**Integrative analysis of multiple autoimmune diseases’ DNA methylation signature**

**Abstract**

**Objective:** Graves’ disease (GD), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and systemic sclerosis (SSc) are all autoimmune diseases with complicated pathogenesis. Multiple lines of evidence have shown that epigenetic modifications, especially DNA methylation is involved in the pathogenesis of these diseases. To further characterize their commonalities, we performed an integrated analysis of the DNA methylation signature of these autoimmune diseases.

**Methods:** We collected Illumina Methylation 450K array raw data in CD4+ and CD8+ T cells of GD, RA, SLE from public databases and SSc from our previous work. Clustering analysis, differential methylation analysis, gene enrichment analysis as well as ROC analysis were performed.

**Results:** Clustering analysis based on the variable CpG sites showed that GD, RA, SLE and SSc patients possess similar DNA methylation signatures. 15289 differential methylation sites were identified between GD/RA/SLE/SSc patients and control individuals in CD4+ T cells, and differentially methylated genes are highly enriched in type I interferon pathway, highlighting the importance of type I interferon in the pathogenesis of multiple autoimmune diseases. Similarly, 9295 differential methylation sites were identified between GD/SSc patients and control individuals in CD8+ T cells, and the differentially methylated genes are mainly enriched in the immune system, especially leukocyte related regulation. Methylation levels of type-I interferon-associated genes demonstrated excellent diagnostic capacity in CD4+ T cells, especially the biomarker panel with all the genes involved (AUC for all patients: 0.90). And Methylation level of IFI44L, a reported biomarker of SLE also showed high potential to distinguish GD, RA, SLE and SSc patients from healthy people in CD4+ T cells (AUC for all patients: 0.80) and fair potential to differentiate GD and SSc patients in CD8+ T cells (AUC for all patients: 0.65). Biomarker panel including all the differential methylation sites found on IFI44L between patients and matched controls presented even better diagnostic ability, with AUC of 0.86 for GD/RA/SLE/SSc patients in CD4+ subset and AUC of 0.75 for GD/SSc patients in CD8+ subset (compared with controls).

**Conclusion:** Hypo-methylation of type I interferon-associated genes is widespread in the methylation signature of GD/RA/SLE/SSc patients in CD4+ T cells, indicating type I interferon-associated genes and pathways might play an important role in the pathogenesis of these four diseases. In addition, the methylation status of type I interferon-associated genes and IFI44L could be promising biomarkers for diagnosis of GD, RA, SLE and SSc.

**Keywords:** autoimmune diseases, DNA methylation, CD4+ T cells, CD8+ T cells, type I interferon, biomarker

**Introduction**

Autoimmune diseases are multifactorial complex diseases characterized by loss of immunologic tolerance to self-antigens, inappropriate activation of autoimmune response and damage of target organ systems [[1](#_ENREF_1" \o "Wandstrat, 2001 #4)]. Currently, over 100 types of autoimmune diseases are affecting 5-10% of the population worldwide and are one of the key causes of morbidity and mortality. Moreover, the economic burden of autoimmune diseases is enormous due to high rates of disability and comorbidity as well as the increasing treatment costs [[2](#_ENREF_2" \o "Shoenfeld, 2008 #5)]. However, the understanding of autoimmune diseases is still limited, with the exact etiology and pathogenesis remains unclear. Recently, multiple studies have suggested that there are a lot of commonalities in the pathogenesis of different kinds of autoimmune diseases. Therefore, the integrative and comparable analysis across different kinds of autoimmune diseases may reveal more clues about the pathogenesis.

Graves’ disease (GD), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and systemic sclerosis (SSc) are all typical autoimmune diseases. They share commonalities of occurring more common in women than men [[2](#_ENREF_2" \o "Shoenfeld, 2008 #5)], producing autoantibodies [[3-7](#_ENREF_3" \o "Genain, 1999 #6)], triggering immune abnormalities in CD4+ or CD8+ T cells [[8-12](#_ENREF_8" \o "McFarland, 2007 #123)] and multiple risk loci of the diseases [[13](#_ENREF_13" \o "Kochi, 2016 #139)]. Thus, research across these autoimmune diseases may figure out the similarity of their pathogenesis.

Epigenetics refers to the study of heritable changes in gene function without alterations in DNA sequence [[14](#_ENREF_14" \o "Bird, 2007 #162)]. A number of environmental risk factors could wield influence over the pathogenesis of autoimmune diseases through epigenetic mechanisms [[15](#_ENREF_15" \o "Zhang, 2015 #163)], which provide a crucial link between environmental and genetic risk factors of the diseases. DNA methylation is one of the major epigenetic mechanisms, and it plays an important role in determining gene function [[16](#_ENREF_16" \o "Hedrich, 2011 #164)]. Recently, multiple lines of evidence have shown that dysregulated DNA methylation plays a critical part in the onset of autoimmune diseases, including GD, RA, SLE and SSc [[16](#_ENREF_16" \o "Hedrich, 2011 #164), [17](#_ENREF_17" \o "Cai, 2015 #210)]. Hence, further exploration of the DNA methylation patterns of these diseases can help to elucidate the pathogenesis behind.

In the present study, we report the integrative study of genome-wide DNA methylome in CD4+ T cells of GD, RA, SLE, SSc patients, as well as in CD8+ T cells of GD, SSc patients. Autoimmune responses mediated by CD4+ and CD8+ T cells play a key role in the diseases mentioned above [[18](#_ENREF_18" \o "O'Garra, 1997 #165), [19](#_ENREF_19" \o "Walter, 2005 #166)]. Commonalities of these autoimmune diseases were identified through our analysis.

