nature medicine

LETTERS

https://doi.org/10.1038/s41591-020-0752-4

An immune-cell signature of bacterial sepsis

Miguel Reyes[1,2,7], Michael R. Filbin[1,3,7], Roby P. Bhattacharyya[1,4], Kianna Billman[1], Thomas Eisenhaure[1], Deborah T. Hung[1,5], Bruce D. Levy[5], Rebecca M. Baron[5], Paul C. Blainey[1,2*], Marcia B. Goldberg[1,4*] and Nir Hacohen[1,6*]

Dysregulation of the immune response to bacterial infection can lead to sepsis, a condition with high mortality. Multiple whole-blood gene-expression studies have defined sepsis-associated molecular signatures, but have not resolved changes in transcriptional states of specific cell types. Here, we used single-cell RNA-sequencing to profile the blood of people with sepsis ($n=29$) across three clinical cohorts with corresponding controls ($n=36$). We profiled total peripheral blood mononuclear cells (PBMCs, 106,545 cells) and dendritic cells (19,806 cells) across all subjects and, on the basis of clustering of their gene-expression profiles, defined 16 immune-cell states. We identified a unique CD14$^+$ monocyte state that is expanded in people with sepsis and validated its power in distinguishing these individuals from controls using public transcriptomic data from subjects with different disease etiologies and from multiple geographic locations (18 cohorts, $n=1,467$ subjects). We identified a panel of surface markers for isolation and quantification of the monocyte state and characterized its epigenomic and functional phenotypes, and propose a model for its induction from human bone marrow. This study demonstrates the utility of single-cell genomics in discovering disease-associated cytologic signatures and provides insight into the cellular basis of immune dysregulation in bacterial sepsis.

Sepsis is a prevalent disease with high mortality that contributes to a large fraction of healthcare spending worldwide[1]. To date, no diagnostic biomarker nor targeted therapeutic agent for sepsis has proven useful or effective. This is probably attributable to substantial heterogeneity of disease due to multiple potential pathogens, sites of infection, individualized host immune responses and manifestations of organ dysfunction[2–4]. Equally, there is limited insight into the cellular and molecular basis of sepsis-induced systemic immune dysregulation[5–8]. Prior host gene-expression profiling studies relied on whole blood to characterize diagnostic or prognostic gene signatures[9–12], an approach that aggregates transcriptomic signals from many different cell types, but may not detect signatures from rarer cells and does not identify cell-type-specific disease signatures[13]. To overcome these limitations, we characterized the spectrum of immune cell states in the blood of people with sepsis using single-cell-resolved gene-expression profiling.

scRNA-seq defines immune cell states in sepsis patients across multiple clinical cohorts

We performed single-cell RNA sequencing (scRNA-seq) on PBMCs from people with sepsis and controls to define the range of cell states present in these subjects, to identify differences in cell-state composition between groups and to detect immune signatures that distinguish sepsis from the normal immune response to bacterial infection (Fig. 1). Our primary cohorts targeted subjects with urinary-tract infection (UTI) early in their disease course, within 12 h of presentation to the emergency department (ED) (Fig. 1b–e and Supplementary Table 1). UTI was selected to minimize heterogeneity introduced by different infectious sites and to maximize diagnostic clarity because a UTI can be reliably confirmed post hoc using a urine culture. We included subjects with UTI (clinical urinalysis with >20 white blood cells per high-power field) as the primary infection both with and without signs of sepsis, and subsequently adjudicated the enrolled subjects into UTI with leukocytosis (blood WBC ≥ 12,000 per mm^3) but no organ dysfunction (Leuk-UTI), UTI with mild or transient organ dysfunction (Int-URO) and UTI with clear or persistent organ dysfunction (Urosepsis, URO) (Methods); subjects with simple UTI without leukocytosis or signs of organ dysfunction were not enrolled. Our schema distinguishes transient versus sustained sepsis-related organ dysfunction, although both meet established criteria (Sepsis-2 criteria) for sepsis[14].

We also profiled subjects from two secondary cohorts from a different hospital: bacteremic individuals with sepsis in hospital wards (Bac-SEP) and those admitted to the medical intensive care unit (ICU) either with sepsis (ICU-SEP) or without sepsis (ICU-NoSEP). Inclusion criteria were the same for primary and secondary cohorts. These secondary cohorts included people later in their disease course, who enrolled at least 24 h after initial hospital presentation and receipt of intravenous antibiotics. For comparison, we analyzed specimens from uninfected, healthy controls (Control). Our multicohort approach, spanning two hospitals and several clinical phenotypes, supports the generalizability of our results across different clinical contexts.

We profiled total CD45$^+$ PBMCs (1,000–1,500 cells per subject) and LIN$^-$CD14$^-$HLA-DR$^+$ dendritic cells (300–500 cells per subject) using a 3′ tag RNA-seq approach. We multiplexed 6–8 samples per experiment using cell hashing, and observed no major batch effects in our data (Extended Data Fig. 1 and Methods). We identified immune-cell states by clustering the cells in two steps: low-resolution clustering to identify the major immune-cell types (Fig. 1f and Extended Data Figs. 2a,b and 3), then subclustering each major cell type separately in a robust manner (Extended Data Fig. 2c-d and Methods). This approach identified 16 cell states that are found across numerous subjects ($n=31–69$ per state) in

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Emergency Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [4]Center for Bacterial Pathogenesis, Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [5]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [6]Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [7]These authors contributed equally: Miguel Reyes, Michael R. Filbin. *e-mail: pblainey@broadinstitute.org; marcia.goldberg@mgh.harvard.edu; nhacohen@broadinstitute.org

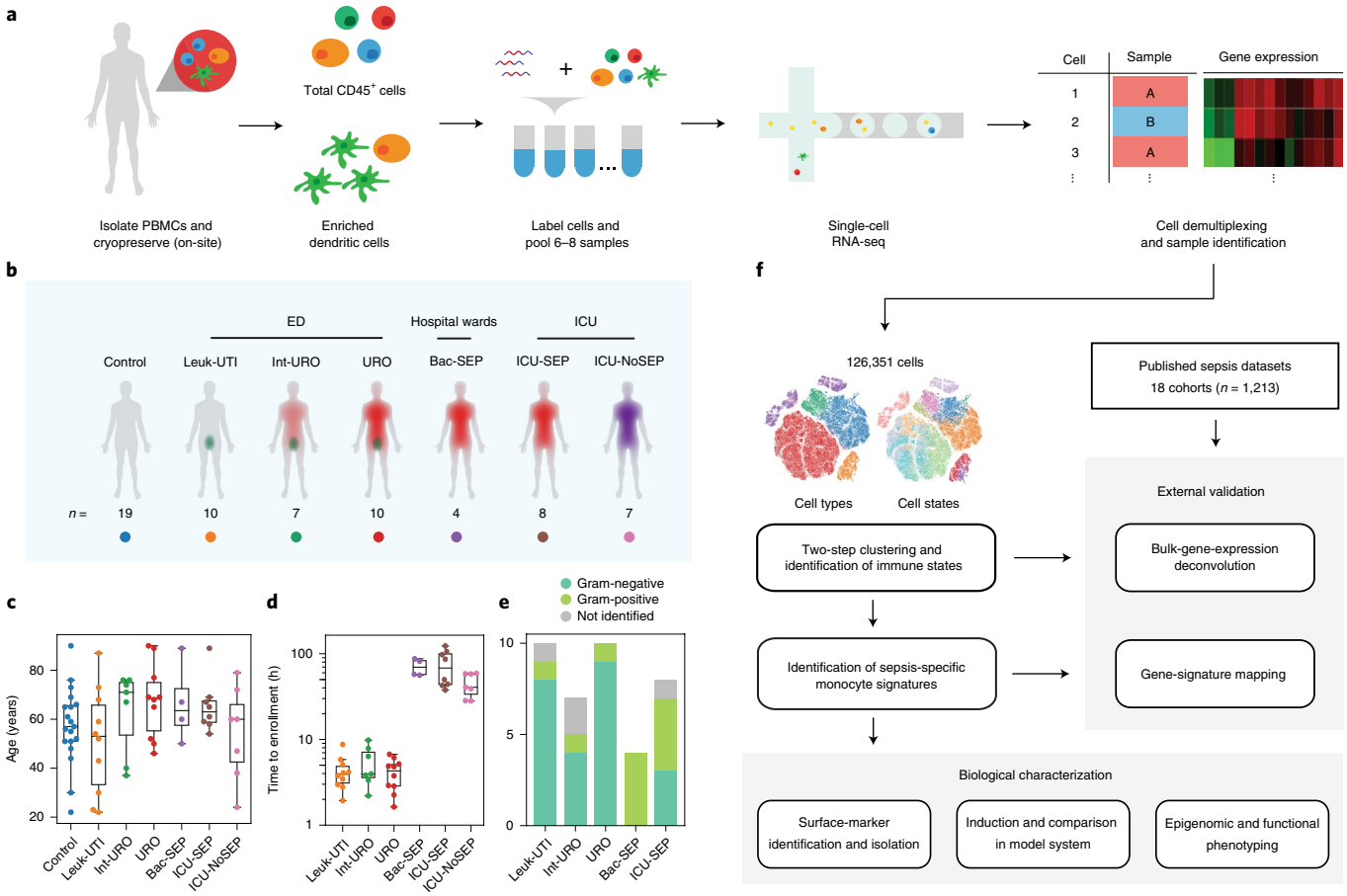


Fig. 1 | Cohort definition and analysis strategy. a, Processing pipeline for blood samples used in this study. Total CD45+ PBMCs and enriched dendritic cells for subject groups were labeled with cell hashing antibodies and loaded on a droplet-based scRNA-seq platform. Cells were demultiplexed and multiplets were removed on the basis of calls for each barcoding antibody. **b**, Schematic and number of subjects for each cohort profiled in this study. **c**, Age distribution of subjects and controls analyzed in this study. **d**, Time to enrollment from hospital presentation for each subject across all cohorts. Boxes show the mean and interquartile range (IQR) for each cohort, with whiskers extending to 1.5× the IQR in either direction from the top or bottom quartile. **e**, Bar plots showing fractions of Gram-positive and Gram-negative pathogens for each cohort. **f**, Analysis pipeline. Cell states were identified via two-step clustering, and fractional abundances thereof were compared to find sepsis-specific states. Further signatures were derived from these states using differential gene-expression and gene-module analysis. These signatures were validated in external sepsis datasets via a combination of bulk-gene-expression deconvolution, direct mapping of gene signatures and meta-analysis. Experiments were performed to identify surface markers, develop a model system for induction, analyze the epigenomic profile and characterize the functional phenotype of the identified cell state.

different cohorts and processing batches (Fig. 2a and Extended Data Fig. 2e,f). Among these are transcriptional states of T, B, natural killer (NK) and dendritic cells[15–17], and importantly, four monocyte states (Extended Data Figs. 4 and 5). We found four distinct monocyte groups: MS1, CD14+cells characterized by high expression of resistin (*RETN*), arachidonate 5-lipoxygenase activating protein (*ALOX5AP*) and interleukin-1 receptor type 2 (*IL1R2*) (Fig. 2b); MS2, characterized by high expression of class II major histocompatibility complex (MHC); MS3, similar to non-classical CD16^{hi} monocytes; and MS4, which is composed of the remaining CD14+ cells that express low levels of both class II MHC and inflammatory cytokines. We noted that some marker genes that characterized the MS1 state (Supplementary Table 2) had been previously associated with sepsis in studies measuring either serum protein or whole-blood messenger RNA levels[18–21].

Expansion of a monocyte state, MS1, in the blood of subjects with sepsis

After defining these clusters using data from all study subjects, we analyzed the differences in abundances of these cell states across different subject phenotypes (Fig. 1f). We found that the fractional abundances of cell states in the blood are strongly associated with the disease status of an individual (Extended Data Fig. 6a,b), whereas absolute abundances are less so (Extended Data Fig. 6c). Whereas the fractions of classical cell types vary substantially among the Control, Leuk-UTI and sepsis (Int-URO, URO, Bac-SEP and ICU-SEP) cohorts, we found more pronounced differences in the relative abundances of particular cell states derived from our clustering, most notably in MS1 (Fig. 2c). MS1 cells constitute a significantly larger fraction of CD45+cells in Int-URO and URO subjects than in Control or Leuk-UTI subjects (FDR < 0.001), and are also enriched in septic subjects in our secondary cohorts (Bac-SEP and ICU-SEP versus Control, FDR < 0.001). Further, MS1 cells are present at a slightly higher fraction in septic subjects (Int-URO, URO, Bac-SEP and ICU-SEP) than severely ill people without infection (ICU-NoSEP, FDR = 0.27).

Given the expansion of MS1 in people with sepsis, we reasoned that analysis of gene-expression signatures within MS1 cells may reveal useful clinical markers for sepsis and further insight into biological mechanisms. We looked for signatures that discriminate sepsis from critical illness without bacterial infection because these cohorts were not significantly distinguished by cell-state

