# Genome-wide DNA methylation analysis in systemic sclerosis reveals hypomethylation of interferon-induced genes in CD4+ and CD8+ T cells

**Abstract**

**Objectives:** Epigenetic modification, especially DNA methylation, plays an important role in the pathogenesis of autoimmune disease. In the present study, we tried to identify the roles of epigenetic variations (DNA methylation) in Systemic sclerosis (SSc).

**Method:** Whole genome DNA methylation analysis to CD4+ and CD8+ T cells from 24 SSc and 24 matched normal individuals were conducted by Illumina HumanMethylation450 microarray. Selected differential methylated regions (DMRs) were validated by NGS-based target bisulfite sequencing.

**Results:** We identified hundreds of T cell subtype (CD4+ and CD8+) specific and consistent differential methylation regions in the discovery stage. Type I Interferon related pathways were significantly enriched by both CD4+ and CD8+ related DMRs indicating the coordination role in the pathogenesis of SSc. In the validation stage, we confirmed the global hypo-methylation signature of the type I interferon (IFN)-associated genes both in CD4+ and CD8+ T cells and we demonstrate that high concentration of IFN-α and IFN-β in serum would lead to the hypomethylation so as to increase its gene expression of the IFN related genes in PBMC.

**Conclusions:**

Our result highlights the dysfunctions of type I IFN pathway in the epigenetic level of SSc patients, and the hypo-methylation of the type I IFN related genes might be of great importance in the pathogenesis of SSc and other autoimmune diseases.

**Key words: DNA methylation, CD4+, CD8+, Systemic sclerosis, Type I Interferon**

**Introduction**

Systemic sclerosis (Scleroderma, SSc) is a complex systemic autoimmune disease caused by complicated interaction between genetic, epigenetic and environment risk factors with a prevalence rate per year is 28-253 per million in USA [[1](#_ENREF_1)]. Limited cutaneous (lcSSc) and diffuse cutaneous (dcSSc) were sub-grouped depending on skin thickening involves areas proximal to or beyond the elbows and knees[[2](#_ENREF_2)]. Similar with other autoimmune disease, such as systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), autoantibodies were also frequently shown in SSc patients which are demonstrated to be related with disease subtype and progress of the therapy. In the past decades, a multitude of classic and molecular epidemiological studies have been conducted and successfully identified larger number of environmental and genetic related susceptibility factors[[3](#_ENREF_3)], genes [[4-9](#_ENREF_4)] as well as copy number variations [[10](#_ENREF_10)]. Meanwhile, evidence shown epigenetic modification, especially DNA methylation, plays an important part in temporal and spatial regulating the gene expression and the pathogenesis of large number of complex diseases, including most frequently autoimmune disease such as RA[[11](#_ENREF_11), [12](#_ENREF_12)], SLE [[13](#_ENREF_13), [14](#_ENREF_14)]. However, epigenetic variations (DNA methylation) in SSc have not been comprehensively investigated, especially for subtype of the immune cells separately. In addition, SSc subtype and autoantibodies related DNA methylation changes also need to be identified to improve the understanding of pathogenesis of different subtype.

In present study, we conducted the whole genome DNA methylation analysis to explore the epigenetic changes in CD4+ and CD8+ T cells of SSc patients (24 SSc and 24 normal) with Illumina 450k HumanMethylation microarray. A comprehensive SSc related differential methylation regions (DMRs) and regions were identified. We also identified immune pathway and other metabolic process was significantly enriched by Gene Ontology analysis. We validated the the most interesting ~60 DMRs derived from microarray result with another independent dataset (43 SSc and 41 normal) by NGS-based targeted bisulfite sequencing and we found a global hypo-methylation signature of the type I interferon (IFN)-associated genes both in CD4+ and CD8+ T cells of SSc patients. In accordance, we also found a significant increase level of IFN-α and IFN-β in serum of SSc patients. *In vitro* studies have confirmed that type I IFN could stimulate and increase the expression of TGF-β in PBMC, which might further contribute to the fibrosis of SSc. These results highlight the dysfunctions of type I IFN pathway in the epigenetic level of SSc patients, and the hypo-methylation of the type I IFN related genes might be of great importance in the pathogenesis of SSc and other autoimmune diseases.

**Materials and methods**

## Study subjects

56 SSc patients (age: 48.6 ± 12.8 years) were recruited from the outpatient dermatology clinic and impatient ward of Shanghai TCM-integrated Hospital. All patients met the criteria for SSc established by the American College of Rheumatology [[2](#_ENREF_2)]. 54 age and gender matched control samples (age: 44.8 ± 10.2 years) were recruited from Shanghai Changning District Tongren Hospital and Affiliated Hospital of Nantong University. None of the control subjects had a family history of autoimmune disorders. The clinical profiles of the SSc patients as well as the normal subjects are shown in Table 1. Study protocol was approved by the human ethics committee of Fudan University. All subjects have signed the informed consent form.