**Results**

In order to recognize the commonalities of different autoimmune diseases, including GD, RA, SLE and SSc, we collected Illumina Methylation450K microarray data of 116 patients with GD/RA/SLE/SSc and 117 controls in CD4+ T cells. Similarly, the methylation data of 61 patients with GD/SSc and 55 control individuals in CD8+ T cells were also collected. The data were acquired from public database and our previous work. The detailed clinical characteristics of patients and healthy controls were described in Table 1-2. Stringent quality filtering, exclusion of cross-hybridizing and SNP-containing CpGs, and normalization for batch (as described in Materials and methods) were conducted before further analysis.

**Hierarchical clustering showed similar methylation patterns of the various diseases**

To compare the methylation patterns of the autoimmune diseases mentioned above, we preformed hierarchical clustering analysis of the CpG sites showing the largest variation (top 50) across all the patient groups based on one-way ANOVA (Supplementary Table 1) in CD4+ samples (analysis was not carried out in CD8+ samples since there were only 2 patient groups). We found that samples from the four diseases cannot be explicitly distinguished from each other (Figure 1), indicating GD, RA, SLE and SSc patients possess similar methylation patterns, which provided further evidence for the commonalities in the pathogenesis of these diseases.

**Genome-wide DNA methylation profiling of the autoimmune diseases**

To further identify the common epigenetic changes associated with the above autoimmune diseases, we performed genome-wide DNA methylation analysis of the patients and controls in both CD4+ and CD8+ T cell subsets. After stringent quality control and differential methylation analysis, we identified 15289 and 9295 differentially methylated sites (DMS) in CD4+ and CD8+ T cells, respectively (Supplementary Table 2 and 3). In the CD4+ T cell methylation dataset, 7889 of these sites were hypo-methylated while 7400 were hyper-methylated. Similarly, 5471 and 3824 probes were hypermethylated and hypomethylated in CD8+ T cells of patients respectively. Meanwhile, the gene region and CpG island location analysis of the DMS showed that they were mainly located inside of genes and outside of CpG islands, which differed significantly from the features of all methylation sites involved in the differential methylation analysis (Figure 2). Over 40% of the DMS were identified within intragenic (body) regions, and more than 30% of which were observed in promoter regions. In relation to CpG islands, only a small proportion (< 10%) of all identified sites were associated with well-defined CpG island regions. In the comparative analysis of methylation patterns between CD4+ and CD8+ T cells, we found 5129 of the DMS were shared, while 10160 DMS were unique to CD4+ T cells and 4166 DMS to CD8+ T cells (Supplementary Figure 1). Altogether, we identified 6027 and 4358 differentially methylated genes in CD4+ and CD8+ T cells, respectively, with 2645 genes were shared in these two subtypes.

To identify common functional characteristics of the differentially methylated genes, we performed a GO analysis on genes annotated to the most significant DMS (top 50) in each cell type. Interestingly, analysis of CD4+ dataset revealed that type I interferon signaling pathway, response to type I interferon and cellular response to type I interferon (p value after FDR = 1.98×10-6) were the most significantly enriched GO terms, indicating that type I IFN-associated genes in CD4+ T cells might play a role in the pathogenesis of autoimmune diseases including GD, RA, SLE and SSc (Table 3). Specifically, we selected the type I IFN-associated genes which differed significantly between diseases and control samples (among top 50 DMS) and presented the methylation status of these genes in each disease alone, indicating that a hypo-methylation signature of type I interferon is shared in these four diseases (Figure 3). Clustering analysis based on the DMS of type I IFN-associated genes showed that most patient samples can be discriminated from controls (Figure 4). In contrast, differential methylation analysis in CD8+ T cells reveal that the most enriched biological processes were more general, including regulation of leukocyte activation, regulation of cell adhesion and regulation of T cell activation (p value after FDR = 0.0667). Meanwhile, most of the top-ranking terms (6 out of 10) were related to immune system processes (Table 4). No shared GO term was identified between CD4+ and CD8+ T cells. These results suggested an independent pattern of differential methylation profiles in CD4+ and CD8+ T cells of the patients.

**Diagnostic value of type I IFN-associated genes for GD, RA, SLE and SSc in CD4+ T cells**

Our study has revealed significant hypomethylation of type-I IFN-associated genes in CD4+ T cells of GD/RA/SLE/SSc patients compared with controls, and we want to further evaluate their ability to distinguish the patients from healthy individuals to see if methylation levels of these genes can be of diagnostic value. Thus, we performed the logistic regression analysis based on the methylation status of type-I IFN-associated genes in CD4+ T cells.

ROC curve analysis were conducted of all the DMS found on type I IFN-associated genes between patients and matched controls. The methylation levels of the 21 DMS showed high AUC value of 0.90, with sensitivity and specificity of 0.82 (Figure 5). In patients with RA, SSc and SLE (compared with healthy controls), the AUC were even higher as 1.00, 1.00 and 0.98 respectively, while the AUC in GD patients (compared with healthy controls) was the same as 0.90 (Supplementary Figure 2A-2D).