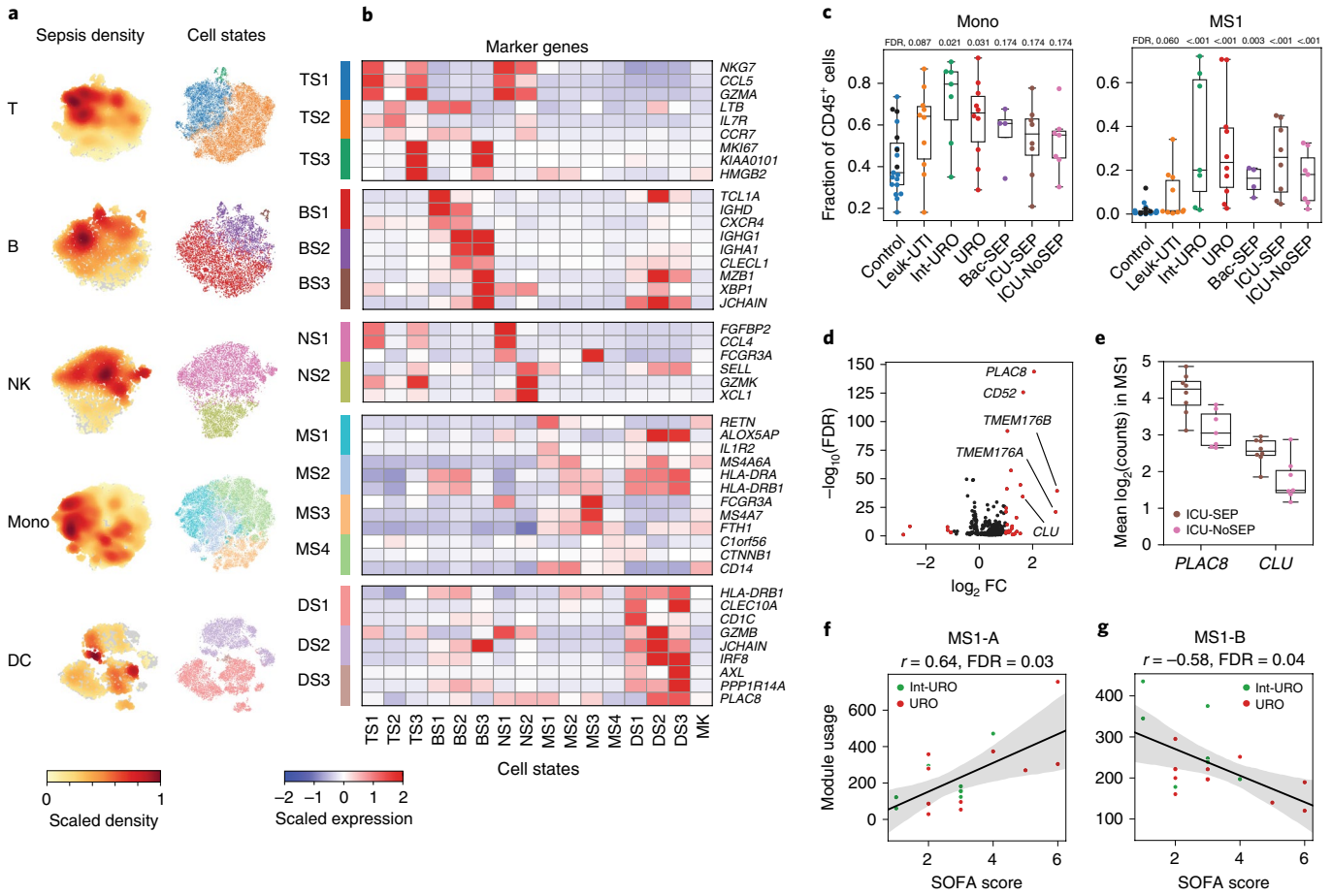


Fig. 2 | scRNA-seq identifies sepsis-specific immune-cell states and gene signatures. a, *t*-SNE plots for each cell type ($n = 32{,}341$, 7,970, 9,390, 58,557 and 14,299 cells for T, B, NK, monocyte (mono) and dendritic (DC) cells, respectively), colored by embedding density of cells from subjects with sepsis (Int-URO, URO, Bac-SEP and ICU-SEP; left) and cell state (right). **b**, Select marker genes that are differentially expressed (false-discovery rate (FDR) < 0.05, two-tailed Wilcoxon rank-sum test) in each cell state, when compared with other cell states within the same cell type. Color scale corresponds to *z*-scored, log-transformed mean gene-expression counts for each cell state. TS, T cell states; BS, B cell states; NS, NK cell states; MS, monocyte states; DS, dendritic cell states; MK, megakaryocytes. **c**, Fraction of total CD45$^+$ cells across each subject type for total monocytes (left) and MS1 cells (right). In the Control group, points for healthy controls that were follow-up samples from enrolled Leuk-UTI, Int-URO and URO subjects are indicated as black symbols, and those for matched healthy control samples from an outside source are indicated as aqua symbols. FDR values are shown when comparing each disease state with healthy controls (two-tailed Wilcoxon rank-sum test, corrected for testing of multiple states). Boxes show the median and IQR for each patient cohort, with whiskers extending to 1.5× the IQR in either direction from the top or bottom quartile. Sample size (n) for each cohort is indicated in Fig. 1b. **d**, Volcano plot showing results from differential expression analysis (two-sided Wilcoxon rank-sum test) between MS1 cells from ICU-SEP and MS1 cells from ICU-NoSEP subjects. Genes with $\log_2(FC) > 1$ are highlighted in red, and the top five genes with the highest positive fold changes are labeled. $n = 2{,}153$ and 1,442 cells from the 8 ICU-SEP and 7 ICU-NoSEP subjects, respectively. **e**, Box and swarm plots showing the mean expression ($\log_2$(unique molecular identifier (UMI) counts) of *PLAC8* and *CLU* in MS1 cells for each subject from the ICU-SEP and ICU-NoSEP cohorts. Boxes show the median and IQR for each patient cohort, with whiskers extending to 1.5× the IQR in either direction from the top or bottom quartile. **f**,**g**, Scatterplots showing correlation between mean gene-module usage in MS1 cells and sequential organ-failure assessment (SOFA) scores for Int-URO and URO subjects. Line and shadow indicate linear regression fit and 95% confidence interval, respectively. Significance of the correlations (Pearson *r*) were calculated with a two-sided permutation test, corrected for testing of multiple modules.

abundance alone. Thus, we identified genes differentially expressed in MS1 cells from ICU-SEP versus ICU-NoSEP subjects (Fig. 2d) and found two genes, placenta-associated 8 (*PLAC8*) and clusterin (*CLU*), that distinguish these two populations of subjects (Fig. 2e and Extended Data Fig. 7c). Whereas *PLAC8* expression has been associated with sepsis in studies analyzing the bulk expression of blood cells[22], *CLU* expression has not, perhaps owing to its specific upregulation in MS1 cells.

We analyzed co-varying genes among MS1 cells using non-negative matrix factorization and found five gene modules that are detected in more than half of our subjects with sepsis (Extended Data Fig. 7d–f and Supplementary Table 3). Of note, the module in MS1 cells corresponding to mitochondrial respiration (MS1-A; *MT-ND4*, *MT-CO3*, *MT-ATP6*) correlated significantly with disease severity in subjects with sepsis from our primary cohort (Int-URO and URO, FDR = 0.03; Fig. 2f), supporting the link between alterations in energy metabolism and immunoparalysis in sepsis[23]. In addition, a module of genes in MS1 related to anti-inflammatory and pro-resolving responses (MS1-B; *S100A8, RETN, ALOX5AP, FPR2*)[24–26] correlates negatively with severity (FDR = 0.04) (Fig. 2g and Extended Data Fig. 7g), consistent with a current model of sepsis wherein people early in their disease course have a heightened inflammatory state, but subsequently switch to an immunosuppressive state[27].

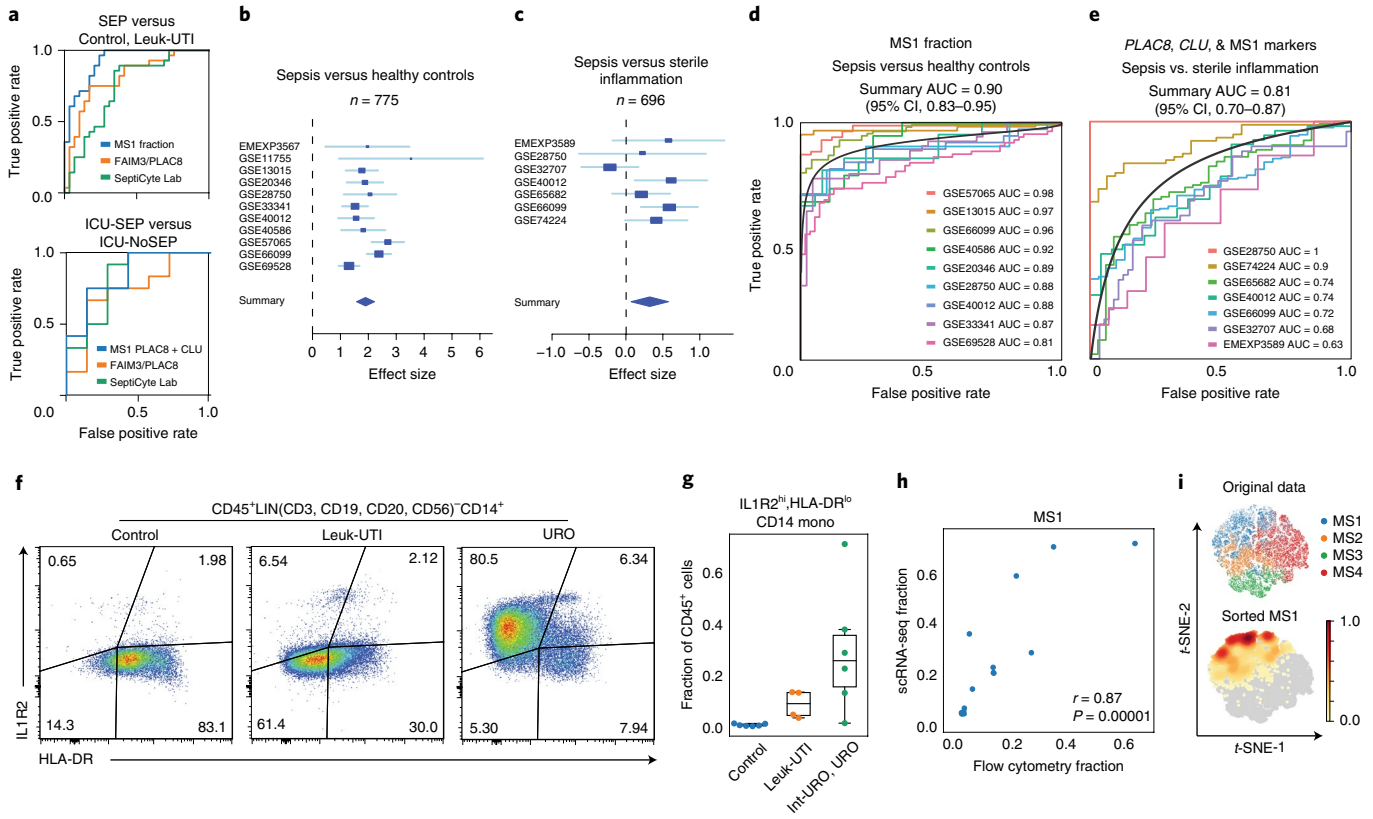


Fig. 3 | Analysis of the MS1 cell state as a sepsis marker. a, Receiver operating characteristic (ROC) curve for subject classification based on MS1 abundance (top) or mean PLAC8 and CLU expression in MS1 cells (bottom), and gene-expression-score-based classifiers (FAIM3/PLAC8, SeptiCyte Lab). MS1 is taken as the fraction of total CD45$^+$cells per subject, as defined by scRNA-seq. Gene-set scores were calculated, as detailed in each corresponding publication, on the pseudo-bulk gene-expression matrix obtained by summing read counts from all cells of each subject. SEP indicates all subjects with sepsis analyzed in this study (Int-URO, URO, Bac-SEP, ICU-SEP). **b**,**c**, Forest plots showing the effect size ($\log_2$(standardized mean difference between indicated patient phenotypes)) of inferred MS1 abundance in each dataset from bulk-gene-expression deconvolution. Accession numbers of the data from each study are listed on the left. Boxes indicate the effect size in an individual study, with whiskers extending to the 95% confidence interval. Size of the box is proportional to the relative sample size of the study. Diamonds represent the summary effect size among the subject groups, determined by integrating the standardized mean differences across all studies. The width of the diamond corresponds to its 95% confidence interval. **d**, Individual ROC curves for sepsis versus uninfected healthy controls; analysis includes each study in **b** for which the number of sepsis subjects and controls were both greater than 5 ($n = 751$ total subjects from 9 cohorts). **e**, ROC curves for classifying sepsis versus sterile inflammation ($n = 696$ total subjects from 7 cohorts) on the basis of the mean expression of *PLAC8*, *CLU* and the top 6 MS1 marker genes (*RETN*, *CD63*, *ALOX5AP*, *SEC61G*, *TXN and MT1X*). Black curves in **d** and **e** indicate the summary ROCs. **f**, Flow cytometry density plots of LIN$^-$CD14$^+$ monocytes (where LIN$^-$ cells are those negative for the indicated lineage markers) gated on surface expression of IL1R2 and HLA-DR. Percentage of the population over total CD14$^+$ monocytes in each quadrant is indicated. Each density plot shows PBMCs from a single subject analyzed in one experiment. **g**, Fractional abundance of CD14$^+$HLA-DR^{lo}IL1R2^{hi} monocytes by flow cytometry in Control, Leuk-UTI, Int-URO and URO subjects ($n = 6$, 4 and 6, respectively). Samples used for this analysis were from the primary cohort (Control, Leuk-UTI, Int-URO, URO). Boxes show the median and IQR for each patient cohort, with whiskers extending to 1.5× the IQR in either direction from the top or bottom quartile. **h**, Correlation of MS1 fractions defined by scRNA-seq (*y* axis) and CD14$^+$HLA-DR^{lo}IL1R2^{hi} monocyte fractions of CD45$^+$ cells (*x* axis) from $n = 4$ Leuk-UTI and $n = 6$ URO subjects from **g**. Significance of the correlation (Pearson *r*) was calculated with a two-sided permutation test. **i**, scRNA-seq of sorted CD14$^+$HLA-DR^{lo}IL1R2^{hi}, monocytes and original MS1 cells visualized with *t*-SNE projection. Top scatterplot ($n = 15{,}021$ cells) shows original classification of cells from the cohorts, and the bottom shows scaled embedding density of sorted cells ($n = 7{,}098$ cells) in the same projection.

Validation of MS1 signatures as markers for sepsis

To compare the performance of our signatures against previously reported classifiers, we quantified the classification accuracy of the MS1 fraction, PLAC8 + CLU expression in MS1 cells, and published gene-based signatures[22,28] in our cohort (Fig. 3a). When classifying all individuals with sepsis (Int-URO, URO, Bac-SEP and ICU-SEP) against Control and Leuk-UTI subjects, the MS1 fraction outperformed two published gene-set signatures (area under the curve (AUC), MS1 fraction = 0.92, FAIM3/PLAC8 ratio = 0.81 and SeptiCyte Lab = 0.74). In addition, PLAC8 + CLU expression in MS1 cells had higher classification accuracy when comparing ICU-SEP with ICU-NoSEP subjects (AUC, MS1 PLAC8 + CLU = 0.85, FAIM3/PLAC8 = 0.74 and SeptiCyte Lab = 0.82). These external gene signatures were derived from whole-blood profiling in varying clinical contexts, which could affect their performance when applied to our PBMC-derived expression data. In addition, the performance of MS1 PLAC8 + CLU may be inflated when applied to a subset of subjects from which MS1 was derived. Nevertheless, our approach provides biological context for these previously derived signature genes, as their expression in our data was specific to certain cell states (Extended Data Fig. 8).

To validate our signatures in external datasets, we analyzed independent cohorts of subjects with bacterial sepsis from published bulk-expression studies of sepsis. First, we validated the use of bulk-gene-expression deconvolution[29,30] on our data to infer the relative fraction of MS1 cells and cells in other states in the

blood (Extended Data Fig. 9a–c and Methods). Upon extending this approach to bulk transcriptional data from 11 sepsis cohorts included in a recent meta-analysis[4], we found the inferred abundance of the MS1 state to be higher in people with sepsis than in controls in each study, with a summary effect size of 1.9 across all cohorts (FDR = 1.75×10^{-30}, Fig. 3b and Supplementary Table 4). Furthermore, the inferred MS1 fraction alone for each subject can be used as a classifier for sepsis in the same datasets, with a summary AUC of 0.90 (range of 0.81–0.98) across all studies (Fig. 3d), performing similarly to reported classifiers that were derived from bulk gene-expression signatures[9] (Extended Data Fig. 9e). In a similar analysis of 7 datasets comparing people with sepsis with ICU controls (people with non-infectious systemic inflammatory response syndrome)[9] (Fig. 3c), MS1 is expanded in sepsis, albeit with a lower but notable effect size of 0.32 (FDR = 0.08), consistent with observations in our own cohorts. Whereas the MS1 fraction alone cannot be used as a sepsis classifier in this context, analyzing the co-expression of *PLAC8*, *CLU* and MS1 marker genes (*RETN*, *CD63*, *ALOX5AP*, *SEC61G*, *TXN* and *MT1X*) in these datasets performs well in classifying subjects with sepsis against sterile inflammation (Fig. 3e), with a summary AUC of 0.81 (range of 0.63–1.00), performing similarly to published signatures (Extended Data Fig. 9d,f). This analysis of published transcriptional data implies that MS1 cells are present in people with sepsis across several geographic locations, genetic backgrounds and clinical contexts, and demonstrates the potential utility of MS1-specific gene signatures for the discrimination of sepsis from sterile inflammation.