## Isolation of CD4+ and CD8+ T cells from peripheral blood mononuclear cells (PBMCs)

Venous peripheral blood (about 10 ml) was collected from all study subjects and preserved with 2 mg/ml EDTA·K2. PBMCs were isolated by density-gradient centrifugation (Ficoll-PaqueTM PLUS; GE Healthcare Bio, Chicago, IL, USA). CD4+ and CD8+ T cells were separated by positive selection, using CD4 and CD8 magnetic micro beads according to the manufacturer’s protocol (Miltonic Biotec, Bergisch Glad Bach, Germany). Extraction of genomic DNA was performed using the Pure Gene Blood Core Kit (Qiagen, Valencia, California, USA) as described in the manufacturer’s protocol.

## Differential methylation analysis for CD4+ and CD8+ T cell

Genome-wide DNA methylation of CD4+ and CD8+ T cells from SSc patients and controls was assessed using the Illumina Infinium HumanMethylation450 Bead Chip microarray, which allows for the interrogation of over 485,000 CpG sites across the genome. This array covers 99% of Refuses genes, with an average of 17 CpG sites per gene across the promoter region, 5’ untranslated region (5’-UTR), first exon, gene body and 3’-UTR. It also covers 96% of CpG islands. Non-CpG-methylated sites recently identified in human stem cells and microRNA promoter regions are also covered [[15](#_ENREF_15)]. DNA methylation analysis was conducted using R software and RnBeads package in Bioconductor [[16](#_ENREF_16), [17](#_ENREF_17)]. In initial quality control and preprocessing steps, we firstly filtered out the probes whose last 3 bases of its target sequence overlap with SNP; then we removed the cross-reactive probes whose sequence maps to multiple genomic coordinates. After that, we conducted the “greedycut” procedure and filtered out the probes whose detection P-value >0.01 in any of the samples. In addition, probes on sex chromosomes, with many missing values and probes of non-CpG context were also filtered out. After initial preprocessing step, we conducted the background subtraction using the “noob” method. The methylation beta values were further normalized using the “BMIQ” normalization method [[18](#_ENREF_18)].

Prior to the finding of differentially methylated sites, we conducted the Combat method using sva package to remove batch effect. After batch effect correction, probes with IQR <= 0.1 were considered as little variation and thus were filtered out. We then computed the p-values of the remaining probes with a linear regression model adding the age and gender as two covariates. Differentially methylated CpG sites were defined as sites with P-value < 0.05. Furthermore, the DMR analysis was conducted with RnBeads package and Gene Ontology analysis was also conducted with GOstats package in R using the top 500 DMRs [[19](#_ENREF_19)].

## DMR validation based on NGS-based target methylation sequencing

CD4+ and CD8+ T cells from another 43 SSc samples as well as 41 normal controls were collected, and genomic DNA were converted with bisulfite treatment. The net-PCR was performed to amplify the targeted DNA sequence involved in DMR. Due to the technical failures, some of the targeted regions were not detected. Then the designed DNA fragments were sequenced by Illumina Hiseq 2000. Bsseeker2 was then utilized for mapping bisulfite treated reads as well as for methylation calling [[20](#_ENREF_20)].

## Gene expression regulation by DNA methylation in CD4+ and CD8+ T cells

Total RNA was extracted from cells using TRizol reagent (Invitrogen, USA) according to the manufacturer's instructions and reverse transcribed using MultiScribe reverse transcriptase kit (ABI, Foster city, CA, USA) according to the manufacturer's instructions. 4 μl of cDNA template was used as for 10 μl Quantitative real-time RT-PCR, using 5 μl of SYBR Green 1x RT-PCR mix (Qiagen) and 1μl of each primer. Reactions were performed on an ABI Q7 Sequence Detector System (ABI), with the following parameters: 94 °C, 3 min; 60 °C, 40 secs, 38 cycles. Melt curves were analyzed from 72 to 88 °C to verify homogeneity.

## T cell culture and stimulation

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood through density-gradient centrifugation (Ficoll-Paque Plus, GE Healthcare), then resuspended in RPMI medium supplemented with 10 % FBS, 100 U/ml penicillin, and 100 μg/ml streptomycin, 2 μl leukocyte Activation Cocktail, with BD GolgiPlug (BD, Franklin Lakes, NJ, USA), transferred onto a 48-well plate (2 × 106 cells/well) and cultured at 37 °C for 4~6 h. To prevent non-specific staining, PBMCs were preincubated with anti-human Fc blocking antibody (0.5 mg/ml; 564219, BD Pharmingen™) for 15 minutes. Subsequently, cells were treated with 50 ng/ml purified human recombinant protein IFN-α4 (rIFN-α4) and/or 100 ng/ml rIFN-β for 6~24 h (Sino biological, Beijing, China).

