nature research

Corresponding author(s): Jeremy M. Boss and Inaki Sanz

Last updated by author(s): May 9, 2019

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-flat.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 (SLE) 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

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Antibody Fluorophore Clone Company Cat# CD19 APC-Cy7 SJ25C1 BD Bioscience 557791 IgD 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 HIT2 BD Bioscience 562445 CXCR5 AX647 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 health care. 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 mls 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) Singlets 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.