

## **Hypomethylation of miR-17-92 cluster in lupus T cells and no significant role for genetic factors in the lupus-associated DNA methylation signature**

Patrick Coit<sup>1,2\*</sup>, Xiavan Roopnarinesingh<sup>3,4\*</sup>, Lourdes Ortiz-Fernandez<sup>1</sup>, Kathleen Maksimowicz-McKinnon<sup>5</sup>, Emily E. Lewis<sup>6</sup>, Joan T. Merrill<sup>4</sup>, W. Joseph McCune<sup>6</sup>, Jonathan D. Wren<sup>4,7</sup>, Amr H. Sawalha<sup>1,8,9,10</sup>

<sup>1</sup> Division of Rheumatology, Department of Pediatrics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>2</sup> Graduate Program in Immunology, University of Michigan, Ann Arbor, Michigan, USA

<sup>3</sup> Graduate Program, Department of Biochemistry and Molecular Biology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

<sup>4</sup> Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

<sup>5</sup> Division of Rheumatology, Henry Ford Health System, Detroit, Michigan, USA

<sup>6</sup> Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

<sup>7</sup> Department of Biochemistry and Molecular Biology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

<sup>8</sup> Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh, Pennsylvania, USA

<sup>9</sup> Lupus Center of Excellence, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

<sup>10</sup> Department of Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

\* These two authors equally contributed to this work

Please address correspondence to Amr H. Sawalha, MD. Address: 7123 Rangos Research Center, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. Phone: (412) 692-8140. Fax: (412) 412-692-5054. Email: [asawalha@pitt.edu](mailto:asawalha@pitt.edu)

**Running title:** T cell epigenetic signature in lupus is not genetically determined.

**Conflict of interest:** The authors have declared that no conflict of interest exists

## Abstract

**Objectives:** Epigenetic dysregulation plays an important role in the pathogenesis of lupus, a systemic autoimmune disease characterized by autoantibody production.

Lupus T cells demonstrate aberrant DNA methylation patterns dominated by hypomethylation of interferon-regulated genes. The objective of this study was to identify additional disease-associated DNA methylation changes in naïve CD4+ T cells from an extended cohort of lupus patients and determine the genetic contribution to epigenetic changes characteristic of lupus.

**Methods:** Genome-wide DNA methylation was assessed in naïve CD4+ T cells isolated from a cohort of 74 lupus patients and 74 age-, sex-, and race-matched healthy controls. We applied a trend deviation analysis approach, comparing methylation data in our cohort to methylation data from over 16,500 samples to characterize lupus-associated DNA methylation patterns. Methylation quantitative trait loci (meQTL) analysis was used to determine genetic contribution to lupus-associated DNA methylation changes.

**Results:** In addition to the previously reported epigenetic signature in interferon-regulated genes, we observed hypomethylation of the promoter regions of microRNA (miRNA) genes in the miR-17-92 cluster in lupus patients. Members of this miRNA cluster play an important role in regulating T cell proliferation and differentiation. Expression of two miRNAs within this cluster, miR-19b1 and miR-18a, showed a significant positive correlation with disease activity in lupus patients. meQTL were identified by integrating genome-wide DNA methylation profiles with genotyping data in lupus patients and controls. Patient meQTL show overlap with genetic risk loci for lupus.

However, less than 1% of differentially methylated CpG sites in lupus patients were associated with an meQTL, suggesting minimal genetic contribution to lupus-associated epigenotypes.

**Conclusion:** The lupus defining epigenetic signature, characterized by robust hypomethylation of interferon-regulated genes, does not appear to be determined by genetic factors. Hypomethylation of the miR-17-92 cluster that plays an important role in T cell activation is a novel epigenetic locus for lupus.

## Introduction

Systemic lupus erythematosus (lupus or SLE) is a heterogeneous autoimmune disease of incompletely understood etiology. The disease is characterized by a loss of immunotolerance and the development of autoantibodies against nuclear antigens. Severe manifestations of lupus have significant impact on quality of life and can lead to organ damage and mortality in affected patients, particularly among patients of non-European genetic ancestry [1, 2]. Genetic risk contributes to the development of lupus, but the estimated heritability of lupus is ~30% [3-5]. Indeed, monozygotic twin studies in lupus suggest a substantial non-genetic contribution to the etiology of lupus [6]. Environmental exposures across the lifespan that can directly impact epigenetic regulation and cellular function are suggested to be involved in the pathogenesis of lupus [7, 8].

DNA methylation is an epigenetic mechanism that regulates gene expression through the enzyme-mediated addition of a methyl group to the cytosine bases in the genome. DNA methylation is heritable across cell generations and can promote gene silencing, making it an important component in regulating the plasticity of immune cell identity and function [9]. Early work demonstrated that adoptive transfer of CD4<sup>+</sup> T cells treated *ex vivo* with DNA methyltransferase (DNMT) inhibitors was sufficient to cause lupus-like disease in mice [10] mimicking the DNA methylation inhibition in patients with drug-induced lupus [11]. Since then, other studies have observed that CD4<sup>+</sup> T cells of lupus patients show a distinct shift in global DNA methylation compared to healthy individuals, potentially in part due to defective MEK/ERK signaling, suppressing DNMT1 activity in

CD4+ T cells, and leading to hypomethylation and overexpression of costimulatory genes [12-16].

We have previously observed a robust hypomethylation signature in interferon-regulated genes defining lupus patients [17, 18]. Our initial findings in CD4+ T cells were subsequently confirmed and expanded to other cell types by our group and others [19-21]. In CD4+ T cells, we observed hypomethylation in interferon-regulated genes at the naïve CD4+ T cell stage, preceding transcriptional activity. This epigenetic “poising” or “priming” of interferon-regulated genes was independent of disease activity [18]. The genetic contribution to this lupus-associated epigenotype is currently unknown.

Methylation quantitative trait loci (meQTL) are genetic polymorphisms that are associated with the methylation state of CpG sites either through direct nucleotide change within the CpG dinucleotide or intermediary mechanisms. Prior studies of lupus patients show an enrichment of meQTL associated with type I interferon genes, genetic risk loci, and specific clinical manifestations in whole blood and neutrophils [22-24]. Furthermore, our previous work suggests that meQTL might at least in part explain differences in DNA methylation between African-American and European-American lupus patients [22].

Herein, we evaluated genome-wide DNA methylation data in naïve CD4+ T cells from a large cohort of lupus patients compared to matched healthy controls. We integrated DNA methylation and genotyping data to better understand the influence of genetic factors upon the DNA methylation changes observed in lupus.

## Methods

**Study Participants and Demographics.** 74 female lupus patients and 74 female healthy age ( $\pm 5$  years), race, and sex-matched controls were recruited as previously described [25, 26] (**Supplementary Table 1**). All patients fulfilled the American College of Rheumatology (ACR) classification criteria for SLE [27]. Institutional review boards at our participating institutions approved this study. All participants signed a written informed consent prior to participation.

**Sample collection, DNA isolation, and data generation.** Genomic DNA samples for this study were collected from naïve CD4+ T cells as previously described [18]. Briefly, magnetic beads and negative selection was used to isolate naïve CD4+ T cells from whole blood samples collected from lupus patients and controls. Genomic DNA was directly isolated from collected cells and bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA). The Illumina HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA) was used to measure DNA methylation levels at over 485,000 methylation sites across the genome.

**Epigenome-wide association study.** Epigenome-wide association study (EWAS) for identifying associations between specific CpG sites and disease status was performed using GLINT [28, 29]. Covariates for age, race, and technical batch were included for the analysis prior to other preprocessing. No outliers beyond four standard deviations were detected in the first two components of PCA space, all 148 samples were included in the analysis. Reference-less cell type composition correction was performed using *ReFACTor*, with six components used in the downstream analysis to account for any cell-type heterogeneity in the samples. An additional covariate was included to account

for effects of genetic admixture using the EPISTRUCTURE algorithm included in GLINT. Cell-type composition covariate components generated by *ReFACTOR* were included at this step to reduce bias from potential cell-type heterogeneity, and polymorphic CpG sites were excluded from this step and the EWAS. Using the initial age, race, and technical batch covariates, along six *ReFACTOR* components and one *EPISTRUCTURE* component, logistic regression for disease status was performed across all CpG sites, excluding the polymorphic and unreliable cross-reactive probes previously described in the literature, as well as CpG sites with low variance (standard deviation <0.01) [30, 31].

