PNAS

Supporting Information for

Transcriptional reprogramming primes CD8+ T cells towards exhaustion in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

David S. Iu[1]*, Jessica Maya[1]*, Luyen T. Vu[1], Elizabeth A. Fogarty[1], Adrian J. McNairn[2], Faraz Ahmed[2], Carl J. Franconi[1], Paul R. Munn[2], Jennifer K. Grenier[2], Maureen R. Hanson[1#], Andrew Grimson[1#]

[1] Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY

[2] Cornell University, Genomics Innovation Hub and TREx Facility, Ithaca, NY

* These authors contributed equally to this work

[#] Corresponding author: Maureen R. Hanson, Andrew Grimson

Email: maureen.hanson@cornell.edu, agrimson@cornell.edu

This PDF file includes:

Other supporting materials for this manuscript includes:

Dataset S1 to S2

Supporting Information Text

Sex as a biological variable. All bulk sequencing and flow cytometry data were collected from female subjects, as ME is diagnosed more frequently in females, and sex differences might confound the analysis. The scRNA-seq data examined subjects of both sexes, with a male:female ratio of 1:2. The downstream analyses, including intercellular communication inference, used only female subjects.

Statistics. RStudio and Microsoft Excel were used to conduct statistical analyses for all survey, sequencing, and flow cytometry data. ME and control subjects were compared with pairwise statistical testing using a Wilcoxon rank-sum test. Likelihood-ratio test was used for determining differential expression in scRNA-seq, while Wald test was used for bulk RNA-seq and ATAC-seq.

Sample collection and blood processing. Whole blood from each subject was drawn from an antecubital fossa vein following 2 maximal-effort cardiopulmonary exercise tests on a stationary cycle, with each test separated by a 24-hour period. Blood was collected into EDTA tubes and processed within 1–2 h into aliquots of whole blood, plasma, and PBMCs, as previously described (1, 2). EDTA tubes were spun at 500× *g* for 5 min, and plasma was separated and stored at −80 °C. Blood was diluted 1:2 in PBS and layered over Histopaque 1077 (Sigma-Aldrich, St. Louis, MO, USA) in 50-mL SepMate tubes (STEMCELL Technologies, Vancouver, Canada). SepMate tubes were spun at 1200× *g* for 10 min, excess plasma was removed, and cells were put into sterile 50-mL conical tubes. Platelets were depleted from the cells by first washing them in PBS at 120× *g* and then again at 300× *g* for 5 min. PBMCs were then resuspended in a freezing medium (60% RPMI 1640, 30% heat-inactivated FBS, 10% DMSO) and stored at −80 in isopropanol-containing freezing containers to slow down freezing. PBMC aliquots from New York City and Los Angeles were transported to the Hanson lab overnight on dry ice. Liquid nitrogen was used to store PBMCs for long-term use.

CD8+ T cells were isolated using STEMCELL EasySep kits on a STEMCELL EasyEights magnet. PBMCs were thawed in a 37 °C water bath and washed in RPMI 1640. To remove cell clumps, PBMCs were strained through a 37-µm cell strainer after being treated with 10 mg/mL DNase I for 10 min at room temperature. After a second wash, cells were isolated following the manufacturer's instructions using the EasySep Human CD8 Positive Selection II Kit. Isolated cells were resuspended in a freezing medium (60% RPMI 1640, 30% heat-inactivated FBS, 10% DMSO) and stored at −80 in isopropanol-containing freezing containers to slow down freezing. Samples were then moved to liquid nitrogen until the day of flow cytometric assays.

Flow cytometric analysis was performed on a BD Biosciences FACSymphony A3 Analyzer at the Cornell Biotechnology Resource Center (BRC) Flow Cytometry Facility. First, antibodies were titrated to the optimal concentration and amount for labeling live cells (3). CD8+ T cells were thawed, moved to 15-mL conical tubes, and incubated with surface marker stains eBioscience Fixable Viability Dye eFluor 506 (65-0866-14; ThermoFisher), anti-CD4 (Pacific Blue, SK3, 344620; BioLegend), anti-CD8 (APC-Cyanine7, RPA-T8, 557760; BD Biosciences), anti-CD27 (PE-Cyanine7, O323, 25-0279-42; ThermoFisher), anti-CD28 (Super Bright 600, CD28.2, 63-0289-42; ThermoFisher), anti-CD45RA (Super Bright 702, HI100, 67-0458-42; ThermoFisher), anti-CD69 (PE-eFluor 610, FN50, 61-0699-42; ThermoFisher), anti-PD-1 (PerCP/Cyanine5.5, NAT105, 367410; BioLegend), and anti-SLAMF6 antibody (EPR22170, ab224201; Abcam) for 20 min on ice in the dark. Cells were next washed with BD Biosciences Stain Buffer (554657; BD Biosciences), resuspended with 2ul of a secondary antibody for conjugation with anti-SLAMF6 (Goat anti-rabbit IgG Secondary Ab Alexa Fluor 488, A-11008; ThermoFisher) in 1 ml of Stain

Buffer, and incubated for 20 minutes on ice. Cells were then washed with Stain Buffer, resuspended, and incubated for 45 minutes on ice in the dark following the ThermoFisher eBioscience Foxp3/Transcription Factor Staining Buffer Set protocol. Next, the cells were resuspended with intracellular nuclear antibody anti-TOX (eFluor660, TXRX10, 50-6502-82; ThermoFisher) using the supplied ThermoFisher permeabilization buffer and incubated for 30 minutes at room temperature in the dark. Finally, cells were washed twice with the permeabilization buffer, resuspended in 200 µL of Stain Buffer, and moved to 5-mL round-bottom polystyrene tubes.

Flow cytometry for CD8+ T cell marker analysis. Flow cytometric analysis was conducted using FlowJo Software (v.10.8.1, Beckman Coulter, Brea, CA, USA). CD8+ T cells were first gated by size, granularity, and viability. They were next gated for naïve (N; CD8+CD27+CD28+CD45RA+PD-1-), central and early effector memory (M; CD8+CD27+CD28+CD45RA-), early effector memory (EM2; CD8+CD27+CD28-CD45RA-), intermediate effector memory (EM4; CD8+CD27-CD28+), late effector memory (EM3; CD8+CD27-CD28-CD45RA-), pre-effector types 1 and 2 that re-express CD45RA (pE1; CD8+CD27+CD28+CD45RA+PD-1+, pE2; CD8+CD27+CD28-CD45RA+) and terminally differentiated effector memory (Eff; CD8+CD27-CD28-CD45RA+) cell populations. Within each of these populations and in total CD8+ T cells, the MFI and frequency of markers were calculated, and were further analyzed for MFI and population frequency within cell subsets positive for PD-1 and/or negative for CD69. FMOs and unstained controls were used to call positive signals. Samples were excluded if the cell count within a cell population was less than 100. Additionally, a compensation matrix was created for both T cell panels using ThermoFisher UltraComp eBeads Plus and single stain samples.

Single cell RNA sequencing pre-processing. The single-cell RNA-seq atlas data was generated by the BRC Genomics Facility as detailed previously (4), consisting of 487,081 cells across 28 ME patients and 30 sedentary controls. Cells were partitioned using monocle3 (v1.3.1; 5) with default settings, and cells in the partition corresponding to T lymphoid cells (336,269 cells) were subsetted for further analysis. Cells expressing HBB (hemoglobin subunit beta), representing likely red blood cell contaminants, and cells with very high mitochondrial transcript percentage (>18%), which reflect poor sample quality, were removed. Samples were preprocessed by Latent Semantic Indexing with 100 dimensions using the preprocess_cds function, then corrected for batch effects using the align_cds function. Dimensional reduction was performed using the reduce_dimension function with umap.fast_gd=TRUE; and clustering was performed using the cluster_cells function with resolution=2.5e-5 and num_iter=20. Marker genes per cluster were identified using top_markers and clusters with extremely similar markers were then manually merged to yield 21 final clusters. Cell counts per cluster were calculated as described previously.

For CD8+ T cell-specific analyses, clusters annotated as CD8+ T cells were selected for reclustering using Seurat (v4.4.0; 6). Briefly, the SCT assay was used to rerun PCA, then the first 15 dimensions were used for the FindNeighbors and RunUMAP functions. FindClusters was used at a resolution of 0.8, yielding 12 clusters, and FindAllMarkers was used to identify marker genes for each cluster. Clusters with extremely similar markers were then manually merged to create the final reclustered object, consisting of 9 clusters.