Surface markers for isolation of MS1 cells

To improve its utility as a cytologic marker, we identified a panel of surface proteins that can be used to define the MS1 cell state by flow cytometry. Among the differentially expressed genes that distinguish it from other CD14$^+$ monocytes, low HLA-DR and high IL1R2 expression can be used to quantify the fraction of MS1 cells (Fig. 3f). Previous studies[31,32] have shown that CD14$^+$ monocytes from people with sepsis have decreased HLA-DR expression; however, we found that monocytes from Leuk-UTI subjects also have this phenotype, signifying that decreased HLA-DR expression alone is insufficient to distinguish patients with sepsis from those with uncomplicated infection. By contrast, HLA-DR^{lo}IL1R2^{hi}CD14$^+$ monocytes were at higher frequencies in Int-URO and URO subjects than in Control or Leuk-UTI subjects (Fig. 3g), and their fractions measured by flow cytometry correlated significantly with fractions determined by scRNA-seq (Pearson $r = 0.87$) (Fig. 3h). Cells sorted with this phenotype (7,098 cells from 5 URO subjects) co-localize by expression profile with MS1 cells in our original dataset when analyzed together and projected on the same *t*-distributed stochastic neighbor embedding (*t*-SNE) plot (Fig. 3i). This combination of cell surface markers could be used to purify the cell state for further molecular and functional characterization, or could potentially be employed as a routine tool for rapid quantification of the MS1 fraction in people at risk of sepsis.

Generation of MS1-like cells from human bone marrow

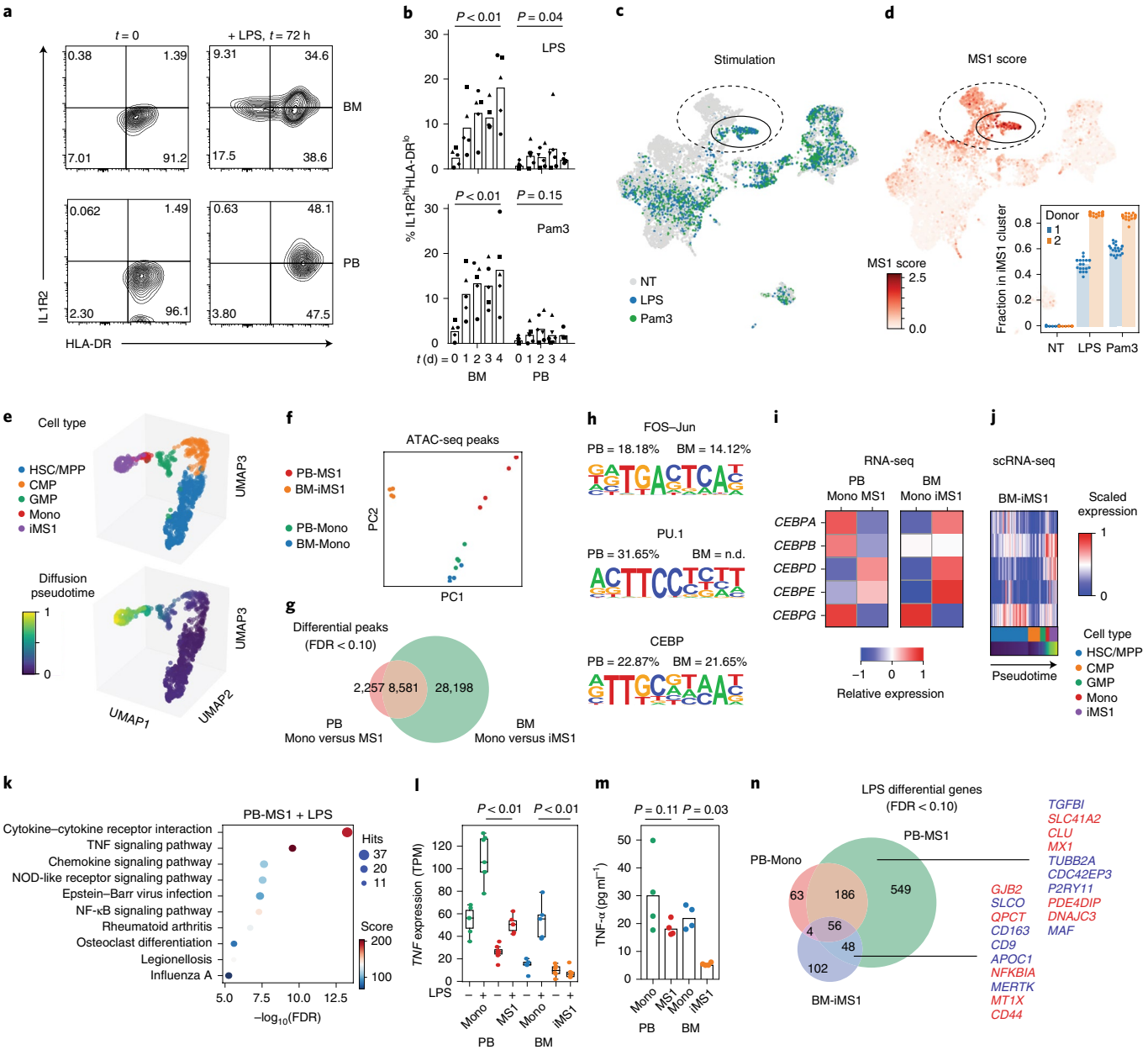
Low HLA-DR expression is associated with monocyte immaturity, resulting in decreased responsiveness to stimuli[33,34]. We hypothesized that MS1 cells might be derived from bone marrow mononuclear cells (BMMCs), which include hematopoietic precursors, rather than from mature immune cells in peripheral blood. We found that chronic stimulation of BMMCs with Pam3CSK4 (Pam3) or lipopolysaccharide (LPS) results in the emergence of a HLA-DR^{lo}IL1R2^{hi}CD14$^+$ population (Fig. 4a). The abundance of this population, as a fraction of total CD14$^+$ cells, increased significantly over time in treated BMMCs, but not in treated PBMCs (Fig. 4b). Furthermore, scRNA-seq profiling of BMMCs treated with LPS or Pam3CSK4 revealed a cluster of cells scoring highly for MS1 signature genes that are absent in the untreated condition (Fig. 4c,d and Extended Data Fig. 10a–e). Trajectory analysis of the myeloid populations suggests that the MS1-like induced population (iMS1, Leiden cluster 14) proceeds initially through a differentiation pathway

Fig. 4 | Induction and characterization of MS1 monocytes. a, Flow cytometry contour plots showing IL1R2 and HLA-DR of cells gated on the CD14$^+$ fraction from either bone marrow (BM, top row) or peripheral blood (PB, bottom row) mononuclear cells. Cells are either freshly thawed (left column) or stimulated with 100 ng ml^{-1} LPS for 3 d in hematopoietic stem cell (HSC) cytokine-rich medium (right column). Each density plot shows cells from a single donor analyzed in one experiment. **b**, Fractional abundance of HLA-DR^{lo}IL1R2^{hi} cells among CD14$^+$ monocytes in PB or BM mononuclear cells stimulated with either 100 ng ml^{-1} LPS (top) or Pam3CSK4 (bottom) over time (0 to 4 d). Different symbols indicate cells obtained from different healthy donors. *P* values are calculated from a two-sided Wilcoxon rank-sum test between day 0 and day 4. **c**,**d**, scRNA-seq of BM mononuclear cells ($n = 8{,}702$ cells) incubated in HSC cytokine-rich medium with no treatment or 100 ng ml^{-1} LPS or Pam3CSK4 for 4 d. Cells are visualized on uniform manifold approximation and projection (UMAP) plots colored by treatment (**c**) or MS1 score (**d**). MS1 scores are given as the ratio of the average expression of the top 15 MS1 marker genes to the average expression of a randomly sampled set of 50 reference genes. In each plot, the cluster with the highest MS1 score is circled. Dotted circles indicate monocyte clusters. Inset in **d** shows the mean fractional abundance of the iMS1 cluster among monocytes across each donor and treatment condition; each individual point is calculated by randomly sampling the data and clustering the subsampled dataset ($n = 20$ iterations). **e**, UMAP projections ($n = 901$ cells) of Pam3CSK4 and LPS-stimulated BM myeloid and progenitor cells (HSC/MPP, CMP, GMP, Mono and iMS1) colored by cell type (top) and diffusion pseudotime (bottom). **f**, Principal-component analysis (PCA) plots of assay for transposase-accessible chromatin using sequencing (ATAC-seq) peak accessibility profiles for four different sorted monocyte populations: PB-Mono, PB-MS1, BM-Mono (CD14$^+$ monocytes from freshly thawed BM cells) and BM-iMS1 (CD14$^+$ monocytes from BM cells stimulated for 4 d with 100 ng ml^{-1} LPS in HSC cytokine-rich medium). Experiments were performed on two donors with two technical replicates each. **g**, Venn diagram showing overlap of differentially accessible peaks (FDR < 0.1, edgeR exact test) from monocyte populations in PB and BM. **h**, Sequence logos showing the top three enriched motifs in the differentially accessible peaks when comparing PB-Mono and PB-MS1. Percentages indicate the number of differential peaks that contain the motif for PB and BM (n.d. indicates that motif was not detected in the enrichment analysis). **i**, Relative expression (normalized $\log_2$(transcripts per kilobase million (TPM))) of the CEBP family of transcription factors across the four monocyte populations. **j**, Scaled expression (normalized log counts) of the CEBP family of transcription factors along the pseudotime trajectory in **e**. **k**, Top 10 enriched pathways in the differentially accessible genes (FDR < 0.1, edgeR exact test) when PB-MS1 cells were rested for 24 h and subsequently stimulated with 100 ng ml^{-1} LPS. RNA-seq experiments were performed on 2 donors, with 3 technical replicates each. Sizes of circles are proportional to the number of gene hits in a set, whereas color represents the enrichment score of each gene set. **l**,**m**, *TNF* expression (**l**) and TNF-α protein levels (**m**) in the supernatant of the indicated four sorted monocyte populations after LPS stimulation. *P* values are calculated from a two-sided Wilcoxon rank-sum test between LPS-stimulated Mono and iMS1 cells. Protein measurements were performed on two donors with two technical replicates each. **n**, Venn diagram showing overlap of differentially accessible genes (FDR < 0.1, edgeR exact test) from the indicated sorted monocyte populations after LPS stimulation. Top ten genes with highest significance are indicated in red for the PB-MS1-exclusive set of genes and the overlap between BM-iMS1 and PB-MS1. HSC/MPP, hematopoietic stem cells and multipotent progenitors; CMP, common myeloid progenitors; GMP, granulocyte-macrophage progenitor.

similar to that of cells from the non-stimulated condition, but that it subsequently deviates from this fate (Fig. 4e and Extended Data Fig. 10f,g). Progenitor populations in the stimulated condition display several differentially expressed genes (Extended Data Fig. 10h); stimulated progenitor cells upregulate several receptors previously associated with inflammation-induced myelopoiesis (for example, *IL3R, IL10R, IFNAR1*)[35], suggesting that an MS1-like population may emerge in the bloodstream as a result of sepsis-induced myelopoiesis. These results demonstrate the potential of human bone-marrow cells as a model for the expansion of the MS1 state in sepsis, and support the hypothesis that the emergence of reprogrammed myeloid cells in the blood stems from dysregulated differentiation of hematopoietic precursors[36,37].

Epigenomic landscape and transcriptional regulators of MS1 cells

We profiled the chromatin accessibility landscapes of monocytes from peripheral blood of healthy controls (PB-Mono), MS1 cells sorted from patients with sepsis (PB-MS1), monocytes from healthy bone marrow (BM-Mono), and monocytes from BMMCs stimulated with LPS and HSC cytokines (BM-iMS1). Principal component analysis of genome-wide ATAC-seq profiles of the four populations showed that PB-Mono and BM-Mono colocalize, whereas PB-MS1 and BM-iMS1 form distinct clusters yet share similar loadings on PC2 (Fig. 4f). Motif enrichment analysis on the differential peaks between PB-Mono and PB-MS1 demonstrates enrichment of the FOS–Jun, PU.1 and CEBP motifs, all of which are families of transcription factors critical to monocyte development[38] (Fig. 4g,h and Supplementary Table 5). Given their important role in inflammation-induced myelopoiesis[38,39], we analyzed the expression of the CEBP transcription factors. Bulk RNA-seq showed an increase in *CEBPD* and *CEBPE* and a decrease in *CEBPG* expression in PB-MS1 compared with PB-Mono, and similarly in BM-iMS1 compared with BM-Mono (Fig. 4i). Analysis of the differentiation trajectory of iMS1 cells from bone-marrow progenitors also showed an increase in *CEBPD* expression after the transition from a GMP state (Fig. 4j). Interestingly, *CEBPD* was among the top genes of the module comprised of transcription-related and housekeeping genes



from our analysis of MS1 cells from people with sepsis (MS1-C, Extended Data Fig. 7f and Supplementary Table 3), suggesting its potential importance in the maintenance of the MS1 program. Altogether, these analyses show that MS1 cells have an epigenomic profile markedly different from that of normal CD14$^+$ blood monocytes, and that these differences are associated with transcription factors involved in monocyte differentiation. Although in vitro-generated BM-iMS1 do not fully recapitulate the epigenomic landscape of MS1 cells, the two populations show significant overlap in accessible peaks and share the upregulation of similar transcriptional regulators.

Functional response of MS1 cells to restimulation

To compare the functional responses of MS1 cells to those of other CD14$^+$ monocytes, we sorted the four monocyte populations and stimulated the cells with 100 ng ml^{-1} LPS after resting for 24 h. As expected, LPS stimulation results in upregulation of genes related to cytokine secretion and activation of the nuclear factor-κB (NF-κB) signaling pathway (Fig. 4k and Supplementary Table 6). However, the magnitude of response is decreased in PB-MS1 cells relative to PB-Mono, and in BM-iMS1 relative to BM-Mono, as evidenced by lower basal and induced expression of the tumor necrosis factor (*TNF*) gene, and less secretion of the TNF-α protein (Fig. 4l–m). Analyzing the overlap in differentially expressed genes upon stimulation reveals a large number of genes that are uniquely upregulated in PB-MS1 (Fig. 4n). This includes *CLU*, one of the genes we propose as a marker for discriminating sepsis from noninfectious inflammation. BM-iMS1 also up-regulated a subset of the genes (12.4%) induced in PB-MS1. Of note, *NFKBIA*, a known inhibitor of inflammatory responses, was upregulated in both PB-MS1 and BM-iMS1, perhaps explaining the blunted response in both populations. This analysis demonstrates that MS1 cells from people with sepsis and those induced from human bone marrow both have a dysregulated response to further bacterial stimulation, recapitulating known phenotypes of monocytes in people with sepsis.