## Detection of cytokines concentration by [enzyme](javascript:void(0);) [linked](javascript:void(0);) [immunosorbent](javascript:void(0);) [assay](javascript:void(0);)

The concentration of IFN-α, IFN-β (PBL Assay Science, Piscataway, NJ, USA), and TGF-β1 (Abcam, San Francisco, USA) in human serum or cell culture supernatant were all determined by [enzyme](javascript:void(0);) [linked](javascript:void(0);) [immunosorbent](javascript:void(0);) [assay](javascript:void(0);) (ELISA) according to the instructions of the manufacturers. The concentration was then calibrated according to the standard dose response curve.

## Statistical analysis

Pearson’ correlation test was conducted to test the associations between gene expression and methylation as well as the serum IFN concentration. In the validation analysis, the p-value was calculated with the wilcoxon rank-sum test. And FDR (false discovery rate) was applied for the multiple test correction. The CpG sites with pvalues <0.05 after FDR adjustment were then considered as significantly differential methylated in the validation dataset. All the statistical analysis was conducted with R version 3.2.1 [[16](#_ENREF_16)].

# Results

## Genome-wide DNA methylation profile of CD4+ and CD8+ in normal and SSc patients

To investigate the genome-wide DNA methylation difference for CD4+ and C8+ T cells between SSc patients and normal individuals. We collected 24 SSc patients as well as 24 age and gender matched normal controls and then CD4+ and CD8+ T cells were isolated from PBMC. Detailed clinical characteristic of the SSc patients included, such as subtype and autoantibody phenotype, were shown in Table 1.CD4+ and CD8+ T cells were purified from each sample and were further confirmed with our PCA analysis incorporating more cell types from additional datasets (Supplementary Figure 1).

After stringent preprocessing and fitering, we identified 1026 and 2613 DMS in CD4+ and CD8+ T cells, respectively (Figure 1A and 1C). Cluster analysis was also conducted based on the DMSs and showed a strong discrimination between SSc patients and controls (Figure 1B and 1D). In the CD4+ T cell methylation dataset, 286 of these sites were hypo-methylated while 740 were hyper-methylated. These probes were distributed across 621 genes, and 440 genes were hyper-methylated and 175 genes were hypo-methylated. In addition, 6 genes showed mixed methylation status. Similarly, in the methylation dataset of CD8+ T cell, 1,830 and 783 probes were hyper-methylated and hypo-methylated respectively in SSc patients (Supplementary Table 1-2). CD8+ derived DMS were distributed across 1,567 genes, including 1,086 genes with only hyper-methylated sites, 412 genes with only hypo-methylated sites and 69 genes with mixed methylation pattern of CpG sites. In addition, we also annotated these DMS and obtaining their genomic coordinates, further compared with the functional elements such as CpG island and TSS (transcription start site) and summarized in Supplementary Table 3.

In the comparative analysis to test the methylation patterns between CD4+ T cells and CD8+ T cells, we found 330 of the DMS were shared, and 696 DMS were unique to CD4+ T cells while 2,283 DMS were unique to CD8+ T cells. 330 shared DMS were all showing the same directions (hyper or hypo) in CD4+ and CD8+ T cells and were distributed across 215 genes. Among them, 138 were hyper-methylated, 75 were hypo-methylated, and 2 genes were displaying mixed methylation pattern (Supplementary Figure 2).

Previous GWAS studies have found several susceptible loci associated with SSc. Due to the tight correlation between the gene functions and its gene methylaiton, it is of importance to explore the methylation status of these susceptible genes. We then found 16 verified genes through all of the SSc GWAS studies and obtained the methylation dataset of the CpG sites located at these genes (Supplementary Table 4). In the CD4+ T cell dataset, we found that the most significantly differential methylated CpG site was located at SOX5 (cg04626868, p=0.00083). While in the CD8+ T cell dataset, the most significantly differential methylated CpG site was located at DNASE1L3 (cg02046562, p=7.09×10-6). However, the fold changes of these DMS in both CD4+ and CD8+ T cell datasets were relatively low, indicating that the exploration of the actual meanings and changes of methylation status of thse susceptible genes need further research with more advanced detection methods.

To identify the biological gene ontologies influenced by the differential methylation pattern in CD4+ and CD8+ T cells between SSc patients and controls, we conducted the differential methylation enrichment analysis based on the top 500 differential methylated regions (DMRs) implemented in RnBeads package. Interestingly, we found that in both CD4+ and CD8+ subgroups, GO terms of response to type I interferon (IFN)and type I interferon signaling pathway were enriched in the significantly hypo-methylated regions, indicating that hypo-methylation status of type I IFN-associated genes were shared in both CD4+ and CD8+ T cells and might play a role in the pathogenesis of SSc (Figure 2 A-B). While in the significantly hyper-methylated regions, no shared GO term was identified in CD4+ and CD8+ T cells.