Single cell RNA sequencing analyses. For single-cell differential expression analysis, scran (v1.22.1; 7; computeSumFactors) was first used to generate per-cell size factors. The samples were binned into 4 age groups. A DESeq2 object (v1.34.0; 8) was created with a formula design to regress out sex and age bins. Calculation of differential

expression and statistical testing of expressed genes per cluster (gene expressed in more than 1% of cells in a cluster) were then performed using DESeq2 and glmGamPoi (v1.6.0; 9), with parameters optimized for single-cell analysis as recommended by the developers and collaborating groups (10). Log2-fold change values from DESeq2 were used for GSEA through clusterProfiler (v4.2.2; 11) with maximum gene set size = 1000 and eps = 1e-20. The Hallmark and C5:BP catalogs were downloaded from the Molecular Signature Database (MSigDB) using the R package msigdbr (v7.5.1; 12, 13) and were used for enrichment testing. CellChat (v1.1.3; 14) was used to infer cell-cell communication probabilities and identify signaling changes across case and control cohorts, as previously described (4). The relative contribution of each ligand-receptor pair to specific pathways was calculated with netAnalysis_contribution.

To compute a CD4+ effector index, we first downloaded a table of effectorness-dependent genes (211 genes; 15), which contains coefficients for effectorness per gene, based on single-cell trajectory analysis. We then calculated row-wise z-scores for each gene included on the effectorness table per CD4+ T cell cluster (clusters 4, 5, 7, 8, 9), which yields a matrix of 211 genes by 176,112 cells. The effectorness index of each cell was then calculated as the product of this z-score matrix and the gene effectorness coefficient. Cells with outlier effectorness index values (top or bottom 1%) were excluded from downstream analysis. Effector indices of each cluster in case and control were compared with pairwise statistical testing by Wilcoxon rank-sum test. Monocle3 (v1.3.1; 5, 16, 17) was used to generate a differentiation trajectory and pseudotime values for CD8+ T cells.

Bulk RNA sequencing. CD3+ cells were enriched using the CD3+ MicroBeads (Miltenyi Biotech Cat#130-097-043) from PBMCs. After enrichment, cells were incubated with Aqua dead cell stain kit (Thermo) with Human TruStain FcX™ Fc Receptor Blocking Solution (Biolegend) for 10 minutes at room temperature. Fluorescently labeled antibodies (CD4-FITC Cat#357406, CD8a-BV785 Cat#301406, CD45RA-BV650 Cat#304136, CCR7-APC-Cy7 Cat#353212, and CD56-PE-Cy7 Cat#362510 all from BioLegend) were subsequently added to the cells and incubated on ice for 30 minutes. 1 mL of FACS buffer (heat inactivated FBS, 0.5M EDTA pH8.0 1x PBS) was added to the cells, mixed well and centrifuged at 300 rcf for 5 minutes at 4 °C. The supernatant was removed and the cells were washed once more. The final cell pellet was resuspended in 100µl of FACS buffer and kept on ice in the dark. Flow sorting was performed on a FACSAria Fusion at the BRC Flow Cytometry Facility. Live and CD56 negative cells were selected for CD4 and CD8 expressions. Naïve cells were sorted as CD45RA+CCR7+. Memory cells were sorted as CD45RA-CCR7+/-. Samples were then submitted to the BRC Transcriptional Regulation and Gene Expression Facility (TREx), where RNA extraction and library preparation were performed as previously reported (4) using up to 20ng total RNA with the NEBNext Ultra II RNA Library Prep Kit. Libraries were sequenced on a NovaSeq6000 (Illumina) at Novogene to a minimum sequencing depth of 30M 2x150bp PE reads.

Tn5 transposome preparation. Tn5 protein and transposomes were produced in the BRC Genomics Innovation Hub. Tn5, encoded on a pET151/D vector, was expressed in *E.coli* (C1301, NEB) and purified via affinity chromatography as described (18). Briefly, a 1L culture was inoculated with an overnight starter culture (1:50 dilution) in 2XYT media with carbencillin and grown to a OD600 of 0.6. The culture was chilled on ice for 10 minutes prior to induction with 1mM IPTG. The culture was then grown for an additional 4-6 hours at room temperature. The culture was centrifuged and the bacterial pellets washed in cold HEGX buffer (20mM HEPES (pH 7.3), 800mM NaCl, 1mM EDTA (pH 8.0), 10% glycerol, 0.2% Triton X-100) and the bacterial pellets frozen and stored at -80°C. Pellets were resuspended in 50mL cold HEGX buffer plus protease inhibitors and sonicated using a

US SOLID sonicator set to 50% power for 4 minutes using a 5sec on/5sec off program and waiting 1 minute between rounds of sonication. After centrifugation to remove insoluble material, genomic DNA was precipitated by adding 2.5% of lysate volume with 10% PEI. The cleared lysate was then added to a chitin resin (NEB) column and allowed to bind. After multiple washes with HEGX, bound protein was eluted using HEGX with 50mM DTT and sealing the column for 48 hours. The elution was then collected, dialyzed (100mM HEPEs (pH 7.3), 200mM NaCl, 0.2mM EDTA (pH 8.0), 20% glycerol, 0.2% Triton X-100, 2mM DTT) and stored at -80°C. ME-oligo duplexes were annealed by slowly cooling equimolar ratios of ME-A or ME-B, and ME-Rev oligos (Table S2). Tn5 transposomes were prepared by combining equimolar amounts of Tn5 protein and ME-oligo duplexes in 1X Tn5 Exchange Buffer (50mM HEPES, pH 7.2, 100mM NaCl, 0.1mM EDTA, 0.1% Triton X-100, 1mM DTT, 50% glycerol) buffer incubating for 21 hours at 25°C with 600rpm shaking in an Eppendorf Thermomixer, followed by storage at -80°C. The functional concentration of active transposomes was determined using a quantitative PCR assay to detect cleavage of a plasmid target.

Nuclei preparation and ATAC-seq. Nuclei were prepared from 12,000-50,000 CD8+ T cells sorted as described above, using a modified Omni-ATAC-seq protocol (19). Briefly, cells were pelleted at 500 xg for 5 minutes at 4°C, supernatant removed, and cells were permeablized in 50 µl cold lysis/permeabilization buffer with 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin for 3 minutes on ice. Lysis was quenched by adding 1mL ice-cold RSB-T (10mM Tris pH 7-8, 10mM NaCl, 3 mM MgCl2, and 0.1% Tween-20). Nuclei were pelleted at 500 xg for 10 minutes at 4°C and supernatant was replaced with 50 µl 1x PBS + 10% DMSO for slow freezing and storage at -80°C. For tagmentation, nuclei were thawed and washed twice with 1x RSB+T before resuspending in 25 µl Tagmentation buffer (1x Cutsmart buffer (NEB), 10% DMF, 0.1% Tween-20, 0.01% digitonin, and 0.2 µM Tn5). Nuclei were tagmented for 30 minutes at 37°C with 500 rpm shaking on an Eppendorf Thermomixer. Following tagmentation, DNA was purified using a Zymo clean and concentrate-5 column and PCR performed using Nextera primers containing unique dual indices and 2X NEBNext High Fidelity master mix for 11-13 cycles (depending on starting input; 72°C for 5min, 98°C 30sec, cycles of 98°C 10 sec, 63°C 30 sec, 72°C 1min). PCR products were purified using SparQ beads (Quanta Bio) and resuspended in 0.1X TE. Libraries were checked on a Fragment Analyzer (Agilent) and size selected to 150-600 bp using a Pippin HT (Sage Sciences). Libraries were shallow sequenced on a MiSeq at Cornell BRC Genomics to check the balance of the pooled libraries. The pool was adjusted and sent for sequencing on two NovaSeq6000 (Illumina) S4 lanes at URMC to a depth of 50M reads per library.

Pre-processing of bulk sequencing data. For RNA-seq, fastq files were trimmed and aligned to GRCh38 as previously described (TrimGalore v0.6 and STAR v2.7; 4, 20, 21). A count table of reads per gene (Ensembl gene annotations; 22) was output by STAR, and the count table was used for downstream analyses. For ATAC-seq, fastq files were trimmed to remove low quality and adaptor sequences as described (23). Trimmed reads were then aligned to hg38 using bowtie2 (v2.4.5; 24). Mitochondrial reads and blacklisted regions (25) were removed, and PICARD (v2.26.1) and samtools (v1.18; 26) were used to sort and remove duplicate reads.

Bulk RNA-seq analysis. DESeq2 (v1.34.0) was used to perform differential expression analysis and statistical testing of expressed genes (more than 10 overlapping reads in at least 25% of samples). Genes were ranked by $\log_2$(case/control) as output by DESeq2 for GSEA. Additional gene sets either comparing naïve, effector, and memory CD8+ T cells (27) or exhausted and non-exhausted effector CD8+ T cells (28) were downloaded from

their respective publications. clusterProfiler (v4.2.2) was used with default parameters to calculate normalized enrichment scores and p-values, followed by Benjamini-Hochberg correction.