To further explore the diagnostic value of each type-I IFN-associated gene involved, we constructed the ROC curve for DMS found on IFIT1, IRF7, MX1, OAS1, USP18 and RSAD2, respectively. In the analysis of IFIT1, AUC value of the 2 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.82 (sensitivity: 0.84, specificity: 0.71) (Figure 6). And in patients with SLE (compared with control individuals), the AUC was much higher as 0.89, while the AUC for GD, SSc and RA patients (compared with control individuals) were relatively lower as 0.79, 0.75 and 0.68 respectively (Supplementary Figure 3A-3D). In the analysis of IRF7, AUC value of the 3 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.78 (sensitivity: 0.74, specificity: 0.75) (Figure 6). And in patients with SLE and RA (compared with control individuals), the AUC were much higher as 0.92 and 0.88 respectively, while the AUC for SSc and GD patients (compared with control individuals) were relatively lower as 0.73 and 0.66 respectively (Supplementary Figure 4A-4D). In the analysis of MX1, AUC value of the 6 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.80 (sensitivity: 0.72, specificity: 0.78) (Figure 6). And in patients with SLE, RA and SSc (compared with control individuals), the AUC were much higher as 0.92, 0.87 and 0.86 respectively, while the AUC for GD patients (compared with control individuals) was relatively lower as 0.73 (Supplementary Figure 5A-5D). In the analysis of OAS1, AUC value of the 3 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.80 (sensitivity: 0.82, specificity: 0.66) (Figure 6). And in patients with SLE, RA and SSc (compared with control individuals), the AUC were higher as 0.89, 0.84 and 0.82 respectively, while the AUC for GD patients (compared with control individuals) was relatively lower as 0.63 (Supplementary Figure 6A-6D). In the analysis of USP18, AUC value of the 4 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.78 (sensitivity: 0.87, specificity: 0.62) (Figure 6). And in patients with RA and SLE (compared with control individuals), the AUC were much higher as 0.92 and 0.89 respectively, while the AUC for SSc and GD patients (compared with control individuals) were the same or relatively lower as 0.78 and 0.62 (Supplementary Figure 7A-7D). In the analysis of RSAD2, AUC value of the 3 DMS in GD/RA/SLE/SSc patients in comparison with healthy controls was 0.81 (sensitivity: 0.83, specificity: 0.69) (Figure 6). And in patients with SLE and RA (compared with control individuals), the AUC were higher as 0.88 and 0.84 respectively, while the AUC for SSc and GD patients (compared with control individuals) were relatively lower as 0.77 and 0.75 (Supplementary Figure 8A-8D).

The results of our analysis imply that DNA methylation levels of type-I IFN-associated genes identified can be useful biomarkers in the evaluation and diagnosis of GD, RA, SLE and SSc. And the biomarker panel with all DMS found on the genes (IFIT1, IRF7, MX1, OAS1, USP18 and RSAD2) involved demonstrated the best diagnostic capacity, which could correctly discriminate between RA or SSc patients and healthy controls.

**Diagnostic value of IFI44L for GD, RA, SLE and SSc**

Previous study has identified that the methylation status of IF44L, a type I IFN-associated gene, could be utilized as the biomarker for diagnosis of autoimmune diseases [[20](#_ENREF_20" \o "Zhao, 2016 #167)]. Moreover, in our integration analysis, we also found the significant hypo-methylation status of IFI44L in the four kinds of autoimmune diseases. Therefore, to further assess the diagnostic value of IFI44L in GD, RA, SLE and SSc, logistic regression analysis based on the methylation status of IFI44L were conducted in CD4+ and CD8+ T cells, respectively.

We constructed the receiver operating characteristic (ROC) curve in both CD4+ and CD8+ T cells to depict the diagnostic ability of cg06872964 for the above autoimmune diseases, which is the hypomethylated CpG site mentioned in the research [[20](#_ENREF_20" \o "Zhao, 2016 #167)] and meanwhile covered by Illumina Human Methylation 450K BeadChip. The result revealed that the AUC value of the site in GD/RA/SLE/SSc patients in comparison with control individuals for CD4+ subset was 0.80 (sensitivity: 0.78, specificity: 0.74) (Supplementary Figure 9A). In patients with SLE (compared with healthy controls), the AUC was much higher as 0.89, while the AUC for SSc, RA and GD patients (compared with healthy controls) were relatively lower as 0.78, 0.78 and 0.58 respectively (Figure 7A-7D). And the result for CD8+ subset showed that the AUC value of the site in GD/SSc patients in comparison with control individuals was 0.65 (sensitivity: 0.93, specificity: 0.34) (Supplementary Figure 9B). In patients with SSc and GD, the AUC were 0.73 and 0.61 respectively (Figure 7E-7F).

To further assess the diagnostic value of IFI44L methylation level in GD, RA, SLE and SSc, we performed the ROC curve analysis of all the DMS found on IFI44L between patients and matched controls in CD4+ and CD8+ T cells, respectively. In ROC curve analysis for GD/RA/SLE/SSc patients as compared with healthy controls in CD4+ subset, the methylation levels of the 6 DMS (including cg06872964) showed high AUC value of 0.86, with sensitivity of 0.81 and specificity of 0.80 (Supplementary Figure 10A). In patients with RA and SLE (compared with control individuals), the AUC were even higher as 0.96 and 0.95 respectively, while the AUC in SSc and GD patients (compared with control individuals) were relatively lower as 0.79 and 0.77 respectively (Figure 8A-8D). And in ROC curve analysis for GD/SSc patients as compared with healthy controls in CD8+ subset, the methylation levels of the 2 DMS (p-value after FDR of cg06872964 is 0.031, which is larger than our definition of DMS, thus not included) showed AUC value of 0.75, with sensitivity of 0.89 and specificity of 0.52 (Supplementary Figure 10B). In patients with GD and SSc (compared with control individuals), the analysis showed the AUC of 0.75 and 0.73 respectively (Figure 8E-8F).

