, C. K. Probst, P. L. Brazee et al., "Uncoupling of the profibrotic and hemostatic effects of thrombin in lung fibrosis," *JCI Insight*, vol. 2, no. 9, 2017.
- [40] L. M. Grove, B. D. Southern, T. H. Jin et al., "Urokinase-type plasminogen activator receptor (upar) ligation induces a raft-localized integrin signaling switch that mediates the hypermotile phenotype of fibrotic fibroblasts," *The Journal of Biological Chemistry*, vol. 289, no. 18, pp. 12791–12804, 2014.
- [41] C. J. Scotton, M. A. Krupiczkoj, M. Königshoff et al., "Increased local expression of coagulation factor X contributes to the fibrotic response in human and murine lung injury," *The Journal of Clinical Investigation*, vol. 119, no. 9, pp. 2550–2563, 2009.
- [42] R. Vepachedu, M. M. Gorska, N. Singhanian, G. P. Cosgrove, K. K. Brown, and R. Alam, "Uncl19 regulates myofibroblast differentiation through the activation of Fyn and the p38 MAPK pathway," *The Journal of Immunology*, vol. 179, no. 1, pp. 682–690, 2007.
- [43] A. Henriques, C. Silva, M. Santiago et al., "Subset-specific alterations in frequencies and functional signatures of gammadelta T cells in systemic sclerosis patients," *Inflammation Research : Official Journal of the European Histamine Research Society*, vol. 65, no. 12, pp. 985–994, 2016.
- [44] S. R. Gilani, L. J. Vuga, K. O. Lindell et al., "CD28 down-regulation on circulating CD4 T-cells is associated with poor prognoses of patients with idiopathic pulmonary fibrosis," *PLoS ONE*, vol. 5, no. 1, 2010.
- [45] H. Miyazaki, K. Kuwano, K. Yoshida et al., "The perforin mediated apoptotic pathway in lung injury and fibrosis," *Journal of Clinical Pathology*, vol. 57, no. 12, pp. 1292–1298, 2004.
- [46] J. C. Choy, A. Kerjner, B. W. Wong, B. M. McManus, and D. J. Granville, "Perforin mediates endothelial cell death and resultant transplant vascular disease in cardiac allografts," *The American Journal of Pathology*, vol. 165, no. 1, pp. 127–133, 2004.
- [47] S. Klunker, M. M. W. Chong, P.-Y. Mantel et al., "Transcription factors RUNX1 and RUNX3 in the induction and suppressive function of Foxp3+ inducible regulatory T cells," *The Journal of Experimental Medicine*, vol. 206, no. 12, pp. 2701–2715, 2009.
- [48] P. Du, W. A. Kibbe, and S. M. Lin, "lumi: a pipeline for processing Illumina microarray," *Bioinformatics*, vol. 24, no. 13, pp. 1547–1548, 2008.
- [49] N. L. Barbosa-Morais, M. J. Dunning, S. A. Samarajiwa et al., "A re-annotation pipeline for Illumina BeadArrays: Improving the interpretation of gene expression data," *Nucleic Acids Research*, vol. 38, no. 3, p. e1713, 2009.
- [50] J. D. Storey and R. Tibshirani, "Statistical significance for genomewide studies," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9440–9445, 2003.
- [51] T. J. Morris, L. M. Butcher, A. Feber et al., "ChAMP: 450k chip analysis methylation pipeline," *Bioinformatics*, vol. 30, no. 3, pp. 428–430, 2014.
- [52] W. S. Noble, "What is a support vector machine?" *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [53] T. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [54] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [55] W. Luo, M. S. Friedman, K. Shedden, K. D. Hankenson, and P. J. Woolf, "GAGE: generally applicable gene set enrichment for pathway analysis," *BMC Bioinformatics*, vol. 10, no. 1, article 161, 2009.
- [56] W. Luo and C. Brouwer, "Pathview: An R/Bioconductor package for pathway-based data integration and visualization," *Bioinformatics*, vol. 29, no. 14, pp. 1830–1831, 2013.
- [57] E. Hannon, H. Spiers, J. Viana et al., "Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci," *Nature Neuroscience*, vol. 19, no. 1, pp. 48–54, 2015.
- [58] F. K. Tan, X. Zhou, M. D. Mayes et al., "Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients," *Rheumatology*, vol. 45, no. 6, pp. 694–702, 2006.
- [59] Z. Brkic, L. Van Bon, M. Cossu et al., "The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis," *Annals of the Rheumatic Diseases*, vol. 75, no. 8, pp. 1567–1573, 2016.
