Epigenetic programming underpins B cell dysfunction in human SLE

Christopher D. Scharer1, Emily L. Blalock2,3, Tian Mi1, Benjamin G. Barwick4, Scott A. Jenks2,3, Tsuneo Deguchi2,3, Kevin S. Cashman2,3, Bridget E. Neary2,3, Dillon G. Patterson1, Sakeenah L. Hicks1, Arezou Khosroshahi2,3, F. Eun-Hyung Lee5, Chungwen Wei6, Iñaki Sanz2,3* and Jeremy M. Boss1*.

Systemic lupus erythematosus (SLE) is characterized by the expansion of extracellular pathogenic B cells derived from newly activated naïve cells. Although these cells express distinct markers, their epigenetic architecture and how it contributes to SLE remain poorly understood. To address this, we determined the DNA methylomes, chromatin accessibility profiles and transcripts of five human B cell subsets, including a newly defined effector B cell subset, from subjects with SLE and healthy controls. Our data define a differentiation hierarchy for the subsets and elucidate the epigenetic and transcriptional differences between effector and memory B cells. Importantly, an SLE molecular signature was already established in resting naïve cells and was dominated by enrichment of accessible chromatin in motifs for AP-1 and EGR transcription factors. Together, these factors act in synergy with T-BET to shape the epigenome of expanded SLE effector B cell subsets. Thus, our data define the molecular foundation of pathogenic B cell dysfunction in SLE.

SLE is characterized by the production of autoantibodies, which places B cells centrally in SLE etiology. Autoreactive B cells are censored through tolerance checkpoints, which can be overcome by the convergence of Toll-like receptor (TLR), cytokine and/or co-receptor malfunction thereby leading to the expansion of pathogenic B cells. Genome-wide association studies (GWAS) of SLE have revealed a striking concentration of disease-susceptibility alleles in the B cell antigen receptor (BCR) signaling and B cell co-stimulation pathways. Experimental evidence supports a role for both germinal center reactions and extracellular B cell activation and differentiation pathways in the generation of autoantibodies and autoimmunity in mice. We have shown previously that SLE is characterized by the expansion of naïve B cells with an activated phenotype and a distinct subset of isotype-switched B cells that harbor somatic hypermutation but lack expression of CD27, a universal marker of memory B cells, termed DN2 (CD27+CD11c+T-BET+CXCR5+) B cells. DN2 cells are poised to differentiate into antibody-secreting cells (ASCs) through unregulated TLR7 and interleukin 21 (IL-21) stimulation and bear resemblance to murine age-associated B cells (ABCs). ABCs require TLR7 signaling for expansion, express CD11c and T-BET and are enriched in BCR-encoding autoreactive B cells. TLR signaling can be modulated by type I interferon (IFN), which is prevalent in SLE, and may epigenetically influence disease flares. The strong linkage of DN2 cells and related subsets to TLR-driven pathways suggests that a unique combination of stimuli influences the fate of autoreactive B cells.

Altered epigenetic states, including DNA hypomethylation in B cells and disease-associated peaks of chromatin accessibility in naïve B cells, have been described for SLE and other autoimmune diseases, suggesting that epigenetic differences may influence B cell responses. Therefore, we examined and compared the DNA methylation epigenomes, chromatin accessibility profiles and transcripts of human B cells from healthy controls and subjects with SLE, analyzing cell subsets that represent resting, activated and memory compartments. A hierarchy in B cell differentiation and multiple SLE disease signatures that already manifested in naïve B cells were found to persist throughout B cell differentiation. Transcription factor networks that were altered in subjects with SLE converged on signaling networks and revealed external environmental cues that contribute to the expansion of pathogenic B cell subsets. Thus, the SLE environment predisposes B cells to a pathogenic phenotype that is epigenetically propagated through B cell differentiation and primes differentiation of extracellular naïve B cells into ASCs.

Results

Molecular relationships of B cell subsets in subjects with SLE and healthy controls. To characterize the epigenetic relationships among human B cell populations in subjects with SLE, a cohort of African-American females (9 subjects with SLE and 12 healthy controls) was recruited (Supplementary Table 1). Peripheral B cells can be divided into subtypes that represent naïve B cells, activated stages and isotype-switched memory cells. Five circulating human B cell subsets were isolated by flow cytometry from each subject using MitoTracker Green dye (MTG) and phenotypic markers: resting naïve (CD19+IgD+CD27−MTG−CD24+CD38+), transitional1 (T3; CD19+IgD+CD27−MTG−CD24+CD38+), activated naïve (CD19+IgD+CD27−MTG+CD24−CD38−), isotype-switched memory (CD19+IgD−CD27+) and double-negative (DN2; CD19+IgD−CD27−CXCR5−) B cells (Supplementary Fig. 1). For comparison, circulating ASCs were also isolated (CD19+IgD−CD27+CD38+) from a subset of the subjects. The DN2...

1Department of Microbiology and Immunology, School of Medicine, Emory University, Atlanta, GA, USA. 2Division of Rheumatology, Department of Medicine, School of Medicine, Emory University, Atlanta, GA, USA. 3Lowance Center for Human Immunology, School of Medicine, Emory University, Atlanta, GA, USA. 4Department of Hematology and Medical Oncology, School of Medicine, Emory University, Atlanta, GA, USA. 5Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, School of Medicine, Emory University, Atlanta, GA, USA.

*e-mail: ignacio.sanz@emory.edu; jmboss@emory.edu
and activated naive populations, both characterized by a CXCR5-phenotype, are expanded in individuals with SLE experiencing flares and are potential disease drivers. Consistent with results from previous cohorts, subjects with SLE had highly significant expansion of the activated naive and DN2 B cell populations, smaller but significant increases in the T3 B cell population and a subtle decrease in the population of resting naive B cells (Supplementary Fig. 1).

An integrated approach was taken that assayed DNA methylation status by reduced-representation bisulfite sequencing (RRBS) and the effect of epigenetic programs on the accessibility of chromatin, as determined by assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Supplementary Fig. 2). The phenotypic readout for epigenetic programs was assessed by RNA sequencing (RNA-seq). Differentially methylated loci (DMLs), differentially accessible regions (DARs) and differentially expressed genes (DEGs) were determined in comparisons of each cell type according to SLE disease status. Principal-component analysis (PCA) of all DML indicated a linear separation of cell types that progressed from resting naive to isotype-switched memory B cells in both subjects with SLE and healthy controls (Fig. 1a). One major epigenetic feature of B cell differentiation is the targeted and progressive global hypomethylation that occurs as B cells differentiate to ASCs, suggesting that the changes observed by PCA may represent a differentiation hierarchy. Indeed, we observed the same phenomena with clear progressive loss of DNA methylation across these cell subsets (Fig. 1b). A phylogenetic analysis was performed using the DMLs identified through comparison of all cell types including ASCs, such that the beginning and differentiation endpoints were represented. For both subjects with SLE and healthy controls, activated naive and T3 B cells were much more similar to resting naive B cells, whereas DN2 and isotype-switched memory B cells were closely related to ASCs (Fig. 1c). Thus, DNA methylation changes separate the various cell subsets and are indicative of a hierarchical relationship.