**Differential DNA methylation analysis of gene promoters.** Raw .idat files were used to generate methylation beta value profiles across all samples using GenomeStudio (Illumina, San Diego, CA, USA) after background subtraction and normalizing to internal control probes. Missing probe values were imputed using *sklearn.impute.KNNImputer*, and *ComBat* was used to correct for batch effects associated with technician sample preparation [32-34]. Ensembl gene loci for hg19 were downloaded using *PyEnsembl* [35]. For each gene, loci for 1500 base pairs upstream of the transcription start site [36] to the TSS were mapped to the overlapping CpG probes using *PyBedtools*, and the mean of the associated probes for each gene was used as the representative methylation value for the resulting 20,437 mapped genes [37]. Differential methylation analysis comparing patients and controls was performed on the mean TSS1500 methylation using *limma*, and false discovery rate adjustment using the Benjamini-Hochberg method was used to correct P-values for multiple testing. Gene Ontology Enrichment for Biological Process terms was performed on the differentially methylated



gene list using *Enrichr* with the mapped promoter gene list used as the background [38, 39].

**Trend deviation analysis.** DNA methylation data derived using the Illumina 450k methylation array from 23,415 samples were downloaded from Gene Expression Omnibus (GEO) [40]. To reduce batch effects, samples from experiments with less than 50 samples were omitted, and the resulting samples were quantile normalized [41]. A matrix of pairwise Pearson's correlation values for DNA methylation levels was computed across TSS1500 gene promoters in 16,541 samples across 37 tissues to create a multi-tissue correlation network (**Supplementary Figure 1**). The differentially methylated genes in lupus naïve CD4+ T cells were clustered by their correlation in global signature created from the GEO data. Hierarchical clustering was performed using *Scipy's* hierarchical linkage. KEGG enrichment analysis was performed using *Enrichr* [42], and P-values were reported after false-discovery rate adjustment.

The goal of a trend deviation analysis is to detect correlation patterns among differentially methylated genes in large DNA methylation datasets. A correlation in methylation among a set of differentially methylated genes between patients and controls suggests a trend is being observed, reinforcing the significance and robustness of the differential DNA methylation detected between patients and controls.

**Genotyping.** Genomic DNA isolated from naïve CD4+ T cells was used as input for the Infinium Global Screening Array-24 v2.0 (Illumina, San Diego, CA, USA). Single nucleotide polymorphisms (SNPs) with a genotyping call rate < 98%, minor allele frequencies (MAF) < 5%, and deviating from Hardy-Weinberg equilibrium (HWE; P-value < 1E-3) were filtered out. Samples were removed if they had a genotyping call

rate < 95%. Sex chromosomes were not analyzed. About 100,000 independent SNPs were pruned and used to perform principal component analysis (PCA) with *Eigensoft* (v.6.1.4) software [43]. Genotyping data were analyzed using *PLINK* (v.1.9) [44]. Genotype profiles were generated for n = 63 patients and n = 68 controls.

**Methylation Quantitative Trait Loci (meQTL) Analysis.** Raw .idat files were used to generate methylation profiles using *minfi* (v.1.32.0) [45, 46] and to check median intensity values and reported sex in the R statistical computing environment (v.3.6.3) [47]. Probes with less than three beads and zero intensity values across all samples were removed using the *DNAMArray* package (v.0.1.1) [48]. Background signal and dye bias were corrected, followed by normalization of signal intensities using functional normalization in the *preprocessFunnorm.DNAMArray* function [48, 49] using the first three principal component values calculated from signal intensities of control probes present on all array spots to correct for technical variation. Any probe with a detection P-value < 0.01 or returned signal intensities in fewer than 98% of samples was removed. This resulted in a total of 476,767 probes used for further analysis. Signal intensities were then converted to M-values with a maximum bound of  $\pm 16$ . M-values were used for meQTL analysis and converted to beta values (0-100% methylation scale) using *minfi* for reporting.

We removed any probe for meeting any of the following technical criteria: A unique probe sequence of less than 30bp, mapping to multiple sites in the genome, polymorphisms that cause a color channel switching in type I probes, inconsistencies in specified reporter color channel and extension base, mapping to the Y chromosome,

and/or having a polymorphism within 5bp of the 3' end of the probe with a minor allele frequency > 1% with the exception of CpG-SNPs with C>T polymorphisms which were retained [50]. Batch correction for chip ID was performed using the *ComBat* function in the *sva* (v.3.34.0) package [51]. After technical filtering, there were a total of 421,214 probes used for meQTL analysis.

We implemented a mixed correspondence analysis with the *PCAmixdata* package (v.3.1) [52] to calculate eigenvalues using patient medication data for prednisone, hydroxychloroquine, azathioprine, mycophenolate mofetil, and cyclophosphamide. The top four components accounted for a cumulative 89.3% of variability in the medication data. Each component value was used as an independent variable in regression analysis to adjust for medication usage across individuals. MeQTL association analysis was performed in patients and controls separately using methylation M-value profiles and corresponding sample genotypes. Age, the top four medication components, and top ten genotype principal components were included as covariates to build a linear model for detecting meQTL using *MatrixEQTL* (v.2.3) [53]. *Cis*-meQTL were defined as CpG sites with methylation values associated with a SNP within a conservative 1000bp of the CpG dinucleotide. We used a Benjamini-Hochberg FDR-adjusted P-value cutoff of < 0.05 as a threshold for significant associations. The above EWAS results were compared with the meQTL results to determine overlap of lupus-associated differentially methylated CpG sites and those CpG sites in an meQTL.

**Gene Set Enrichment Analysis (GSEA).** ToppGene Suite was used for functional gene ontology enrichment analysis [54] of Molecular Function and Biological Process Gene Ontologies and KEGG Pathways in meQTL loci. P-values were derived using a

hypergeometric probability mass function, and a Benjamini-Hochberg FDR-adjusted P-value cutoff of  $< 0.05$  was used as a threshold of significance. A minimum membership of 3 genes and maximum of 2000 genes in each term was used as a threshold for inclusion. IFN-regulated genes were identified using the set of genes associated with the CpG site in each meQTL as input using Interferome (v.2.01) [55]. The type I interferon response genes were defined as genes with an expression fold change of 1.5 or greater between type I interferon-treated and untreated samples using gene expression datasets from all available CD4+ T cell experiments in the Interferome database.

**MicroRNA Expression Microarray.** MicroRNA (miRNA) expression was measured in naïve CD4+ T cells from a subset of lupus patients and healthy matched controls ( $n = 16$ ). Cells were immediately lysed with TRIzol Reagent (ThermoFisher Scientific, NY, USA) followed by storage at  $-80^{\circ}\text{C}$ . Total RNA was isolated using the Direct-zol RNA MiniPrep Kit (Zymo Research, CA, USA) following the manufacturer's directions. The Affymetrix miRNA 4.1 Array Strip (Affymetrix, CA, USA) was used to measure expression of over 2,000 premature and 2,500 mature human miRNA sequences. RNA sequences were polyadenylated and ligated to a biotin-labeled oligomer using the FlashTag Biotin HSR RNA Labeling Kit (Affymetrix, CA, USA). Biotin-labeled sequences were hybridized to array probes and washed then stained with streptavidin-PE. The Affymetrix Expression Console & Transcriptome Analysis Console 2.0 software (Affymetrix, CA, USA) was used to analyze biotin/streptavidin-PE fluorescence measurements. All samples passed signal intensity, polyadenylation, and ligation quality controls. Signal intensities were background adjusted and normalized. Log<sub>2</sub>-

transformed expression values for each probeset was calculated using a robust multi-array average model (23). The Pearson  $r$  correlation coefficient for median expression values of probes for miR-17, miR-18a, miR-19a, miR-19b1, and miR-20a and patient SLEDAI score were calculated using GraphPad Prism (v9.3.0) (GraphPad Software, CA, USA).

## Results

### Differential methylation of gene promoters in naïve CD4+ T cells isolated from

**lupus patients.** A comparison of DNA methylation profiles from circulating naïve CD4+

T cells isolated from 74 lupus patients and 74 age, sex and race matched healthy

controls revealed a total of 2,627 CpGs out of 334,337 total CpG sites included in the

EWAS with a significant differential methylation. Significant hypomethylation in

interferon-regulated genes was observed, consistent with previous reports

**(Supplementary Table 2)**. Average promoter methylation for each gene was calculated

by including all CpG sites on the array within 1500bp of the associated gene's

transcription start site (TSS). A total of 51 genes showed a significant difference in

average promoter methylation between lupus patients and controls (17 hypomethylated

and 34 hypermethylated in patients compared to controls) **(Table 1) (Figure 1)**.