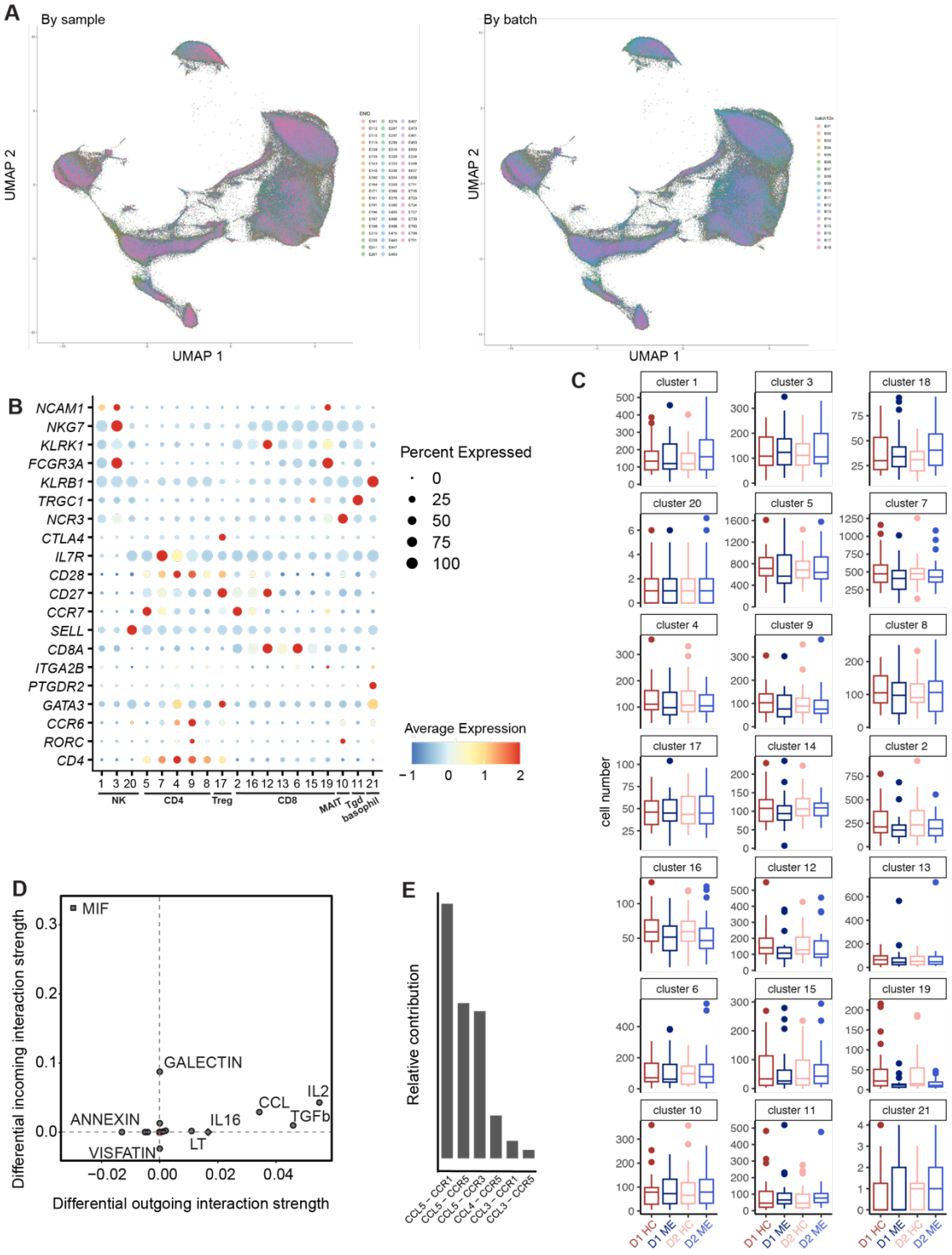
Annotations for ATAC-seq analysis. Promoter annotations were downloaded from GENCODE release H43 (29). Putative TF binding sites were collected from JASPAR2022 using its companion R package (JASPAR2022 v0.99.7; 30).

ATAC-seq analysis. Peaks were called using MACS3 (v3.0.0b1; 31; callpeak –f BAMPE –g hg –B –q 0.01). Consensus peaks from all cell types and biological conditions were then merged by muMerge (v1.1.0; 32) to create a unified set of chromatin accessible regions (ChARs). featureCounts under subread (v2.0.3; 33) was used to calculate raw read counts per ChAR. For differential accessibility analysis, these regions were further filtered such that only ChARs with at least 5 overlapping reads in at least 25% of samples were kept, resulting in 67,189 ChARs. DESeq2 was used to estimate size factors for data normalization, and for statistical testing of accessible regions between biological conditions. For gene-level analyses, ChARs were associated with their nearest protein-coding genes based on the shortest distance between the region and the gene's promoter, defined as 1kb upstream and 500bp downstream of the annotated transcription start site, then reads were aggregated per gene. DESeq2 was used for calculating differential aggregate accessibility of genes across biological conditions, using size factors previously generated in the peak-level analysis. GSEA was performed as described above. For epigenetic scarring analysis, we first downloaded a list of ChARs in HCV-specific CD8+ T cells which remain refractory to curing of the chronic infection (34). We then used a permutation test to determine whether ChARs upregulated in the case cohort significantly overlapped these regions.

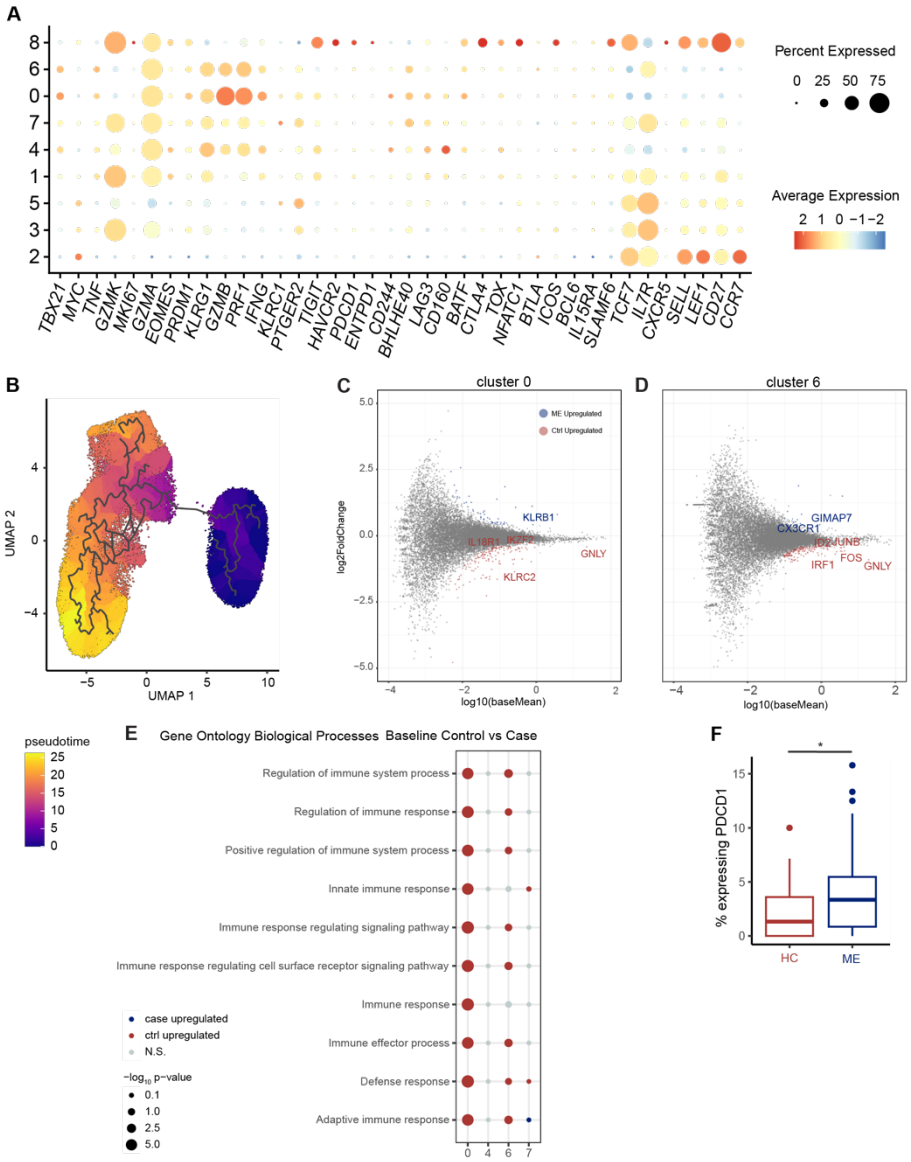
Motif analysis. Only TFs expressed in CD8+ T cells according to our RNA-seq data were kept, resulting in 525 motifs. monaLisa (v1.0.0; 35) with recommended parameters was used for binned motif enrichment analysis for TF binding sites, with the bins being ChARs significantly differentially accessible in either case or control (1571 control upregulated, 517 case upregulated), ChARs that were differentially but not statistically significantly accessible in either case or control (6334 control upregulated, 11080 case upregulated), and an unchanged control consisting of ChARs with $\log_2$(case/control) < 0.1 (9381).