PCA of DARs and DEGs indicated that there was a progression from resting naive cells through T3, activated naive and isotype-switched memory to DN2 cells. Whereas the DNA methylation analysis separated the cell types along both principal components, chromatin accessibility and gene expression PCA identified SLE B cells as distinct from B cells from healthy controls with principal component 1 (PC1) separating the populations by disease status and PC2 separating the populations by cell types (Fig. 1a). For phylogenetic analysis of chromatin accessibility data, promoters and distal elements were considered separately as the latter have been shown to better separate cell types. Accessible promoters had a small overall phylogenetic distance between cell types in SLE and control samples, indicating similar promoter accessibility architecture, although DN2 and isotype-switched memory cells were closer in SLE samples than in healthy controls (Fig. 1c). In contrast, when considering accessibility at distal elements, an expanded phylogenetic distance was observed, indicating that cis elements such as enhancers are distinct between SLE and control cell types. Similarly to previously reported data, analysis of the RNA-seq, ATAC-seq and RRBS data indicated that DN2 cells were most closely related to ASCs by phylogenetic tree position and distance (Fig. 1c). As with the accessibility data, the transcriptomes of DN2 and activated naive cells showed a much closer relationship in SLE samples than in control samples, a finding that was supported by fewer DARs and DEGs in comparison of these populations in SLE samples than in control HC samples (Supplementary Fig. 3). Thus, distinct epigenomes differentiate critical B cell subsets in subjects with SLE and healthy controls.

Progressive upregulation of the B cell differentiation transcriptional program. To better understand the phenotypic relationships of the cell subsets, the transcriptional programs were analyzed using gene set enrichment analysis (GSEA) comparing B cell subsets from healthy controls and subjects with SLE to the respective resting naive B cells. Across cell subsets, a progressive enrichment was observed in gene sets associated with metabolism (for example, ALDOA), cell cycle (E2F1), unfolded protein response (UPR; XBP1) and activated B cells (ZBTB32) as well as genes expressed in plasma cells (PRDM1 and SLAMF7) (Fig. 1d,e and Supplementary Fig. 3). The overall expression levels of the gene sets were increased in T3 and activated naive cells and peaked in isotype-switched memory and DN2 cells. For example, consistent with the transcriptional upregulation of PRDM1, the PRDM1 promoter demonstrated increased accessibility and a decrease in DNA methylation that was most pronounced in DN2 and isotype-switched memory B cells. Additionally, transcriptional activity, as well as promoter and distal element accessibility, progressively increased at the SLAMF7 and XBP1 loci. Together, these data provide epigenetic evidence for progressive upregulation of genes associated with B cell and plasma cell differentiation and a differentiation hierarchy among the B cell subsets.

Resting naive B cells are epigenetically distinct in SLE. As expected, resting naive B cells from subjects with SLE and healthy controls expressed many hallmark B cell genes such as PAX5, IRF8, CD22, IL21R and CD19, and lacked expression of activation markers such as CD86 (Fig. 1d,e and Supplementary Fig. 3), suggesting that core features of B cell development are not grossly altered by SLE. In healthy immune systems, resting naive cells represent the earliest unperturbed mature B cells available to mount primary immune responses. However, we have previously reported that resting naive B cells in SLE have distinct differences in accessible chromatin in comparison to cells at the corresponding stage from healthy controls, indicating that the unique environment in SLE may alter the molecular machinery and properties of B cells conventionally defined as resting naive cells. Comparison of the epigenetic and transcriptional states of resting naive B cells from subjects with SLE and healthy controls revealed clear differences in each of the assays (Fig. 2a). Correlation of the DNA methylation changes with gene expression identified a group of genes that were upregulated and demethylated in SLE B cells, including IFI44, PDCD1 and SPRY2 (Fig. 2b). Additionally, a subset of genes were upregulated and contained DARs that had increased accessibility in resting naive B cells from subjects with SLE, including IFI44, STAT4, FOSL2, NR4A1, NR4A3 and SPRY2 (Fig. 2b–d and Supplementary Fig. 4). Expression of IFI44 is induced in SLE and in response to IFNγ. Expression of NR4A3 and NRA1 (Nur77) is induced in response to TLR stimulation and signaling through the BCRγ, respectively, suggesting that resting naive B cells in SLE may have received stimulation through these pathways. SPRY2 encodes a negative regulator of receptor tyrosine kinase (RTK) signaling. Thus, these data indicate that resting naive B cells in SLE are epigenetically primed and distinct from their counterparts in healthy controls.

Intriguingly, differences observed in resting naive B cells in SLE were also present in all the downstream B cell subsets (Fig. 2c,d), suggesting that the SLE epigenetic B cell program is established as early as the resting naive stage. Notably, however, transcriptional upregulation of other gene sets at significant levels was only initiated at the activated naive stage and levels of regulation did not differ from those observed in control samples. This dichotomy was illustrated by two conventional activation markers: CD69 already highly overexpressed by resting naive B cells from subjects with SLE, whereas expression of CD86 was upregulated in activated naive B cells and peaked in DN2 cells (Supplementary Fig. 3). The second pattern was also found for other immunologically relevant genes, including TBL2 (encoding T-BET), ITGAX (encoding CD11c), AICDA (Supplementary Fig. 5) and PDCD1 (encoding PD-1), with the latter two also suggesting BCR engagement. This pattern reflects
Epigenetic states of B cell subsets identify cell type relationships and differentiation hierarchies. a, PCA of DMLs (left), DARs (middle) and DEGs (right) identified in comparisons of cell types. Each cell type is represented by a point and the 99% confidence interval for each cell subset is denoted by a circle. Sample sizes for each cell type can be found in Supplementary Table 5. HC, healthy control; rN, resting naive; aN, activated naive; SM, switched memory. b, Mean CpG methylation for each cell type is plotted. Data represent mean ± s.d. Significance was determined by two-way analysis of variance (ANOVA) with Tukey’s post-hoc test. *P < 0.05. c, Phylogenetic dendrograms for healthy control (top) and SLE (bottom) subsets for the indicated data. (ANOVA) with Tukey’s post-hoc test. * vs. rN * vs. rN

DNA methylation stratifies cells from subjects with SLE and healthy controls. Analysis of DNA methylation profiles identified a distinct molecular SLE disease signature consisting of 6,664 DMLs that stratified all control and SLE samples (Fig. 3a). DMLs were classified into three distinct modules by k-means clustering. CpGs in module 1 lost DNA methylation as the cells progressed through differentiation but were hypermethylated in all SLE B cells subsets as compared to their counterparts in healthy controls (Fig. 3b). The mean CpG methylation for each cell type was quantified and revealed that SLE B cells were hypermethylated as compared to B cells from healthy controls. CpGs in modules 2 and 3 corresponded to sub-sets of DMLs that were hypo- and hypermethylated, respectively, in all SLE B cells. Because DNA methylation may provide a useful biomarker for disease, we identified the CpGs in modules 2 and 3 that best stratified SLE samples from healthy control samples regardless of individuals’ age, which is known to affect DNA methylation21. Using a minimum DNA methylation change of 40%, we identified 111 CpGs that stratified subjects by SLE status and included hypomethylated sites near the IFITM1, IFITM1, YBX1 and TAF8 genes and hypermethylated CpGs surrounding SOX12, ARFGAP3, CCDC81 and MEG3 (Fig. 3c, d and Supplementary Table 2). EPSTI1, which encodes a positive regulator of NF-κB signaling22, is induced by IFN and is associated with SLE susceptibility23, had a highly predictive cluster of five CpGs in its first intron (Fig. 3c, f). Each of the CpGs was hypomethylated and EPSTI1 expression was upregulated in all of the SLE B cell subsets (Fig. 3g). To confirm these findings, we performed a quantitative PCR (qPCR) assay24 on resting naive B cells from an independent cohort and found that CpGs at the EPSTI1, IFITM1 and MXI loci were consistently demethylated in a very high degree of overall transcriptional similarity between activated naive and DN2 cells. Together, these data suggest that there are unique and shared disease-specific features that may impact cellular differentiation in SLE.
subjects with SLE as compared to healthy controls (Fig. 3h). Thus, clear differences in the DNA methylation patterns of B cells from subjects with SLE and healthy controls exist and identify DMLs that could be used as potential SLE biomarkers.