To further confirm our findings, we then used the targeted bisulfite sequencing method based on the next generation sequencing (NGS) to validate the 62 differentially methylated CpG sites between SSc patents and controls in CD4+ and CD8+ T cells based on the p-values and its functions (Supplementary Table 5). 43 SSc patients and 41 normal controls were obtained for validation, including 13 patients and 12 controls from the first stage to assess the consistency between the two stages. Among these CpG sites, 57 were successfully investigated while the remainders failed due to technical limitations. Due to the advantages of targeted bisulfite sequencing, we identified 154 additional CpG sites close to the selected significantly methylated CpG sites (Supplementary Table 6). In order to evaluate the consistency between the microarray and targeted bisulfite sequencing results, we conducted the correlation analysis with the shared samples in two stages. It turned out that the methylation ratios from two different platforms were almost linear and the Pearson’s correlation coefficients were 0.86 and 0. 90 for CD4+ T cells and CD8+ T cells respectively (Figure 2C), showing a good consistency between the two platforms. Of these 57 CpG sites, we found that 7 and 9 CpG sites were validated to be significantly differential methylated between SSc patients and normal controls in CD4+ T and CD8+ T cells respectively (Table 2). In addition to the selected 57 CpG sites, we also investigated the 157 nearby sites and identified 1 and 10 significantly differential methylated sites between SSc patients and controls in CD4+ and CD8+ T cells, respectively. And all of the 11 CpG sites were located at promoter region of MX1 gene (Figure 2D). Based on the validation dataset, it turned out that the CpG sites located at interferon-related genes were all validated while the methylation status of the other CpG sites were not consistent between the discovery and validation stage, and we then focused our further analysis on sites and genes related to the type I interferon pathway.

## DNA methylation biomarker in SSc subtype and autoantibody phenotype

SSc patients with different subtype and autoantibody phenotype will have different prognosis and clinical manifestations. As a result, we would like to explore the differential methylation status in different subtype and autoantibody. We then conducted the differential methylation analysis with the same protocol as described above (Supplementary Table 7-8). Based on our criteria, we found that there were 307 and 382 DMS in comparing the dcSSc and lcSSc patients in CD4+ and CD8+ T cells, respectively. Moreover, we found that four of the top five DMS (located at SDF4, COLEC11, FAM132A, ZBTB46) were shared in CD4+ and CD8+ T cells. In order to evalute its role in discrimating dcSSc and lcSSc, we performed a logistic regression based on these four DMS and found that the AUC of the prediction model is 0.98 and 0.99 for CD4+ and CD8+ T cells, indicating that these CpG sites might be of potential to be the biomarkers for SSc (Supplementary Figure 3).

In addition to the subtype, we also compared the different methylation status in different autoantibodies. In the comparison between ACA+ and ACA- patients, we found 460 and 304 DMS in CD4+ and CD8+ T cells, respectively. Moreover, the CpG site located at IL17RA showd strongest difference both in CD4+ and CD8+ T cells. While comparing the methylation status of ATA+ and ATA- patients, we also found 848 and 304 DMS in CD4+ and CD8+ T cells, respectively. The CpG sites located at NEGR1 and GPR85 showed strongest difference in CD4+ and CD8+ T cells, respectively. Interestingly, few of the top DMS in CD4+ and CD8+ T cells were shared which is different from the result of comparison between dcSSc and lcSSc as well as the comparison between ACA+ and ACA-.

## DNA methylation and type I interferon-induced gene expression as well as its correlation

It is widely acknowledged that there is a strong correlation between gene expression and methylation. To assess the gene expression of these differential methylated genes, we performed q-PCR experiments in CD4+ and CD8+ T cells from 23 SSc patients and 20 healthy controls. The mRNA levels of type I IFN-associated genes, IFI44L, IFITM1, MX1, and PARP9 showed significantly increased expression compared to those in controls in both CD4+ and CD8+ T cells, while EIF2AK2 showed increased expression in cases only in CD4+ T cells (Figure 3A-B). Pearson’s correlation test was conducted to assess the correlation between gene expression and methylation (Supplementary Table 9). In the combined analysis of CD4+ and CD8+ T cells datasets, we found that the methylation status of three CpG sites located at IFI44L gene showed strong inverse correlation with its expression (Figure 3C-D). And other 5 CpG sites located at PARP9, MX1 and IFI44L showed marginal significant correlations. Based on these datasets, we concluded that the expression of type I IFN-associated genes were significantly higher in SSc patients both in CD4+ and CD8+ T cells, and the correlation analysis between DNA methylation and the expression of IFN-associated genes suggested that the DNA methylation change might be pivotal to the regulation of IFN-associated genes.

In addition to the methylation and expression profile detection, we also measured the concentration of type I interferon (mainly refer to IFN-α and -β) in the serum of SSc patients and normal controls with ELISA. Concordantly, significantly increased concentration of IFN-α (Figure 4A) and IFN-β (Figure 4B) were shown in SSc patients.