Track visualization. RNA-seq and ATAC-seq bam files were indexed (samtools index), then normalized to counts per million (CPM) using bamCoverage from deeptools (v3.5.1; 36) to generate bigWig files for visualization in igv (v2.9.1; 37).

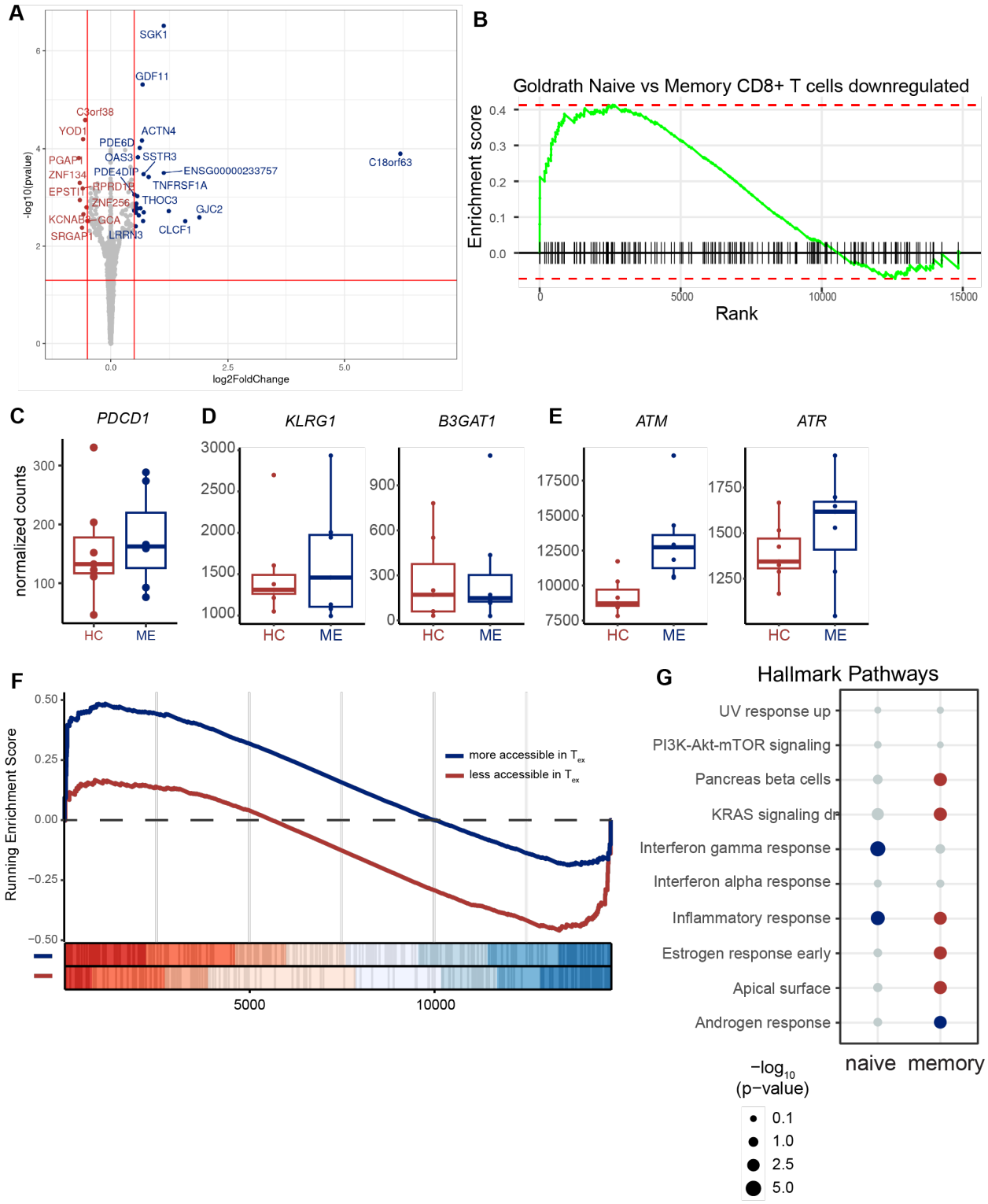
Supplemental Figures



Supplemental Figure 1. Analysis of ME atlas T lymphoid cells
(A) UMAP of T cell clusters from Vu, et al. 2024, colored by sequencing batch (left) or original sample (right).
(B) Relative expression of selected marker genes (y-axis) per cluster (x-axis). Dots represent average expression (color) and percentage of expressing cells per cluster and condition (size).
(C) Box plots showing cell numbers in each biological condition in each cluster.
(D) Scatter plot of differential incoming versus outgoing interaction strength in γδ T cells. Positive values indicate increased signaling strength in patients and vice versa. Surprisingly, we also observed increased outgoing signaling in the C-C motif chemokine (CCL) pathway, despite downregulation of *CCL3* and *CCL3L1*.
(E) Bar plot of the ratio of aggregate communication probability per ligand-receptor pair in the CCL pathway to that of the overall pathway (relative contribution). The increased outgoing CCL pathway signal derives from elevated *CCL5*, another chemokine responsible for localization of lymphocytes to inflamed tissues and previously implicated in ME pathology.