DN2 B cells in SLE have a distinct chromatin landscape from DN2 and isotype-switched memory B cells in healthy controls. In the DNA methylation phylogenetic analysis, DN2 and isotype-switched memory B cells were in a similar position with respect to their differentiation state and hierarchy. However, the PCA and phylogenetic analyses of accessible chromatin and gene expression datasets indicated that distinct differences existed between DN2 and isotype-switched memory cells. A comparison of DARs identified thousands of loci that were distinct between DN2 and isotype-switched memory B cells in both SLE and healthy control samples (Fig. 4a). Annotation of DARs to genes revealed that approximately 50% of the accessible chromatin in DN2 cells was shared by healthy controls and subjects with SLE with samples from both groups displaying uniquely accessible gene promoters or regulatory regions (Supplementary Table 3). Irrespective of disease status, a highly significant correlation was observed between DARs and DEGs for isotype-switched memory and DN2 B cells (Fig. 4b), suggesting that events that generate DARs may influence gene expression. Moreover, direct comparison of DN2 B cells from subjects with SLE and healthy controls identified significant numbers of DARs and DEGs (Fig. 4c). Therefore, DN2 B cells in SLE are distinct from DN2 and isotype-switched memory B cells in healthy controls.

Functional annotation of genes surrounding accessibility patterns in DN2 B cells, in comparison to isotype-switched memory B cells (from healthy control and subjects with SLE) was determined by Gene Ontology (GO) analysis. This identified genes involved in signaling from the TCR and BCR, signaling from co-stimulatory...
molecules and the response to lipopolysaccharide (Fig. 4d). As described for ABCs\textsuperscript{3,4} and consistent with our initial functional and transcriptional results\textsuperscript{5}, these data suggest that TLR stimulation may be unique to the differentiation history of DN2 B cells, a contention that is supported by the significantly lower expression of TRAF5, a negative regulator of TLR activation, in B cells, including in DN2 B cells\textsuperscript{31}. Additional genes associated with ABCs that were unique to both DN2 and activated naive B cells from healthy controls and subjects with SLE (as compared to isotype-switched memory B cells) included TBX21, AICDA and GAS7 (Supplementary Fig. 5). We further characterized the similarity of DN2 B cells with previously published datasets from ABCs\textsuperscript{2,3}. DN2 B cells from both healthy controls and subjects with SLE displayed a significant enrichment of multiple ABC gene signatures\textsuperscript{3,5} (Supplementary Fig. 6). However, no significant enrichment of ABC signatures was observed when comparing DN2 B cells from subjects with SLE to healthy controls, indicating that the signature is common to DN2 B cells irrespective of disease status. DN2 B cells also had unique accessibility patterns indicating that the signature is common to DN2 B cells irrespective of disease status. DN2 B cells from subjects with SLE were specifically enriched individuals with chronic lymphocytic leukemia with poor clinical outcomes\textsuperscript{3}. DN2 B cells from healthy controls were uniquely enriched for the distinct signaling pathways, transcriptional repression and regulation of the cell cycle (Fig. 4d and Supplementary Fig. 3). DN2 B cells from subjects with SLE were specifically enriched for genes associated with tyrosine kinase signaling as compared to isotype-switched memory B cells (Fig. 4d). For example, FGFR1,
DN2 B cells in SLE have a chromatin conformation driven by TLR and RTK signaling pathways. a, Volcano plots of the DARs between DN2 and isotype-switched memory cell types for healthy controls (top) and subjects with SLE (bottom). DN2 DARs were annotated to the nearest gene, and the overlap of healthy control and SLE samples is shown as a Venn diagram. DARs represent features with at least twofold change and FDR < 0.05 as determined by edgeR. b, Correlation of DARs from a and DEGs for the same comparisons for healthy control (top) and SLE (bottom) subjects. Genes with a positive correlation between DARs and DEGs (top right quadrant) that overlap between healthy control and SLE samples are shown in a Venn diagram. Significance was determined by one-way ANOVA. c, Volcano plots of DARs (top) and DEGs (bottom) comparing DN2 B cells in SLE to DN2 B cells in healthy controls. The number of differential features is indicated. DARs represent features with at least twofold change and FDR < 0.05 as determined by edgeR. d, Bar plots of the top enriched GO biological processes for healthy-control-specific, shared and SLE-specific genes from the Venn diagram in a. Significance was determined by Fisher's exact test. e, Bar plots of gene expression levels for PDCD1 ± s.d. An asterisk indicates DEGs between SLE and healthy control samples (at least twofold change and FDR < 0.05) as determined by edgeR. f, Genome plot of the PDCD1 locus showing the chromatin accessibility levels for each cell type. DARs between DN2 and isotype-switched memory B cells are boxed. Data represent the mean for each cell type from one experiment. g, Flow cytometry analysis of the percentage of PD-1+ cells for the indicated cell type. Data are plotted as mean ± s.d. An asterisk indicates significance determined by two-tailed Student's t-test. See also Supplementary Fig. 5.
Chromatin accessibility in DN2 B cells is driven by T-BET, AP-1 and EGR transcription factors. a. Heat map depicting normalized enrichment of transcription factor motifs in DN2 and isotype-switched memory B cell types. Enrichment P values were normalized to the minimum value for each cell type. Motif grouping is indicated on the left. Significance was determined by binomial distribution using HOMER. b. Histograms (left) and box plots (right) of accessibility at the indicated motifs and surrounding sequences for each cell type. Data represent the mean for each cell type. The locations of motifs are indicated for the histograms. In box plots, the center line indicates the median, the lower and upper bounds of boxes indicate the first and third quartiles and whiskers indicate the upper and lower limits of the data. Significance was determined by two-tailed Student’s t test. See also Supplementary Fig. 7. c. Genome plots showing accessibility and T-BET binding in GM12878 B cells36 for the indicated loci. DARs between DN2 and isotype-switched memory B cells with T-BET-binding peaks are boxed. Data represent the mean for each cell type from one experiment.

and cis-regulatory elements were highly accessible in DN2 B cells as compared to isotype-switched memory B cells and PDCD1 was upregulated at both the mRNA and protein levels in SLE B cell subsets (Fig. 4e–g). Consistent with the phylogenetic analysis, highly similar patterns of gene expression and accessibility were observed for DN2 and activated naive B cells in SLE samples. For example, GAS7 was only expressed in activated naive and DN2 B cells. These data suggest a common epigenetic programming relationship between these cell subsets and a mechanistic link for their prominent expansion in active SLE.

DN2 B cell accessibility is programmed by T-BET, AP-1 and EGR transcription factors. To identify the transcription factor networks driving the accessibility patterns in DN2 and isotype-switched memory B cells, the DARs were searched for enriched transcription factor-binding motifs. Enriched motifs were grouped, revealing common and unique sets of motifs for accessible loci in DN2 and isotype-switched memory B cells. Shared motifs included those for lineage factors important for B cell identity, such as PU.1, RUNX1 and E2A as well as the ETS:IRF composite motif (Fig. 5a). The top enriched motifs in DN2 B cells were for T-BET, ISGF3, and the AP-1 and EGR transcription factor families. T-BET and AP-1 motifs have been found to be enriched in the accessible chromatin of ABCs that accumulate in SWEF-family-deficient mice that develop systemic autoimmunity6. Conversely, loci in isotype-switched memory B cells were uniquely enriched for EBF, NF-xB and OCT motifs.