As it is widely acknowledged that high level of type I IFN in serum will lead to the higher expression of its induced genes, we conducted the correlation analysis to test the association between serum type I IFN level and the gene expression status of their associated genes. We found that the expression of IFITM1 as well as the IFI44L gene in CD4+ T cell showed significantly positive correlation between serum IFN α level and its expression, while the expression of IFITM1, EIF2AK2, PARP9 and MX1 gene showed strong positive correlation between IFN β and its expression (Figure 4 C-E). However, in the CD8+ T cell datasets, we didn’t found any significant association between gene expression and the serum IFN α/β level of these 5 genes (Supplementary Table 10).

Correlation analysis between the gene methylation and serum concentration of type I IFN was also conducted. In CD4+ T cells, we found the site located at IFITM1 gene was inversely correlated with serum IFN- α level with a marginal significant p-value. Two sites in PARP9 and one site located at EIF2AK2 were found to be inversely correlated with serum IFN- β level in CD8+ T cells (Supplementary Table 11).

Based on our experiments, we proposed that higher level of serum IFN-α and -β may affect the methylation as well as the expression of its induced genes, further affecting the functions of T cells and other immune cells. In addition to the increase numbers and activity of Treg cells, it is of great importance to test if type I IFN stimulus could affect the expression of TGF-β, further affecting the fibrosis of SSc. It is shown that in the IFN -α and -β treatment groups, significantly higher level of TGF-β mRNA and higher concentration of TGF-β protein was observed (Figure 5).

**Discussion**

Previous study had detected the global methylation status in SSc fibroblasts and found some of the abnormalties [[21](#_ENREF_21)]. However, the methylation status of immune cells, which are of great importance for the pathogenesis of autoimmune diseases, still remains unknown. Here, we integrated whole-genome DNA methylation microarray and targeted bisulfite sequencing methods to determine the global DNA methylation status in CD4+ and CD8+ T cells of SSc patients. We found that the global hypo-methylation of type I IFN signaling pathway associated genes was shared in both CD4+ and CD8+ T cells of SSc patients. Increased expression of type I IFN induced genes and serum concentration of type I IFN were both identified in the SSc patients. It is reasonable to presume that the elevated type I interferon could be the cause for the hypo-methylation as well as the increased expression of its induced genes. And we further confirmed that type I IFN stimulus could promote the secretion of TGF-β, which might finally contribute to the pathogenesis of SSc.

We noticed W Lei et al found a global hypo-methylation in CD4+ T cells from SSc patients with the Methylamp™ Global DNA Methylation Quantification Kit, which quantified the global methylation level with ELISA method [[22](#_ENREF_22)]. Interestingly, in our study, we found no significant hyper/hypo-methylation in CD4+ T cells from SSc patients using Illumina HumanMethylation 450K microarry. This discordance might be caused by the different methods of methylation measurement.. In addition, the sporadic coverage of the whole genome CpG sites of HumanMethylation 450K microarry may also lead to this discrepancy. More comprehensive detection methods like RRBS and WGBS are urgently needed for drawing a better conclusion.

Type I Interferon has been recently considered to contriture largely to the pathogenesis of several autoimmune diseases through the genetic and epigenetic studies. Patrick Coit et al and Devin M. Absher et al both found a hypo-methylation signature in SLE patients, suggesting its importance in the pathogenesis of SLE[[13](#_ENREF_13), [14](#_ENREF_14)]. And Juliana Imgenberg-Kreuz et al also found the hypo-methylation status of interferion related genes in pSS [[23](#_ENREF_23)]. Moreover, Brandon W Higgs et al found a common activated type I interferon pathway in five autoimmune diseases with gene expression microarray [[24](#_ENREF_24)]. Furthermore, many of the susceptible genes identified in SSc, SLE and RA were belong to the type I interferon pathway, such as the STAT4, IFR5 etc [[25-30](#_ENREF_25)]. Another study also demonstrated that the activated type I IFN system in several systemic autoimmune diseases including SSc and Sjogren syndrome may contribute to the vascular pathology and involve in the profibrotic process [[31](#_ENREF_31)]. As a result, it is necessary to examine if the methylation alteration of type I IFN associated genes were also shown in other kinds of autoimmune diseases. We then downloaded the datasets of SLE and RA from GEO database (GSE42861 and GSE59250). And we analyzed the significantly differential methylated genes of SSc in these two autoimmune diseases. As expected, we found that these genes were shown to be strongly hypo-methylated in the CD4+ T cells of SLE as well as in the peripheral blood leukocyte of RA (Supplementary Table 12). These consistent hypo-methylation of type I IFN associated genes in three autoimmune diseases indicated that type I IFN hypo-methylation might be a common event in some of the pathogenesis of autoimmune diseases, and might be of importance for the drug discovery of autoimmune diseases.