- [60] M. Okrój, M. Johansson, T. Saxne, A. M. Blom, and R. Hesselstrand, "Analysis of complement biomarkers in systemic sclerosis indicates a distinct pattern in scleroderma renal crisis," *Arthritis Research & Therapy*, vol. 18, no. 1, article no. 267, 2016.
- [61] C. Scambi, S. Ugolini, T. Sakari Jokiranta et al., "The local complement activation on vascular bed of patients with systemic sclerosis: A hypothesis-generating study," *PLoS ONE*, vol. 10, no. 2, 2015.
- [62] J. Jarczák, E. M. Kościuczuk, P. Lisowski et al., "Defensins: natural component of human innate immunity," *Human Immunology*, vol. 74, no. 9, pp. 1069–1079, 2013.
- [63] N. Sakamoto, T. Kakugawa, A. Hara et al., "Association of elevated  $\alpha$ -defensin levels with interstitial pneumonia in patients with systemic sclerosis," *Respiratory Research*, vol. 16, no. 1, 2015.
- [64] J. Imgenberg-Kreuz, J. K. Sandling, J. C. Almlöf et al., "Genome-wide DNA methylation analysis in multiple tissues in primary Sjögren's syndrome reveals regulatory effects at interferon-induced genes," *Annals of the Rheumatic Diseases*, vol. 75, no. 11, pp. 2029–2036, 2016.
- [65] S. Kato, I. Kishiro, N. Ohnuma et al., "Suppressive effect of saprorelate hydrochloride on Raynaud's phenomenon and respiratory failure in patients with systemic sclerosis," *Respirology*, vol. 5, no. 1, pp. 27–32, 2000.
- [66] O. Kowal-Bielecka, K. Kowal, A. Lewszuk, A. Bodzenta-Lukaszyk, J. Walecki, and S. Sierakowski, " $\beta$  thromboglobulin and platelet factor 4 in bronchoalveolar lavage fluid of patients with systemic sclerosis," *Annals of the Rheumatic Diseases*, vol. 64, no. 3, pp. 484–486, 2005.
- [67] R. F. Macko, A. C. Gelber, and B. A. Young, "Increased circulating concentrations of the counteradhesive proteins SPARC and thrombospondin-1 in systemic sclerosis (scleroderma). Relationship to platelet and endothelial cell activation," *The Journal of Rheumatology*, vol. 29, no. 12, pp. 2565–2570, 2002.
- [68] J. D. Tozkır, H. Tozkır, H. Gürkan et al., "The investigation of killer cell immunoglobulin-like receptor genotyping in patients with systemic lupus erythematosus and systemic sclerosis," *Clinical Rheumatology*, vol. 35, no. 4, pp. 919–925, 2016.
- [69] T. A. Rauch, X. Wu, X. Zhong, A. D. Riggs, and G. P. Pfeifer, "A human B cell methylome at 100-base pair resolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 3, pp. 671–678, 2009.

- [70] M. P. Ball, J. B. Li, Y. Gao et al., "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells," *Nature Biotechnology*, vol. 27, no. 4, pp. 361–368, 2009.
- [71] H. Zhu, W. Mi, H. Luo et al., "Whole-genome transcription and DNA methylation analysis of peripheral blood mononuclear cells identified aberrant gene regulation pathways in systemic lupus erythematosus," *Arthritis Research & Therapy*, vol. 18, p. 162, 2016.
- [72] P.-S. Tsou, J. D. Wren, and M. A. Amin, "Histone deacetylase 5 is overexpressed in scleroderma endothelial cells and impairs angiogenesis via repression of proangiogenic factors," *Arthritis & Rheumatology*, vol. 68, pp. 2975–2985, 2016.
- [73] C. J. Wagner, C. Schultz, and M. A. Mall, "Neutrophil elastase and matrix metalloproteinase 12 in cystic fibrosis lung disease," *Molecular and Cellular Pediatrics*, vol. 3, no. 1, 2016.
- [74] L. Akkari, V. Gocheva, J. C. Kester et al., "Distinct functions of macrophage-derived and cancer cell-derived cathepsin Z combine to promote tumor malignancy via interactions with the extracellular matrix," *Genes & Development*, vol. 28, no. 19, pp. 2134–2150, 2014.
- [75] N. Iwamoto, S. Vettori, B. Maurer et al., "Downregulation of miR-193b in systemic sclerosis regulates the proliferative vasculopathy by urokinase-type plasminogen activator expression," *Annals of the Rheumatic Diseases*, vol. 75, no. 1, pp. 303–310, 2016.
- [76] N. Legány, G. Toldi, J. H. W. Distler et al., "Increased plasma soluble urokinase plasminogen activator receptor levels in systemic sclerosis: Possible association with microvascular abnormalities and extent of fibrosis," *Clinical Chemistry and Laboratory Medicine*, vol. 53, no. 11, pp. 1799–1805, 2015.