In comparison to all other cell subsets, DN2 B cells had the highest levels of accessibility at T-BET, EGR- and AP-1 motifs (Fig. 5b and Supplementary Fig. 7). In line with the findings above, activated naive B cells also had high accessibility at these motifs with activated naive B cells from subjects with SLE displaying higher accessibility than activated naive B cells from healthy controls. For each of the three motifs, minimal accessibility was observed in T3, resting naive and isotype-switched memory B cells. In contrast, NF-xB motifs displayed the highest levels of accessibility in isotype-switched memory B cells and the lowest accessibility in activated naive and DN2 B cells. T-BET binding profiles from human B cells8 displayed enrichment at loci accessible in activated naive and DN2 B cells at the GAS7, TNFRSF1B, ITGAX and ZAP70 loci (Fig. 5c and Supplementary Fig. 5). Interestingly the TBX21 locus, which encodes T-BET, also contained an upstream distal element that was bound by T-BET and showed enhanced accessibility in activated naive and DN2 B cells. Thus, isotype-switched memory and DN2 B cells express distinct transcription factors that correlate with unique accessibility patterns, transcription factor networks and gene expression programs.

Coordination of the SLE transcriptome and transcription factor signatures. For genes such as STAT4 (which harbors a major SLE
risk allele\(^1\) and \(JIF44\), the expression and accessibility differences observed in resting naive B cells from subjects with SLE versus the corresponding cells in healthy controls were also shared among the other cell subsets, indicating the presence of a common SLE disease signature. Indeed, 5,090 DEGs were identified across all cell types (Fig. 6a). GSEA identified pathways that were upregulated in all SLE B cells including gene sets related to the inflammatory, IFN-\(\gamma\) and IFN-\(\alpha\) responses that have been previously identified in SLE samples\(^2\) (Fig. 6b and Supplementary Fig. 8). Additionally, we identified the WNT, Notch and estrogen signaling pathways, cytokine signaling from IL-6 and IL-2 receptors, and the p53 and apoptosis pathways as enriched in SLE B cells. By contrast, DN2 B cells from subjects with SLE showed negative enrichment for several pathways, including those driving the UPR and G2/M checkpoint.

Analysis of the SLE disease signature manifested in the accessible chromatin landscape identified thousands of SLE-specific DARs (Fig. 6c). The PageRank algorithm\(^3\) was used to identify transcription factor-binding motifs that correlated with changes in gene expression at target genes and to determine the overall importance of each transcription factor to the SLE transcriptional

---

**Fig. 6 | SLE transcription factor networks correlate with disease-specific transcriptomes.** a. Heat map of 5,090 DEG that stratify SLE and healthy control cell types. DEGs represent features with at least twofold change and FDR < 0.05 as determined by edgeR. b, GSEA plots for the indicated gene sets showing the enrichment score for each SLE cell type as compared to healthy control samples. See also Supplementary Fig. 8. c, Volcano plot of DARs between all healthy control and SLE cell types. DARs represent features with at least twofold change and FDR < 0.05 as determined by edgeR. d, Heat map of the log2-transformed fold change in PageRank statistic for each cell type between SLE and healthy control samples. See also Supplementary Fig. 8. e, Bar plot of gene expression levels for selected factors from b that are regulated by EGR factors. Line thickness is scaled to significance as determined by Fisher’s exact test. See also Supplementary Fig. 8. f, RT-qPCR analysis of select genes in resting naive B cells from an independent cohort of healthy controls (\(n = 15\)) and subjects with SLE (\(n = 8\)). Data represent mean ± s.d. and are normalized to percentage of ACTB expression. Significance was determined by two-tailed Student’s t-test. g, Network diagram depicting gene sets from b that are regulated by EGR factors. Line thickness is scaled to significance as determined by Fisher’s exact test. See also Supplementary Fig. 8. EMT, epithelial–mesenchymal transition.
ATF3 response pathways are activated in SLE DN2 B cells. To specifically identify transcription factors and networks associated with the expanded DN2 B cells in SLE, PageRank enrichment in SLE DN2 B cells versus the change in gene expression was analyzed and revealed a positive correlation between the datasets (Fig. 7a). As described above, all four EGR family members were upregulated and highly enriched in the SLE DN2 B cell network. Additional significant regulators constituted the ATF/CREB family of factors, including CREM, CHOP (encoded by DDIT3) and ATF3. ATF3 is induced in response to BCR and TLR stimulation and cellular stress and can mediate the responses of p53 and various hormone receptors. ATF3 was upregulated in all SLE B cell subsets, with maximal expression in SLE DN2 B cells (Fig. 7b). To validate these findings, we quantified ATF3 mRNA levels in resting naive B cells by RT–qPCR and found that ATF3 was significantly upregulated in resting naive B cells from SLE samples as compared to healthy control samples (Fig. 7c). Consistent with mRNA levels, ATF3 protein was more abundant in resting naive, activated naive, isotype-switched memory and DN2 B cell subsets from subjects with SLE as compared to healthy controls (Fig. 7d). Using the predicted target genes from the PageRank analysis, we identified 98 ATF3 target genes that were disease DEGs, of which 87% were upregulated in SLE (Fig. 7e and Supplementary Table 4). DNA footprinting analysis revealed a large difference at ATF3 motifs and the surrounding chromatin in SLE B cells as compared to B cells from healthy controls (Fig. 7g). Quantification of accessibility of the ATF3 motif in each cell subset identified SLE DN2 B cells as having the highest levels of accessibility (Fig. 7h). Additionally, ATF3 peaks identified by ChIP–seq significantly overlapped SLE DARS (Supplementary Fig. 7). ATF3 heterodimerizes with Jun family members, of which three (JUN, JUNB and JUND) were enriched in the DN2 networks and upregulated in SLE (Fig. 7a and Supplementary Fig. 8). Additionally, for these ATF3 motifs we did not find any co-occurrence with IRF4 in the context of AICE composite motif, and only 7 of the 98 genes (7%) were predicted to be regulated by IRF4 by PageRank. ATF3 target genes were mapped to functional gene sets enriched in SLE samples and found to have roles in cell cycle and metabolism through the MTORC1 pathway, G2/M checkpoint, TNF signaling, hypoxia, UPR and apoptosis (Fig. 7i). Thus, ATF3 and EGR together with their respective family members regulate DN2-specific B cell transcriptional pathways that are important for SLE epigenetic programming.

Discussion
Here we used an integrated transcriptional and epigenetic approach to unravel the nature of B cell populations, their differentiation hierarchy and the molecular programs that underpin their abnormal behavior in SLE. We focused on finely discriminated subsets, representing the human extrafollicular naive B cell differentiation pathway responsible for the generation of DN2 cells. Our analysis was performed on an African-American cohort with high disease activity. Going forward it will be important to determine how these epigenetic programs evolve during the progression of SLE, in quiescent versus active disease, and whether this epigenetic imprint exists in other ancestry groups.

The data suggest a differentiation order with T3 and activated naive B cells being derived from activation of resting naive cells with DN2 and isotype-switched memory B cells placed further down the differentiation path to ASCs. T3 cells have been previously defined as transitional cells in human studies of repopulating B cells after rituximab treatment. Our study suggests that in active SLE T3 cells represent an early phase of naive B cell activation. In SLE, consistent with their phenotypic similarity, activated naive and DN2 B cells shared features of accessible chromatin, were transcriptionally related to each other, yet distinct from the other three B cell subsets; and were closer to ASCs than to any other subset, including isotype-switched memory B cells. In combination, these data support the notion that DN2 B cells represent a distinct effector subset that is expanded in SLE. Our results are also consistent with separate differentiation pathways in which ASCs are derived from either activated naive–DN2 cells or isotype-switched memory cells.

Consistent with our initial findings, resting naive B cells in subjects with SLE were distinctly primed at an early differentiation stage. That some of the SLE transcriptional and epigenetic signatures were shared by all SLE B cell subsets demonstrates that SLE B cells are under the control of a unique transmissible epigenome that regulates the effect of the SLE environment on all mature B cells and modulates their activation threshold. Elucidating the earliest stage of development at which B cell precursors may be epigenetically modified in SLE could guide optimization of B cell therapies.