In our study, we used the Illumina HumanMethylation 450K microarray for whole-genome methylation status detection. And it only covers about 2% of the total CpG sites in human genome. Due to the sparse density of the detected CpG sites, we failed to determine the methylation status for all the CpG sites of the type I IFN associated genes. Meanwhile, type I IFN is a key component for innate immunity and plays an importance role in the immune system, which means that it may affect many other cell types, such as NK cells, B cells and monocytes etc. In our analysis, we focused only on type I IFN effect on T cells, while the other kinds of cells may also play a role in the pathogenesis of SSc. As a result, further studies may concentrate on the effect of type I IFN on the other cell types. Further, in our analysis, we found that the type I interferon stimulus could change the methylation status as well as the expression level of its induced genes in several cell types. And we found the up-regulated TET and down-regulated DNMT occurred after the type I interferon stimulus. We speculated that the change of methylation status by type I IFN is regulated by the balance of TET and DNMT. Further work is needed to answer the mechanism between them.

# Acknowledgement

Thanks to Pro. Guang-jie Chen of Shanghai Jiaotong University for technical help in T cells separation. This work was supported by the National Natural Science Foundation of China (grant No)

**References**

1. Mayes MD, Lacey JV, Jr., Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, Schottenfeld D. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum. 2003; 48: 2246-55. doi: 10.1002/art.11073.

2. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, Matucci-Cerinic M, Naden RP, Medsger TA, Jr., Carreira PE, Riemekasten G, Clements PJ, Denton CP, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis. 2013; 72: 1747-55. doi: 10.1136/annrheumdis-2013-204424.

3. Mora GF. Systemic sclerosis: environmental factors. J Rheumatol. 2009; 36: 2383-96. doi: 10.3899/jrheum.090207.

4. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, Shin ES, Reveille JD, Mayes MD, Kim JH, Song R, Choi JY, Park JA, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum. 2009; 60: 3807-14. doi: 10.1002/art.24982.

5. Radstake TRDJ, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, Van Riel PLCM, Van T Slot R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nature Genetics. 2010; 42: 426-9. doi:

6. Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, Matucci-Cerinic M, Riemekasten G, Airo P, Melchers I, Hachulla E, Cusi D, Wichmann HE, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet. 2011; 7: e1002091. doi: 10.1371/journal.pgen.1002091.

7. Bossini-Castillo L, Martin JE, Broen J, Gorlova O, Simeon CP, Beretta L, Vonk MC, Callejas JL, Castellvi I, Carreira P, Garcia-Hernandez FJ, Castro MF, Coenen MJH, et al. A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. Human Molecular Genetics. 2012; 21: 926-33. doi:

8. Martin JE, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, van 't Slot R, Rodriguez-Rodriguez L, Vicente E, Fonollosa V, Ortego-Centeno N, Gonzalez-Gay MA, Garcia-Hernandez FJ, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Hum Mol Genet. 2012; 21: 2825-35. doi: 10.1093/hmg/dds099.

9. Yi L, Wang JC, Guo XJ, Gu YH, Tu WZ, Guo G, Yang L, Xiao R, Yu L, Mayes MD, Assassi S, Jin L, Zou HJ, et al. STAT4 is a genetic risk factor for systemic sclerosis in a Chinese population. Int J Immunopathol Pharmacol. 2013; 26: 473-8. doi:

10. Guo S, Li Y, Wang Y, Chu H, Chen Y, Liu Q, Guo G, Tu W, Wu W, Zou H, Yang L, Xiao R, Ma Y, et al. Copy Number Variation of HLA-DQA1 and APOBEC3A/3B Contribute to the Susceptibility of Systemic Sclerosis in the Chinese Han Population. The Journal of Rheumatology. 2016. doi: 10.3899/jrheum.150945.

11. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotech. 2013; 31: 142-7. doi: 10.1038/nbt.2487

<http://www.nature.com/nbt/journal/v31/n2/abs/nbt.2487.html#supplementary-information>.

12. Guo S, Zhu Q, Jiang T, Wang R, Shen Y, Zhu X, Wang Y, Bai F, Ding Q, Zhou X. Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis. Modern Rheumatology. 2016: 1-7. doi:

13. Coit P, Jeffries M, Altorok N, Dozmorov MG, Koelsch KA, Wren JD, Merrill JT, McCune WJ, Sawalha AH. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naive CD4+ T cells from lupus patients. J Autoimmun. 2013; 43: 78-84. doi: 10.1016/j.jaut.2013.04.003.

14. Absher DM, Li X, Waite LL, Gibson A, Roberts K, Edberg J, Chatham WW, Kimberly RP. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet. 2013; 9: e1003678. doi: 10.1371/journal.pgen.1003678.

15. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011; 3: 771-84. doi: 10.2217/epi.11.105.