The results of our study suggest that aberrant DNA methylation of IFI44L is a common signature of autoimmune diseases showing similar DNA methylation signatures, which are GD, RA, SLE, SSc in CD4+ subset and GD, SSc in CD8+ subset. IFI44L methylation could be a promising diagnostic biomarker for GD, RA, SLE and SSc patients, with biomarker panel including all the DMS found on IFI44L between patients and matched controls presenting better diagnostic ability than the single CpG site (cg06872964) reported in the previous research.

**Discussion**

In our study, we integrated hierarchical clustering analysis and differential methylation analysis to determine the commonalities across various autoimmune diseases including GD, RA, SLE and SSc in both CD4+ and CD8+ T cells. We identified that GD, RA, SLE and SSc patients possess similar DNA methylation signatures. Furthermore, 15289 and 9295 DMS were identified between GD/RA/SLE/SSc patients and controls in CD4+ and CD8+ T cells, respectively. Moreover, we identified widespread hypo-methylation status of CpG sites located at genes involved in type I IFN signaling in CD4+ T cells, indicating that GD/RA/SLE/SSc patients might be hypersensitive to type I interferon. In addition, methylation status of type I IFN-associated genes and IFI44L were figured out to be promising diagnostic biomarkers for diagnosis of GD, RA, SLE and SSc.

Aberrant type I interferon function has recently been implicated in several autoimmune diseases, including GD, RA, SLE and SSc [[21](#_ENREF_21" \o "Chen, 2017 #169), [22](#_ENREF_22" \o "Bohbot, 2006 #170)]. A study by Henault et al. reported that increased amounts of IFN-α derived from self-reactive immunoglobulin E-activated plasmacytoid dendritic cells and the excessive IFN responses triggered were capable of exacerbating self-destructive autoimmune responses in SLE patients [[23](#_ENREF_23" \o "Henault, 2016 #171)]. Another study demonstrated that expression of interferon type I regulated genes in RA patients were significantly elevated compared with healthy individuals [[24](#_ENREF_24" \o "van der Pouw Kraan, 2007 #172)]. In addition, more than half of SSc patients were found to present up-regulation of type I IFN-associated genes in both whole blood and PBMC [[25](#_ENREF_25" \o "Tan, 2006 #173), [26](#_ENREF_26" \o "York, 2007 #175)]. And 3 hepatitis C patients with no known history of familial or personal thyroid disease developed GD after treatment with IFN-α therapy, implicating the strong correlation between type I IFN and GD [[22](#_ENREF_22" \o "Bohbot, 2006 #170)]. Furthermore, multiple type I IFN-associated genes were found to be related to the pathogenesis of one or more of the autoimmune diseases mentioned above from genetic association studies, including STAT4, IRF5, IFIH1 and PLZF [[13](#_ENREF_13" \o "Kochi, 2016 #139), [27](#_ENREF_27" \o "Ghodke-Puranik, 2015 #177), [28](#_ENREF_28" \o "Stefan, 2014 #178)]. In summary, dysfunctional type I IFN-associated pathway is involved in the pathogenesis and progression of GD, RA, SLE and SSc. In accordance with the previous studies, we identified the hypomethylation of type I IFN-associated genes including IFIT1, IRF7, MX1, OAS1, USP18 and RSAD2 at the epigenetic level in CD4+ T cells of GD/RA/SLE/SSc patients, indicating the aberration of DNA methylation profile of these genes might play a role in the pathogenesis of the autoimmune diseases mentioned above.

IFIT1 is one of the most strongly induced interferon stimulated genes, which preferentially recognizes 2’-O unmethylated RNA or uncapped non-self viral mRNA and inhibits translation initiation or blocks the RNA from the actively replicating pool [[29](#_ENREF_29" \o "Diamond, 2014 #179)]. An increased expression of IFIT1 was found in established RA patients, suggesting its potential association with viral infections in autoimmune diseases [[30](#_ENREF_30" \o "Castaneda-Delgado, 2017 #180)]. IRF7 is a crucial regulator of type I IFN against pathogenic infections, it is predominantly activated by TLR7 in plasmacytoid dendritic cells with pathogenic nucleic acids recognized by pathogen recognition receptors [[31](#_ENREF_31" \o "Ning, 2011 #185)]. A microarray study revealed up-regulation of IRF7 mRNA level in peripheral blood cells of SSc patients with early diseases [[32](#_ENREF_32" \o "Wu, 2013 #187)]. MX1 is a key downstream gene of type I interferon and plays a part in mediating the interferon-induced antiviral response against a large variety of viruses [[33](#_ENREF_33" \o "Haller, 2011 #189)]. Previous study and our previous work have reported the up-regulation of MX1 in SLE and SSc [[34](#_ENREF_34" \o "Coit, 2013 #194), [35](#_ENREF_35" \o "Pu, 2017 #191)]. OAS1 is an interferon-induced antiviral enzyme which can recognize dsRNA derived by virus and then defense against the viral infections [[36](#_ENREF_36" \o "Kristiansen, 2011 #182)]. It is shown to be upregulated in peripheral blood of SLE patients and overexpressed in SSc skin compared with normal individuals [[37](#_ENREF_37" \o "Ye, 2007 #183), [38](#_ENREF_38" \o "Assassi, 2015 #184)]. USP18 is an IFN-inducible protein, which can deconjugate the ubiquitin-like IFN-stimulated gene 15 (ISG15) from target proteins [[39](#_ENREF_39" \o "Potu, 2010 #192)], and was found to inhibit IFN-induced JAK–STAT signaling at the level of the IFN receptor [[40](#_ENREF_40" \o "Malakhova, 2006 #193)]. Significant overexpression of USP18 was observed in CD4+ T cells from SLE patients, indicating its role in the pathogenesis of autoimmune diseases [[34](#_ENREF_34" \o "Coit, 2013 #194)]. RSAD2 is also an interferon-inducible protein that inhibits the replication of a broad spectrum of DNA and RNA viruses[[41](#_ENREF_41" \o "Seo, 2011 #211)]. Expression level of RSAD2 is dramatically elevated in patients with RA and SLE compared with healthy individuals according to some research before [[24](#_ENREF_24" \o "van der Pouw Kraan, 2007 #172), [34](#_ENREF_34" \o "Coit, 2013 #194)].