Analysis of changes in epigenetic modifications identified programs responsible for the abnormal behavior of individual B cell subsets in SLE and, in particular, programs associated with extrafollicular effector cells. The accessible chromatin of DN2 B cells was specifically enriched for T-BET, EGR and AP-1 transcription factor motifs. Our data provide insight into a T-BET-driven program that is shared by healthy control and SLE DN2 B cells. T-BET was upregulated in DN2 B cells irrespective of disease status, was enriched in DN2-specific accessible chromatin, and regulated other markers characteristic of ABCs and DN2 cells, such as CD11c. Moreover, T-BET bound to its own gene, suggesting that it may function in an autoregulatory manner in these cells. This indicates that DN2 B cells are a normal product of B cell differentiation, but in abnormal immune environments, such as in SLE, their formation is enhanced...
We previously postulated a developmental link between activated naive and DN2 B cells and ASCs in SLE. This model is supported by the fact that activated naive and DN2 B cells in SLE share distinct epigenetic and transcriptional profiles and have close relationships in phylogenetic trees. This close relationship was mirrored for many genes, such as ITGAX, GAS7 and PDCD1. Additionally, the enhanced accessibility of core SLE AP-1 and EGR transcription factor motifs was present in activated naive cells (relative to resting naive and T3 cells) and increased further in DN2 cells, but was largely absent in isotype-switched memory B cells. Thus, these data provide a mechanistic link between dysregulation of central SLE networks and transcription factors and the skewing of B cell differentiation towards the activated naive–DN2 cell fate.

An informative example of the SLE programs shared across B cell populations that are already engaged in resting naive cells is illustrated by the NR4A family of nuclear receptors. NR4A1 is induced specifically by BCR stimulation, while NR4A2 and NR4A3 are induced by other stimuli. The involved genes are typically expressed in a cell type-specific manner, suggesting that these receptors likely play a role in the regulation of B cell activation and differentiation.
NR4A3 are induced in response to TLR stimulation\(^5\). This is consistent with mouse lupus models induced by BCR and TLR co-stimulation leading to tolerance breakdown\(^6\) and our previous data showing that activated naive and DN2 B cells are hyper-responsive to TLR7 stimulation and contain clonally expanded autoreactive B cells\(^3\). Of note, DN2 B cells demonstrated unique chromatin accessibility patterns surrounding BCR and TLR signaling components, suggesting that initial engagement of both these pathways in naive B cells may be propagated to later stages of differentiation. These data fit a model of autoimmunity in which BCR and TLR signals can converge in extracellular settings to stimulate and expand autoreactive B cells\(^4,44,45\). Our data indicate that in SLE, resting naive B cells are subjected to positive antigenic stimulation and therefore provide an important window to understand early antigenic triggers in SLE.

The expression of ATF/CREB, AP-1 and EGR family transcription factors was maximally enriched in the DN2 subset in SLE. EGR proteins have been linked to autoimmunity, are induced by BCR stimulation\(^6\) and regulate plasma cell differentiation\(^14,26\). EGR1 was highly induced in SLE in multiple cohorts and demonstrated connectivity to many dysregulated transcriptional networks. Similarly, ATF3 was highly enriched in the DN2 B cell network. Each of these factors was predicted to impact metabolism, cell cycle, apoptosis and the response to cellular stress. The upregulation of oxidative phosphorylation enhances plasma cell differentiation\(^14\). Interestingly, the G2/M checkpoint and apoptosis pathways were induced in all SLE B cell subsets except DN2 B cells, which might explain their expansion in SLE. These data suggest that the balance between pathways that promote or oppose differentiation may be altered in SLE DN2 B cells. Mechanistically, this may occur by switching transcription factor partners. ATF3 for example, functions as a repressor when bound as a homodimer but can heterodimerize with Jun family members to function as an activator\(^16\). The coordinated upregulation of Jun family members with ATF3 suggests that the equilibrium of ATF3 activity shifts toward activation in SLE, particularly in the DN2 B cell subset. Moreover, the main T-BET inducer, IFN-γ, can synergize with TLR stimulation to enhance the expression of ATF3 (ref. \(^7\)). Thus, the coordination of stimuli with known pathogenic relevance may offer a mechanistic link between molecular and cellular abnormalities in SLE.

SLE B cell subsets had higher overall levels of DNA methylation than their counterparts in healthy controls. This may reflect more recent differentiation from a highly methylated progenitor B cell\(^1\). Alternatively, the SLE program may depend on the maintenance of methylation at the loci observed. Nonetheless, while the DNA methylation patterns did not specifically identify disease pathways, the presence or absence of DNA methylation at different loci could potentially serve as biomarkers of disease and disease progression as illustrated by 111 CpGs that discriminated between SLE and healthy control B cells. As these differential DNA methylation modifications occurred in all B cell subsets, including in resting naive B cells, it is not known when such modifications are first imprinted into the SLE epigenome and these modifications may represent an epigenetic record of differentiation history and disease state.

In summary, we have defined the transcriptomes and epigenomes of B cell subsets that are expanded in SLE. Our data indicate a disease signature across all cell subsets, and importantly, in mature resting B cells, suggesting that such cells may have been exposed to disease-inducing signals. Together, these data define the imprint of SLE on the epigenome of specific B cell populations and highlight transcription factor networks and programming of normal and pathogenic B cell subsets.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at [https://doi.org/10.1038/s41590-019-0419-9](https://doi.org/10.1038/s41590-019-0419-9).