16. Team RC. R: A language and environment for statistical computing. 2013. doi:

17. Assenov Y, Müller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nature methods. 2014; 11: 1138-40. doi:

18. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013; 29: 189-96. doi:

19. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007; 23: 257-8. doi:

20. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen P-Y, Pellegrini M. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC genomics. 2013; 14: 1. doi:

21. Lei W, Luo Y, Lei W, Luo Y, Yan K, Zhao S, Li Y, Qiu X, Zhou Y, Long H. Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. Scandinavian journal of rheumatology. 2009; 38: 369-74. doi:

22. Lei W, Luo Y, Lei W, Luo Y, Yan K, Zhao S, Li Y, Qiu X, Zhou Y, Long H, Zhao M, Liang Y, Su Y, et al. Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. Scand J Rheumatol. 2009; 38: 369-74. doi: 10.1080/03009740902758875.

23. Imgenberg-Kreuz J, Sandling JK, Almlöf JC, Nordlund J, Signér L, Norheim KB, Omdal R, Rönnblom L, Eloranta M-L, Syvänen A-C. 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. 2016: annrheumdis-2015-208659. doi:

24. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, Brohawn P, Kiener PA, Richman L, Fiorentino D. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Annals of the rheumatic diseases. 2011; 70: 2029-36. doi:

25. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PIW, Le JM, Lee H-S, Batliwalla F, Li W, Masters SL, Booty MG, et al. STAT4 and the Risk of Rheumatoid Arthritis and Systemic Lupus Erythematosus. New England Journal of Medicine. 2007; 357: 977-86. doi: doi:10.1056/NEJMoa073003.

26. Han J-W, Zheng H-F, Cui Y, Sun L-D, Ye D-Q, Hu Z, Xu J-H, Cai Z-M, Huang W, Zhao G-P, Xie H-F, Fang H, Lu Q-J, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009; 41: 1234-7. doi: <http://www.nature.com/ng/journal/v41/n11/suppinfo/ng.472_S1.html>.

27. Abelson A-K, Delgado-Vega AM, Kozyrev SV, Sanchez E, Velazquez-Cruz R, Eriksson N, Wojcik J, Reddy ML, Lima G, D’Alfonso S. STAT4 associates with systemic lupus erythematosus through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk. Annals of the rheumatic diseases. 2009; 68: 1746-53. doi:

28. Carmona FD, Martin J-E, Beretta L, Simeón CP, Carreira PE, Callejas JL, Fernández-Castro M, Sáez-Comet L, Beltrán E, Camps MT. The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. PLoS One. 2013; 8: e54419. doi:

29. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FA, Zhernakova A, Hinks A. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature genetics. 2010; 42: 508-14. doi:

30. Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Rönnelid J, Klareskog L, Huizinga TW. Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. Arthritis & Rheumatism. 2007; 56: 2202-10. doi:

31. Eloranta ML, Franck-Larsson K, Lovgren T, Kalamajski S, Ronnblom A, Rubin K, Alm GV, Ronnblom L. Type I interferon system activation and association with disease manifestations in systemic sclerosis. Ann Rheum Dis. 2010; 69: 1396-402. doi: 10.1136/ard.2009.121400.

# Figure legends

**Figure 1** Manhattan plots and PCA plots of the genome-wide DNA methylation analysis in CD4+ and CD8+ T cells from patients with SSc and control individuals. The −log10 (p) values of the association of each CpG site against the chromosomal position of the investigated sites were shown in Manhattan plots. Some of the most significantly differential methylated CpGsites located at interferon-related genes were highlighted by their gene symbol annotation. PCA plot was conducted with the methylation status of the DMS in CD4+ and CD8+ T cells.

**Figure 2** The Gene Ontology analysis and the validation of the most significant differential methylated CpGsites. (A-B) represents the 10 of the most significantly enriched gene ontologies in CD4+ and CD8+ T cells based on the DMRs, respectively. (C) represents the correlation analysis between the results of 450k methylation microarray and the targeted bisulfite sequencing in validation. (D) represents the differential methylation status of MX1 gene in CD8+ T cells.

**Figure 3** The expression analysis of the interferion-related genes and its correlation with gene methylaiton. (A-B) represents the gene expression profiles of the five interferion-related genes in CD4+ and CD8+ T cells, respectively. (C) showed the significantly inverse correlation between the methylation of a CpGsite located at IFI44L (chr:79085586) and the expression of IFI44L. (D) showed the significantly inverse correlation between the methylation of a CpGsite located at PARP9(chr:122281975) and the expression of PARP9.

**Figure 4** The elevated expression of IFN-α and -β in serum of the SSc patients and its correlation with the gene expression in CD4+ T cells. (A-B) shows the serum concentrations of the IFN-α and -β in SSc patients as well as in the normal controls. (C-E) represents the positively correlation between the serum concentrations of IFN-β and the gene expression of IFITM1, EIF2AK2 and PARP9 in CD4+ T cells.