Moreover, our study implied that methylation of type-I IFN-associated genes may be helpful in the evaluation and diagnosis of GD, RA, SLE and SSc. AUC of all DMS found on the genes involved was particularly high, and RA and SSc patients could even be totally distinguished from healthy controls via methylation signature. Methylation levels of DMS on single type-I IFN-associated gene also showed high diagnostic potential, with AUC all around 0.80.

Besides the genes mentioned above, IFI44L was also found to be hypo-methylated in these four kinds of autoimmune diseases. Our analysis suggested that DNA methylation level of IFI44L may facilitate to evaluate and diagnose GD, RA, SLE and SSc, with higher diagnostic ability in CD4+ T cells than CD8+ T cells. AUC, sensitivity and specificity of IFI44L methylation (especially with all DMS found) for RA and SLE compared with matched controls in CD4+ T cells are particularly high, manifesting that RA and SLE patients can be well discriminated from healthy controls through the methylation difference of IFI44L. And our findings of SLE is in line with the biomarker study conducted in SLE peripheral blood previously [[20](#_ENREF_20" \o "Zhao, 2016 #167)], which provides further proof of the diagnostic value of IFI44L.

In conclusion, we performed the integrative study of multiple autoimmune diseases’ genome-wide DNA methylation patterns with GD, RA, SLE and SSc involved for the first time. We identified the similarity of methylation profiles across these diseases and the common hypo-methylation signature of type I IFN-associated genes in CD4+ T cells of GD/RA/SLE/SSc patients. However, since the Illumina Human Methylation450K BeadChip covers only about 2% of all the CpG sites in the human genome, we could just quantify the methylation profiles of the limited number of CpG sites presented for each gene, with a mass of genes unable to be detected. Nevertheless, the identification of commonalities among various autoimmune diseases’ DNA methylation patterns will help to facilitate the understanding of the pathogenesis of these diseases, especially the effect of type I IFN, and will potentially improve the diagnosis and medication of the autoimmune diseases through epigenetic level.

**Materials and methods**

**Sample and data collection**

DNA methylation microarray data of GD/RA/SLE patients and corresponding control individuals were acquired from previous publications [[42-44](#_ENREF_42" \o "Limbach, 2016 #199)], while raw data of SSc patients and matched controls were sourced from our own lab. All data collected were based on Illumina Methylation-450 array platform. The clinical and demographic information of the patients and healthy subjects are summarized in Table 1 and 2.

**Preprocessing of methylation data**

Initial quality control and preprocessing of raw data in the form of IDAT files were conducted in R [[45](#_ENREF_45" \o "Team, 2013 #202)] with the Bioconductor package RnBeads [[46](#_ENREF_46" \o "Assenov, 2014 #203)]. Probes whose last 3 bases of the target sequence overlap with SNP and cross-hybridizing probes whose sequence maps to multiple genomic coordinates were removed before further analysis. “Greedycut” procedure was conducted to filter out the probes with a detection p-value >0.01 in any of the samples. In addition, probes located on sex chromosomes and probes with many missing values were also filtered out. Then, “noob” method was used to perform background subtraction. Finally, beta-mixture quantile normalization (BMIQ) of the methylation beta values were conducted to adjust the deviation caused by different types of probes (type I and type II).

**Data integration and batch normalization**

Preprocessed data of the diseases were adjusted to the unified format. Methylation beta values of samples from multiple diseases were then integrated together in both CD4+ and CD8+ sample subset. Batch normalization was conducted with the SVA package [[47](#_ENREF_47" \o "Leek, 2012 #206)] to remove batch effect among arrays used for samples of different diseases.

**ANOVA and clustering analysis**

One-way analysis of variance (ANOVA) was used for multiple comparisons of methylation beta values among patients with different diseases. Hierarchical clustering analysis was performed on the top 50 CpG sites showing the largest variation across all the patient groups based on one-way ANOVA to examine the relatedness among methylation patterns of the diseases. Heat maps were generated to visualize the results of clustering analysis.