Biological Process Gene Ontology enrichment analysis of differentially methylated

promoter regions did not show significant enrichment compared to the background of all

gene promoters after adjusting for multiple testing **(Supplementary Table 3)**.

The pairwise correlation of the 51 gene promoters identified above was calculated

across a collection of 16,541 samples from 37 tissues available in GEO. Hierarchical

clustering of correlations showed that 21 out of the 51 gene promoters were highly

correlated. KEGG Pathway enrichment analysis showed a significant enrichment for

three pathways among the 21 correlated gene promoters: "microRNAs in cancer" (P-

value = 3.86E-04), "cytokine-cytokine receptor interaction" (P-value = 4.34E-02), and

"rheumatoid arthritis" (P-value = 4.34E-02) **(Table 2) (Figure 2)**. The microRNAs in

"cancer" pathway included genes encoding miR-17, miR-18a, miR-19a, miR-19b1, and

miR-20a. Four of seven CpG sites used to calculate the average promoter methylation (TS1500) in this locus showed a significant reduction in median methylation in lupus patients compared to healthy controls (**Figure 3A**). These sites: cg17799287 ( $\Delta\beta = -5.5\%$ ; P-value =  $2.05E-03$ ), cg07641807 ( $\Delta\beta = -4.4\%$ ; P-value =  $1.71E-02$ ), cg23665802 ( $\Delta\beta = -5.8\%$ ; P-value =  $1.19E-02$ ), and cg02297838 ( $\Delta\beta = -4.9\%$ ; P-value =  $3.48E-02$ ) were all hypomethylated in lupus patients compared to healthy controls and overlapped with enhancers and regions flanking TSSs in peripheral naïve CD4+ T cells using data collected from the Epigenome Roadmap [56] and visualization using the WashU Epigenome Browser [57]. We examined expression levels of the microRNAs included in the “microRNAs in cancer” pathway (miR-17, miR-18a, miR-19a, miR-19b1, and miR-20a) in naïve CD4+ T cells of a subset of our lupus patients (n = 16) and healthy matched controls (n = 16). We did not observe a difference in expression between patients and control. However, two miRNAs, miR-18a-5p and miR-19b1-5p, showed a significant positive correlation (hsa-miR-18a-5a P-value = 0.038 & hsa-miR-19b1-5p P-value = 0.042) between median expression level and SLEDAI scores in lupus patients (**Figure 3B**) (**Supplementary Table 4**).

#### **Naïve CD4+ T cell methylation quantitative trait loci (meQTL) in lupus patients.**

Global genotype profiles were generated in a subset of patients and controls and compared with global DNA methylation profiles to identify CpG sites with allele-specific methylation associations. There was no significant difference in the average age between the patient (n = 63) and control (n = 68) subsets (patient average age = 41.6; patient age SD = 12.8; control average age = 40.8; control age SD = 12.5; t-test statistic = 0.3811; two-tailed P-value = 0.7038). Allele-specific DNA methylation associations

were measured as meQTL where the CpG site was within 1000bp of the measured SNP separately in patients and controls. After adjusting for age, genetic background, and medication use in patients, we identified 5,785 meQTL present in the naïve CD4+ T cells of lupus patients with an FDR-adjusted P-value < 0.05 (**Supplementary Table 5**). These meQTLs represented 4,649 (80.4%) unique CpG sites and 4,120 (71.2%) unique polymorphisms. Of the 4,791 meQTL with a CpG-associated gene annotation, 2,356 (49.2%) were unique.

A linear model adjusting for age and genetic background was fit to controls separately. We identified a total of 7,331 meQTL with an FDR-adjusted P-value < 0.05 in controls (**Supplementary Table 6**). These meQTLs represented 5,885 (80.3%) unique CpG sites and 5,138 (70.1%) unique polymorphisms. Of the 6,061 meQTL with a CpG-associated gene annotation, 2,846 (47.0%) were unique.

We compared meQTL in lupus patients and healthy controls with the 2,627 CpG sites differentially methylated between the two groups. Of these, we identified 17 (0.6%) and 34 (1.3%) unique CpG sites with a significant change in DNA methylation in lupus patients and healthy controls, respectively (**Figure 4A and B**). We examined the overlap of meQTL in lupus patients and healthy controls and identified a total of 3,957 meQTL (68.4% of lupus patient meQTL and 54.0% of healthy control meQTL) shared between both patients and control meQTL sets (**Supplementary Table 7**). This shared set of meQTL contained 8 (0.3% of differentially methylated CpG sites) unique CpG sites that we identified as differentially methylated between lupus patients and controls (**Figure 4C**).



Gene set enrichment analysis was performed using genes associated with CpG sites in our meQTL shared between patients and controls. GSEA revealed multiple ontologies and pathways for cell adhesion (“cell-cell adhesion”; P-value = 1.04E-12, “biological adhesion”; P-value = 6.80E-12, and “cell adhesion”; P-value = 8.25E-12, “Cell adhesion molecules (CAMs)”; P-value = 2.25E-06), transporter associated with antigen processing (TAP) proteins and antigen presentation (“TAP binding”; P-values = 1.59E-7, “peptide antigen binding”; P-value = 4.40E-5), immune disorder pathways (“Type I diabetes mellitus; P-value = 1.92E-8, and “Graft-versus-host disease”; P-value = 4.38E-7) (**Supplementary Table 8**).

There were 1,828 meQTL detected only in lupus patients but not in controls. These were enriched in gene ontologies and pathways related to tissue growth and development (“animal organ morphogenesis”; P-value = 8.44E-10, “urogenital system development”, P-value = 1.05E-07) and gene silencing (“negative regulation of gene silencing by miRNA”; P-value = 2.54E-6, “negative regulation of posttranscriptional gene silencing”; P-value = 5.41E-6) (**Supplementary Table 9**).

We compared our list of meQTL in lupus patients to previously identified lupus susceptibility loci from genome-wide association studies[4, 58-61] and found 41 meQTL contained CpG site-associated genes that overlapped with 20 unique risk loci genes (**Supplementary Table 10**). This included interferon regulatory factor genes *IRF5* and *IRF7*. We examined the overlap of our meQTL-associated genes identified in lupus patient naïve CD4+ T cells and genes that respond to type I interferon treatment in CD4+ T cells to better understand the association between patient genetics and type I

interferon-response gene methylation differences in lupus. A total of 101 unique type I interferon-response genes were meQTL loci in our data (**Supplementary Table 11**).

## Discussion

We generated genome-wide DNA methylation data in naïve CD4+ T cells from a large cohort of lupus patients and matched healthy controls. Implementing an innovative trend deviation analysis, we identified a cluster of microRNAs (miRNAs) (miR-17, miR-18a, miR-19a, miR-19b1, miR-20a) among differentially methylated loci in lupus patients. Promoter methylation analysis revealed significant hypomethylation in this microRNA cluster in lupus patients compared to controls. Trend deviation analysis suggested a coordinated, disease-associated change in promoter methylation for these miRNAs. Indeed, the expression of miR-18a and miR-19b1 included within this cluster positively correlated with disease activity, as measured using SLEDAI score, in our lupus patients. MiRNAs play an important role in post-transcriptional gene regulation by targeting specific complementary gene transcripts for degradation [62]. Peripheral blood cells in lupus patients show altered expression of miRNAs [63]. Some deregulated miRNAs in lupus target DNA methyltransferase 1 (DNMT1), and as a result, contribute to altered DNA methylation patterns in lupus CD4+ T cells [64-66]. MiR-17, miR-18a, and miR-20a form the “miR-17 family” while miR-19a and miR-19b1 form the “miR-19 family”, which are grouped by sequence homology and encoded in a single polycistronic miRNA gene called the “miR-17-92 cluster”. This cluster has been well-studied as an oncogene and an immune regulator [67]. Average promoter methylation of miR-17, miR-18a, miR-19a, miR-19b1, and miR-20a was reduced by ~5% in lupus patients compared to controls, which has not been previously described in immune cells of lupus patients. Enterovirus 71 infection has been observed to suppress miR-17-92 cluster expression by increasing DNMT-mediated promoter methylation [68], and chemical inhibition of DNMT1 activity in

bleomycin-induced lung fibrosis mouse model increases miR-17-92 cluster expression in lung fibroblasts [69]. This suggests that miR-17-92 cluster promoter methylation plays an important role in regulating the expression of its members.