**Figure 5** The expression of TGF-β with or without the stimulus of IFN-α and -β in PBMC of the healthy donors. HD represents the expression level of TGF-β without any stimulus in healthy donors, and PMA is the stimulus for synthesizing TGF-β.

**Supplementary Figure 1** The PCA analysis for validating the cell types. The datasets of CD4+ and CD8+ T cells were obtained from the dataset of GSE59065. And PBL represents the peripheral blood leucocytes, which is downloaded from the normal subsets of the GSE42861 dataset. WB represents the whole blood. WB and CD16 datasets were all originated from the normal samples of GSE63499 dataset. SSc-CD4 and SSc-CD8 represents the methylation data of CD4+ and CD8+ T cell of our own datasets.

**Supplementary Figure 2** The Venn diagram showing the distribution of DMS in CD4+, CD8+ as well as the shared between them.

**Supplementary Figure 3** The AUC curve of the logistic regression models to discriminate dcSSc and lcSSc with the methylation status of 4 DMS in CD4+ and CD8+ T cells. Figure A and B represent the AUC curve of the CD4+ and CD8+ methylation datasets, respectively.

# Tables

**Table 1 Characteristics of patients with systemic sclerosis (SSc) and control individuals**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Discovery Stage** | |  | **Validation Stage** | |
|  | **Patients with SSc** | **Controls** |  | **Patients with SSc** | **Controls** |
| Individuals number | 24 | 24 |  | 43 | 41 |
| Women | 17 (70.8%) | 19 (79%) |  | 32 (74.4%) | 10 (24.4%) |
| Age, year, (±SD) | 47.42±12.47 | 46.91±11.00 |  | 46.94±10.70 | 44.29±10.40 |
| Subtype (%) |  |  |  |  |  |
| dcSSc | 11 (45.8%) | - |  | 17 (40%) | - |
| lcSSc | 13 (54.2%) | - |  | 26 (60%) | - |
| Autoantibody (%) |  |  |  |  |  |
| ANA | 20 (83.3%) | - |  | 39 (90.7%) | - |
| ATA | 16 (66.7%) | - |  | 25 (58.1%) | - |
| ACA | 8 (33.3%) | - |  | 12 (27.9%) | - |

‘-’ denotes not available. Methylation status of samples in discover stage was obtained by HM450 microarray while methylation status of samples in discover stage was obtained by NGS-based methylation target sequencing.

**Table 2 The methylation status of validated CpG sites of Interferon-related genes between SSc patients and control individuals in CD4+ and CD8+ T cells.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Differential methylated sites (DMS)** | | | |  | **CD4+** | | |  | **CD8+** | | |
| **Genomic Coordinates** | **Gene** | **CpG**  **island** | **Gene property** |  | **McaM** | **McoM** | **Pval** |  | **McaM** | **McoM** | **Pval** |
| chr3: 122281975 | PARP9 | N\_Shore | 5'UTR |  | 0.43 | 0.53 | **6.93×10-5** |  | 0.45 | 0.60 | **3.05×10-9** |
| chr1: 79085586 | IFI44L | - | TSS1500 |  | 0.67 | 0.74 | **9.10×10-5** |  | 0.71 | 0.76 | **1.79×10-2** |
| chr2: 37383568 | EIF2AK2 | N\_Shore | 5'UTR |  | 0.12 | 0.18 | **3.15×10-5** |  | 0.14 | 0.19 | **9.71×10-5** |
| chr11: 313354 | IFITM1 | S\_Shore | TSS1500 |  | 0.26 | 0.31 | **1.11×10-3** |  | 0.27 | 0.28 | 1.44×10-1 |
| chr3: 122281939 | PARP9 | N\_Shore | 5'UTR |  | 0.72 | 0.79 | **1.60×10-3** |  | 0.73 | 0.83 | **1.00×10-7** |
| chr21: 42798747 | MX1 | Island | 5'UTR |  | 0.06 | 0.07 | **2.50×10-3** |  | 0.07 | 0.08 | 1.10×10-1 |
| chr1: 79085765 | IFI44L | - | TSS1500 |  | 0.45 | 0.50 | **3.11×10-2** |  | 0.52 | 0.61 | **1.20×10-3** |
| chr1: 79085713 | IFI44L | - | TSS1500 |  | 0.64 | 0.68 | 1.22×10-1 |  | 0.69 | 0.74 | **1.17×10-2** |
| chr11: 312518 | IFITM1 | S\_Shore | TSS1500 |  | 0.05 | 0.06 | 3.58×10-1 |  | 0.06 | 0.06 | 5.33×10-1 |

McaM represents the mean methylation level of the cases while McoM represents the mean methylation status of the control samples. Pvals were conducted with wilcoxon rank-sum test.