Received: 28 September 2018; Accepted: 9 May 2019; Published online: 1 July 2019

References
1. Rawlings, D. J., Metzler, G., Wray-Dutra, M. & Jackson, S. W. Altered B cell signaling in autoimmunity. *Nat. Rev. Immunol.* 17, 421–436 (2017).
2. Langefeld, C. D. et al. Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* 8, 16021 (2017).
3. Chung, S. A. et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PloS Genet.* 7, e1001323 (2011).
4. Williams, J., Euler, C., Christensen, S. & Shlomchik, M. J. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. *Science* 297, 2066–2070 (2002).
5. Tipton, C. M. et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell populations expansions in acute systemic lupus erythematosus. *Nat. Immunol.* 16, 755–765 (2015).
6. Jenks, S. A. et al. Distinct effector B cells induced by unregulated Toll-like receptor 7 contribute to pathogenic responses in systemic lupus erythematosus. *Immunity* 49, 725–739 (2018).
7. Hao, Y., O'Neill, P., Naradzikian, M. S., Scholz, I. L. & Cancro, M. P. A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood* 118, 1294–1304 (2011).
8. Rubtsov, A. V. et al. Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c+ B-cell population is important for the development of autoimmunity. *Blood* 118, 1305–1315 (2011).
9. Manni, M. et al. Regulation of age-associated B cells by IFRS in systemic autoimmunity. *Nat. Immunol.* 19, 407–419 (2018).
10. Poovarsayi, J. S. & Bishop, G. A. Type I IFN receptor and the B cell antigen receptor regulate TLR7 responses via distinct molecular mechanisms. *J. Immunol.* 189, 1757–1764 (2012).
11. Ulf-Moller, C. J. et al. Twin DNA methylation profiling reveals flare-dependent interferon signature and B cell promoter hypermethylation in systemic lupus erythematosus. *Arthritis Rheumatol.* 70, 878–890 (2018).
12. Banchereau, R. et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell* 165, 551–565 (2016).
13. Scharer, C. D. et al. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naive SLE B cells. *Sci. Rep.* 6, 27030 (2016).
14. Javierre, B. M. et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 20, 170–179 (2010).
15. Hewagama, A. & Richardson, B. The genetics and epigenetics of autoimmune diseases. *J. Autoimmun.* 33, 3–11 (2009).
16. Sanz, I., Wei, C., Lee, F. E. & Anolik, J. Phenotypic and functional heterogeneity of human memory B cells. *Semin. Immunol.* 20, 67–82 (2008).
17. Wirths, S. & Lanzavecchia, A. ABCB1 transporter discriminates human resting naive B cells from cycling transitional and memory B cells. *Eur. J. Immunol.* 35, 3433–3441 (2005).
18. Barwick, B. G., Schärer, C. D., Bally, A. P. R. & Boss, J. M. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. *Nat. Immunol.* 17, 1216–1225 (2016).
19. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218 (2013).
20. Kulis, M. et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* 47, 746–756 (2015).
21. Conges, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* 48, 1193–1203 (2016).
22. Schärer, C. D., Barwick, B. G., Guo, M., Bally, A. P. R. & Boss, J. M. Plasma cell differentiation is controlled by multiple cell division-coupled epigenetic programs. *Nat. Commun.* 9, 1698 (2018).
23. Schärer, C. D., Barwick, B. G., Guo, M., Bally, A. P. R. & Boss, J. M. Plasma cell differentiation is controlled by multiple cell division-coupled epigenetic programs. *Nat. Commun.* 9, 1698 (2018).
24. Schärer, C. D., Barwick, B. G., Guo, M., Bally, A. P. R. & Boss, J. M. Plasma cell differentiation is controlled by multiple cell division-coupled epigenetic programs. *Nat. Commun.* 9, 1698 (2018).
25. Nagaoka, M. et al. The human nuclear receptor NR4A3 is involved in the function of dendritic cells. *J. Immunol.* 199, 2958–2967 (2017).
26. Ashouri, J. F. & Weiss, A. Endogenous Nur77 is a specific indicator of antigen receptor signaling in human T and B cells. *J. Immunol.* 198, 657–668 (2017).
27. Gross, I., Bassit, B., Benezra, M. & Licht, J. D. Mammalian Sprouty proteins inhibit cell growth and differentiation by preventing Ras activation. J. Biol. Chem. 276, 46460–46468 (2001).
28. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. Cell 153, 1194–1217 (2013).
29. Li, T. et al. Identification of epithelial stromal interaction 1 as a novel effector downstream of Kruppel-like factor 8 in breast cancer invasion and metastasis. Oncogene 33, 4746–4755 (2014).
30. Mo, J. S. & Chae, S.-C. EPSTI1 polymorphisms are associated with systemic lupus erythematosus. Genes Genomes Genet. 39, 445–451 (2017).
31. Buchta, C. M. & Bishop, G. A. TRAF5 negatively regulates TLR signaling in B lymphocytes. J. Immunol. 192, 145–150 (2014).
32. Russell Knode, L. M. et al. Age-associated B cells express a diverse repertoire of V\\_\\_\\_\\_\\_\\_\\_V\\_\\_\\_\\_\\_\\_\\_ and V\\_\\_\\_\\_\\_\\_\\_ genes with somatic hypermutation. J. Immunol. 198, 1921–1927 (2017).
33. Wiestner, A. et al. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood 101, 4944–4951 (2003).
34. Veleeparambil, M. et al. Constitutively bound EGFR-mediated tyrosine phosphorylation of TLR9 is required for its ability to signal. J. Immunol. 200, 2809–2818 (2018).
35. Chattopadhyay, S. et al. EGFR kinase activity is required for TLR4 signaling and the septic shock response. EMBO Rep. 16, 1535–1547 (2015).
36. ENCODE Project Consortium An integrated encyclopedia of DNA elements with functional annotations. Nature 497, 171–202 (2013).
37. Yu, B. et al. Identification of epithelial stromal interaction 1 as a novel effector downstream of Kruppel-like factor 8 in breast cancer invasion and metastasis. Oncogene 33, 4746–4755 (2014).
38. Myouzen, K. et al. Regulatory polymorphisms in EGR2 are associated with CD8+ T cell differentiation. Nat. Immunol. 18, 573–582 (2017).
39. Oh, Y. K., Jang, E., Paik, D. J. & Youn, J. Early growth response-1 plays a non-redundant role in the differentiation of B cells into plasma cells. Immune Netw. 15, 161–166 (2015).
40. Gururajan, M. et al. Early growth response genes regulate B cell development, proliferation, and immune response. J. Immunol. 181, 4590–4602 (2008).
41. Price, M. J., Patterson, D. G., Scharer, C. D. & Boss, J. M. Progressive upregulation of oxidative metabolism facilitates plasminblast differentiation to a T-independent antigen. Cell Rep. 23, 3152–3159 (2018).
42. Ho, H. H., Antoniv, T. T., Ji, J. D. & Ivashkiv, L. B. Lipopolysaccharide-induced expression of matrix metalloproteinases in human monocytes is suppressed by IFN-γ via superinduction of ATF-3 and suppression of AP-1. J. Immunol. 181, 5089–5097 (2008).

**Acknowledgements**

We thank the members of the Boss and Sanz laboratories for critical reading of the manuscript, the New York University Genome Technology Center for Illumina sequencing, the Yerkes Genomics Core for RNA-seq library preparation, the Emory Pediatrics Flow Cytometry core for flow cytometry isolation of cell subsets and the Emory Integrated Genetics and Computational Core for Bioanalyzer and sequencing library quality control. This work was supported by NIH grants U19 AI110483 to J.M.B. and I.S., P01 AI125180 to I.S., F.E.-H.L. and J.M.B., RO1 AI113021 to J.M.B., F31 AI112261 to B.G.B., and T32 GM008490 to J.M.B.

**Author contributions**

C.D.S. and E.L.B. designed and performed experiments, analyzed the data and wrote the manuscript; B.G.B. and T.M. analyzed data; D.G.P. performed ATAC-seq; S.A.J. performed PD-1 and ATF3 phenotyping; T.D., K.S.C. and S.L.H. sorted and prepared cDNA for validation cohorts; B.E.N., F.E.-H.L. and C.W. provided cell sorting and biobanking expertise and performed sample preparation; A.K. evaluated cohort clinical data; and I.S. and J.M.B. designed experiments, wrote the manuscript and oversaw the project.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-019-0419-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to I.S. or J.M.B.

**Peer review information.** Laurie Dempsey was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Human subjects. Four independent cohorts were used for this study and are detailed in Supplementary Table 1. Whole-blood samples were obtained with informed consent from all participants in accordance with Emory University School of Medicine Institutional Review Board protocols.

Flow cytometry isolation of human B cell subsets. Whole blood (50–100 ml) was isolated from each subject via venous puncture utilizing cell preparation tubes containing sodium heparin and Ficoll-Hypaque solution (Becton Dickinson). Washed PBMCs were resuspended in 1 ml of prewarmed complete medium (RPMI 1640, 10% FBS and 1% l-glutamine) containing 20 mM of MTG and incubated at 37 °C in 5% CO2, for 30 min. Next, PBMCs were centrifuged at 1,300 r.p.m. for 10 min at room temperature, resuspended in 10 ml of prewarmed complete medium and incubated at 37 °C in 5% CO2 for 30 min to allow for excess MTG to be pumped out of B cells possessing the ABCB1 transporter. PBMCs were centrifuged at 1,300 r.p.m. for 10 min at 4 °C and resuspended at 10^6 RPMCs per 100 μl in PBS with 0.5% BSA, 0.5% normal mouse serum and 5% rat serum containing the following fluorochrome-conjugated human-specific monoclonal antibodies: anti-CD3 and anti-CD24 (Invitrogen); and anti-CD19, anti-IgD, anti-CD27 and anti-CD38 (BD Biosciences). Cells were incubated in the dark for 30 min at 4 °C, washed with 5 ml of PBS and centrifuged at 1,300 r.p.m. for 10 min at 4 °C. PBMCs were resuspended in 1 ml of PBS and subjected to live/dead Aqua staining (Invitrogen) according to the manufacturer's protocol. Following live/dead incubation, PBMCs were washed in 5 ml of PBS containing 0.5% BSA and centrifuged at 1,300 r.p.m. for 10 min at 4 °C. Pelleted PBMCs were resuspended in 1 ml of complete medium and filtered using a prewetted 35-μm cell strainer FACS tube (Corning). B cell subsets were purified using a FACS Aria II. To ensure sort purity, the FACS Aria II was calibrated to 99.9% purity using fluorescent beads before each sort.