Discussion

Our work identified a specific monocyte state in the blood, MS1, as a signature of sepsis in people with bacterial infection in multiple cohorts. This state is also expanded in people in the ICU with sterile inflammation, but particular gene signatures within the MS1 cell state further distinguish these two populations. Associated surface markers can be used to detect and purify the cell state, enabling routine monitoring in settings where people are at risk for sepsis. Additional profiling of the granulocyte compartment, which we were not able to capture using the approach employed here, will be needed to comprehensively survey the host immune landscape in sepsis. Nevertheless, these findings highlight the utility of patient phenotyping based on underlying immune responses, and provide a new level of resolution for patient stratification and therapeutic development in sepsis.

By characterizing the epigenomic and functional phenotype of the MS1 state, our findings provide insight into its biology and support a model in which MS1-like cells can be induced from bone-marrow precursors. Whereas the α and β members of the CEBP family of transcription factors are associated with steady-state and sepsis-induced myelopoiesis[39,40], we identified a different paralog, CEBPδ, that may drive the MS1 state. These data suggest a new mechanism through which hematopoietic adaptation plays a role in the pathophysiology of sepsis, although analysis of hematopoietic precursors from people with sepsis is needed. In addition, studies in vitro and in model organisms are needed to test whether MS1 cells and their transcriptional regulators play a direct role in the induction of sepsis and to determine their potential utility as targets for therapeutic manipulation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-0752-4.

Received: 16 December 2019; Accepted: 3 January 2020;
Published online: 17 February 2020

References

1. Rudd, K. E. et al. The global burden of sepsis: barriers and potential solutions. *Crit. Care* **22**, 232 (2018).
2. Filbin, M. R. et al. Presenting symptoms independently predict mortality in septic shock: importance of a previously unmeasured confounder. *Crit. Care Med.* **46**, 1592–1599 (2018).
3. Seymour, C. W. et al.Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* **321**, 915–925 (2019).
4. Sweeney, T. E. et al. Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit. Care Med.* **46**, 915–925 (2018).
5. Opal, S. M., Dellinger, R. P., Vincent, J.-L., Masur, H. & Angus, D. C. The next generation of sepsis clinical trial designs: what is next after the demise of recombinant human activated protein C?* *Crit. Care Med.* **42**, 1714–1721 (2014).
6. Scicluna, B. P. & Baillie, J. K. The search for efficacious new therapies in sepsis needs to embrace heterogeneity. *Am. J. Respir. Crit. Care Med.* **15**, 936–938 (2018).
7. Prescott, H. C., Calfee, C. S., Thompson, B. T., Angus, D. C. & Liu, V. X. Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. *Am. J. Respir. Crit. Care Med.* **194**, 147–155 (2016).
8. Coopersmith, C. M. et al. Surviving sepsis campaign: research priorities for sepsis and septic shock. *Intensive Care Med.* **44**, 1400–1426 (2018).
9. Sweeney, T. E. & Khatri, P. Benchmarking sepsis gene expression diagnostics using public data. *Crit. Care Med.* **45**, 1–10 (2017).
10. Sweeney, T. E. et al. A community approach to mortality prediction in sepsis via gene expression analysis. *Nat. Commun.* **9**, 694 (2018).
11. Davenport, E. E. et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *Lancet Respir. Med.* **4**, 259–271 (2016).
12. Sweeney, T. E., Shidham, A., Wong, H. R. & Khatri, P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci. Transl. Med.* **7**, 287ra71 (2015).
13. Reyes, M. et al. Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signatures. *Sci. Adv.* **5**, eaau9223 (2019).
14. Levy, M. M. et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med.* **29**, 530–538 (2003).
15. Villani, A.-C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
16. Crinier, A. et al. High-dimensional single-cell analysis identifies organ-specific signatures and conserved NK cell subsets in humans and mice. *Immunity* **49**, 971–986 (2018).
17. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
18. Sundén-Cullberg, J. et al. Pronounced elevation of resistin correlates with severity of disease in severe sepsis and septic shock. *Crit. Care Med.* **35**, 1536–1542 (2007).
19. Lang, Y. et al. Interleukin-1 receptor 2: a new biomarker for sepsis diagnosis and gram-negative/gram-positive bacterial differentiation. *Shock* **47**, 119–124 (2017).
20. Schaack, D., Siegler, B. H., Tamulyte, S., Weigand, M. A. & Uhle, F. The immunosuppressive face of sepsis early on intensive care unit-A large-scale microarray meta-analysis. *PLoS One* **13**, e0198555 (2018).
21. Bauer, M. et al. A transcriptomic biomarker to quantify systemic inflammation in sepsis — a prospective multicenter phase II diagnostic study. *EBioMedicine* **6**, 114–125 (2016).
22. McHugh, L. et al. A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: discovery and validation in independent cohorts. *PLoS Med.* **12**, e1001916 (2015).
23. Cheng, S.-C. et al. Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis. *Nat. Immunol.* **17**, 406–413 (2016).
24. Ulas, T. et al. S100-alarmin-induced innate immune programming protects newborn infants from sepsis. *Nat. Immunol.* **18**, 622–632 (2017).

25. Jang, J. C. et al. Human resistin protects against endotoxic shock by blocking LPS–TLR4 interaction. *Proc. Natl. Acad. Sci. USA* **114**, E10399–E10408 (2017).
26. Orr, S. K. et al. Gene expression of proresolving lipid mediator pathways is associated with clinical outcomes in trauma patients. *Crit. Care Med.* **43**, 2642–2650 (2015).
27. Hotchkiss, R. S., Monneret, G. & Payen, D. Immunosuppression in sepsis: a novel understanding of the disorder and a new therapeutic approach. *Lancet Infect. Dis.* **13**, 260–268 (2013).
28. Scicluna, B. P. et al. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am. J. Respir. Crit. Care Med.* **192**, 826–835 (2015).
29. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
30. Roy Chowdhury, R. et al. A multi-cohort study of the immune factors associated with *M. tuberculosis* infection outcomes. *Nature* **560**, 644–648 (2018).
31. Gossez, M. et al. Proof of concept study of mass cytometry in septic shock patients reveals novel immune alterations. *Sci. Rep.* **8**, 17296 (2018).
32. Landelle, C. et al. Low monocyte human leukocyte antigen-DR is independently associated with nosocomial infections after septic shock. *Intensive Care Med.* **36**, 1859–1866 (2010).
33. Veglia, F., Perego, M. & Gabrilovich, D. Myeloid-derived suppressor cells coming of age. *Nat. Immunol.* **19**, 108–119 (2018).
34. Schrijver, I. T., Théroude, C. & Roger, T. Myeloid-derived suppressor cells in sepsis. *Front. Immunol.* **10**, 327 (2019).
35. Chavakis, T., Mitroulis, I. & Hajishengallis, G. Hematopoietic progenitor cells as integrative hubs for adaptation to and fine-tuning of inflammation. *Nat. Immunol.* **20**, 802–811 (2019).
36. Rodriguez, S. et al. Dysfunctional expansion of hematopoietic stem cells and block of myeloid differentiation in lethal sepsis. *Blood* **114**, 4064–4076 (2009).
37. Zhang, H. et al. Sepsis induces hematopoietic stem cell exhaustion and myelosuppression through distinct contributions of TRIF and MYD88. *Stem Cell Reports* **6**, 940–956 (2016).
38. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* **7**, 105–117 (2007).
39. McPeak, M. B. et al. Frontline Science: myeloid cell-specific deletion of Cebpb decreases sepsis-induced immunosuppression in mice. *J. Leukoc. Biol.* **102**, 191–200 (2017).
40. Manz, M. G. & Boettcher, S. Emergency granulopoiesis. *Nat. Rev. Immunol.* **14**, 302–314 (2014).

Methods

Study samples and clinical adjudication. Primary cohorts comprised subjects with UTI and urosepsis who presented to the ED at the Massachusetts General Hospital (MGH), and secondary cohorts were hospitalized subjects with and without sepsis on inpatient services at the Brigham and Women's Hospital (BWH); both hospitals are located in Boston, MA. Informed consent was obtained from subjects or their surrogates. Blood samples from these subjects and healthy controls were drawn with EDTA Vacutainer tubes (BD Biosciences) and processed within 3 h of collection. De-identified BMMC samples were purchased from AllCells or Stemcell Technologies. This study was approved by the Institutional Review Boards at the Broad Institute of MIT and Harvard and at Partners HealthCare, under protocols 2017P001681/PHS, 2008P000495/PHS and 2017P000367/PHS.

The primary cohorts were enrolled in the ED at the Massachusetts General Hospital (MGH) from December 2017 to November 2018. They consisted of people with UTI, defined by a urine white blood cell count of >20 per high-power field on clinical urinalysis. Study samples were collected within 12 h of subject arrival to the ED. Individuals with UTI were initially enroled into one of two categories: (1) those with leukocytosis (blood WBC ≥ 12,000 per mm^3) without another cause, indicating systemic inflammation from the UTI, but without organ dysfunction (cohort Leuk-UTI), and (2) those with organ dysfunction, which defines urosepsis. For the urosepsis group, we recruited subjects meeting UTI criteria in the presence of organ dysfunction, as specified in national quality measure definitions[41] that are adapted from Sepsis-2 consensus definitions[14], specifically systolic blood pressure < 90 mmHg, lactate > 2.0 mg dl^{-1}, requirement for vasopressor medication, new glascow coma score (GCS) < 15 denoting altered mental status, new creatinine > 2.0 mg dl^{-1} or need for mechanical ventilation. SOFA scores were calculated, but they were not a specific criterion for enrollment or adjudication.

Once the results of initial diagnostics sent in the course of routine clinical care, including cultures, were available and the subsequent clinical course during hospitalization was known (that is, at least 48 h after initial presentation), clinical adjudication of each enrolled subject was independently performed by three investigators (M.R.F, R.P.B, M.B.G.), blinded to research analysis outcomes. Each enrolled subject who was found to meet criteria for the study was adjudicated to one of three clinical categories: Leuk-UTI, Int-URO and URO. Given the spectrum of organ dysfunction severity among enrolled patients, we differentiated between mild or transient organ dysfunction (intermediate urosepsis, or Int-URO) and sustained infection-related organ dysfunction (urosepsis, or URO). Int-URO includes subjects with physiologic perturbations that qualify as sepsis in the setting of infection per national quality measure and Sepsis-2 consensus definitions[14,41], but for whom observed organ dysfunction was isolated and relatively mild, and resolved quickly with initial therapies; examples include hypotension that resolved with fluid resuscitation, isolated mild elevation in creatinine that normalized within 24 h or elevated initial lactate or alteration in mental status that improved within 4–6 h. URO includes subjects with organ dysfunction that persisted or worsened despite initial therapy; examples include refractory hypotension requiring vasopressor support, persistent renal dysfunction >24 h after enrollment, lactate increasing despite adequate volume resuscitation or multiple-organ-system dysfunction. Discrepancies in adjudication among the three clinicians were resolved as a group.

For the category Leuk-UTI, we specifically targeted enrollment of subjects with UTI with systemic response but without sepsis so as to provide the most appropriate comparison for our urosepsis cohorts, as a comparison with subjects with simple UTI without evidence of a systemic response might highlight host signature differences attributable to a systemic response to localized infection rather than being specific to sepsis. To obtain as pure an immune signature for infection as possible, we excluded those with immunodeficiencies, including HIV, concurrent immunomodulatory drug therapy (including prednisone or steroid equivalent, chemotherapy, or biologic immunomodulators), recipients of bone-marrow or solid-organ transplantation and individuals with autoimmune disease. Of note, two subjects in the Leuk-UTI cohort were asplenic. For all these primary cohorts (Leuk-UTI, Int-URO and URO), we excluded patients who had received their first intravenous antibiotic >12 h prior to enrollment. Of the 27 people enrolled in these cohorts, 7 were enrolled prior to antibiotic initiation, and 20 were enrolled within 7 h of antibiotic initiation, with the median time to enrollment from antibiotic initiation for all enrolled patients 1.1 h (IQR, 0.2–2.4 h).

Uninfected control samples for the primary cohorts were obtained from two sources. First, follow-up blood samples were obtained from four primary cohort patients at 2–3 months after index enrollment (2 Leuk-UTI and 2 URO subjects). For all other primary cohort subjects, uninfected control samples consisted of blood samples from age-, gender- and ethnicity-matched healthy controls obtained from Research Blood Components (Watertown, MA).

Secondary cohorts consisted of hospitalized subjects identified as having bacteremia and sepsis but not requiring ICU admission (Bac-SEP), subjects with sepsis requiring ICU care (ICU-SEP) and subjects in the ICU for conditions other than sepsis (ICU-NoSEP). These cohorts were enrolled in the Brigham and Women's Hospital (BWH) as part of the Registry of Critical Illness; the criteria for subject recruitment for this cohort are described elsewhere[42,43]. The Bac-SEP subjects were recruited between December 2017 and September 2018 from hospital inpatient floors (not ICU) and had positive blood cultures within 24 h of sample collection (excluding those blood cultures that grew coagulase-negative *Staphylococcus* species, which was considered likely to be a contaminant). The ICU-SEP and ICU-NoSEP subjects were enrolled in the BWH ICU between November 2017 and October 2018.

In contrast to the primary cohorts enrolled in the MGH ED, most subjects in the secondary cohorts were enrolled 2–3 d after initial presentation of disease and initiation of therapy, with all subjects enrolled > 24 h from hospital presentation; most had therefore received antibiotics for > 24 h prior to enrollment (median, 70 h for Bac-SEP, IQR 61–79 hrs; median, 49 h for ICU-SEP, IQR 44–65 h). The sources of infection for secondary-cohort subjects included pulmonary, urinary, intraabdominal and endovascular sites. To ensure consistency of adjudication among cohorts, secondary cohorts were adjudicated for the presence of sepsis by the three adjudicators who adjudicated the primary cohort (M.R.F, R.P.B, M.B.G.), employing the same methods used for the primary cohorts.

During the index illnesses and/or hospitalizations, there were no deaths among subjects in the Leuk-UTI, Int-URO, BAC-SEP and ICU-NoSEP cohorts, and there was one death in each group among subjects in the URO and ICU-SEP cohorts. Given the small numbers of deaths, we did not specifically analyze their potential significance.

Isolation and cryopreservation of PBMCs from whole blood. Cells were isolated from whole-blood samples using density-gradient centrifugation, as described in a previous study[13]. Briefly, whole blood was diluted 1:1 with 1× PBS, layered on top of Ficoll-Paque Plus (GE Healthcare), and centrifuged at 1,200*g* for 20 min. The PBMC layer was resuspended in 10 ml RPMI-1640 (Gibco), and centrifuged again at 300*g* for 10 min. The cells were counted, resuspended in Cryostor CS10 (StemCell Technologies) and aliquoted in 1.5-ml cryopreservation tubes at a concentration of 2×10^6 cells per milliliter. The tubes were kept at −80 °C overnight, then transferred to liquid nitrogen for long-term storage. The plasma layer from density gradient separation was also collected, aliquoted in 1-ml tubes and stored at −80 °C.