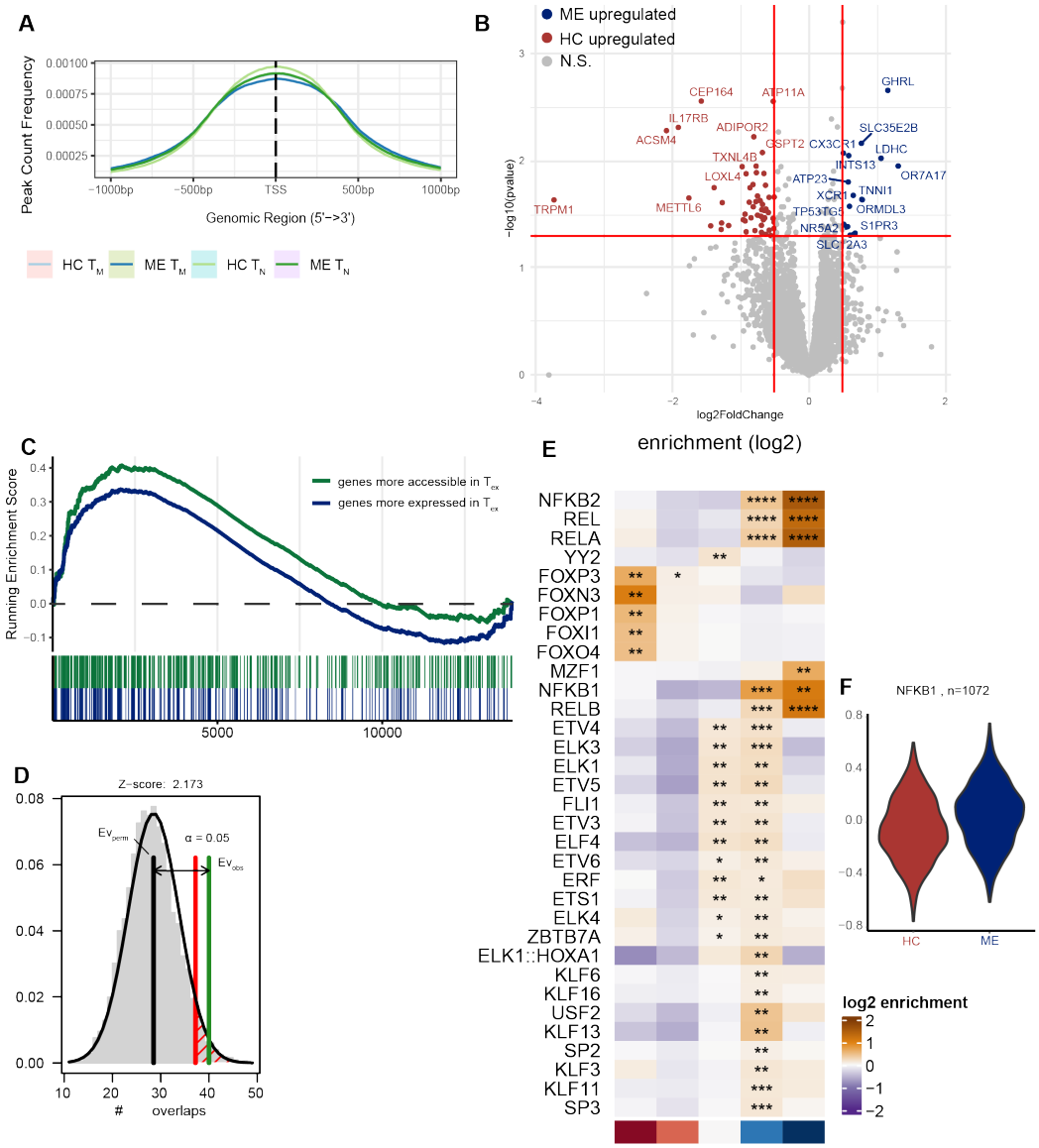


Supplemental Figure 2. Additional analysis of ME atlas CD8+ T cell subsets.
(A Relative expression of effector, memory and exhaustion marker genes (y-axis) per cluster, ordered from least differentiated (bottom), to most differentiated (top). Dots represent average expression (color) and percentage of expressing cells (size).**)**
(B) UMAP of CD8+ T cells colored by pseudotime values. Lines connecting nodes indicate edges of the predicted trajectory. Darker colors (low pseudotime values) indicate earlier stages in the trajectory. This trajectory originates from T_N, transitions to T_{EM}^{early} (cluster 5), and branches to another T_{EM}^{early} subset (cluster 3) then the TIM-3+ T_{EM} subset, or into T_{EM}^{int} (clusters 1 and 4) and T_{EM}^{late} subsets (clusters 0, 6, and 7).
(C) MA plot of differentially expressed genes in CD8+ T cell cluster 0 and in **(D)** cluster 6 ($p \leq 0.05$, colored dots). **(E)** GSEA dot plot of effector memory CD8+ T cells against hallmark pathways from MSigDB. Size indicates statistical significance, while color indicates whether the pathway is enriched in case (blue) or control (red).
(F) Box plot showing the percentage of cells expressing *PDCD1* in cluster 0 (p=0.016) in each biological condition; 2.2% in HC vs 3.9% in ME; p=0.0164. Statistical significance was calculated by Wilcoxon rank-sum test. *$p \leq 0.05$



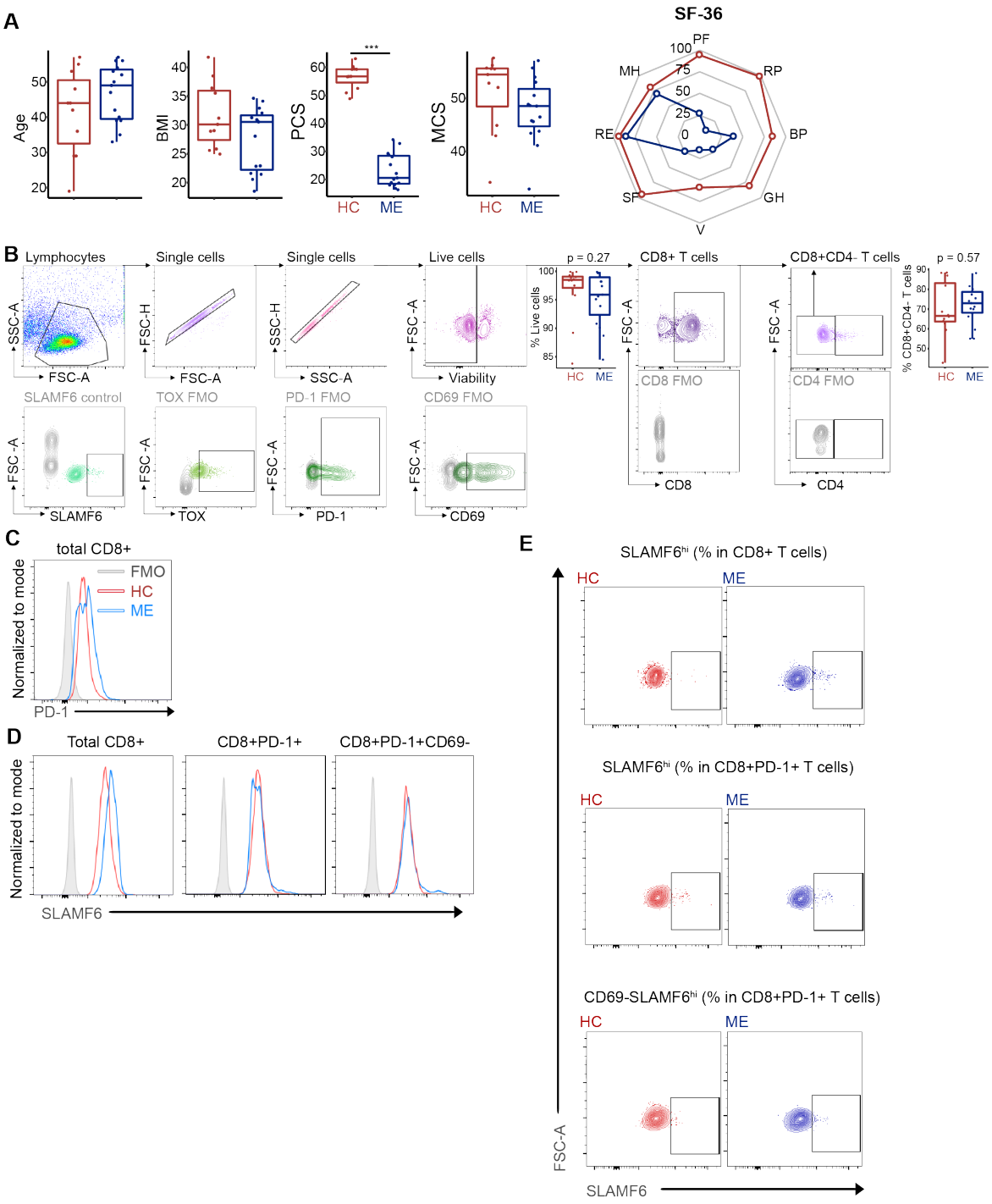
Supplemental Figure 3. Bulk transcriptional profiles of ME CD8+ T cells.
(A) Volcano plot highlighting differentially expressed genes using RNA-seq in naïve CD8+ T cells ($p \leq 0.05$, colored dots).

(B) GSEA enrichment plot of naïve CD8+ T cells against genes downregulated in naïve CD8+ T cells when compared with memory CD8+ T cells *(42)*.
(C) Box plot of normalized counts of *PDCD1* (p=0.75), **(D)** *KLRG1* (p=0.22) and *B3GAT1* (p=0.94), and **(E)** *ATM* (p=0.0005) and *ATR* (p=0.18), comparing between case and control.
(F) GSEA enrichment plot of effector memory CD8+ T cells against genes differentially accessible in *in vivo* exhausted T cells *(47)*.
(G) GSEA dot plot of naïve and effector memory CD8+ T cells against hallmark pathways from Molecular Signature Database (MSigDB). Size indicates statistical significance, while color indicates whether the pathway is enriched in case (blue) or control (red).



Supplemental Figure 4. Chromatin accessibility of ME CD8+ T cells.
(A) Combined ATAC-seq signal across all transcriptional start sites for each biological condition in naïve and effector memory CD8+ T cells.
(B) Volcano plot highlighting differentially accessible genes using ATAC-seq in naïve CD8+ T cells ($p \leq 0.05$, colored dots).
(C) GSEA enrichment plot of effector memory CD8+ T cells against genes with increased expression or accessibility in *in vivo* exhausted T cells.
(D) Permutation test histogram showing single-dataset enrichment for epigenetically scarred regions detailed in Yates et al. 2021.
(E) Heatmap of TF binding motifs significantly enriched in at least one bin of ChARs in T_N. ChARs were grouped into five bins by differential accessibility in cases versus controls: significantly decreased (dark red), decreased (light red), unchanged (white), increased (light blue), and significantly increased (dark blue) accessibility in ME vs. controls. **** $q \leq 1\times10^{-10}$, *** $q \leq 1\times10^{-6}$, ** $q \leq 0.0001$, * $q \leq 0.05$. Enrichment scores normalized row-wise.

(F) Distribution of peak accessibility z-scores in T_N across all ChARs that contain at least one NFKB1 motif.



Supplemental Figure 5. Analysis of in total CD8+ T cell populations

(A) Age (p=0.34), BMI (p=0.18), and SF-36 spider plot comparing ME (n=15) and control (HC) cohorts (n=11), including box plots of the physical (PCS, $p=2.6\times10^{-7}$) and mental component score (MCS, p=0.16). Comparisons performed by Wilcoxon rank-sum test. *$p\leq0.05$, ***$p\leq0.001$.