Quantification of PD-1 and ATF3 expression. PBMCs were isolated as described above and stained in the presence of 10% normal mouse serum with the following fluorochrome-conjugated antibodies: anti-CD19, anti-CD27, anti-CD38 and anti-IgD (BD Biosciences) and anti-CD11c, anti-CD19, anti-CXCR5 and anti-CD279 (PD-1) (BioLegend). For intracellular staining, cells were fixed, permeabilized and stained using the Tru-nuclear Transcription Buffer set (BioLegend). Cells were blocked with 50 μg/ml rabbit IgG and stained with 2 μg/ml rabbit anti-human ATP3 (USBio) or rabbit IgG isotype control (Invitrogen). Single-color-stained anti-mouse Ig beads (Bangs) were used as compensation controls. Flow cytometry data were collected with BD FACSDiva v.6.2 and data analysis was performed in FlowJo (TreeStar) v.10.4.

RNA sequencing. RNA from sorted cell populations was isolated using the AllPrep DNA/RNA Mini kit (Qiagen) according to the manufacturer's instructions. Total RNA (50 pg) was used as input for the SMART-seq v3 cDNA synthesis kit (Takara) using ten cycles of PCR amplification. cDNA (1 ng) was used as input for the NexteraXT kit (Illumina) using nine cycles of PCR amplification. cDNA (50 pg) was used as input for the Methylated DNA/RNA Mini kit (Qiagen) according to the manufacturer's instructions. Total RNA sequencing data analysis. Raw fastq files were mapped to the hg19 version of the human genome using Bowtie v.1.1.1 (ref. 17) with the default parameters and duplicate reads were removed with PICARD v.1.127 (http://broadinstitute.github.io/picard/). Accessible peaks were called using MACS2 v.2.1 (ref. 18) and the SINE read count for each peak was normalized using the GenomicRanges v.1.22.4 package (ref. 19). The resulting data were reads per peak per million (RPPM) normalized for the fraction of reads in peaks and sequencing depth as previously described 20. DARs were identified using edgeR v.3.18.1 (ref. 21) and a generalized linear model. To determine disease differences covariates included patient and cell type; for cell type differences covariates included disease status and patient. Peaks with a greater than twofold change and FDR < 0.05 in comparisons were termed significant. To identify motifs enriched in DARs, the findMotifsGenome.pl function of HOMER v.4.8.2 (ref. 22) and the ‘de novo’ output were used for downstream analysis. For the motif heat map analysis, each motif with a maximum P value in one cell type. To generate motif footprints the motifs occurring in peaks were annotated with the HOMER v.4.8.2 (ref. 22) annotatePeaks.pl function using the option ‘-size given’. Next, the read depth at the motif and surrounding sequence was computed using the GenomicRanges v.1.22.4 package and custom scripts in R and Bioconductor and the coverage at each position was plotted as a histogram or summarized as a box plot. GO biological process enrichment analysis of DARs was performed using DAVID 23. The resulting list of GO terms was filtered using REVIGO 24 to generate a list of terms containing unique sets of genes.

Reduced-representation bisulfite sequencing. DNA was isolated from the same fraction as RNA using the AllPrep DNA/RNA Mini kit (Qiagen) according to the manufacturer’s instructions. A dual-restriction enzyme RRBS assay was performed by splitting the DNA into two aliquots and digesting over night with either MspI or TaqI (NEB). The resulting DNA was purified and digestion efficiency was quantified by qPCR using primers spanning sequences that contained one, both or no restriction site. For bisulphite conversion controls, ΦX174 phage DNA (NEB) was in vitro methylated with M.SssI methylase (NEB). Methylated φX174 phage DNA and unmethylated lambda phage DNA (NEB) were sonicated to generate random 200- to 400-bp fragments. Each DNA (10 pg, 1:10,000 ratio of phage to genomic DNA) was spiked into the purified digested genomic DNA to use as a bisulphite conversion control. Samples were ligated using the HyperPrep kit (KAPA BioSystems) and fully methylated short adaptors 25. Adaptor-ligated DNA was bisulfite converted using the EpiTect Bisulfite kit (Qiagen) according to the manufacturer’s protocol except that the denaturation temperature was increased to 99 °C for 10 min. Final libraries were PCR amplified (using the SeqxF (X1 and X2 left and right flank indexed primers and Hifi indexed and polyCN) primers) and the resulting RRBS library was quality checked by Bioanalyzer, quantified by qPCR, pooled at equimolar ratio and sequenced on a HiSeq2500 using 50-bp paired-end chemistry.

RRBS data analysis. Raw fastq files were mapped to a composite genome index containing hg19, ΦX174 and lambda phage genomes using Bismark v.0.16.3 (ref. 26). CGp calls were determined using the Bismark tools v.1.22.0 and data table v.11.0.4.3 packages and custom R scripts as previously described 27. All CGps with at least 10X coverage per group (9,803,151) were determined and differential methylation analysis was performed using the DSS v.2.10.0 package 28 and a generalized linear model. To determine disease differences, covariates included patient and cell type; for cell type differences, covariates included disease status and patient. CGps with FDR < 0.05 and a difference in methylation greater than 20% in comparisons were called significant. 4-Lmeans clustering of the disease DMLs was performed using the Biganalytics package (https://CRAN.R-project.org/package=biganalytics). To determine predictive CGps from modules 2 and 3, we built a linear regression model at each CGp to find associations between SLE status and DNA methylation using a two-sided t test with age as a covariate. The resulting P values were FDR corrected using the Benjamini–Hochberg method and CGps with a minimum change of 40% between healthy controls and subjects with SLE and FDR < 5 × 10^-5 were considered highly predictive.

Meta-analysis. For PCA and phylogeographic analysis, all significant differences (DMLs, DEGs and DARs) for comparisons between all cell types were selected pairwise and the data were z score normalized. PCA was performed using the vegan v.2.4.3 package in R and Bioconductor. Phylogeographic analysis was performed by computing a Euclidean distance matrix between samples, which
were clustered using the ‘hclust’ function in R and Bioconductor. Phylogenetic trees were plotted with APE v3.4 (ref. 3) as ‘unrooted’ trees. The reproducibility of each phylogenetic tree was assessed by bootstrapping using the ‘boot.phylol’ function with 10,000 permutations. To correlate epigenetic and transcriptome datasets, DMLs and DARs were annotated to the nearest transcription start site using the annotatePeaks.pl function in HOMER v.4.8.2 (ref. 5). The hg38 bigWig file was converted to hg19 using the ‘lift’ function of bwtool v.1.0 (ref. 65). Genes displaying FDR <0.05 and fold change >4 for the indicated comparisons were used as gene sets. PageRank analysis of ATAC-seq and RNA-seq data was performed using the raw ATAC-seq data and normalized differential expression files comparing each SLE and healthy control cell type. The significance of transcription factor target genes within GSEA gene sets was calculated by Fisher’s exact test and network diagrams were plotted using Cytoscape v.3.4.0, scaling the network connections to $-\log_{10} P$.

RT-qPCR DNA methylation analysis. DNA from sorted resting naïve B cells was split into three equal aliquots and digested with the methyl-sensitive HpaII or the methyl-insensitive isoschizomer MspI or mock digested. Equal portions of each reaction were subject to qPCR using primers covering target CpGs (primers are detailed in Supplementary Table 6). DNA methylation was determined as the ratio of HpaII to mock digested signal and quantitated using a standard curve.