MiR-17-92 cluster genes play a vital role in regulating T cell activities including proliferation and differentiation. Overexpression of miR-17-92 cluster genes promotes lymphoproliferative disease and autoimmunity in mice by targeting critical immunotolerance regulators Bim and PTEN [70]. Conditional knock out of miR-17-92 cluster in a murine model of chronic graft-versus-host disease (cGVHD) reduced disease-associated T cell infiltration and IgG deposition in the skin [71]. In cGVHD mice, miR-17-92 cluster expression in CD4<sup>+</sup> T cells supports Th1, Th17, and Tfh cell differentiation. Loss of miR-17-92 cluster expression leads to a corresponding reduction in Tfh-dependent germinal center formation and plasma cell differentiation [71]. MiR-17, miR-18a, miR-19a, and miR-20a are overexpressed in splenic T cells of MRL/*lpr* mice [72]. Similarly, miR-17, miR-17a, miR-18a, miR-19a, miR-19b1, and miR-20a are overexpressed in circulating CD4<sup>+</sup> T cells of lupus patients [73]. MiR-19b1 expression, specifically, has a significant positive correlation with disease activity as measured by SLEDAI score [73]. MiR-17 and miR-20 are downregulated in circulating PBMCs [74], B cells [75], as well as circulating free miRNAs [76] of lupus patients compared to healthy controls, suggesting tissue-specific and miR-specific expression patterns. Of the miR-17-92 cluster miRNAs identified as differentially methylated in our analysis, only miR-18a and miR-19b1 showed a significant positive correlation between median expression and disease activity in naïve CD4<sup>+</sup> T cells of lupus patients, consistent with these prior observations. MiR-19b1 promotes proliferation of mature CD4<sup>+</sup> T cells, Th1

differentiation and IFN- $\gamma$  production, and suppresses inducible Treg differentiation [77]. MiR-18a expression increases rapidly early on in CD4+ T cell activation [78, 79], and suppresses Th17 cell differentiation through direct targeting of critical Th17 transcription factor transcripts including *Smad4*, *Hif1a*, and *Rora* in human CD4+ T cells *in vitro* and *in vivo* murine airway inflammation models [78]. We did not observe a difference in the expression of members in the miR-17-92 cluster between lupus patients and controls in naïve CD4+ T cells, likely because these miRNAs are upregulated upon T cell activation. Evidence for hypomethylation in lupus in naïve CD4+ T cells suggests epigenetic priming of this locus, similar what we previously observed in interferon-regulated gene loci in lupus [18]. Further study is needed to determine if altered DNA methylation at these miRNA promoter sites is associated with expression changes in miRNAs that play a role in T cell development and lupus pathogenesis and their potential use as a biomarker for monitoring disease activity.

We used analysis of meQTL to identify allele-specific DNA methylation associations across the genome of naïve CD4+ T cells from lupus patients and healthy controls. Our primary objective was to understand to what extent are DNA methylation changes associated with lupus (the lupus-defining epigenetic profile), explained by genetic factors. We found that < 1% of differentially methylated sites in lupus patients compared to healthy controls were associated with a *cis*-meQTL. This suggests that almost all of the DNA methylation alterations observed in lupus are not associated with local allelic differences in the genome, suggesting a greater contribution from non-genetic and possibly environmental factors to epigenetic dysregulation in lupus. A previous study of meQTL in whole blood of lupus patients found that a majority of meQTLs were shared

between patients and controls [24]. We observed 68% of meQTL in lupus patients and 54% of meQTL in healthy controls were shared by both groups, supporting this observation. Gene ontology and pathway analysis of the meQTL-associated genes unique to lupus patients were enriched for KEGG pathways related to type I diabetes, viral myocarditis, and graft-versus-host disease all of which were primarily driven by the presence of meQTLs in HLA genes. Gene ontologies were related to the development of various tissues without an apparent relationship to disease pathogenesis. Of note, we did not observe meQTL effect involving the miR-17-92 cluster.

Our prior analysis of neutrophils from a cohort of lupus patients identified overlap in meQTL genes and lupus genetic risk loci [22]. MeQTL pairs including *ARID5B* (cg13344587-rs10821936), *HLA-DQB1* (cg13047157-rs9274477), and *IRF7* (cg16486109-GSA-rs1131665) were found in both neutrophils and naïve CD4+ T cells from lupus patients. Risk loci genes unique to naïve CD4+ T cell meQTLs included *CD80* (cg06300880-GSA-rs3915166), *TYK2* (cg06622468-rs280501), *IKBKE* (cg22577136-GSA-rs17020312), and *CTLA4* (cg05092371-GSA-rs16840252, cg05092371-rs4553808). Naïve CD4+ T cell-specific meQTL risk loci genes are all related to signal response and activation in CD4+ T cells compared to the more general DNA repair and type I interferon signaling seen in the shared meQTL risk loci genes. Disease-relevant meQTL show tissue-specific patterns which should be considered when teasing apart their potential impact.

In summary, we investigated genome-wide DNA methylation changes in naïve CD4+ T cells from an extended cohort of lupus patients and controls, and using a methylation trend deviation analysis method, we showed promoter hypomethylation of the miR-17-

92 cluster that has a significant regulatory function in T cells growth, function, and differentiation. Combining genome-wide DNA methylation and genotyping data, we were able to determine genetic contribution to the lupus-defining epigenotype. Our data indicate that epigenetic changes characteristic of lupus are not under direct genetic influence. This suggests a more important role for non-genetic factors in the epigenetic dysregulation observed in lupus patients, including the robust demethylation of interferon-regulated genes.

### **Acknowledgements**

This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) grant number R01 AI097134 to Dr. Sawalha. Dr. Wren is supported by NIH grants number P30 AG050911 and P20 GM103636.

## References

1. Olesinska M, Saletta A: **Quality of life in systemic lupus erythematosus and its measurement.** *Reumatologia* 2018, **56**(1):45-54.
2. Lewis MJ, Jawad AS: **The effect of ethnicity and genetic ancestry on the epidemiology, clinical features and outcome of systemic lupus erythematosus.** *Rheumatology (Oxford)* 2017, **56**(suppl\_1):i67-i77.
3. Kwon YC, Chun S, Kim K, Mak A: **Update on the Genetics of Systemic Lupus Erythematosus: Genome-Wide Association Studies and Beyond.** *Cells* 2019, **8**(10).
4. Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tombleson P, Chen L, Cunninghame Graham DS, Bentham J, Roberts AL *et al*: **Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus.** *Nat Genet* 2016, **48**(8):940-946.
5. Sun C, Molineros JE, Looger LL, Zhou XJ, Kim K, Okada Y, Ma J, Qi YY, Kim-Howard X, Motghare P *et al*: **High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry.** *Nat Genet* 2016, **48**(3):323-330.
6. Generali E, Ceribelli A, Stazi MA, Selmi C: **Lessons learned from twins in autoimmune and chronic inflammatory diseases.** *J Autoimmun* 2017, **83**:51-61.
7. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG *et al*: **Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.** *Genome Res* 2010, **20**(2):170-179.
8. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J *et al*: **Epigenetic differences arise during the lifetime of monozygotic twins.** *Proc Natl Acad Sci U S A* 2005, **102**(30):10604-10609.
9. Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Pérez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S *et al*: **A Critical Role for Dnmt1 and DNA Methylation in T Cell Development, Function, and Survival.** *Immunity* 2001, **15**(5):763-774.
10. Qudus J, Johnson KJ, Gavalchin J, Amento EP, Chrisp CE, Yung RL, Richardson BC: **Treating activated CD4+ T cells with either of two distinct DNA methyltransferase inhibitors, 5-azacytidine or procainamide, is sufficient to cause a lupus-like disease in syngeneic mice.** *J Clin Invest* 1993, **92**(1):38-53.
11. Cornacchia E, Golbus J, Maybaum J, Strahler J, Hanash S, Richardson B: **Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity.** *J Immunol* 1988, **140**(7):2197-2200.
12. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M: **Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis.** *Arthritis Rheum* 1990, **33**(11):1665-1673.