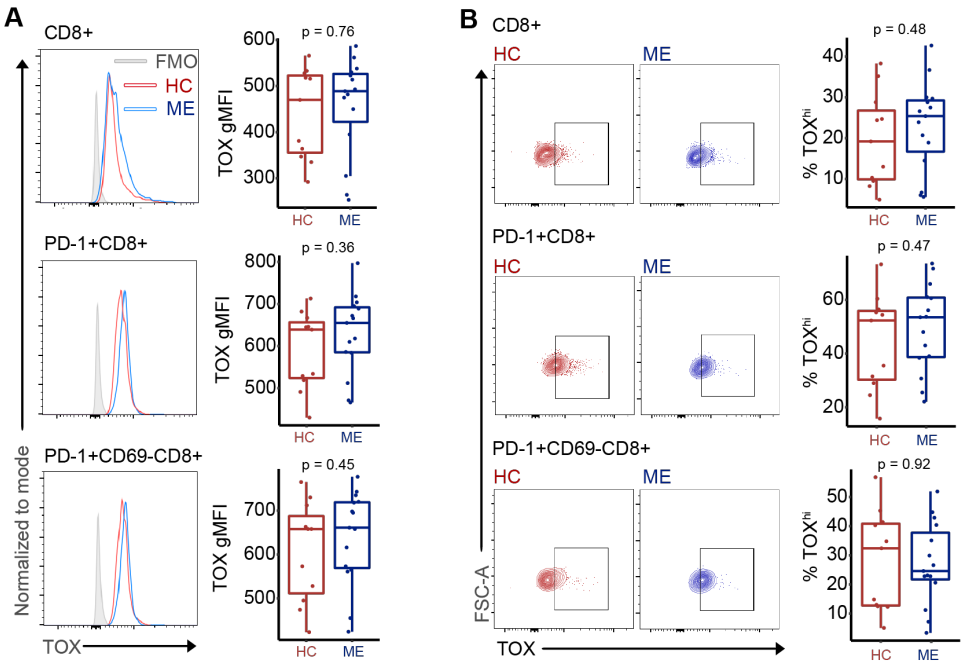
(B) Top: Representative scatter plots of gating strategy for single live CD8+CD4- T cells, including a boxplot for the frequency of live cells in single cell population (p=0.27, n=11 healthy control samples; n=15 ME samples (n=11/15)) and a

boxplot for the frequency of CD8+CD4- T cells in the live cell population (p=0.57, n=11/15). Bottom: Representative scatter plots of total CD8+ T cells SLAMF6 negative control, TOX fluorescent minus one control (FMO), PD-1 FMO, CD69 FMO, CD8 FMO, and CD4 FMO, (grey) with respective full panel scatterplots (green) to show positive gating calls. Boxes in SLAMF6 and TOX FMO plots represent gates for SLAMF6^{hi} and TOX^{hi}.

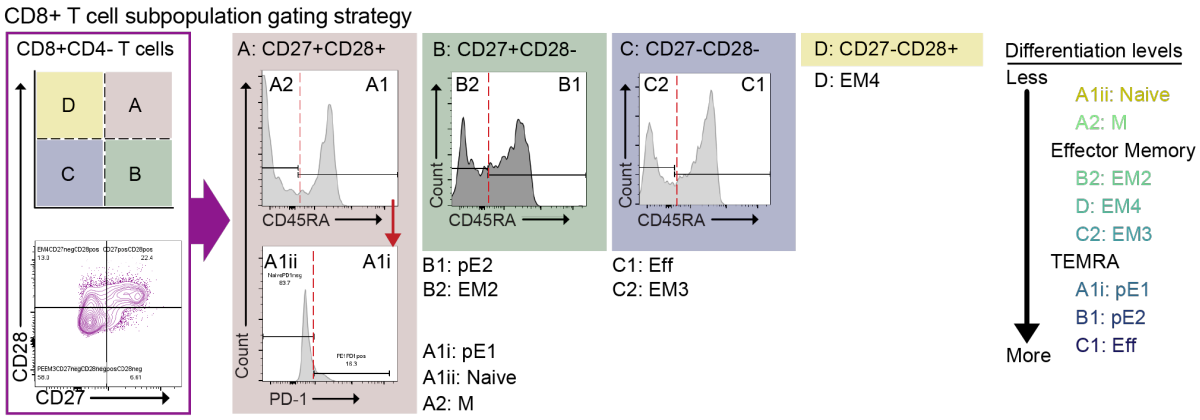
(C) Representative histogram for PD-1 expression on CD8+ T cells.

(D) Representative histograms for SLAMF6 expression in total CD8+ (left), CD8+PD-1+ (middle), and CD8+PD-1+CD69- (right) T cells.

(E) Representative scatter plots of healthy control and ME SLAMF6^{hi} populations in total CD8+ (top), CD8+PD-1+ (middle) and CD8+PD-1+CD69-(bottom) T cells.

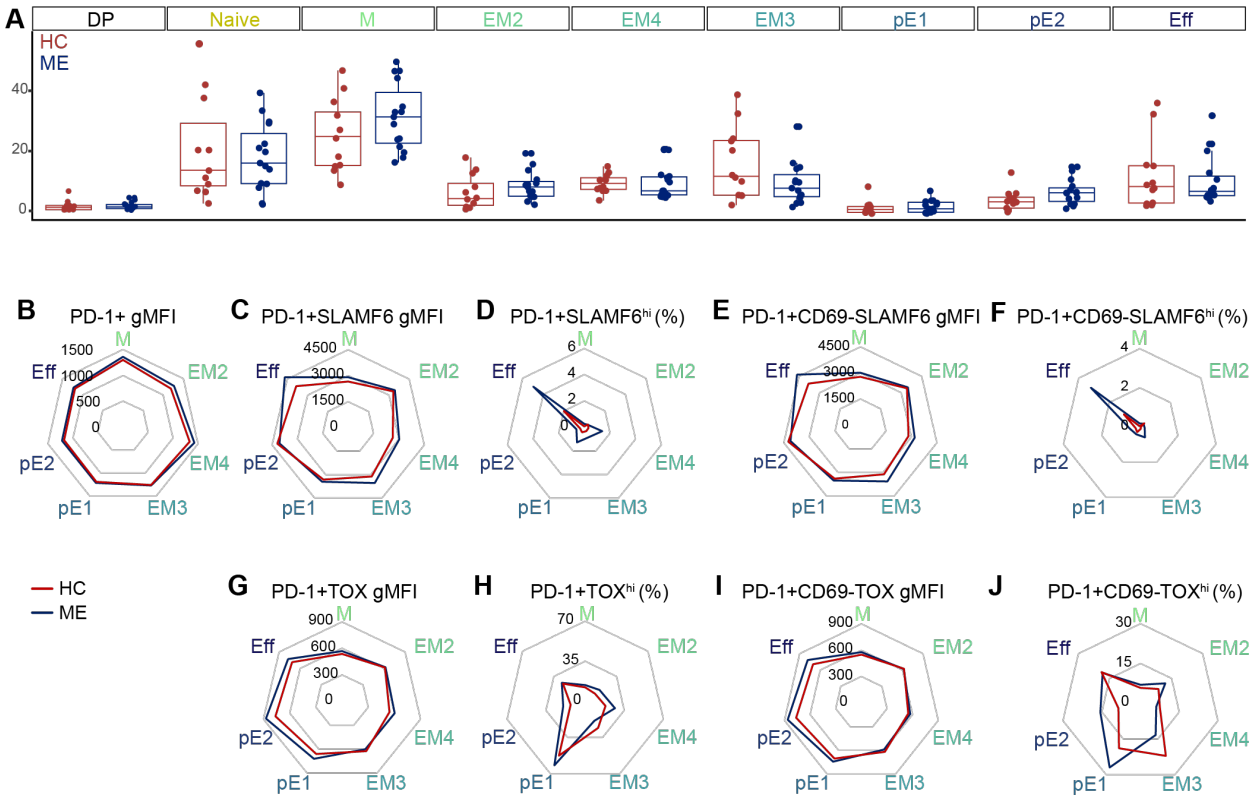


Supplemental Figure 6. TOX is unaltered in total ME T cells.
(A) Left: representative histograms for TOX expression in total CD8+, CD8+PD-1+, and CD8+PD-1+CD69- T cells. Right: TOX expression levels in respective cell subsets (*n*=11/15).
(B) Left: representative scatter plots of healthy control and ME TOX^{hi} populations in total CD8+, CD8+PD-1+, and CD8+PD-1+CD69- T cells. Right: TOX^{hi} frequencies in respective cell subsets. Significance was calculated by Wilcoxon rank-sum test (* p<0.05) (n =11/15).

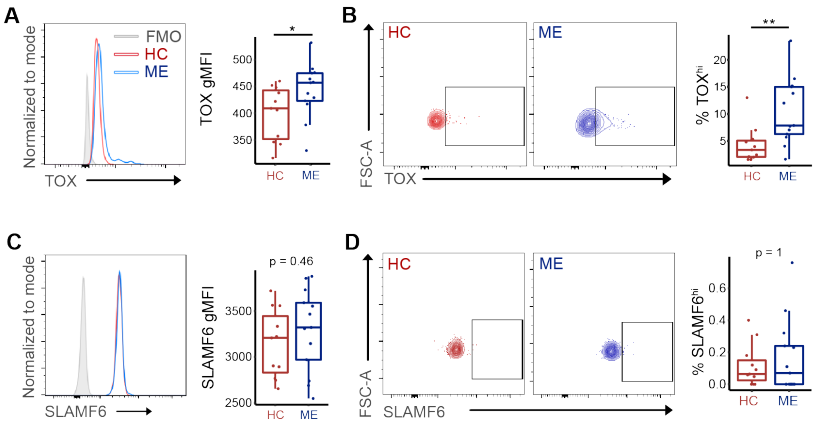


Supplemental Figure 7. Gating strategy for flow cytometric analysis of CD8+ T cell subpopulations.