**Differential methylation analysis**

P-values of the probes involved in the integrated methylation data were computed with a linear regression model, including gender, age and ethnicity as covariates. Benjamini and Hochberg false discovery rate (FDR) [[48](#_ENREF_48" \o "Benjamini, 1995 #208)] was employed to adjust the raw p-values for multiple testing. CpG sites with p-value after FDR < 0.01 were defined as differentially methylated sites (DMS). Further gene ontology analysis was conducted with g:Profiler [[49](#_ENREF_49" \o "Reimand, 2016 #212)] using CpGs with the most significant methylation difference (top 50).

**Gene and CpG island features of methylation sites**

Location of the methylation sites in relation to gene and CpG island subregions were annotated based on the annotation files provided by Illumina. Gene associated features TSS1500, TSS200, 5’ UTR and 1st exon were collectively referred to as promoter region. CpG sites that were located in different parts of these genes were counted multiple times.

**ROC analysis**

Binary logistic regression model was selected to fit the receiver operating characteristic (ROC) curve where sensitivity and (1-specificity) were transformed into complex logarithmic variables. The area under the ROC curve (AUC) was calculated.

**Conflict of interest**

The authors declare no competing interests.

**Acknowledgements**

Left to be added

**Funding**

Left to be added

**References**

1. Wandstrat, A. and E. Wakeland, *The genetics of complex autoimmune diseases: non-MHC susceptibility genes.* Nat Immunol, 2001. **2**(9): p. 802-9.

2. Shoenfeld, Y., et al., *The autoimmunologist: geoepidemiology, a new center of gravity, and prime time for autoimmunity.* J Autoimmun, 2008. **31**(4): p. 325-30.

3. Genain, C.P., et al., *Identification of autoantibodies associated with myelin damage in multiple sclerosis.* Nat Med, 1999. **5**(2): p. 170-5.

4. Weetman, A.P., *Graves' disease.* N Engl J Med, 2000. **343**(17): p. 1236-48.

5. van Boekel, M.A., et al., *Autoantibody systems in rheumatoid arthritis: specificity, sensitivity and diagnostic value.* Arthritis Research & Therapy, 2001. **4**(2): p. 87.

6. Sherer, Y., et al., *Autoantibody explosion in systemic lupus erythematosus: More than 100 different antibodies found in SLE patients.* Seminars in Arthritis and Rheumatism, 2004. **34**(2): p. 501-537.

7. Steen, V.D., *Autoantibodies in systemic sclerosis.* Semin Arthritis Rheum, 2005. **35**(1): p. 35-42.

8. McFarland, H.F. and R. Martin, *Multiple sclerosis: a complicated picture of autoimmunity.* Nat Immunol, 2007. **8**(9): p. 913-9.

9. Covas, M.I., et al., *Peripheral blood T-lymphocyte subsets in autoimmune thyroid disease.* J Investig Allergol Clin Immunol, 1992. **2**(3): p. 131-5.

10. Firestein, G.S., *Evolving concepts of rheumatoid arthritis.* Nature, 2003. **423**(6937): p. 356-61.

11. Crispin, J.C., et al., *Pathogenesis of human systemic lupus erythematosus: recent advances.* Trends Mol Med, 2010. **16**(2): p. 47-57.

12. Hussein, M.R., et al., *Alterations of mononuclear inflammatory cells, CD4/CD8+ T cells, interleukin 1beta, and tumour necrosis factor alpha in the bronchoalveolar lavage fluid, peripheral blood, and skin of patients with systemic sclerosis.* J Clin Pathol, 2005. **58**(2): p. 178-84.

13. Kochi, Y., *Genetics of autoimmune diseases: perspectives from genome-wide association studies.* Int Immunol, 2016. **28**(4): p. 155-61.

14. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.

15. Zhang, Z. and R. Zhang, *Epigenetics in autoimmune diseases: Pathogenesis and prospects for therapy.* Autoimmun Rev, 2015. **14**(10): p. 854-63.

16. Hedrich, C.M. and G.C. Tsokos, *Epigenetic mechanisms in systemic lupus erythematosus and other autoimmune diseases.* Trends Mol Med, 2011. **17**(12): p. 714-24.

17. Cai, T.T., et al., *Genome-wide DNA methylation analysis in Graves' disease.* Genomics, 2015. **105**(4): p. 204-10.

18. O'Garra, A., L. Steinman, and K. Gijbels, *CD4+ T-cell subsets in autoimmunity.* Curr Opin Immunol, 1997. **9**(6): p. 872-83.

19. Walter, U. and P. Santamaria, *CD8+ T cells in autoimmunity.* Curr Opin Immunol, 2005. **17**(6): p. 624-31.

20. Zhao, M., et al., *IFI44L promoter methylation as a blood biomarker for systemic lupus erythematosus.* Annals of the Rheumatic Diseases, 2016. **75**(11): p. 1998-2006.

21. Chen, K., J. Liu, and X. Cao, *Regulation of type I interferon signaling in immunity and inflammation: A comprehensive review.* J Autoimmun, 2017. **83**: p. 1-11.

22. Bohbot, N.L., et al., *Interferon-α-induced hyperthyroidism: a three-stage evolution from silent thyroiditis towards Graves’ disease.* European journal of endocrinology, 2006. **154**(3): p. 367-372.

23. Henault, J., et al., *Self-reactive IgE exacerbates interferon responses associated with autoimmunity.* Nat Immunol, 2016. **17**(2): p. 196-203.

24. van der Pouw Kraan, T.C., et al., *Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients.* Ann Rheum Dis, 2007. **66**(8): p. 1008-14.

25. Tan, F.K., et al., *Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients.* Rheumatology, 2006. **45**(6): p. 694-702.