13. Lu Q, Wu A, Tesmer L, Ray D, Yousif N, Richardson B: **Demethylation of CD40LG on the inactive X in T cells from women with lupus.** *J Immunol* 2007, **179**(9):6352-6358.
14. Lu Q, Wu A, Richardson BC: **Demethylation of the same promoter sequence increases CD70 expression in lupus T cells and T cells treated with lupus-inducing drugs.** *J Immunol* 2005, **174**(10):6212-6219.
15. Lu Q, Kaplan M, Ray D, Ray D, Zacharek S, Gutsch D, Richardson B: **Demethylation of ITGAL (CD11a) regulatory sequences in systemic lupus erythematosus.** *Arthritis Rheum* 2002, **46**(5):1282-1291.
16. Sawalha AH, Jeffries M, Webb R, Lu Q, Gorelik G, Ray D, Osban J, Knowlton N, Johnson K, Richardson B: **Defective T-cell ERK signaling induces interferon-regulated gene expression and overexpression of methylation-sensitive genes similar to lupus patients.** *Genes Immun* 2008, **9**(4):368-378.
17. Jeffries MA, Dozmorov M, Tang Y, Merrill JT, Wren JD, Sawalha AH: **Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus.** *Epigenetics* 2011, **6**(5):593-601.
18. 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.
19. Coit P, Yalavarthi S, Ognenovski M, Zhao W, Hasni S, Wren JD, Kaplan MJ, Sawalha AH: **Epigenome profiling reveals significant DNA demethylation of interferon signature genes in lupus neutrophils.** *J Autoimmun* 2015, **58**:59-66.
20. Hedrich CM, Mabert K, Rauen T, Tsokos GC: **DNA methylation in systemic lupus erythematosus.** *Epigenomics* 2017, **9**(4):505-525.
21. Ballestar E, Sawalha AH, Lu Q: **Clinical value of DNA methylation markers in autoimmune rheumatic diseases.** *Nat Rev Rheumatol* 2020, **16**(9):514-524.
22. Coit P, Ortiz-Fernandez L, Lewis EE, McCune WJ, Maksimowicz-McKinnon K, Sawalha AH: **A longitudinal and transancestral analysis of DNA methylation patterns and disease activity in lupus patients.** *JCI Insight* 2020, **5**(22).
23. Lanata CM, Paranjpe I, Nititham J, Taylor KE, Gianfrancesco M, Paranjpe M, Andrews S, Chung SA, Rhead B, Barcellos LF *et al*: **A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus.** *Nat Commun* 2019, **10**(1):3902.
24. Imgenberg-Kreuz J, Carlsson Almlof J, Leonard D, Alexsson A, Nordmark G, Eloranta ML, Rantapaa-Dahlqvist S, Bengtsson AA, Jonsen A, Padyukov L *et al*: **DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus.** *Ann Rheum Dis* 2018, **77**(5):736-743.
25. Coit P, Ognenovski M, Gensterblum E, Maksimowicz-McKinnon K, Wren JD, Sawalha AH: **Ethnicity-specific epigenetic variation in naive CD4+ T cells and the susceptibility to autoimmunity.** *Epigenetics Chromatin* 2015, **8**:49.
26. Mok A, Solomon O, Nayak RR, Coit P, Quach HL, Nititham J, Sawalha AH, Barcellos LF, Criswell LA, Chung SA: **Genome-wide profiling identifies associations between lupus nephritis and differential methylation of genes**

- regulating tissue hypoxia and type 1 interferon responses.** *Lupus Sci Med* 2016, **3**(1):e000183.
27. Hochberg MC: **Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus.** *Arthritis Rheum* 1997, **40**(9):1725.
  28. Flanagan JM: **Epigenome-wide association studies (EWAS): past, present, and future.** *Methods Mol Biol* 2015, **1238**:51-63.
  29. Rahmani E, Yedidim R, Shenhav L, Schweiger R, Weissbrod O, Zaitlen N, Halperin E: **GLINT: a user-friendly toolset for the analysis of high-throughput DNA-methylation array data.** *Bioinformatics* 2017, **33**(12):1870-1872.
  30. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics* 2013, **8**(2):203-209.
  31. Rahmani E, Shenhav L, Schweiger R, Yousefi P, Huen K, Eskenazi B, Eng C, Huntsman S, Hu D, Galanter J *et al*: **Genome-wide methylation data mirror ancestry information.** *Epigenetics Chromatin* 2017, **10**:1.
  32. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis.** *PLoS Genet* 2007, **3**(9):1724-1735.
  33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V: **Scikit-learn: Machine learning in Python.** *the Journal of machine Learning research* 2011, **12**:2825-2830.
  34. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2007, **8**(1):118-127.
  35. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T *et al*: **The Ensembl genome database project.** *Nucleic Acids Res* 2002, **30**(1):38-41.
  36. Consortium F, the RP, Clst, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T *et al*: **A promoter-level mammalian expression atlas.** *Nature* 2014, **507**(7493):462-470.
  37. Dale RK, Pedersen BS, Quinlan AR: **Pybedtools: a flexible Python library for manipulating genomic datasets and annotations.** *Bioinformatics* 2011, **27**(24):3423-3424.
  38. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A *et al*: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Res* 2016, **44**(W1):W90-97.
  39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**(7):e47.
  40. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: **NCBI GEO: archive for functional genomics data sets--update.** *Nucleic Acids Res* 2013, **41**(Database issue):D991-995.

41. Wang T, Guan W, Lin J, Boutaoui N, Canino G, Luo J, Celedon JC, Chen W: **A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data.** *Epigenetics* 2015, **10**(7):662-669.
42. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J *et al*: **SciPy 1.0: fundamental algorithms for scientific computing in Python.** *Nat Methods* 2020, **17**(3):261-272.
43. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**(8):904-909.
44. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: **Second-generation PLINK: rising to the challenge of larger and richer datasets.** *Gigascience* 2015, **4**:7.
45. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.** *Bioinformatics* 2014, **30**(10):1363-1369.
46. Fortin JP, Triche TJ, Jr., Hansen KD: **Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.** *Bioinformatics* 2017, **33**(4):558-560.
47. Team RC: **R: A Language and Environment for Statistical Computing.** 2020.
48. Sinke L, van Iterson, Maarten, Cats, Davy, Slieker, Roderick & Heijmans, Bas: **DNAMArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data.** *Zenodo* 2019.
49. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD: **Functional normalization of 450k methylation array data improves replication in large cancer studies.** *Genome Biol* 2014, **15**(12):503.
50. Zhou W, Laird PW, Shen H: **Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes.** *Nucleic Acids Res* 2017, **45**(4):e22.
51. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments.** *Bioinformatics* 2012, **28**(6):882-883.
52. Chavent M, Kuentz-Simonet V, Labenne A, Saracco J: **Multivariate analysis of mixed data: The R Package PCAmixdata.** *arXiv preprint arXiv:14114911* 2014.
53. Shabalin AA: **Matrix eQTL: ultra fast eQTL analysis via large matrix operations.** *Bioinformatics* 2012, **28**(10):1353-1358.
54. Chen J, Bardes EE, Aronow BJ, Jegga AG: **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W305-311.
55. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, Chapman R, Hertzog PJ: **Interferome v2.0: an updated database of annotated interferon-regulated genes.** *Nucleic Acids Res* 2013, **41**(Database issue):D1040-1046.

56. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al*: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**(7539):317-330.
57. Li D, Hsu S, Purushotham D, Sears RL, Wang T: **WashU Epigenome Browser update 2019**. *Nucleic Acids Res* 2019, **47**(W1):W158-W165.
58. Bentham J, Morris DL, Graham DSC, Pinder CL, Tomblinson P, Behrens TW, Martin J, Fairfax BP, Knight JC, Chen L *et al*: **Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus**. *Nat Genet* 2015, **47**(12):1457-1464.
59. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, Xu JH, Cai ZM, Huang W, Zhao GP *et al*: **Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus**. *Nat Genet* 2009, **41**(11):1234-1237.
60. Alarcon-Riquelme ME, Ziegler JT, Molineros J, Howard TD, Moreno-Estrada A, Sanchez-Rodriguez E, Ainsworth HC, Ortiz-Tello P, Comeau ME, Rasmussen A *et al*: **Genome-Wide Association Study in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture**. *Arthritis & rheumatology (Hoboken, NJ)* 2016, **68**(4):932-943.
61. Lee YH, Bae SC, Choi SJ, Ji JD, Song GG: **Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis**. *Mol Biol Rep* 2012, **39**(12):10627-10635.
62. Gebert LFR, MacRae IJ: **Regulation of microRNA function in animals**. *Nat Rev Mol Cell Biol* 2019, **20**(1):21-37.
63. Shen N, Liang D, Tang Y, de Vries N, Tak PP: **MicroRNAs--novel regulators of systemic lupus erythematosus pathogenesis**. *Nat Rev Rheumatol* 2012, **8**(12):701-709.
64. Pan W, Zhu S, Yuan M, Cui H, Wang L, Luo X, Li J, Zhou H, Tang Y, Shen N: **MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4+ T cells by directly and indirectly targeting DNA methyltransferase 1**. *J Immunol* 2010, **184**(12):6773-6781.
65. Zhao S, Wang Y, Liang Y, Zhao M, Long H, Ding S, Yin H, Lu Q: **MicroRNA-126 regulates DNA methylation in CD4+ T cells and contributes to systemic lupus erythematosus by targeting DNA methyltransferase 1**. *Arthritis Rheum* 2011, **63**(5):1376-1386.
66. Qin H, Zhu X, Liang J, Wu J, Yang Y, Wang S, Shi W, Xu J: **MicroRNA-29b contributes to DNA hypomethylation of CD4+ T cells in systemic lupus erythematosus by indirectly targeting DNA methyltransferase 1**. *J Dermatol Sci* 2013, **69**(1):61-67.
67. Mogilyansky E, Rigoutsos I: **The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease**. *Cell Death Differ* 2013, **20**(12):1603-1614.

68. Fu Y, Zhang L, Zhang R, Xu S, Wang H, Jin Y, Wu Z: **Enterovirus 71 Suppresses miR-17-92 Cluster Through Up-Regulating Methylation of the miRNA Promoter.** *Front Microbiol* 2019, **10**:625.
69. Dakhlallah D, Batte K, Wang Y, Cantemir-Stone CZ, Yan P, Nuovo G, Mikhail A, Hitchcock CL, Wright VP, Nana-Sinkam SP *et al*: **Epigenetic regulation of miR-17~92 contributes to the pathogenesis of pulmonary fibrosis.** *Am J Respir Crit Care Med* 2013, **187**(4):397-405.
70. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, Henderson JM, Kutok JL, Rajewsky K: **Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes.** *Nat Immunol* 2008, **9**(4):405-414.
71. Wu YX, Schutt S, Paz K, Zhang MM, Flynn RP, Bastian D, Sofi MH, Nguyen H, Dai M, Liu C *et al*: **MicroRNA-17-92 is required for T-cell and B-cell pathogenicity in chronic graft-versus-host disease in mice.** *Blood* 2018, **131**(17):1974-1986.
72. Dai R, Zhang Y, Khan D, Heid B, Caudell D, Crasta O, Ahmed SA: **Identification of a common lupus disease-associated microRNA expression pattern in three different murine models of lupus.** *PLoS One* 2010, **5**(12):e14302.
73. Qin HH, Zhu XH, Liang J, Wu JF, Yang YS, Xu JH: **The expression and significance of miR-17-92 cluster miRs in CD4+ T cells from patients with systemic lupus erythematosus.** *Clin Exp Rheumatol* 2013, **31**(3):472-473.
74. Dai Y, Huang YS, Tang M, Lv TY, Hu CX, Tan YH, Xu ZM, Yin YB: **Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients.** *Lupus* 2007, **16**(12):939-946.
75. Te JL, Dozmorov IM, Guthridge JM, Nguyen KL, Cavett JW, Kelly JA, Bruner GR, Harley JB, Ojwang JO: **Identification of unique microRNA signature associated with lupus nephritis.** *PLoS One* 2010, **5**(5):e10344.
76. Carlsen AL, Schetter AJ, Nielsen CT, Lood C, Knudsen S, Voss A, Harris CC, Hellmark T, Segelmark M, Jacobsen S *et al*: **Circulating microRNA expression profiles associated with systemic lupus erythematosus.** *Arthritis Rheum* 2013, **65**(5):1324-1334.
77. Jiang S, Li C, Olive V, Lykken E, Feng F, Sevilla J, Wan Y, He L, Li QJ: **Molecular dissection of the miR-17-92 cluster's critical dual roles in promoting Th1 responses and preventing inducible Treg differentiation.** *Blood* 2011, **118**(20):5487-5497.
78. Montoya MM, Maul J, Singh PB, Pua HH, Dahlstrom F, Wu N, Huang X, Ansel KM, Baumjohann D: **A Distinct Inhibitory Function for miR-18a in Th17 Cell Differentiation.** *J Immunol* 2017, **199**(2):559-569.
79. Teteloshvili N, Smigielska-Czepiel K, Kroesen BJ, Brouwer E, Kluiver J, Boots AM, van den Berg A: **T-cell Activation Induces Dynamic Changes in miRNA Expression Patterns in CD4 and CD8 T-cell Subsets.** *Microna* 2015, **4**(2):117-122.