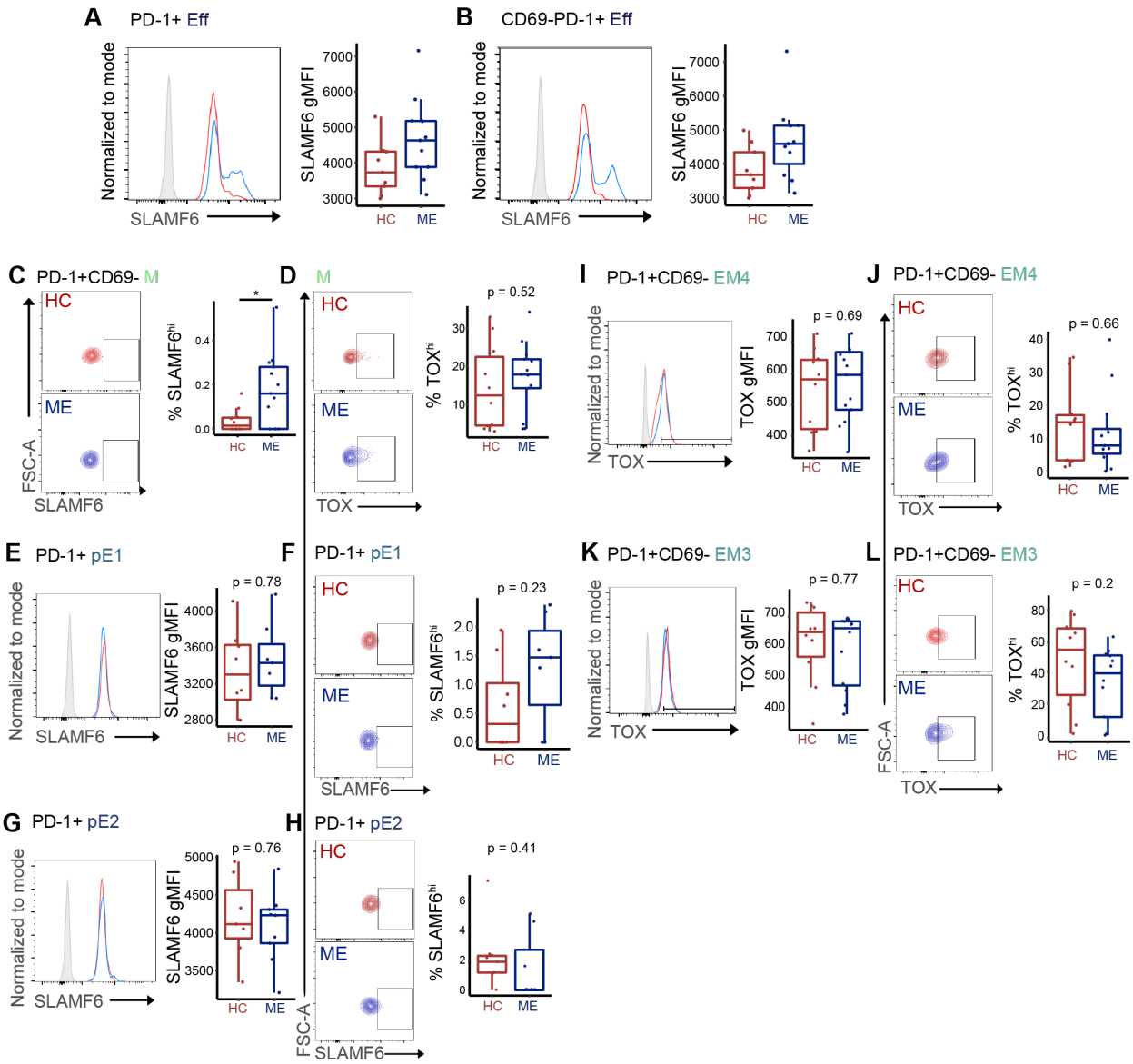
CD8+ T cells were gated on CD27, CD28, and CD45RA (along with PD-1 for T_N and pE1 populations) for naïve (N; CD8+CD27+CD28+CD45RA+PD-1-), central and early effector memory (M; CD8+CD27+CD28+CD45RA-), early intermediate (EM2; CD8+CD27+CD28-CD45RA-), intermediate effector memory (EM4; CD8+CD27-CD28+), late effector memory (EM3; CD8+CD27-CD28-CD45RA-), effector memory re-expressing CD45RA (pE1; CD8+CD27+CD28+CD45RA+PD-1+, pE2; CD8+CD27+CD28-CD45RA+) and terminally differentiated effector memory (Eff; CD8+CD27-CD28-CD45RA+). Levels of differentiation status are noted on the right (Koch et al., 2008, Chikuma et al. 2009). Cells/events in Figure 7 are from an ME representative sample. Statistical significance was calculated by Wilcoxon rank-sum test (* p<0.05).



Supplemental Figure 8. Additional T cell marker frequencies and abundances measured between ME and healthy sedentary controls.
(A) Box and whisker plots for the frequencies of each CD8+CD4- T subpopulation: CD4+CD8+ double positive (p=0.31, n=11 HC samples; n=15 ME samples (n=11/15)), naïve (T_N) (p=0.96, n=11/15), central and early effector memory (M) (p=0.22, n=11/15), early intermediate (EM2) (p=0.18, n = 11/15), intermediate effector memory (EM4) (p=0.55, n=11/15), late effector memory (EM3) (p=0.10, n=11/15), late effector memory re-expressing CD45RA (pE1) (p=0.38, n=11/15), late effector memory re-expressing CD45RA (pE2) (p=0.077, n=11/15), and terminally differentiated effector memory (Eff) T cells (p=1, n=11/15) compared between HC and ME.
(B) Spider plots of PD-1+ M, EM2, EM4, EM3, pE2, and Eff T cell subsets depicting PD-1+ gMFI, **(C)** SLAMF6 gMFI, **(D)** SLAMF6^{hi} frequency, and PD-1+CD69- **(E)** SLAMF6 gMFI, and **(F)** SLAMF6^{hi} frequency between HC and ME. Each red (HC) or blue (ME) line represents the mean % or gMFI per T cell subset.
(G) Spider plots of PD-1+ M, EM2, EM4, EM3, pE2, and Eff T cell subsets depicting TOX gMFI, **(H)** TOX^{hi} frequency, and PD-1+CD69- **(I)** TOX gMFI, and **(J)** TOX^{hi} frequency between HC and ME.



Supplemental Figure 9. Exhaustion marker analysis in naïve CD8+ T cells. **(A)** TOX MFI in naïve CD8+ T cells (p=0.04) and **(B)** TOX^{hi} frequency in naïve CD8+ T cells (p=0.003, n =11/13), compared between HC and ME.
(C) SLAMF6 MFI in naïve CD8+ T cells (p=0.46) and **(D)** SLAMF6^{hi} frequency in naïve CD8+ T cells (p=1, n =11/13), compared between HC and ME. Statistical comparisons were performed with Wilcoxon rank-sum test. *$p \leq 0.05$, **$p \leq 0.01$.



Supplemental Figure 10. Additional analyses of exhaustion marker frequencies and abundances in effector memory populations.
(A) PD-1+ Eff (p=0.08, n=9/11), and **(B)** PD-1+CD69- Eff T cells (p=0.08, n=9/11), compared between HC and ME.
(C) Representative scatter plots of SLAMF6^{hi} populations in PD-1+CD69- M T cells, and box plot of the percentage of SLAMF6^{hi} cells in PD-1+CD69- M T cells (p=0.52, n =10/13), compared between HC and ME.
(D) Representative scatter plots and box plot of the percentage of TOX^{hi} cells in M T cells (p=0.52, n =10/13), compared between HC and ME.
(E) Representative histogram and box plot of SLAMF6 expression in PD-1+ pE1 T cells (p=0.78, n =7/7), compared between HC and ME.
(F) Representative scatter plots and box plot of the percentage of SLAMF6^{hi} cells in PD-1+ pE1 T cells (p=0.23, n =7/7), compared between HC and ME.
(G) Representative histogram and box plot of SLAMF6 expression in PD-1+ pE2 T cells (p=0.76, n =7/9).