26. York, M.R., et al., *A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type I Interferons and Toll-like receptor agonists.* Arthritis and Rheumatism, 2007. **56**(3): p. 1010-1020.

27. Ghodke-Puranik, Y. and T.B. Niewold, *Immunogenetics of systemic lupus erythematosus: A comprehensive review.* J Autoimmun, 2015. **64**: p. 125-36.

28. Stefan, M., et al., *Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(34): p. 12562-12567.

29. Diamond, M.S., *IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation.* Cytokine Growth Factor Rev, 2014. **25**(5): p. 543-50.

30. Castaneda-Delgado, J.E., et al., *Type I Interferon Gene Response Is Increased in Early and Established Rheumatoid Arthritis and Correlates with Autoantibody Production.* Front Immunol, 2017. **8**: p. 285.

31. Ning, S., J.S. Pagano, and G.N. Barber, *IRF7: activation, regulation, modification and function.* Genes Immun, 2011. **12**(6): p. 399-414.

32. Wu, M. and S. Assassi, *The role of type 1 interferon in systemic sclerosis.* Front Immunol, 2013. **4**: p. 266.

33. Haller, O. and G. Kochs, *Human MxA Protein: An Interferon-Induced Dynamin-Like GTPase with Broad Antiviral Activity.* Journal of Interferon and Cytokine Research, 2011. **31**(1): p. 79-87.

34. Coit, P., et al., *Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naive CD4+T cells from lupus patients.* Journal of Autoimmunity, 2013. **43**: p. 78-84.

35. Pu, W.L., et al., *Genome-Wide DNA Methylation Analysis in Systemic Sclerosis Reveals Hypomethylation of Interferon-Associated Genes in CD4+and CD8+T Cells.* Arthritis & Rheumatology, 2017. **69**.

36. Kristiansen, H., et al., *The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities.* J Interferon Cytokine Res, 2011. **31**(1): p. 41-7.

37. Ye, S., et al., *Could 2'5'-oligoadenylate synthetase isoforms be biomarkers to differentiate between disease flare and infection in lupus patients? A pilot study.* Clin Rheumatol, 2007. **26**(2): p. 186-90.

38. Assassi, S., et al., *Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis.* Arthritis Rheumatol, 2015. **67**(11): p. 3016-26.

39. Potu, H., A. Sgorbissa, and C. Brancolini, *Identification of USP18 as an important regulator of the susceptibility to IFN-alpha and drug-induced apoptosis.* Cancer Res, 2010. **70**(2): p. 655-65.

40. Malakhova, O.A., et al., *UBP43 is a novel regulator of interferon signaling independent of its ISG15 isopeptidase activity.* Embo Journal, 2006. **25**(11): p. 2358-2367.

41. Seo, J.Y., R. Yaneva, and P. Cresswell, *Viperin: A Multifunctional, Interferon-Inducible Protein that Regulates Virus Replication.* Cell Host & Microbe, 2011. **10**(6): p. 534-539.

42. Limbach, M., et al., *Epigenetic profiling in CD4+ and CD8+ T cells from Graves' disease patients reveals changes in genes associated with T cell receptor signaling.* J Autoimmun, 2016. **67**: p. 46-56.

43. Guo, S., et al., *Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis.* Mod Rheumatol, 2017. **27**(3): p. 441-447.

44. Absher, D.M., et al., *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**(8): p. e1003678.

45. Team, R.C., *R: A language and environment for statistical computing.* 2013.

46. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.

47. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.

48. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.

49. Reimand, J., et al., *g:Profiler-a web server for functional interpretation of gene lists (2016 update).* Nucleic Acids Res, 2016. **44**(W1): p. W83-9.

**Tables**

**Table 1. Characteristics of patients and healthy controls for CD4+ T cells**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **GD samples** | |  | **RA samples** | |  | **SLE samples** | |  | **SSc samples** | |
|  |  | **Patients** | **Controls** |  | **Patients** | **Controls** |  | **Patients** | **Controls** |  | **Patients** | **Controls** |
| **Individuals, n** |  | 36 | 31 |  | 12 | 12 |  | 44 | 50 |  | 24 | 24 |
| **Women, n (%)** |  | 32 (89) | 28 (90) |  | 9 (75) | 9 (75) |  | 37 (84) | 39 (78) |  | 17 (71) | 19 (79) |
| **Age, years, (mean±SD)** |  | 56.2±15.5 | 53.7±11.9 |  | 40.8±11.5 | 40.6±11.4 |  | 31.0±8.4 | 29.2±6.3 |  | 47.4±14.1 | 46.8±10.0 |
| **Ethnicity** |  | Caucasian | Caucasian |  | Chinese | Chinese |  | Caucasian/  African | Caucasian/  African |  | Chinese | Chinese |

**Table 2. Characteristics of patients and healthy controls for CD8+ T cells**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **GD samples** | |  | **SSc samples** | |
|  |  | **Patients** | **Controls** |  | **Patients** | **Controls** |
| **Individuals, n** |  | 37 | 31 |  | 24 | 24 |
| **Women, n (%)** |  | 33 (89) | 28 (90) |  | 17 (71) | 19 (79) |
| **Age, years, (mean±SD)** |  | 55.0±16.3 | 53.7±11.9 |  | 47.4±14.1 | 46.8±10.0 |
| **Ethnicity** |  | Caucasian | Caucasian |  | Chinese | Chinese |