Figure 1

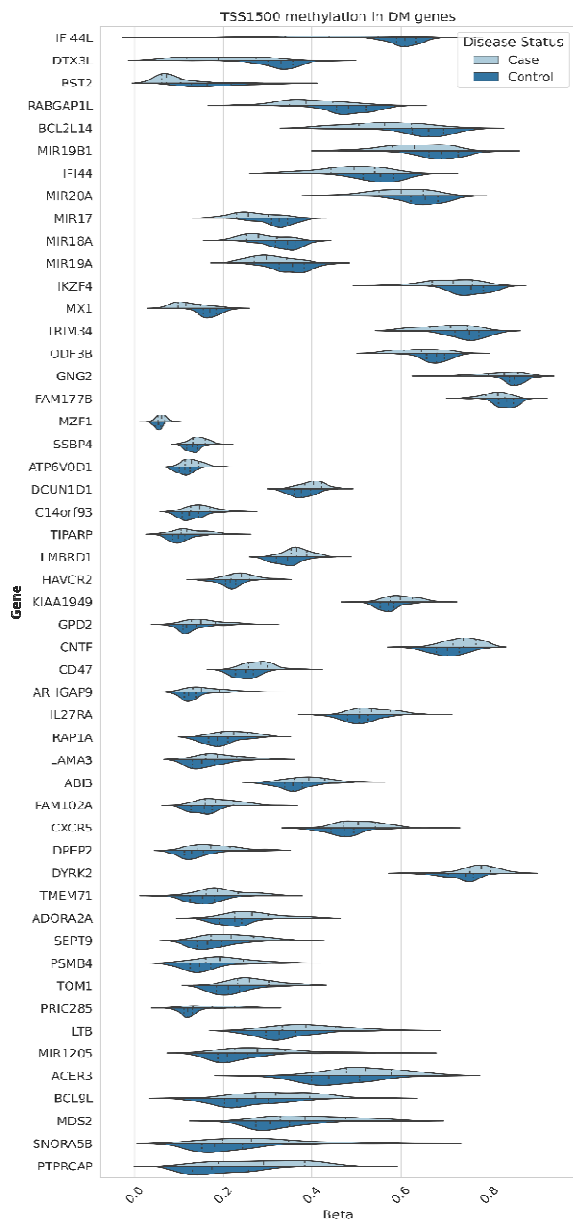
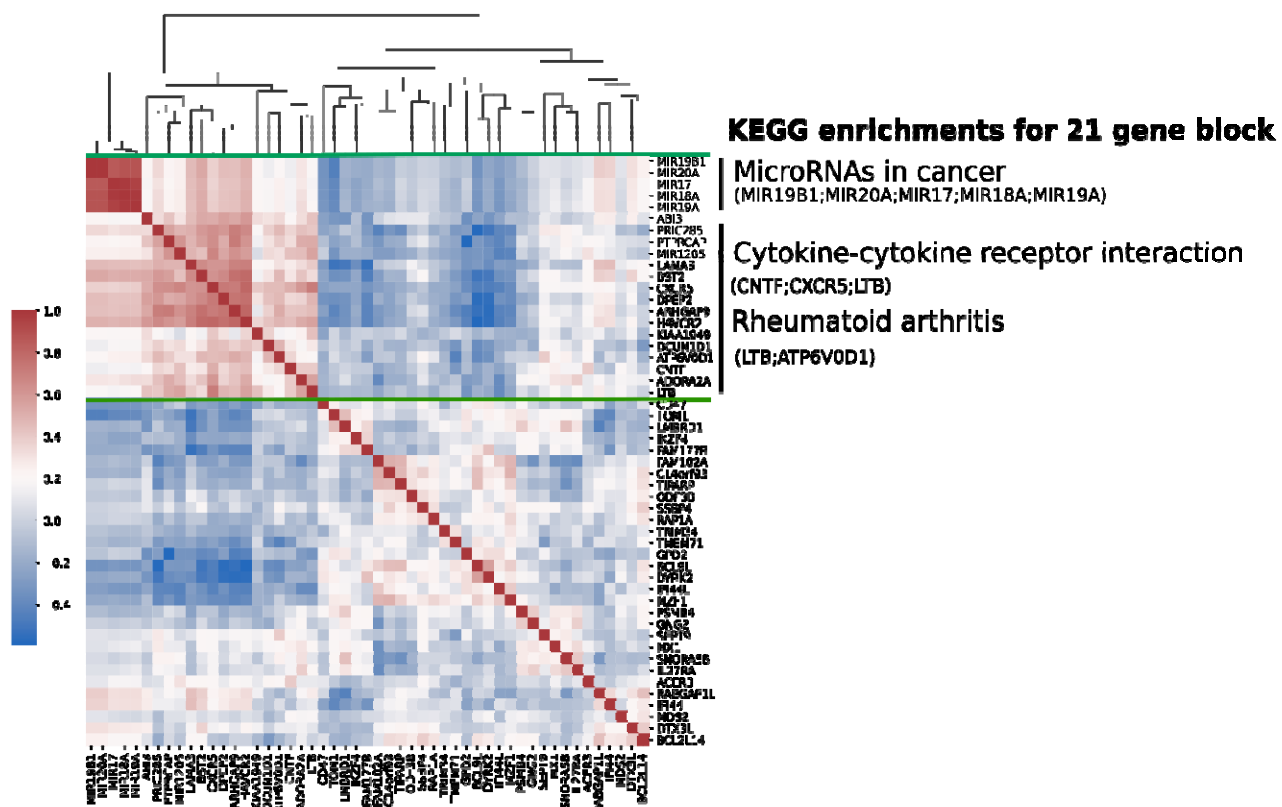
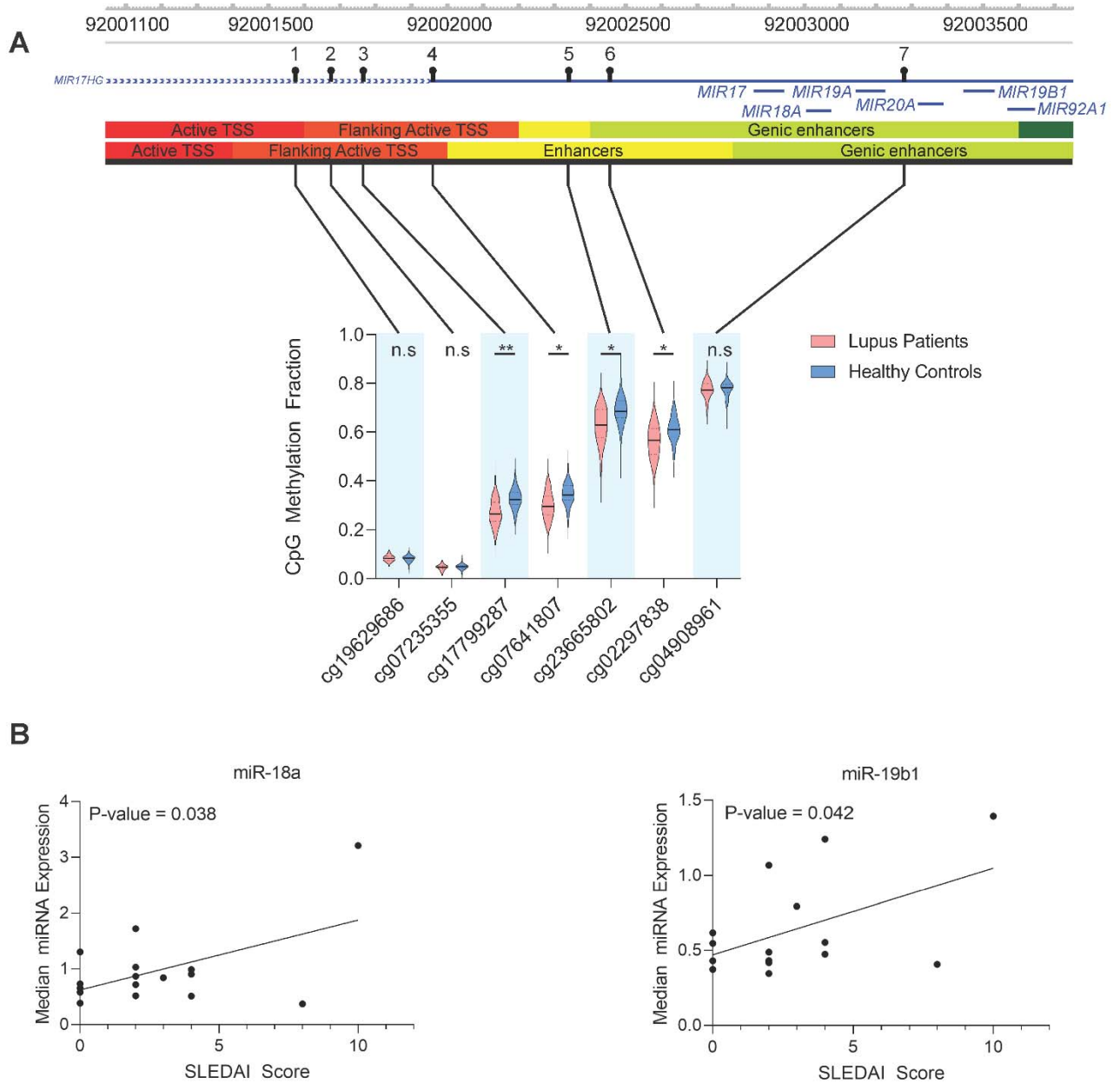


Figure 2

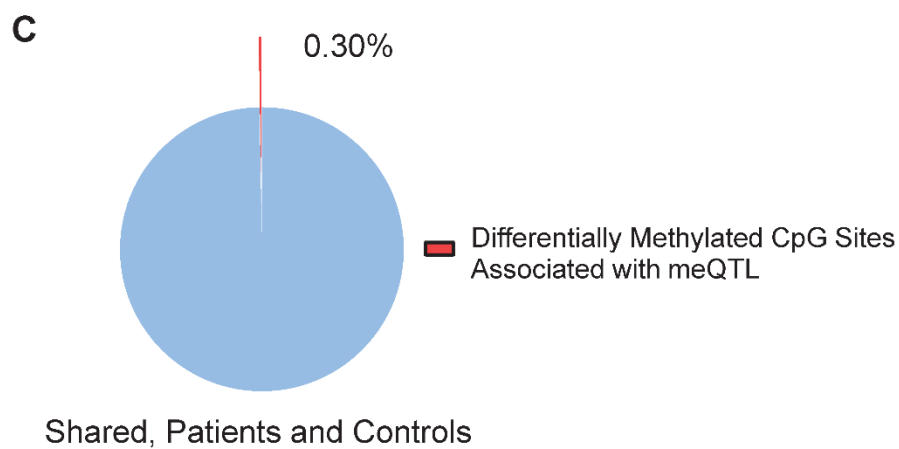
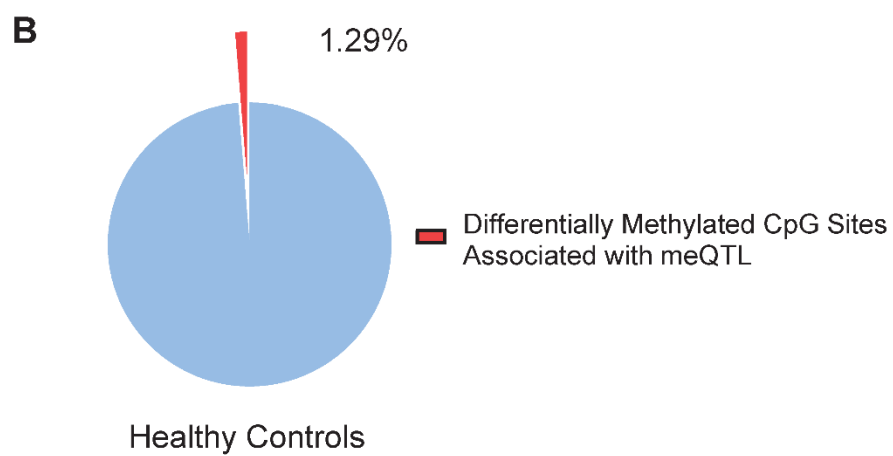
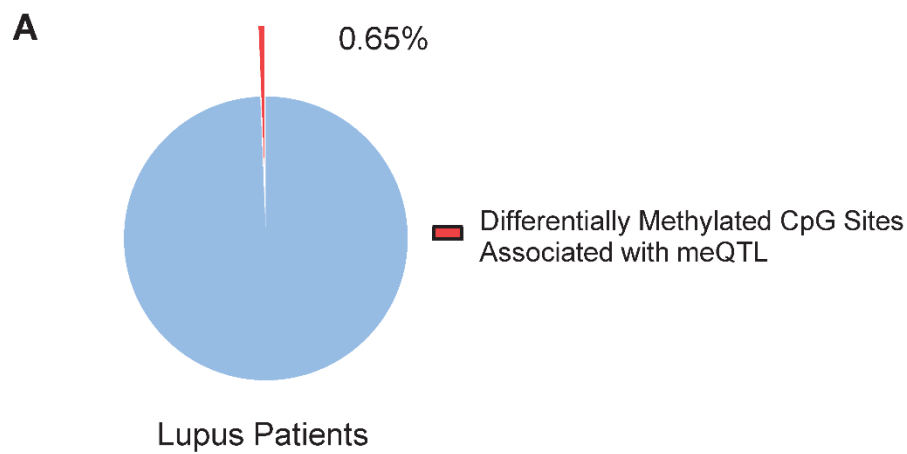