Staining, flow cytometry and dendritic-cell enrichment. Samples were processed in batches of six or eight for pooling in single-cell RNA sequencing runs. All cells were stained with a general panel: DAPI, CD3–APC (HIT3a), CD19–APC (HIB19), CD20–APC (2H7), CD56–APC (5.1H11), CD14–FITC (M5E2), CD16–AF700 (B73.1), CD45–PE-Cy7 (HI30) and HLA-DR–PE (L243) (BioLegend). At the same time, 1 µl of cell-hashing antibody (HTO) was added to each sample (BioLegend). Samples were run on a SH800 cell sorter (Sony) to obtain flow-cytometry data and sort both live $CD45^+$ cells and dendritic cells. For samples from subjects enrolled in the MGH ED, dendritic cells were enriched separately with a MACS human pan-DC enrichment kit (Miltenyi Biotec). For sorting MS1 cells, the following panel was used: DAPI, CD3–APC (HIT3a), CD19–APC (HIB19), CD20–APC (2H7), CD56–APC (5.1H11), CD14–FITC (M5E2), CD45–AF700 (HI30), HLA-DR–PE-Cy7 (L243) (BioLegend) and IL1R2–PE (34141, ThermoFisher Scientific).

Single-cell RNA-seq and analysis. Single-cell RNA-seq was performed on the Chromium platform, using the single cell expression 3′ v2 profiling chemistry (10X Genomics) combined with cell hashing[44]. HTO-labeled cells from 6–8 donors were pooled equally then washed twice with RPMI-1640 immediately before loading on the 10X controller. Complementary DNA amplification and library construction were conducted following the manufacturer's protocol, with additional steps for the amplification of HTO barcodes. Libraries were sequenced to a depth of ~50,000 reads per cell on a Novaseq S2 (Illumina). The data were aligned to the GRCh38 reference genome using cellranger v2.1 (10X Genomics), and the hashed cells were demultiplexed using the CITE-seq count tool (https://github.com/Hoohm/CITE-seq-Count).

Single-cell data analysis was performed using scanpy[45]. Count matrices from the cellranger output were preprocessed by filtering for cells and genes (minimum cells per gene, 10; minimum UMI per cell, 100). Before clustering, the full dataset or a subset thereof was filtered for highly variable genes (minimum mean, 0.0125 and dispersion , 0.5 per gene) and scaled. Clustering was performed on the top 50 principal components of the data using the Leiden algorithm with varying resolution. To quantify the robustness of each clustering solution, the data were subsampled without replacement (90% of cells, 20 iterations) and re-clustered, and an adjusted Rand index was then computed between the solutions for the original and subsampled data. The highest resolution at which the robustness began to decrease was chosen for further analysis. To ensure that no subject- or batch-specific clusters were included in the data, small clusters (<500 cells) were combined with the next closest cluster on the basis of their similarity in gene-expression profiles. Differentially expressed genes were determined for each state by a Wilcoxon rank-sum test, with an FDR cutoff of 0.01. For visualization, *t*-SNE projections were computed on the top 10 principal components of the dataset or subsets thereof. To specifically find genes that distinguish between ICU-SEP and ICU-NoSEP populations, differentially expressed genes were filtered for those that have an in-group fraction >0.4 and out-group fraction <0.6. Consensus non-negative matrix factorization analysis was performed as detailed in a previous publication[46]. To ensure that no subject- or batch-specific modules were analyzed, only gene programs with a mean usage >50 across all subjects were included for further analysis.

Subject classification and comparison with published predictors. All comparison of abundances were tested for significance by a Wilcoxon rank-sum test. Benjamini–Hochberg FDR correction was applied to the calculated *P* values for multiple testing of either cell types or states. To compare against published gene-based predictors, we summed UMI counts for each gene from all cells for each subject, scaled to the total number of UMI counts per patient, and calculated the FAIM-to-PLAC8 ratio[28], SeptiCyte Lab[22] and Sepsis Metascore[12] following published protocols. ROC curves were plotted on the basis of these absolute scores, as well as the fraction of MS1 for each subject.

Bulk-data deconvolution, gene-signature mapping and meta-analysis. We constructed a reference signature matrix for our cell states by generating bulk profiles from single-cell references, and ranking the genes based on effect size[47]. We first optimized the number of genes in the signature matrix by finding the minimum number of genes where the reduction in prediction error is saturated. We found this value to be at >50 genes and selected 100 genes per state and lineage (1,201 total, union of all genes) in the final matrix. To construct the signature matrix, we summed UMI counts for each state, normalized to the number of total UMIs per state and quantile-normalized the resulting matrix.

Datasets comparing sepsis and healthy controls were obtained as outlined in two published studies[4,9]. Datasets with gene-expression matrices that were not publicly available were not included in our analysis. Gene-expression deconvolution was performed using CIBERSORT[29]. Noting that our state signatures only capture PBMC states and exclude high-density cells in whole blood, the data were deconvolved with a no-sum-to-one constraint and absolute scoring. The resulting score matrix was then used as an input to MetaIntegrator[48]. The effect size of each state was visualized using forest plots, and the classification performance of MS1 cells was quantified by generating a summary ROC plot.

Stimulation of bone marrow and peripheral blood cells. For MS1-induction experiments, bone marrow or peripheral mononuclear cells were cultured in SFEM II supplemented with 1× CC110 (StemCell Technologies) with or without the presence of 100 ng ml^{-1} LPS or Pam3CSK4 (Invivogen) for up to 4 d. For restimulation experiments, sorted monocytes were rested for 24 h in RPMI-1640 supplemented with 10% heat-inactivated FBS and 1× penicilin–streptomycin (Gibco), before adding 100 ng ml^{-1} LPS (Invivogen).

ATAC-seq processing and data analysis. ATAC-seq was performed on 25,000 sorted cells, as described in a published protocol[49]. Libraries were sequenced on a NextSeq (Illumina) with 38×38 paired-end reads and at least 10 million reads per sample. Sequencing data were aligned using the ENCODE Project ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/), and further analyzed using custom scripts. To generate a peak count matrix, we first identified a consensus peak set using the 'multiinter' function, and then analyzed the number of counts for each sample using the function 'coverageBed' from bedtools v2. Differential peak analysis was performed using edgeR, using the peak count matrix as input. Peak motifs were analyzed using the 'findMotifsGenome' function in Homer v4.1, with a window size of 200 bp.

Bulk RNA-seq processing and data analysis. Bulk RNA-seq was performed using Smart-Seq2 (ref. [50]) with minor modifications, as described in a previous study[13]. Briefly, 5,000 sorted or cultured cells were resuspended in 15 µl of Buffer TCL (Qiagen), and their RNA was purified by a 2.2× SPRI cleanup with RNAClean XP magnetic beads (Agencourt). After reverse transcription, amplification and cleanup, libraries were quantified using a Qubit fluorometer (Invitrogen), and their size distributions were determined using an Agilent Bioanalyzer 2100. Amplicon concentrations were normalized to 0.1 ng ml^{-1} and sequencing libraries were constructed using a Nextera XT DNA Library Prep Kit (Illumina), following the manufacturer's protocol. All RNA-seq libraries were sequenced with 38×38 paired-end reads using a NextSeq (Illumina). RNA-seq libraries were sequenced to a depth of >2 million reads per sample. STAR was used to align sequencing reads to the UCSC hg19 transcriptome and RSEM was used to generate an expression matrix for all samples. Both raw count and transcripts per million data were analyzed using edgeR and custom python scripts. The list of identified receptor–ligand pairs was obtained from a previous publication[51].

Cytokine measurements. Culture supernatants were diluted 2× in PBS and frozen at −80 °C before processing. Samples from multiple experiments were thawed and analyzed in parallel using the Legendplex Human Inflammation Panel, TNF-α (BioLegend). Flow cytometry data were acquired on a Cytoflex LX (Beckman Coulter) and analyzed using FlowJo v10.1.

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

Data availability

scRNA-seq data is available for visualization and secondary analysis through the Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell): SCP548 (subject PBMCs) and SCP550 (bone-marrow stimulation). Expression matrices are also deposited in SCP. Raw sequencing data are available for download from the controlled-access data repository Broad DUOS (https://duos.broadinstitute.org/).

Code availability

Code used for data analysis are available at https://github.com/reyes-m/sepsis_signature.

References

41. Center for Medicare & Medicaid Services. *Sepsis Bundle Project (SEP), National Hospital Inpatient Quality Measures*. http://www.qualitynet.org/ (accessed 28 April 2019).
42. Nakahira, K. et al. Circulating mitochondrial DNA in patients in the ICU as a marker of mortality: derivation and validation. *PLoS Med.* **10**, e1001577 (2013). discussion e1001577.
43. Dolinay, T. et al. Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am. J. Respir. Crit. Care Med.* **185**, 1225–1234 (2012).
44. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
45. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
46. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. *Elife* **8**, 310599 (2019).
47. Vallania, F. et al. Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nat. Commun.* **9**, 4735 (2018).
48. Haynes, W. A. et al. Empowering multi-cohort gene expression analysis to increase reproducibility. *Pac. Symp. Biocomput.* **22**, 144–153 (2017).
49. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
50. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
51. Ramilowski, J. A. et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).

Acknowledgements

We thank the subjects in our clinical cohorts for participating and donating blood for our study. We are also grateful to the clinical research coordinators and fellows who conducted patient enrollment and performed blood collections: B. Parry, K. Gohres, J. Lee, J. Hwang and K. Richards (MGH); K. Cummins, A. Higuera-Moreno, M. Pinilla-Vera and M. Benson (BWH). We thank A.-C. Villani, A. Arazi, J. Wenger and other members of the Blainey and Hacohen Labs (Broad Institute) for helpful discussions. We thank the Broad Flow Cytometry core for assistance in cell-sorting experiments, and the Broad Genomics Platform for assistance in sequencing. We also thank P. Khatri, S. Schaffert and M. Donato (Stanford University) for advice on bulk gene expression deconvolution and meta-analysis of published datasets. This work was supported by a SPARC award from the Broad Institute (M.B.G.), awards from the Ellison Foundation and the Allergan Foundation (M.B.G.), BARDA 1DSEP160030-01-00 (D.T.H.) and NIH NIAID U24 AI118668 (N.H. and P.C.B.). N.H. was supported by the David P. Ryan, MD Endowed Chair in Cancer Research, R.P.B. was supported by a Mentored Clinical Scientist Research Career Development Award from the NIH (1K08AI119157-04) and P.C.B. was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

Author contributions

M.R. designed and performed experiments, and analyzed the data with clinical input from M.R.F., R.P.B., R.M.B. and M.B.G.. K.B. and T.E. assisted in sorting and sample processing for single-cell RNA-seq. M.R.F., R.P.B. and M.B.G. designed the MGH clinical cohorts and supervised enrollment and specimen collection, and performed clinical adjudications. D.T.H., B.D.L. and R.M.B. supervised enrollment and specimen collection at the BWH. M.B.G., M.R.F., N.H., P.C.B., D.T.H. and R.P.B. conceived the study. M.B.G., N.H. and P.C.B. supervised the study. M.R., P.C.B. and N.H. prepared the manuscript; all authors reviewed and edited the final manuscript.

Competing interests

The Broad Institute and MIT may seek to commercialize aspects of this work, and related applications for intellectual property have been filed. In addition, P.C.B. is a consultant to and equity holder in a company, 10X Genomics, whose products were used in this study.

Additional information

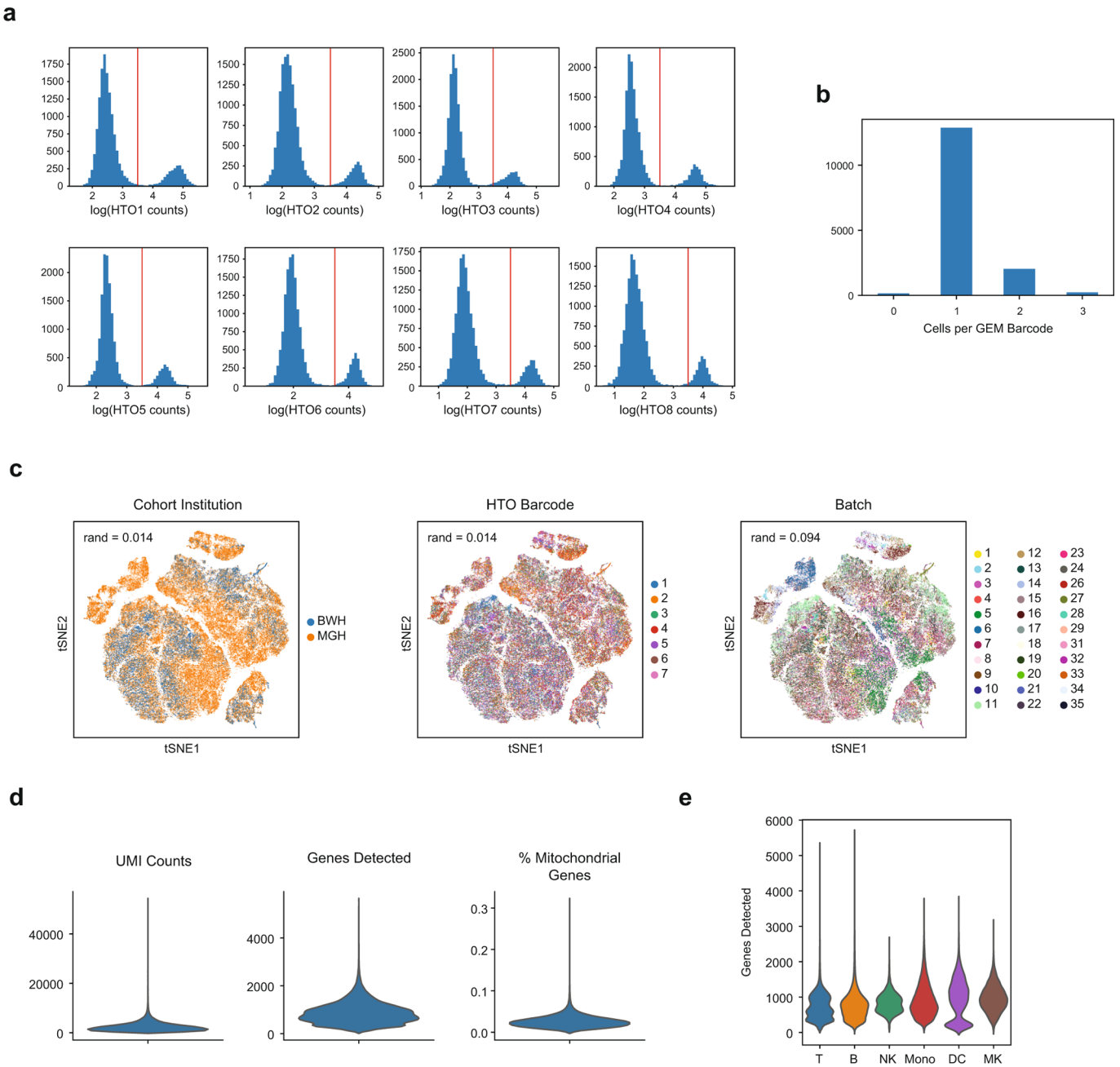
Extended data is available for this paper at https://doi.org/10.1038/s41591-020-0752-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-0752-4.