RT-qPCR analysis. RNA from 1,000 resting naïve B cells was isolated using the Quick-RNA MicroPrep kit (Zymo Research) and the reaction was used as input for the SMART-seq v4 cDNA synthesis kit (Takara) using ten cycles of PCR amplification. cDNA was diluted 1:25 and qPCR was performed to quantify the target gene expression using SybrGreen incorporation on a CFX96 instrument (BioRad) with CFX Manager v.3.1 software. Data were normalized to the percentage of ACTB expression. All primers are detailed in Supplementary Table 6.

Statistics and reproducibility. All statistical analyses were performed with R or Excel and $P<0.05$ was considered significant. Exact statistical tests are noted in the figure legends and the Methods for specific experiments and analyses. The statistical analysis of RNA-seq and ATAC-seq data was performed using edgeR (ref. 5) and DSS for RRBS (ref. 5). The sample sizes for each healthy control and SLE cell type profiled by RRBS, ATAC-seq and RNA-seq can be found in Supplementary Fig 2 and Supplementary Table 5. The genomics data represent one experiment from a single cohort. The expression of select genes and DNA methylation status at select CpGs were validated in independent cohorts as noted.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the NCBI Gene Expression Omnibus (GEO) under accession GSE118256 and are detailed in Supplementary Table 5.

Code availability
Code and data processing scripts are available from the corresponding author upon request and at https://github.com/cdschar.

References
51. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
52. Hsu, F. et al. The UCSC known genes. Bioinformatics 22, 1036–1046 (2006).
53. Lawrence, M. et al. Software for computing and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118 (2013).
54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
55. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
56. Zhang, H., et al. Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137 (2008).
57. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
58. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831 (2012).
59. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
60. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6, e21800 (2011).
61. Barwick, B. G. et al. B cell activation and plasma cell differentiation are inhibited by de novo DNA methylation. Nat. Commun. 9, 1900 (2018).
62. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572 (2011).
63. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res. 42, e69 (2014).
64. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290 (2004).
65. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. Bioinformatics 30, 1618–1619 (2014).
66. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. Bioinformatics 26, 2363–2367 (2010).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

☐ n/a
☐ Confirmed
☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ The statistical test(s) used AND whether they are one- or two-sided
   Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ A description of all covariates tested
☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
   Give P values as exact values whenever suitable.
☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Flow cytometry data was collected with BD FACSDiva v6.2 and qPCR data collected with BioRad CFX Manager v3.1.

Data analysis
Flow cytometry data analysis: FlowJo (TreeStar) v10.4
RNA-seq data analysis: TopHat2 v2.0.13, PICARD v1.127, GenomicRanges v1.22.4, edgeR v3.18.1, GSEA v3.0
ATAC-seq data analysis: Bowtie v1.1.1, PICARD v1.127, GenomicRanges v1.22.4, edgeR v3.18.1, HOMER v4.8.2, MACS2 2.1
RRBS data analysis: Bismark v0.16.3, Rsamtools v1.22.0, data.table v1.10.4.3, DSS v2.10.0
Other tools: R v3.4.3, vegan v2.4-3, APE v3.4, bwtool v1.0, oligo v1.40.2, Cytoscape v3.4.0, DAVID, REVIGO, Biganalytics
Custom data plotting scripts can be found on GitHub at https://github.com/cdschar

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the NCBI Gene Expression Omnibus (GEO) under accession GSE118256 and are detailed in Supplementary Table 5. Code and data processing scripts are available from the corresponding author upon request and at https://github.com/cdschar.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
A power analysis was performed and identified a sample size of 8 or more to ensure 80% power to detect a 2-fold change with a false positive rate of 0.05%.

Data exclusions
No data was excluded from the analysis.

Replication
Four replication cohorts were used to successfully validate a subset of the DNA methylation and gene expression results. One cohort was used to validate the DNA methylation findings at 3 CpG surrounding EPST1, IFI44, and MX1. A second cohort was used for qRT-PCR analysis to validate EGR family and ATF3 transcription factors. A third and fourth cohort were used to validate the expression of ATF3 and PD1 protein expression.

Randomization
To control for covariates such as age and gender we recruited only female subjects with a median age of 32 [healthy] and 32 [SLC] for the genomics portion of the study. The other validation cohorts included expanded ethnicity groups. All patient ethnicity, age, and gender information is included in Supplementary Table 1.

Blinding
There was no risk of bias in this study from knowing the sample details so blinding was not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a | Involved in the study |
| ☐ | X Antibodies |
| ☐ | X Eukaryotic cell lines |
| ☐ | X Palaeontology |
| ☐ | X Animals and other organisms |
| ☐ | X Human research participants |
| ☑ | X Clinical data |
| ☑ | X ChiP-seq |
| ☐ | ☐ Flow cytometry |
| ☑ | ☑ MRI-based neuroimaging |

Antibodies

Antibodies used

Antibody Fluorophore Clone Company Cat#
CD19 APC-Cy7 SJ25C1 BD Bioscience 557791
IgG FITC IA6-2 BD Bioscience 562023
CD27 BV395 L128 BD Bioscience 563815
CD11c BV605 3.9 Biolegend 301635
CD279(PD-1) PE EH12.2H7 Biolegend 329905
CD38 BV421 H122 BD Bioscience 562445
CXCR5 AK647 J252D4 Biolegend 356905
CD3 BV711 OKT3 Biolegend 317327
MitoTracker Green™ Cell Signaling Technology 9074
CD24 PE-A610 SN3 ThermoFisher Scientific MHCD2422
Live/Dead Aqua™ ThermoFisher Scientific L34957
ATF3 AF488™ USBiological Life Sciences 032160 FITC
Rabbit IgG FITC ThermoFisher Scientific 11-4614-80

Validation

Validation is described on the company website for each product.
Human research participants

Policy information about studies involving human research participants

Population characteristics
For the genomics assays in this study nine African American female subjects who were diagnosed with SLE with a mean age of 32 (21-51) were recruited. As a control cohort, 12 sex and race matched healthy controls with a mean age of 32 (21-48) were also recruited. Four additional cohorts were recruited for validation experiments and detailed information is included in Supplementary Table 1.

Recruitment
By design we limited recruitment to women of African American ancestry for both SLE patients and healthy controls. Although this is a highly relevant but understudied population in lupus, our results may not be generalizable to male patients or patients of non-African American ancestry. SLE patients were recruited predominately from a large public hospital serving the city of Atlanta and we focused recruitment on patients with moderate to high disease activity. This is a clinically relevant population but the B cell abnormalities we observed may be less apparent in patients with well controlled disease and greater access to healthcare. Healthy control donors were recruited predominately from the Emory University Community with autoimmunity, chronic illness, and recent infection or vaccination used as exclusion criteria. In addition to not having an autoimmune disease, healthy control donors also likely have better overall health than patients. Given that we consistently observe changes in genes previously linked to B cell activation and lupus in the SLE patients, we believe that SLE is the primary driver with any other comorbidities playing only a minor role.

Ethics oversight
This study was approved by the Emory University School of Medicine Institutional Review Board protocols.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:
☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
☒ All plots are contour plots with outliers or pseudocolor plots.
☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
50-100 ml s of whole blood were isolated from each subject via venous puncture utilizing cell preparation tubes (CPT) containing sodium heparin and ficoll-hypaque solution (Becton Dickinson).

Instrument
BD FACS Aria II

Software
Flowjo (TreeStar) v10.4

Cell population abundance
Purity was not determined for these samples as cell numbers were limiting. The FACS Aria was calibrated to 99.9% purity using fluorescent beads prior to each sort.

Gating strategy
All samples were pre-gated using the following strategy: 1) Lymphocytes based on FSC-A/SSC-A; 2) Singletons based on FSC-H/FSC-W, then SSC-H/SSC-W; 3) Live cells based on L/D Aqua negative; 4) CD19+ CD3.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.