**Figure 3**





**Figure 4**



## Figure legends

**Figure 1.** Distribution of average CpG methylation levels within 1500bp of the TSS for the respective genes differentially methylated in naïve CD4+ T cells of lupus patients compared to healthy controls.

**Figure 2.** Heatmap of hierarchical clustering of pairwise Pearson correlation coefficient values of 51 differentially methylated gene promoters (TSS1500) in global tissue signature derived from 16,541 samples. Range from +1 (red) to -1 (blue), represent a greater to lower correlation in global tissue, respectively. KEGG pathways are significantly enriched (FDR-adjusted P-value < 0.05) in a block of 21 genes (green bars).

**Figure 3. (A)** Violin plots of the seven CG probes in lupus patients and healthy controls used to calculate the average promoter methylation (TSS1500) for the miR-17-92 cluster. The solid black bar represents the median value and the dashed lines the first and third quartiles. Genomic visualization and annotation are from WashU Epigenome Browser using AuxillaryHMM tracks from peripheral naïve CD4+ T cell tissues (E038 and E039, top and bottom tracks, respectively). For P-values: n.s. = not significant, \* =  $P < 0.05$ , \*\* =  $P < 0.01$ . **(B)** Correlation of median miRNA expression in naïve CD4+ T cells of a subset ( $n = 16$ ) of lupus patients with SLEDAI score. Hsa-miR-18a-5p and hsa-miR-19b1-5p had a Pearson correlation ( $r$ ) of 0.52 ( $P$ -value = 0.038) and 0.51 ( $P$ -value = 0.042), respectively.

**Figure 4.** Proportion of differentially methylated CpG sites in naïve CD4+ T cells of lupus patients compared to healthy controls associated with an meQTL in **(A)** lupus patients, **(B)** healthy controls, and **(C)** the subset of meQTL shared between lupus patients and healthy controls.

**Table 1.** Genes with differentially methylated promoter regions in naive CD4+ T cells of lupus patients compared to healthy controls.

<b>Gene</b>	<b><math>\Delta\beta</math></b>	<b><math>-\log_{10}</math> (FDR-adjusted P-value)</b>	<b>t-statistic</b>
<i>IFI44L</i>	-0.177	infinity	-10.757
<i>DTX3L</i>	-0.130	infinity	-11.566
<i>BST2</i>	-0.089	11.323	-9.285
<i>RABGAP1L</i>	-0.088	9.165	-8.421
<i>BCL2L14</i>	-0.086	5.520	-6.908
<i>MIR19B1</i>	-0.059	3.169	-5.846
<i>IFI44</i>	-0.059	2.057	-5.304
<i>MIR20A</i>	-0.055	3.088	-5.807
<i>MIR17</i>	-0.054	6.882	-7.487
<i>MIR18A</i>	-0.051	6.537	-7.342
<i>MIR19A</i>	-0.049	4.771	-6.579
<i>IKZF4</i>	-0.048	3.289	-5.902
<i>MX1</i>	-0.046	10.624	-9.004
<i>TRIM34</i>	-0.045	2.184	-5.367
<i>ODF3B</i>	-0.034	1.712	-5.128
<i>GNG2</i>	-0.033	2.138	-5.344
<i>FAM177B</i>	-0.025	1.897	-5.223
<i>MZF1</i>	0.008	1.493	5.014
<i>SSBP4</i>	0.015	1.344	4.934
<i>ATP6V0D1</i>	0.018	2.594	5.569
<i>DCUN1D1</i>	0.025	2.068	5.309
<i>C14orf93</i>	0.025	1.922	5.236
<i>TIPARP</i>	0.026	2.069	5.310
<i>LMBRD1</i>	0.027	2.211	5.381
<i>HAVCR2</i>	0.027	2.574	5.560
<i>KIAA1949</i>	0.030	3.158	5.841
<i>GPD2</i>	0.032	1.953	5.251
<i>CNTF</i>	0.033	1.705	5.124
<i>CD47</i>	0.034	4.259	6.350
<i>ARHGAP9</i>	0.036	3.339	5.926
<i>IL27RA</i>	0.036	1.367	4.946
<i>RAP1A</i>	0.036	2.573	5.559
<i>LAMA3</i>	0.037	1.445	4.988
<i>ABI3</i>	0.037	1.436	4.983
<i>FAM102A</i>	0.038	3.161	5.842
<i>CXCR5</i>	0.039	1.439	4.985
<i>DPEP2</i>	0.040	1.889	5.219

<i>DYRK2</i>	0.041	3.924	6.197
<i>TMEM71</i>	0.044	2.757	5.649
<i>ADORA2A</i>	0.046	2.234	5.392
<i>SEPT9</i>	0.047	2.036	5.293
<i>PSMB4</i>	0.052	2.935	5.734
<i>TOM1</i>	0.055	5.415	6.862
<i>PRIC285</i>	0.057	9.934	8.729
<i>LTB</i>	0.062	2.036	5.293
<i>MIR1205</i>	0.067	1.698	5.121
<i>ACER3</i>	0.073	2.612	5.578
<i>BCL9L</i>	0.079	4.034	6.248
<i>MDS2</i>	0.080	3.149	5.836
<i>SNORA5B</i>	0.083	1.712	5.128
<i>PTPRCAP</i>	0.091	3.620	6.057

---

FDR correction was performed using the Benjamini-Hochberg method with an FDR-adjusted P value threshold of  $< 0.05$ .  $\Delta\beta$ : methylation difference in median methylation value of CpG sites within 1500bp upstream of the associated gene's transcription start site (TSS1500) between lupus patients and healthy controls.

**Table 2.** KEGG Pathway gene enrichment of 21 gene promoters highly correlated with each other in multi-tissue DNA methylation data constructed from 16,541 samples available through Gene Expression Omnibus.

<b>Pathway (KEGG_2019_Human)</b>	<b>P-value</b>	<b>FDR- adjusted P-value</b>	<b>Odds Ratio</b>	<b>Genes</b>
MicroRNAs in cancer	1.21E-05	0.00039	20.92	<i>MIR19B1;MIR20A;MIR17;MIR18A;MIR19.</i>
Cytokine-cytokine receptor interaction	0.0034	0.043	11.28	<i>CNTF;CXCR5;LTB</i>
Rheumatoid arthritis	0.0041	0.043	23.52	<i>LTB;ATP6V0D1</i>

**Supplementary Figure 1.** Sample distribution of 16,541 Infinium HumanMethylation450 array samples across 37 tissues sourced from the Gene Expression Omnibus (GEO). These samples were used to develop a multi-tissue correlation network used for trend deviation analysis.