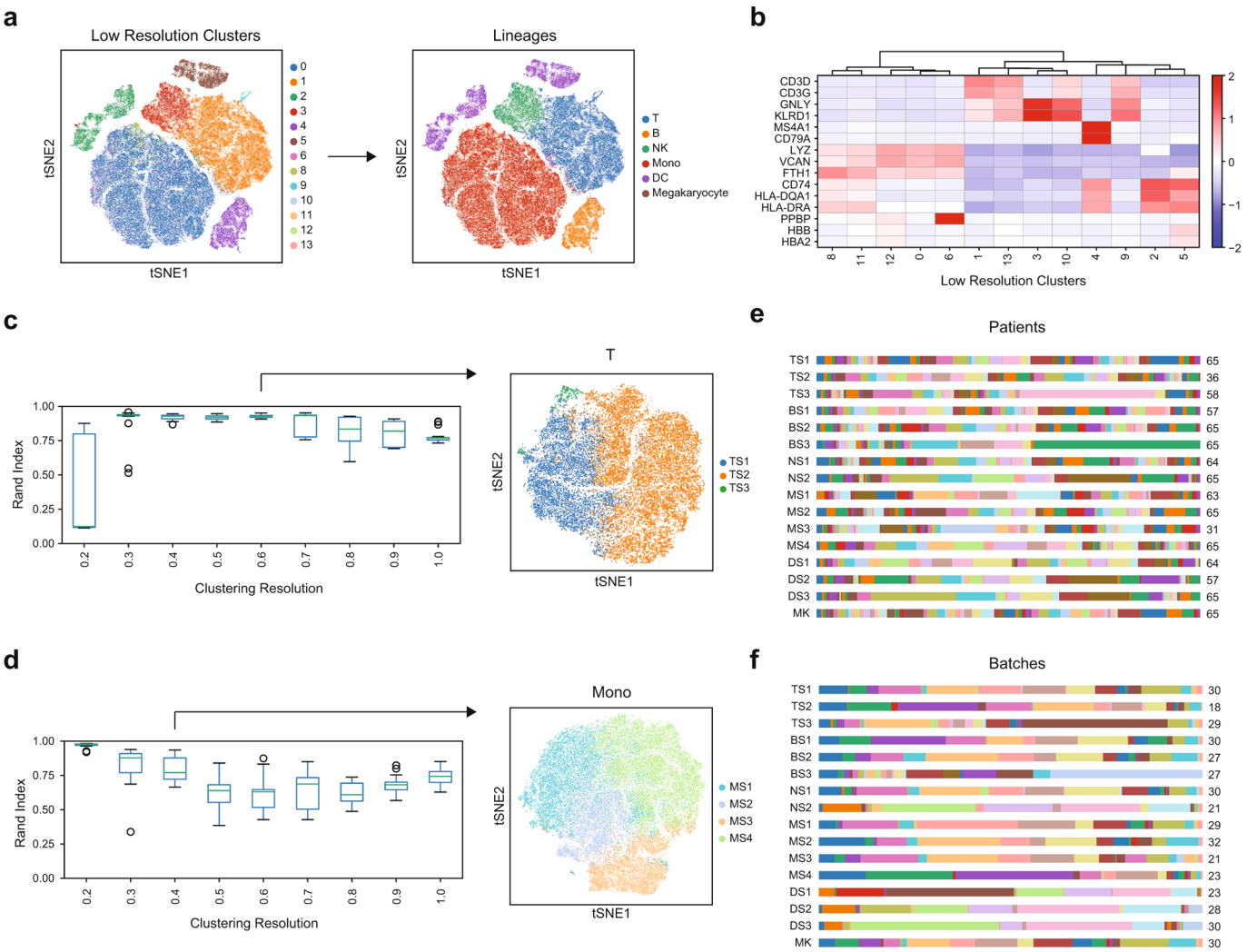
Correspondence and requests for materials should be addressed to P.C.B., M.B.G. or N.H.

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

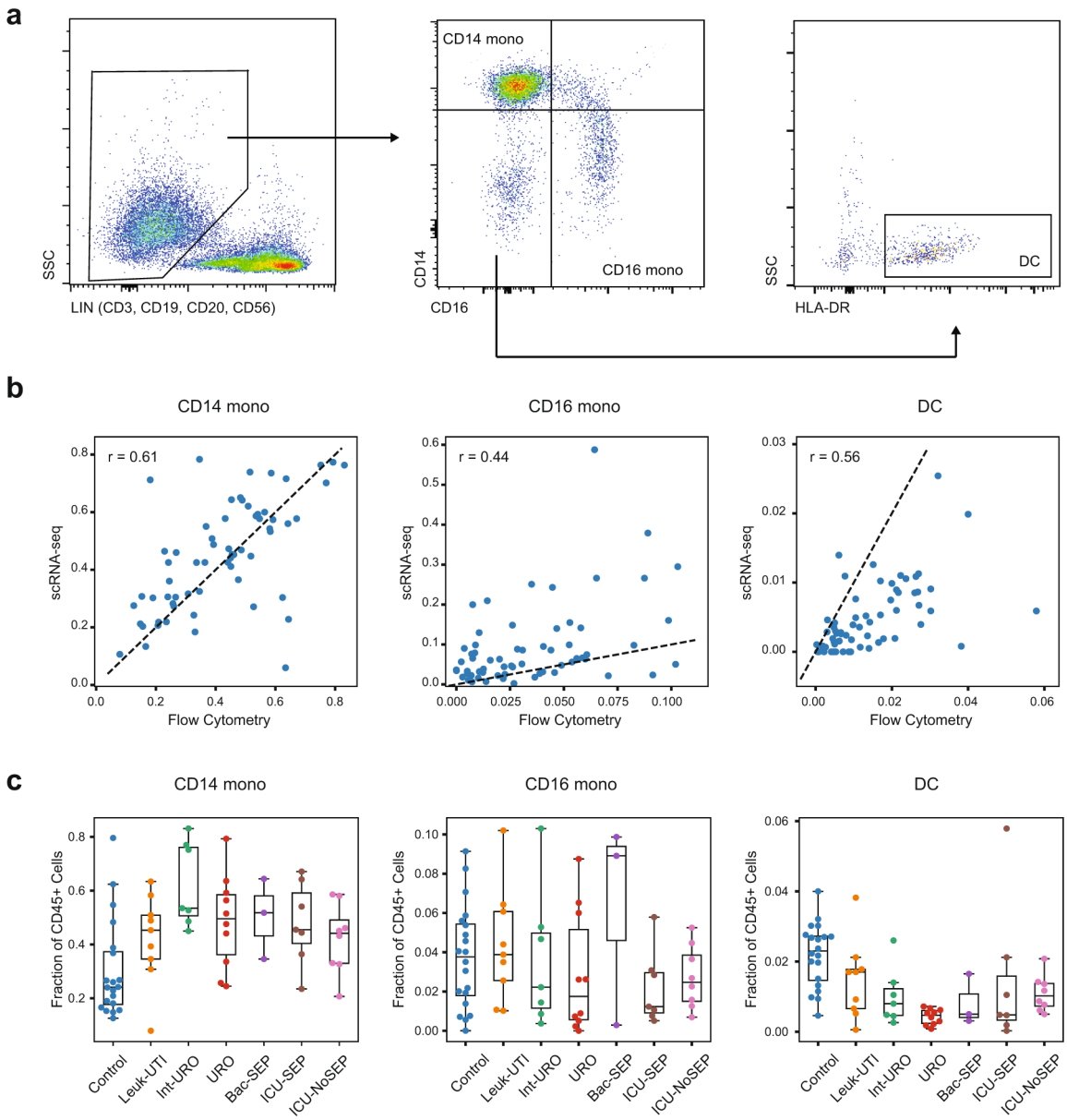
Reprints and permissions information is available at www.nature.com/reprints.



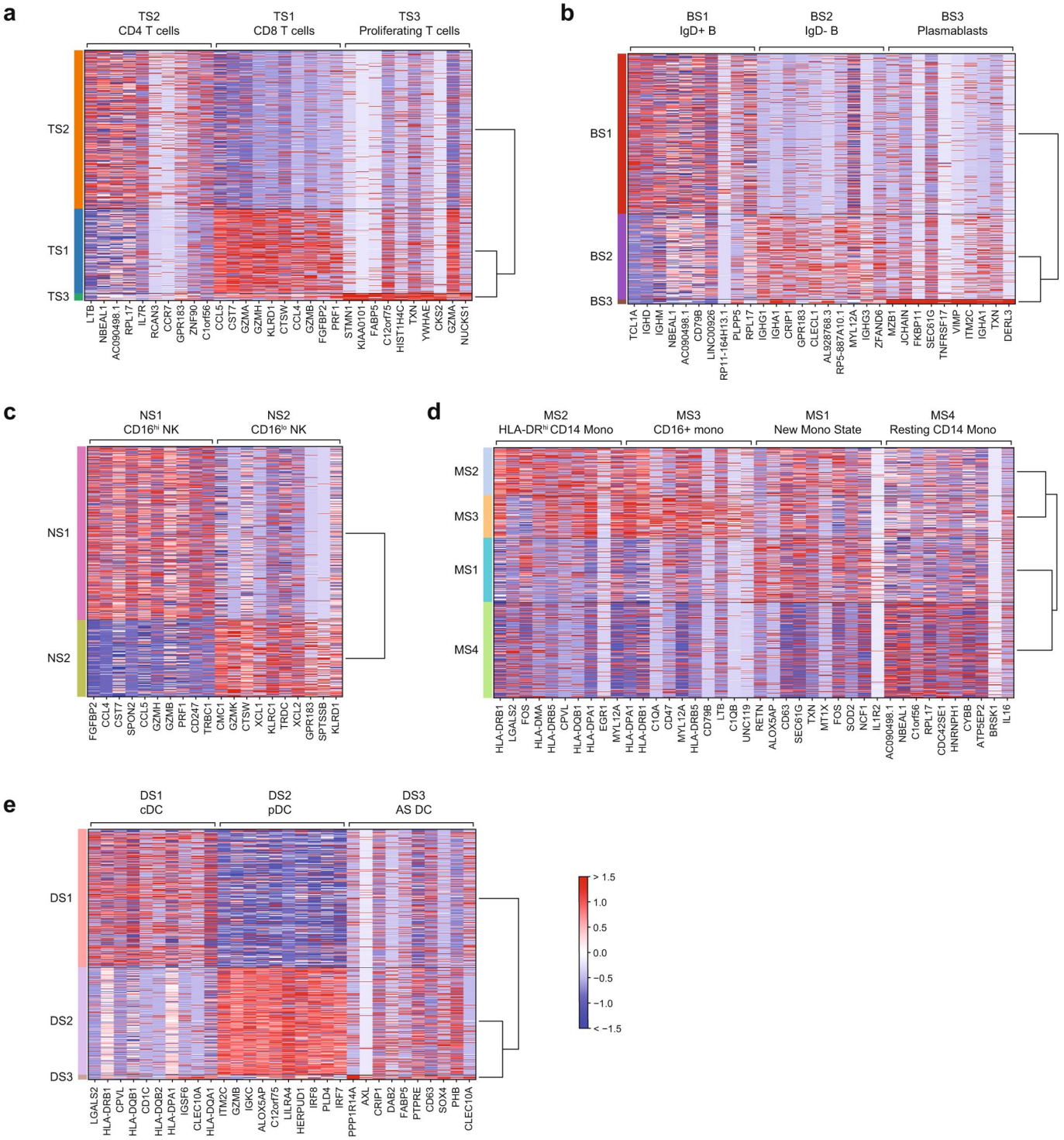
Extended Data Fig. 1 | scRNA-seq demultiplexing and quality assessment. (**a**) Sample strategy for gating for hashtag oligo (HTO) positive cells based on UMI tag counts of each barcode. (**b**) Histogram of cells per 10X gel beads in emulsion (GEM) barcode for one representative channel. Data are shown for one channel with $n = 15{,}304$ detected GEMs. (**c**) tSNE plots of all cells ($n = 126{,}351$ cells total from 65 individuals) in the study colored by institution of origin of the cohort, hashtag barcode, and processing batch. Adjusted Rand index is shown for each when compared with cell state assignments. (**d**) Violin plots ($n = 126{,}351$ cells total from 65 individuals) of various quality control metrics for the full scRNA-seq dataset generated in this study. (**e**) Violin plots of genes detected across different cell-types ($n = 32{,}341$, 7,970, 9,390, 58,557, 14,299, 3,794 cells for T, B, NK, Mono, DC, and MK, respectively). Violin plots show a kernel density estimate of the data, using Scott's rule to calculate the appropriate kernel bandwidth. The violin extends to 2x the bandwidth in both directions.