(H) Representative scatter plots of SLAMF6^{hi} populations in PD-1+ pE2 T cells, and box plot of the percentage of SLAMF6^{hi} cells in PD-1+CD69- pE2 T cells (p=0.41, n =7/9), compared between HC and ME.
(I) Representative histogram and box plot of TOX expression in PD-1+CD69- EM4 T cells (p=0.69, n=10/13)
(J) Representative scatter plots of TOX^{hi} populations in PD-1+CD69- EM4 T cells, and box plot of TOX^{hi}+ frequency in PD-1+CD69- EM4 T cells (p=0.66, n=10/13), compared between HC and ME.
(K) Representative histogram and box plot of TOX expression in PD-1+CD69- EM3 T cells (p=0.77, n=10/12).
(L) Representative scatter plots of TOX^{hi} populations in PD-1+CD69- EM3 T cells, and box plot of TOX^{hi}+ frequency inPD-1+CD69- EM3 T cells (p=0.2, n=10/12), compared between HC and ME.

Supplemental data

Table S1

Subject demographics for participant data from Figures 2-5 (ME = 21; HC = 18). Duration of ME, BMI, and Age are shown as ranges placed into bins.

ME Duration Bins: Bin 1: <5, Bin 2: 5-12, Bin 3: 13-20, and Bin 4: >20.

BMI Bins: Bin 1: ≤18.5; Bin 2: >18.5 to ≤25; Bin 3: >25 to ≤27; Bin 4: >27 to ≤30; Bin 5: >30

Age Bins: Bin 1: >18 to ≤35; Bin 2: >35 to ≤45; Bin 3: >45 to ≤55; Bin 4: >55 to ≤70

BAS Score: Bell activity scale score.

SF-36 survey: PF=physical functioning, RP=role: physical, BP=bodily pain, GH=general health, V=vitality, SF=social functioning, RE=role: emotional, MH=mental health, PCS=physical component score, MCS=mental component score.

ME onset: 1=sudden, 2=gradual.

Assay: whether the sample was included in the 7 ME and 7 HC sequencing analysis (T) or the 15 ME and 11 HC flow cytometry analysis (F).

ME Duration	1	1	1	2	1	3	1	1	N/A	2	4	2	4	1	4	1	1	1
BMI	2	2	5	2	4	2	5	5	5	5	2	5	2	5	5	1	2	2
Age	3	1	3	3	2	4	2	3	4	2	1	3	2	3	4	4	1	2
BAS Score	50	30	30	30	30	40	55	20	20	20	20	20	20	20	50	N/A	N/A	N/A
MCS	49	33	49	43	46	56	46	51	54	53	49	41	57	44	46	42	27	49
PCS	29	29	18	25	19	28	34	20	18	17	22	21	16	19	33	30	38	16
MH	70	55	75	80	60	80	69	50	90	80	70	60	90	50	70	64	20	68
RE	100	33	92	75	75	100	58	100	100	92	100	75	100	100	83	67	100	100
SF	13	25	13	0	25	50	63	50	13	25	13	0	25	0	50	38	13	13
VT	25	13	6	0	13	31	38	25	6	19	13	0	13	6	25	10	10	15
GH	35	10	15	40	20	42	20	20	30	5	10	25	20	20	10	57	10	15
BP	100	21	12	22	22	62	62	N/A	31	10	62	31	51	21	41	31	74	41
RP	6	0	0	19	6	19	31	0	0	25	0	0	0	0	50	0	0	0
PF	15	55	35	35	5	30	45	25	25	20	25	10	5	15	60	55	80	15
ME Onset	1	2	1	2	1	1	1	1	2	2	1	1	1	1	1	2	2	2
Assay	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T	T	T
Phenotype	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME

ME Duration	2	3	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
BMI	5	4	2	5	3	4	5	5	4	5	3	5	5	2	2	3	5	2	3	2
Age	4	2	1	2	1	2	3	2	1	3	1	2	3	4	4	4	1	1	3	1
BAS Score	N/A	6	7	90	100	90	80	100	100	100	100	100	90	90	100	90	90	90	70	100
MCS	56	4	0	56	54	53	34	56	56	58	56	43	45	52	60	49	22	50	58	54
PCS	12	6	1	51	57	57	49	57	59	60	59	63	59	52	49	58	60	59	53	58
MH	84	29	35	85	85	85	45	85	85	90	90	60	70	80	92	60	40	80	85	80
RE	100	21	24	100	100	100	100	92	100	100	100	100	67	83	100	100	0	92	100	100
SF	13	34	35	100	100	100	25	100	100	100	100	100	100	100	100	100	50	100	100	100
VT	15	21	37	56	63	44	6	75	75	81	63	19	56	50	80	70	40	50	75	63
GH	25	31	35	62	62	82	55	97	100	95	90	97	72	77	62	82	40	72	72	100
BP	22	75	85	62	100	84	51	62	84	100	100	100	84	84	72	100	100	100	74	74
RP	0	100	100	100	100	100	88	100	100	100	100	100	100	88	100	100	100	100	100	100
PF	10	25	38	95	100	100	80	100	100	100	100	100	95	75	95	90	85	100	90	95
ME Onset	1	2	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Assay	T	T	T	F	F	T	F	F	F	F	F	F	F	F	T	T	T	T	T	T
Phenotype	ME	ME	ME	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC

Table S2
List of oligonucleotides used for ATAC-seq library preparation.

oligo	sequence
ME-A	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′
ME-B	5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′
ME-Rev	5'PHOS/CTGTCTCTTATACACATCT

Dataset S1
Related to Figure 1. Differential expression analysis in scRNA-seq data analyzed per cluster.

Dataset S2
Related to Figure 3. List of dysregulated genes showing a loss of poising, with decreased accessibility and no change in expression level. Positive values indicate increased accessibility or expression in ME.

SI References

1. A. Germain, *et al.*, Plasma metabolomics reveals disrupted response and recovery following maximal exercise in myalgic encephalomyelitis/chronic fatigue syndrome. *JCI Insight* **7**, e157621 (2022).

2. A. H. Mandarano, *et al.*, Myalgic encephalomyelitis/chronic fatigue syndrome patients exhibit altered T cell metabolism and cytokine associations. *J Clin Invest* **130**, 1491–1505 (2020).

3. R. Hulspas, Titration of fluorochrome-conjugated antibodies for labeling cell surface markers on live cells. *Curr Protoc Cytom* **Chapter 6**, Unit 6.29 (2010).

4. L. T. Vu, *et al.*, Single-cell transcriptomics of the immune system in ME/CFS at baseline and following symptom provocation. *Cell Reports Medicine* **5**, 101373 (2024).

5. J. Cao, *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

6. Y. Hao, *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

7. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).

8. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

9. C. Ahlmann-Eltze, W. Huber, glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* **36**, 5701–5702 (2021).

10. A. Zhu, J. G. Ibrahim, M. I. Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

11. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

12. I. Dolgalev, msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.5.1.9001. (2022). Available at: https://igordot.github.io/msigdbr/ [Accessed 12 May 2024].

13. A. Subramanian, *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).

14. S. Jin, *et al.*, Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

15. E. Cano-Gamez, *et al.*, Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat Commun* **11**, 1801 (2020).

16. X. Qiu, *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979–982 (2017).

17. C. Trapnell, *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).

18. S. Picelli, *et al.*, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

19. M. R. Corces, *et al.*, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).

20. F. Krueger, *et al.*, FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. (2023). https://doi.org/10.5281/zenodo.7598955. Deposited 2 February 2023.

21. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

22. F. J. Martin, *et al.*, Ensembl 2023. *Nucleic Acids Research* **51**, D933–D941 (2023).

23. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

24. B. Langmead, C. Wilks, V. Antonescu, R. Charles, Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).

25. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).

26. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

27. C. J. Luckey, *et al.*, Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. *Proceedings of the National Academy of Sciences* **103**, 3304–3309 (2006).

28. B. Bengsch, *et al.*, Epigenomic-Guided Mass Cytometry Profiling Reveals Disease-Specific Features of Exhausted CD8 T Cells. *Immunity* **48**, 1029-1045.e5 (2018).

29. A. Frankish, *et al.*, GENCODE 2021. *Nucleic Acids Res* **49**, D916–D923 (2021).

30. J. A. Castro-Mondragon, *et al.*, JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **50**, D165–D173 (2022).

31. Y. Zhang, *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

32. J. D. Rubin, *et al.*, Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment. *Commun Biol* **4**, 1–15 (2021).

33. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

34. K. B. Yates, *et al.*, Epigenetic scars of CD8+ T cell exhaustion persist after cure of chronic infection in humans. *Nat Immunol* **22**, 1020–1029 (2021).

35. D. Machlab, *et al.*, monaLisa: an R/Bioconductor package for identifying regulatory motifs. *Bioinformatics* **38**, 2624–2625 (2022).

36. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–W191 (2014).

37. J. T. Robinson, *et al.*, Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).