**Table 3. Gene ontology analysis of DMS (top 50) for** **GD/RA/SLE/SSc CD4+ T cell dataset**

|  |  |  |
| --- | --- | --- |
| **Term** | **p-value.fdr** | **Counta** |
| type I interferon signaling pathway | 1.98E-06 | 6 |
| response to type I interferon | 1.98E-06 | 6 |
| cellular response to type I interferon | 1.98E-06 | 6 |
| negative regulation of viral genome replication | 6.76E-06 | 5 |
| defense response to virus | 1.01E-05 | 7 |
| negative regulation of viral life cycle | 2.83E-05 | 5 |
| regulation of viral genome replication | 4.02E-05 | 5 |
| negative regulation of viral process | 4.42E-05 | 5 |
| response to cytokine | 4.42E-05 | 11 |
| response to virus | 4.63E-05 | 7 |

aCount represents number of genes involved in the GO term.

**Table 4. Gene ontology analysis of DMS (top 50) for GD/SSc CD8+ T cell dataset**

|  |  |  |
| --- | --- | --- |
| **Term** | **p-value.fdr** | **Counta** |
| regulation of leukocyte activation | 0.0667 | 5 |
| positive regulation of leukocyte apoptotic process | 0.0667 | 2 |
| regulation of cell adhesion | 0.0667 | 6 |
| positive regulation of leukocyte cell-cell adhesion | 0.0667 | 4 |
| regulation of T cell activation | 0.0667 | 4 |
| positive regulation of cell activation | 0.0667 | 5 |
| regulation of leukocyte cell-cell adhesion | 0.0667 | 5 |
| postsynaptic membrane organization | 0.0667 | 2 |
| leukocyte cell-cell adhesion | 0.0667 | 5 |
| localization within membrane | 0.0667 | 3 |

aCount represents number of genes involved in the GO term.

**Figures**

****

**Figure 1. Heatmap of CpG sites showing the largest variation (top 50) across patient groups based on ANOVA in CD4+ T cells.** Each column represents a sample, each row represents the methylation level of all the samples involved on one CpG site, sample clustering tree appears at the top.

****

**Figure 2. Location of methylation sites in relation to gene and CpG island subregions.** Features of all the methylation sites involved in the differential methylation analysis (red) and the differential methylation sites identified (blue) are both shown in a single figure. (A, B) Proportion of methylation sites related to various gene and CpG island subregions in CD4+ T cells. (C, D) Proportion of methylation sites related to various gene and CpG island subregions in CD8+ T cells.

****

**Figure 3. Methylation level of CpG sites enriched to be relevant to type I IFN in GD, RA, SLE and SSc CD4+ T cell dataset.** Methylation statuses of patients (red) and controls (blue) are shown simultaneously in a single disease panel.

****

**Figure 4. Heatmap of DMS on type I IFN-associated genes among all samples in CD4+ T cells.** Each column represents a sample, each row represents the methylation level of all the samples involved on one CpG site, sample clustering tree appears at the top.

****

**Figure 5. ROC curve of the DNA methylation levels at DMS found on all type I IFN-associated genes in GD/RA/SLE/SSc patients compared with matched controls in CD4+ T cells.**

****

**Figure 6. ROC curves of the DNA methylation levels at DMS found on each type I IFN-associated gene in GD/RA/SLE/SSc patients compared with matched controls in CD4+ T cells.** (A-F) Curves for IFIT1, IRF7, MX1, OAS1, USP18 and RSAD2 respectively.

****

**Figure 7. ROC curves of the DNA methylation level at cg06872964 in patients with various diseases compared with matched controls.** (A-D) Curves for patients with GD, RA, SLE and SSc in CD4+ T cells, respectively. (E, F) Curves for patients with GD and SSc in CD8+ T cells, respectively.

****

**Figure 8. ROC curves of the DNA methylation levels at DMS found on IFI44L in patients with various diseases compared with matched controls.** (A-D) Curves for patients with GD, RA, SLE and SSc in CD4+ T cells, respectively. (E, F) Curves for patients with GD and SSc in CD8+ T cells, respectively.

**Supplementary**

****

**Supplementary Figure 1. Venn diagram of the DMS distribution.** This venn diagram shows the distribution of both the unique and the shared DMS between CD4+ and CD8+ T cells.



**Supplementary Figure 2. ROC curves of the DNA methylation levels at DMS found on all type I IFN-associated genes in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 3. ROC curves of the DNA methylation levels at DMS found on IFIT1 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 4. ROC curves of the DNA methylation levels at DMS found on IRF7 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 5. ROC curves of the DNA methylation levels at DMS found on MX1 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 6. ROC curves of the DNA methylation levels at DMS found on OAS1 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 7. ROC curves of the DNA methylation levels at DMS found on USP18 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 8. ROC curves of the DNA methylation levels at DMS found on RSAD2 in patients with various diseases compared with matched controls in CD4+ T cells.** (A-D) Curves for patients with GD, RA, SLE and SSc respectively.



**Supplementary Figure 9. ROC curves of the DNA methylation level at cg06872964 in all patients compared with matched controls.** (A) Curve for GD/RA/SLE/SSc patients in CD4+ T cells. (B) Curve for GD/SSc patients in CD8+ T cells.



**Supplementary Figure 10. ROC curves of the DNA methylation levels at DMS found on IFI44L in all patients compared with matched controls.** (A) Curve for GD/RA/SLE/SSc patients in CD4+ T cells. (B) Curve for GD/SSc patients in CD8+ T cells.