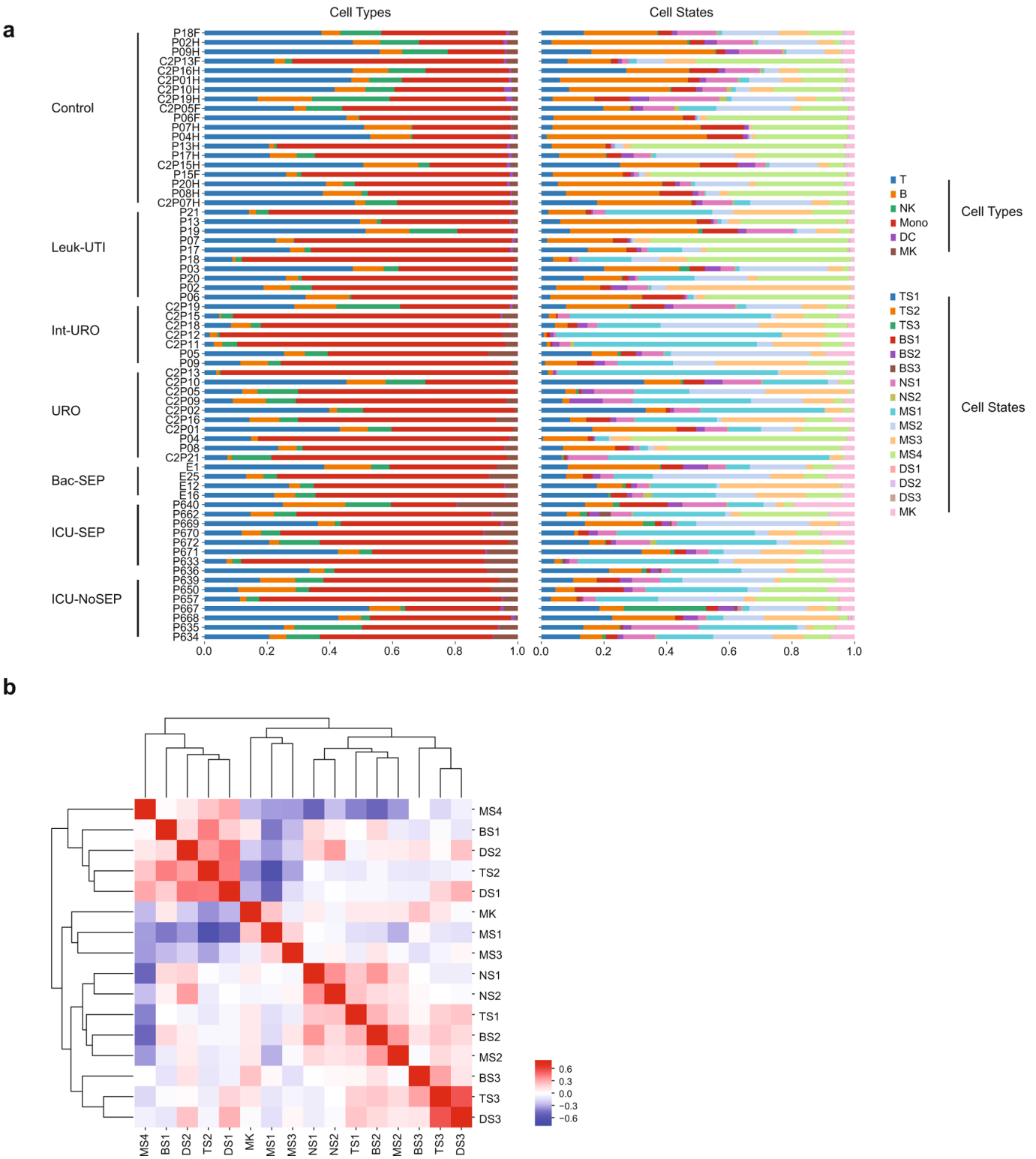
Extended Data Fig. 2 | Robust identification of cell states with two-step clustering. (**a**-**b**) Identification of immune cell types based on marker genes of low-resolution clusters. Color scale in (**b**) corresponds to z-scored, log2-transformed mean gene expression counts across all cells ($n = 126{,}351$ cells total from 65 individuals). (**c**-**d**) Assessment of cluster robustness for T-cells (T) (**c**) and monocytes (Mono) (**d**) ($n = 32{,}341$ and 58,557 cells for T and Mono, respectively). Boxplots show distributions of Rand indices when comparing clustering solutions with subsampled data (20 iterations). Boxes show the median and IQR for each resolution, with whiskers extending to 1.5 IQR in either direction from the top or bottom quartile. tSNE plots show final assigned states for each cell type. (**e**-**f**) Barplots showing the fraction of each patient (**e**) and batch (**f**) in each of the 16 cell states (number of patients or batches with each state is indicated).



Extended Data Fig. 3 | Flow cytometry abundances of classical myeloid cell states. (**a**) Gating strategy for determination of CD14 + mono, CD16 + mono, and dendritic cell abundance. (**b**) Correlation of fractional abundances defined by flow cytometry and scRNA-seq for each patient ($n = 65$ individuals). (**c**) Fractional abundance of the three cell types based on flow cytometry, grouped by disease state. Sample size (n) for each cohort is indicated in Fig. 1b. Boxes show the median and IQR for each patient cohort, with whiskers extending to 1.5 IQR in either direction from the top or bottom quartile.

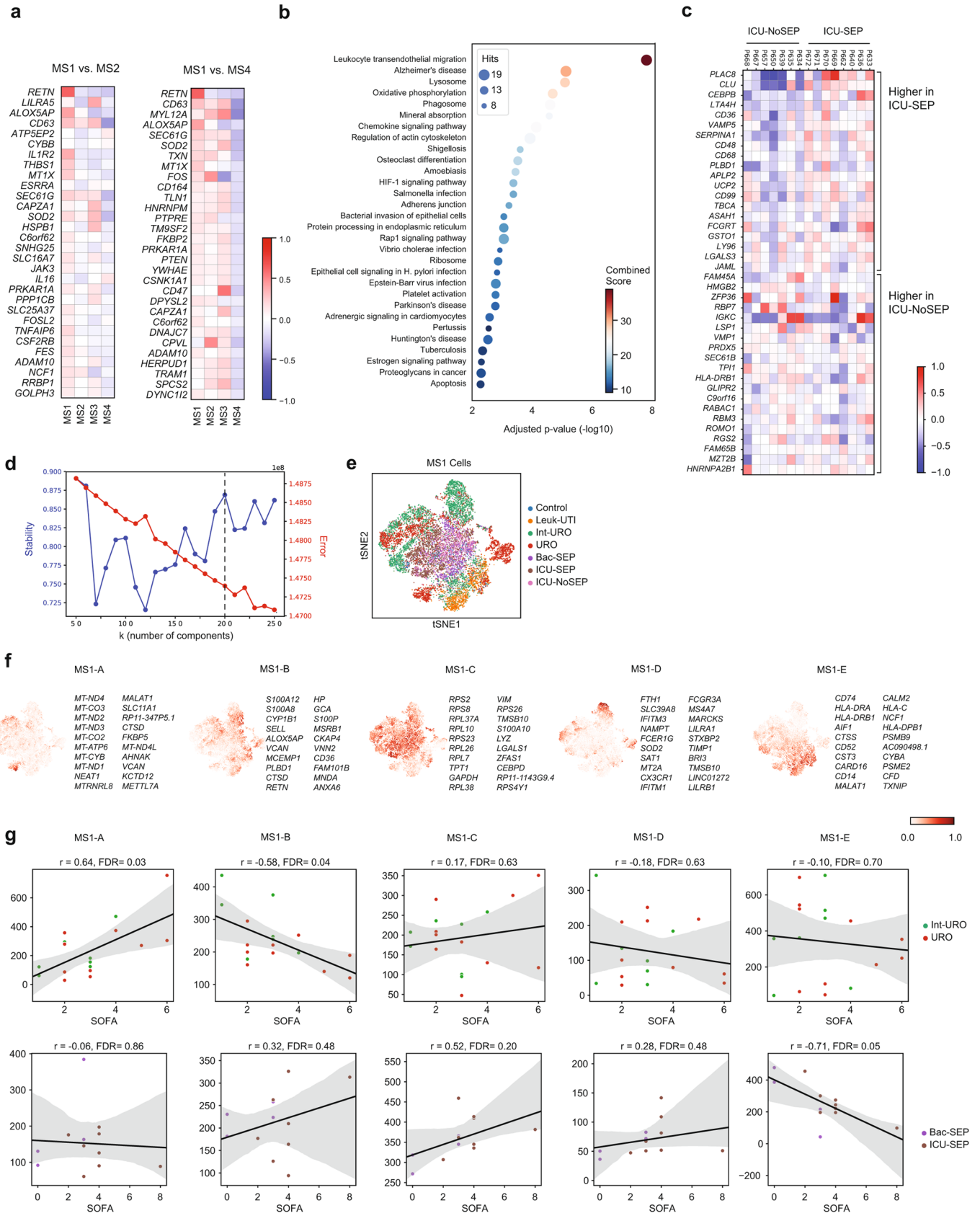


Extended Data Fig. 4 | Differentially expressed genes in immune cell states. Top 10 differentially expressed genes (FDR < 0.05, two-tailed Wilcoxon rank-sum test) for each cell state when compared with other cells within the same cell type. Heatmaps are grouped according to the parent cell type of the different states: (**a**) T cells, (**b**) B cells, (**c**) NK cells, (**d**) monocytes, and (**e**) dendritic cells. $n = 32{,}341$, 7,970, 9,390, 58,557, 14,299, 3,794 cells for T, B, NK, Mono, DC, and MK, respectively. cDC, conventional dendritic cells; pDC, plasmacytoid dendritic cells; AS DC, AXL-SIGLEC6 dendritic cells.



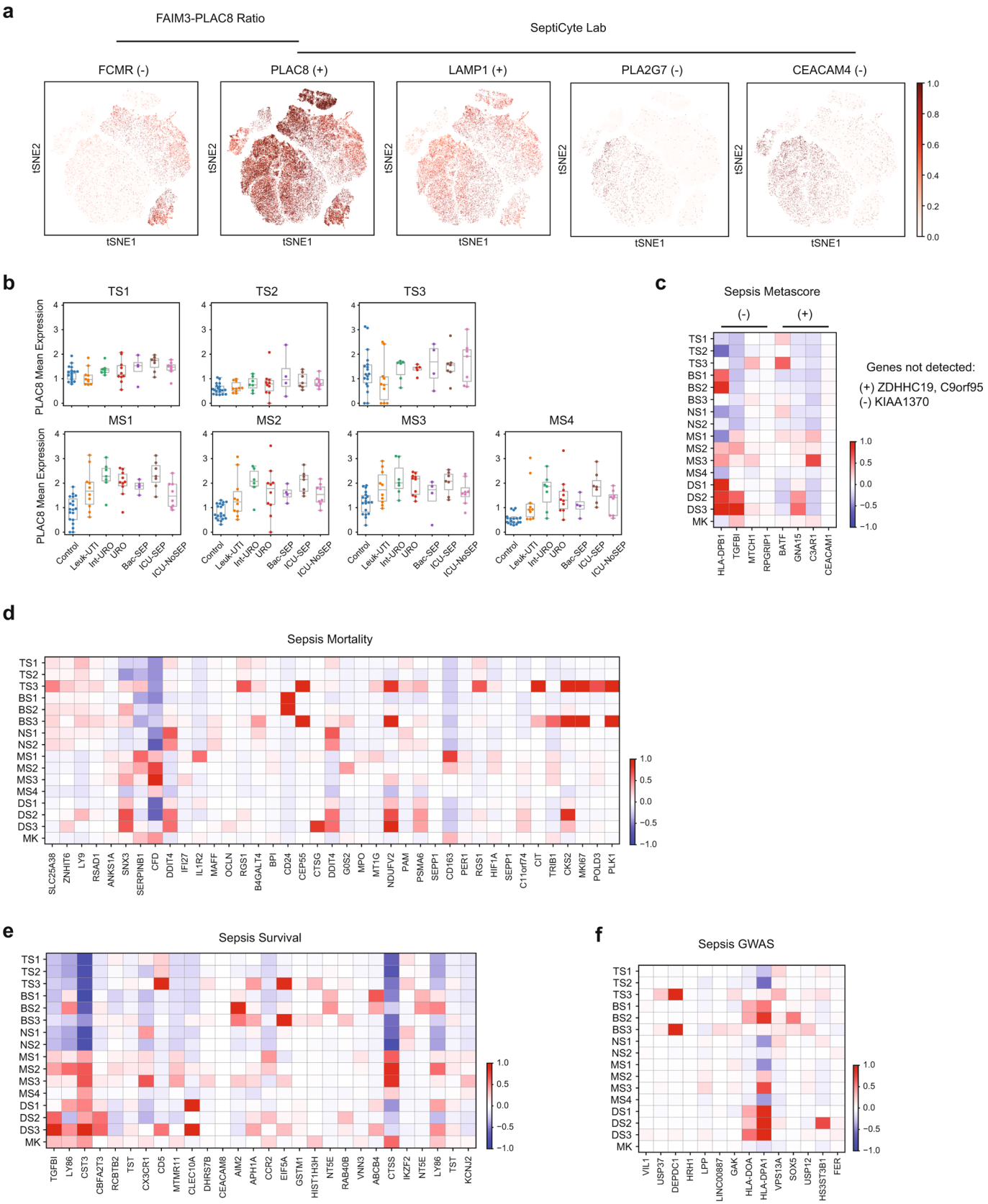
Extended Data Fig. 5 | Fractional abundance of states defined by scRNA-seq. (**a**) Cell type and state composition for each patient in each cohort. (**b**) Pearson correlation matrix of cell states across all patients ($n = 65$ patients).

Extended Data Fig. 6 | Disease-specific abundance of cell types and states. Boxplots showing fractional abundances of (**a**) cell types and (**b**) states among patients grouped by patient cohort. FDR values are shown when comparing each disease state with healthy controls (two-tailed Wilcoxon rank-sum test, corrected for multiple testing of states). Sample size (*n*) for each cohort is indicated in Fig. 1b. (**c**) Boxplots showing absolute abundances of cell states among patients (for which leukocyte counts were available), grouped by patient cohort. Boxes show the median and IQR for each patient cohort, with whiskers extending to 1.5 IQR in either direction from the top or bottom quartile. $n = 10, 6, 10, 3, 6$, and 4 patients, for Leuk-UTI, Int-URO, URO, BAC-SEP, ICU-SEP and ICU-NoSEP, respectively.