- [77] B. Ma, H. Cheng, R. Gao et al., "Zyxin-Siah2-Lats2 axis mediates cooperation between Hippo and TGF- $\beta$  signalling pathways," *Nature Communications*, vol. 7, article no. 11123, 2016.
- [78] A. Bianchi-Smiraglia, D. Kunnev, M. Limoge, A. Lee, M. C. Beckerle, and A. V. Bakin, "Integrin- $\beta$ 5 and zyxin mediate formation of ventral stress fibers in response to transforming growth factor  $\beta$ ," *Cell Cycle*, vol. 12, no. 21, pp. 3377–3389, 2013.
- [79] Y. Asano, H. Ihn, K. Yamane, M. Jinnin, Y. Mimura, and K. Tamaki, "Increased expression of integrin  $\alpha$ v $\beta$ 3 contributes to the establishment of autocrine TGF- $\beta$  signaling in scleroderma fibroblasts," *The Journal of Immunology*, vol. 175, no. 11, pp. 7708–7718, 2005.
- [80] F. Cevikbas, S. Seeliger, M. Fastrich et al., "Role of protease-activated receptors in human skin fibrosis and scleroderma," *Experimental Dermatology*, vol. 20, no. 1, pp. 69–71, 2011.
- [81] P. F. Mercer, X. Deng, and R. C. Chambers, "Signaling pathways involved in proteinase-activated receptor1-induced proinflammatory and profibrotic mediator release following lung injury," *Annals of the New York Academy of Sciences*, vol. 1096, pp. 86–88, 2007.
- [82] V. Riccieri, G. Parisi, A. Spadaro et al., "Reduced circulating natural killer T cells and gamma/delta T cells in patients with systemic sclerosis," *The Journal of Rheumatology*, vol. 32, no. 2, pp. 283–286, 2005.
- [83] L. Parolin Ercole, M. Malvezzi, A. C. Boaretti, S. Ramos da Rosa Utiyama, and A. Rachid, "Analysis of lymphocyte subpopulations in systemic sclerosis," *Journal of Investigational Allergology and Clinical Immunology*, vol. 13, no. 2, pp. 87–93, 2003.
- [84] T. Gambichler, C. Tigges, B. Burkert, S. Hoxtermann, P. Altmeyer, and A. Kreuter, "Absolute count of T and B lymphocyte subsets is decreased in systemic sclerosis," *European Journal of Medical Research*, vol. 15, no. 1, pp. 44–46, 2010.
- [85] M. Horikawa, M. Hasegawa, K. Komura et al., "Abnormal natural killer cell function in systemic sclerosis: altered cytokine production and defective killing activity," *Journal of Investigative Dermatology*, vol. 125, no. 4, pp. 731–737, 2005.
- [86] T. Momot, S. Koch, N. Hunzelmann et al., "Association of Killer Cell Immunoglobulin-Like Receptors with Scleroderma," *Arthritis & Rheumatology*, vol. 50, no. 5, pp. 1561–1565, 2004.
- [87] M. Mahmoudi, F. Fallahian, S. Sobhani et al., "Analysis of killer cell immunoglobulin-like receptors (KIRs) and their HLA ligand genes polymorphisms in Iranian patients with systemic sclerosis," *Clinical Rheumatology*, vol. 36, no. 4, pp. 853–862, 2017.
- [88] S. Kulkarni, Y. Qi, C. O'hUigin et al., "Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 51, pp. 20705–20710, 2013.
- [89] H. Li, R. Yang, X. Fan et al., "MicroRNA array analysis of microRNAs related to systemic scleroderma," *Rheumatology International*, vol. 32, no. 2, pp. 307–313, 2012.
- [90] N. Vince, H. Li, V. Ramsuran et al., "HLA-C Level Is Regulated by a Polymorphic Oct1 Binding Site in the HLA-C Promoter Region," *American Journal of Human Genetics*, vol. 99, no. 6, pp. 1353–1358, 2016.
- [91] R. White, C. Gonsior, N. M. Bauer, E.-M. Krämer-Albers, H. J. Luhmann, and J. Trotter, "Heterogeneous nuclear ribonucleoprotein (hnRNP) F is a novel component of oligodendroglial RNA transport granules contributing to regulation of myelin basic protein (MBP) synthesis," *The Journal of Biological Chemistry*, vol. 287, no. 3, pp. 1742–1754, 2012.
- [92] J. Kamholz, J. Toffenetti, and R. A. Lazzarini, "Organization and expression of the human myelin basic protein gene," *Journal of Neuroscience Research*, vol. 21, no. 1, pp. 62–70, 1988.