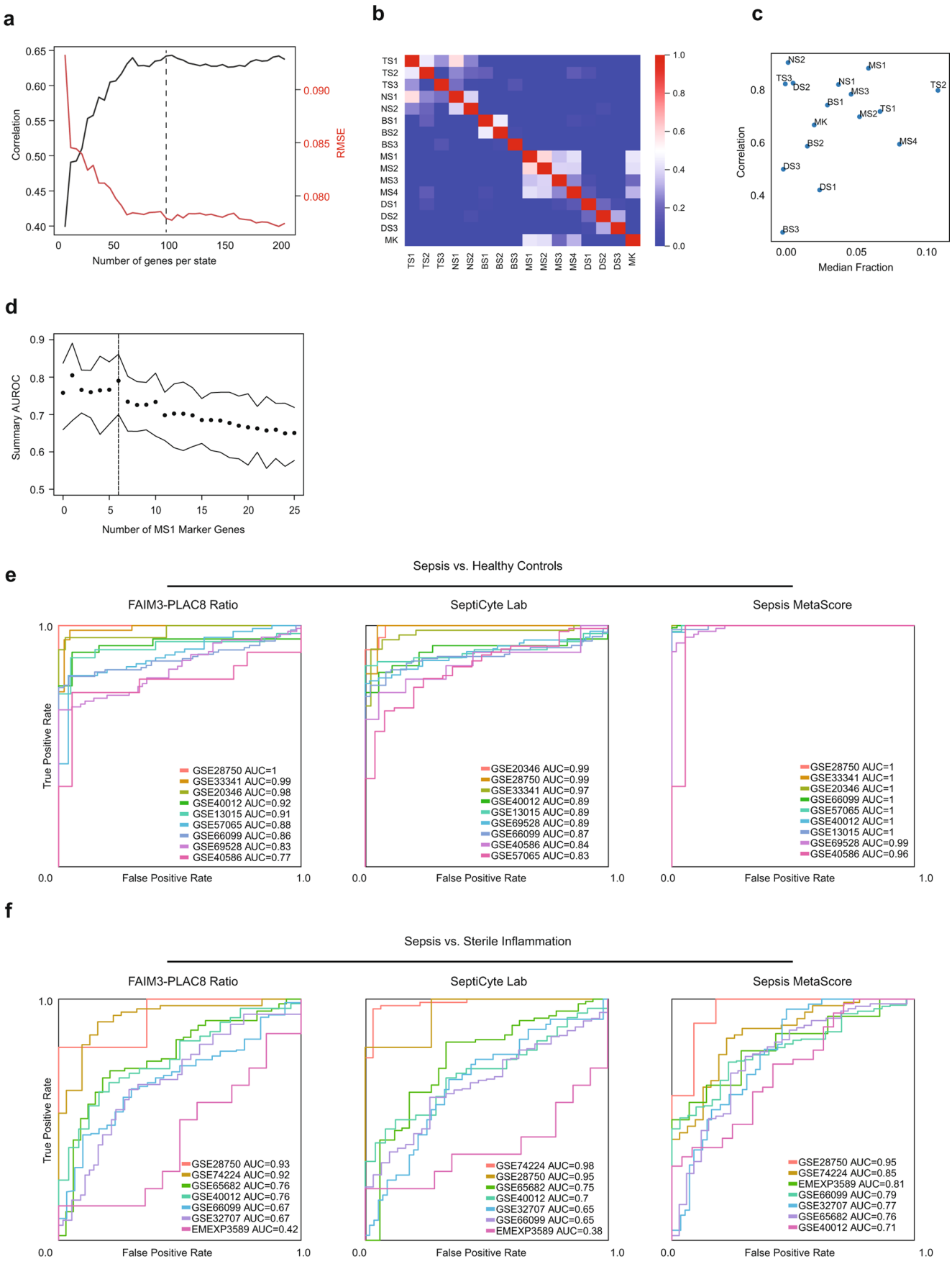


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | In-depth analysis of the gene expression profile of MS1. (**a**) Top 30 differentially expressed genes (among highly variable genes) when comparing MS1 against other CD14 + monocyte states (MS4 and MS2). (**b**) Dotplot showing enrichment of pathways (KEGG database) for upregulated genes in MS1 vs. MS2 (FDR < 0.1, edgeR exact test). $n = 15{,}021$ and 11,439 cells for MS1 and MS2, respectively. Sizes of circles are proportional to the number of gene hits in a set, whereas color represents the enrichment score of each gene set. (**c**) Heatmap showing the average expression of genes that are differentially expressed (FDR < 0.1, two-sided Wilcoxon rank-sum test) between all MS1 cells from each patient in the ICU-SEP cohort and all MS1 cells from each patient in the ICU-NoSEP cohort ($n = 2{,}153$ and 1,442 cells from 8 and 7 ICU-SEP and ICU-NoSEP patients, respectively). To specifically identify genes that discriminate the two patient populations, genes are filtered for expression in-group fraction > 0.4 and out-group fraction < 0.6. (**d**) k-selection plot to determine the number of components for non-negative matrix factorization (NMF). Dotted line indicates selected number of components for further analysis. (**e**) tSNE plot of MS1 cells ($n = 15{,}021$ cells) colored by patient cohort of origin. (**f**) Scaled TPM usage values of each gene module derived from NMF analysis. The top 20 genes in each module are shown. (**g**) Scatterplots showing correlation between mean gene module usage in MS1 cells and sequential organ failure assessment (SOFA) scores for Int-URO and URO patients (top row), and Bac-SEP and ICU-SEP patients (bottom row). Sample size (n) for each cohort is indicated in Fig. 1b. Significance of the correlation (Pearson r) was calculated with a two-sided permutation test. Line and shadow indicate linear regression fit and 95% confidence interval, respectively.

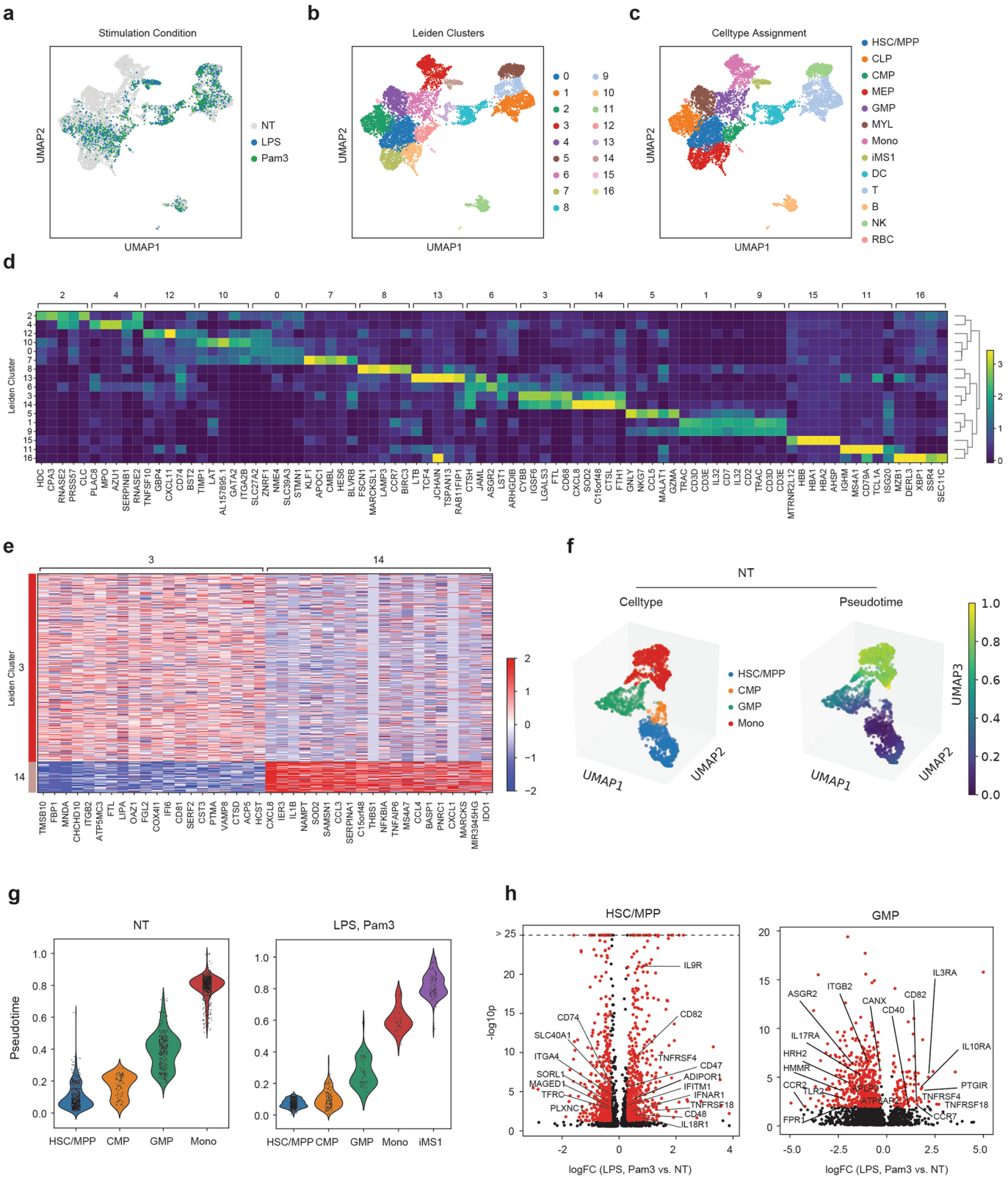


Extended Data Fig. 8 | State-specific expression of sepsis signature genes. (**a**) tSNE plots showing scaled gene expression counts across all cells ($n = 126,351$ total from 65 individuals) for FAIM3-PLAC8[28] and SeptiCyte Lab[22] genes (+or - indicates that a gene is up- or down-regulated, respectively, in sepsis). (**b**) Mean expression of *PLAC8* in T cell (top row) and monocyte (bottom row) states across patients grouped by cohort. Sample size (n) for each cohort is indicated in Fig. 1b. (**c**-**f**) Heatmaps show state-specific expression of Sepsis Metascore[12] genes (**c**), genes previously associated with sepsis mortality (**d**) or survival[10,11] (**e**), and sepsis-linked GWAS genes (**f**). Color scale corresponds to z-scored, log2-transformed mean gene expression counts for cell state.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | State matrix generation and performance comparison of gene-based signatures. (**a**) Optimization of the number of marker genes per cell state in the basis matrix for deconvolution. Mean deconvolution accuracy is shown for pseudo-bulk gene expression data generated for each patient in our study ($n = 65$ patients). Accuracy is measured as high correlation or low root mean-squared error (RMSE) between predicted and true values. Dotted line indicates number of genes used for downstream analysis. (b) Gene expression correlation of all states using the signature matrix with 100 genes per cell state (1,201 total, union of all genes). (**c**) Scatterplot showing deconvolution accuracy (measured by Pearson correlation between true and inferred fractions) increases with median fractional abundance of cell states. (**d**) Summary area under the receiver operating characteristic curve (AUROCs) of the mean expression of *PLAC8*, *CLU*, and the indicated number of MS1 marker genes when classifying sepsis patients against sterile inflammation in published datasets. Top and bottom lines indicate the 95% confidence interval of the summary AUROC. Dotted line indicates number of MS1 marker genes used for downstream analysis. (**e**-**f**) Individual ROC curves of FAIM3-PLAC8 Ratio, SeptiCyte Lab, and Sepsis MetaScore on published datasets comparing sepsis vs. healthy controls (e; $n = 751$ total patients from 9 cohorts) or sepsis vs. sterile inflammation (f; $n = 696$ total patients from 7 cohorts).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | scRNA-seq characterization of stimulated bone-marrow mononuclear cells. BM mononuclear cells incubated in HSC cytokine-rich media with no treatment (NT) or 100 ng/mL LPS or Pam3CSK4 (Pam3) for 4 days. Cells ($n = 8,702$) are visualized on a UMAP projection and colored by (**a**) treatment, (**b**) Leiden clusters, and (**c**) cell-type annotations. (**d**) Matrixplot showing the mean log-transformed UMI counts of the top 5 differentially expressed genes (FDR < 0.01, two-tailed Wilcoxon rank-sum test) for each cluster in (**b**). (**e**) Heatmap showing differentially expressed genes (FDR < 0.01, two-tailed Wilcoxon rank-sum test) between clusters 3 (CD14 monocytes, $n = 786$ cells) and 14 (iMS1 cluster, $n = 130$ cells). (**f**) UMAP projections of non-stimulated BM myeloid and progenitor cells (HSC/MPP, CMP, GMP, Mono; $n = 1,976$ cells total) colored by cell type (left) and diffusion pseudotime (right). (**g**) Violin plots showing pseudotime values for each cell type in each stimulation condition. $n = 1,976$ and 901 cells for NT and LPS or Pam3 treatments, respectively. Violin plots show a kernel density estimate of the data, using Scott's rule to calculate the appropriate kernel bandwidth. The violin extends to 2x the bandwidth in both directions. (**h**) Volcano plots showing differentially expressed genes between LPS or Pam3CSK4 and untreated cells for the HSC/MPP ($n = 1,168$ cells) and GMP populations ($n = 519$ cells). Differentially expressed genes (logFC > 0.3, $p < 0.05$; two-sided Wilcoxon rank-sum test) are shown in red. Known receptors (based on a previously published database[51]) that are differentially expressed are labelled. HSC/MPP, hematopoietic stem cells and multipotent progenitors; CMP, common myeloid progenitors; GMP, granulocyte-macrophage progenitor; MEP, megakaryocyte-erythroid progenitors; MYL, myeloblasts; RBC, red blood cells.

nature research

Corresponding author(s): Nir Hacohen, Marcia B. Goldberg, Paul C. Blainey

Last updated by author(s): Dec 22, 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	Cellranger v2.02 (10X Genomics) was used to align and generate expression matrices for downstream analysis.
Data analysis	Cell hashing reads were aligned using the CITE-seq count tool v1.4 (https://github.com/Hoohm/CITE-seq-Count). Scanpy v1.4 was used to perform analysis of single cell data. scikit-learn v0.20.3 was used to perform regression analysis and assess classifier performance. CIBERSORT source code was obtained from the original publication (https://cibersort.stanford.edu/index.php) and used as is. MetaIntegrator v2.00 was used to performed meta-analysis. bedtools v2 and Homer v4.1 were used to analyze ATAC-seq data. edgeR v3.1 was used to perform differential expression and peak accessibility analysis.

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

scRNA-seq data is available for visualization and secondary analysis through the Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell): SCP548 (patients PBMCs) and SCP550 (bone marrow stimulation). Expression matrices are also deposited in SCP. Raw sequencing data are available for download from the controlled-access data repository, Broad DUOS (https://duos.broadinstitute.org/)

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	No sample-size calculations were performed. Sample size was determined to be adequate based on the degree and consistency of differences between groups. 1. Single cell RNA-sequencing and cell state identification - A total of 65 patients were profiled. Cells from all study subjects were used to determine cell states. 2. Sorting and FACS Analysis of MS1 - a number of HC, UTI, and URO patients were ran with the MS1 panel based on the availability of aliquots of frozen samples. 3. Publicly available datasets - the number of datasets and samples included therein were used as outlined in Sweeney, et. al. 2017 and Sweeney, et. al. 2018.
Data exclusions	Datasets outlined in Sweeney, et. al. 2018 for which the processed gene expression data matrix was not publicly available were not analyzed.
Replication	No assays were repeated on the same patient sample due to limited availability of cells.
Randomization	Single cell RNA-sequencing - For the primary cohort, patient samples were batched such that each multiplex scRNA-seq has 1/2 samples from UTI or URO and 1/2 HC. For the secondary cohort, patient samples were processed in the order they were received.
Blinding	Single cell RNA-sequencing - cells from all patients were initially analyzed blind to which patient they originated from to define the 16 cell states.

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	CD3-APC (HIT3a), CD19-APC (HIB19), CD20-APC (2H7), CD56-APC (5.1H11), CD14-FITC (M5E2), CD16-AF700 (B73.1), CD45-PE-Cy7 (HI30), HLA-DR-PE (L243), CD45-AF700 (HI30), HLA-DR-PE-Cy7 (L243) and cell hashing antibodies (TotalSeq-A) were obtained from BioLegend. IL1R2-PE (34141) was obtained from ThermoFisher Scientific
Validation	For FACS analysis, the antibodies were used as prescribed by the manufacturer (5 uL per 100 uL of cells). Antibodies used are validated with immunofluorescent staining with flow cytometry analysis as indicated by the manufacturer. For cell hashing and scRNA-seq, 0.2 ug of each HTO antibody was used per sample, the concentration was determined by titrating the antibody with scRNA-seq analysis of pooled PBMC samples.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Median ages of the clinical cohorts range from 53-71. 49% of the the individuals profiled are male, and 86% are white, non-Hispanic. Further details on disease status, age, gender, and race of all participants in this study are provided in Supplementary Table 1.

Recruitment	Primary cohort- patients entering the Emergency Department at the Massachusetts General Hospital were screened for inclusion criteria pertinent to a urinary tract infection, with and without organ damage. Secondary cohorts - all patients entering the Medical Intensive Care Unit at the Brigham and Women's Hospital were recruited in the study before a diagnosis. Patients in the hospital wards are screened for positive bacteraemia cultures. Enrollment in the cohorts were based on voluntary participation or through surrogate consent, no selection biases were present.
Ethics oversight	Research on the samples was approved by the Institutional Review Boards at the Broad Institute of MIT and Partners Health Care, (MA, USA).

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	Samples were from cryopreserved PBMCs (Methods)
Instrument	All FACS analysis were performed on a Sony SH800
Software	Data were analyzed using FlowJo v10.2
Cell population abundance	The post-sort purity of the sorted CD45+ and dendritic cells were not measure via flow cytometry. Based on scRNA-seq profiles, the sort purities are > 90% for CD45+ cells and > 70% for dendritic cells.
Gating strategy	Gating strategies are described in Supplementary Figure 3, Figure 3f, and Figure 4a.

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