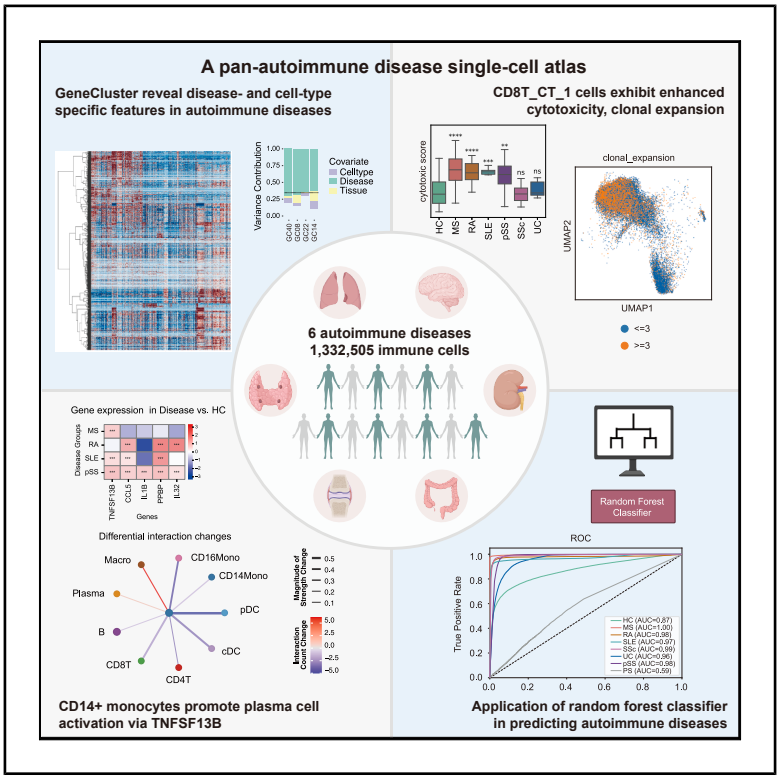
iScience

Article

Single-cell landscape of immune cells in multiple autoimmune diseases

Graphical abstract



Highlights

- A comprehensive single-cell immune atlas for autoimmune diseases is constructed
- 41 gene clusters reveal disease- and cell-type-specific molecular signatures
- GC40 and GC08 drive CD8^{+} T cytotoxicity and monocyte-plasma crosstalk
- A single-cell-based classifier accurately classifies autoimmune diseases

Authors

Lin Zhu, Wujianan Sun, Kun Chen, ..., Xiaomei Li, Xu Liu, Kun Qu

Correspondence

lxu1016@ustc.edu.cn (X.L.), qukun@ustc.edu.cn (K.Q.)

In brief

Immunology; Bioinformatics; Genomics

Zhu et al., 2026, iScience *29*, 114515
January 16, 2026
https://doi.org/10.1016/j.isci.2025.114515

CellPress

iScience

CellPress
OPEN ACCESS

Article

Single-cell landscape of immune cells in multiple autoimmune diseases

Lin Zhu,[1,2] Wujianan Sun,[1] Kun Chen,[1] Zitian Liu,[1] Meifang Tang,[1] Chen Jiang,[1,2] Yu Shi,[1] Qian Liu,[1] Wen Zhang,[1] Junyu Wang,[1] Ao Zheng,[1] Jun Lin,[1] Guosheng Wang,[1] Xiaomei Li,[1] Xu Liu,[1,*] and Kun Qu[1,2,3,4,*]

[1]Department of Rheumatology and Immunology, The First Affiliated Hospital of USTC, State Key Laboratory of Eye Health, School of Basic Medical Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230021, China

[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

[3]School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei 230026, China

[4]Lead contact

*Correspondence: lxu1016@ustc.edu.cn (X.L.), qukun@ustc.edu.cn (K.Q.)

https://doi.org/10.1016/j.isci.2025.114515

SUMMARY

Autoimmune pathologies arise from dysregulated immune activation, yet conserved molecular programs across autoimmune contexts remain incompletely characterized. Here, we analyzed integrative single-cell RNA sequencing data of over 1.3 million cells from 239 samples spanning six autoimmune diseases, revealing disease-specific and cell type-specific transcriptional programs. Systematic immune profiling identified 41 functionally annotated gene clusters (GCs) with cross-disease activation. We identified a cytotoxic *CD8*$^+$ T cell-enriched gene cluster (GC40) that drives enhanced cytotoxic function and clonal expansion across four autoimmune diseases. In parallel, we identified a secretory granule lumen-associated GC08 that is highly expressed in *CD14*$^+$ monocytes and promotes plasma cell activation through upregulation of TNFSF13B secretion. Furthermore, we leveraged a disease classifier to discriminate autoimmune disease types and healthy states. Our study provides both a resource and a predictive framework at the single-cell level, supporting future targeted therapies for autoimmune diseases.

INTRODUCTION

The immune system, comprising specialized organs, cells, and signaling molecules, serves as a robust defense against pathogenic threats.[1] Autoimmune diseases represent a spectrum of immune dysregulatory conditions arising from compromised self-tolerance mechanisms, clinically manifesting through autoantibody production and heterogeneous pathological presentations.[2] Currently, researchers have discovered over 150 distinct autoimmune disorders affecting multiple organ systems, including prototypical conditions such as systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), inflammatory bowel disease, multiple sclerosis (MS), systemic sclerosis (SSc), and primary Sjögren's syndrome (pSS).[3] As central mediators of autoimmune pathogenesis, immune cells exhibit dual functionality in maintaining immunological homeostasis and driving pathological responses.[4–6] While emerging therapies targeting immune cell pathways are advancing through preclinical and clinical development, systematic comparative analyses of their pathogenic mechanisms across different autoimmune diseases remain critically underexplored.[7–9]

Autoimmune pathogenesis involves multiple immune cell lineages, including T lymphocytes, NK cells, B cells, myeloid populations, and platelets. Recent studies have highlighted critical roles of immune cells and their interactions in driving disease pathogenesis and progression.[2] Among these, *CD8*$^+$ cytotoxic T lymphocytes (CTLs) emerge as central pathogenic mediators through targeted cytolysis of autoantigen-presenting cells.[4] In type 1 diabetes and MS, CTLs mediate tissue-specific destruction of pancreatic β-cells and oligodendrocytes, respectively, leading to progressive functional impairment.[10,11] Sustained CTL activation perpetuates chronic inflammatory cycles and cumulative tissue damage, thereby amplifying disease progression. The cytotoxic specificity of CTLs positions them as key therapeutic targets for interventional strategies mitigating autoimmune tissue injury.[12] Additionally, myeloid cells, such as monocytes, macrophages, and dendritic cells (DCs), play a pivotal role in the development and progression of system in autoimmune diseases like SLE, RA, and SSc.[6] In these diseases, myeloid cells contribute to immune responses by secreting pro-inflammatory cytokines and chemokines, as well as promoting the activation of autoreactive T and B cells.[13] Overexpression of *TNFSF13B* (*BAFF*) is critical to autoimmune pathology, as it allows autoreactive B cells to escape deletion, thereby promoting autoantibody production and correlating with increased disease activity in conditions such as SLE and MS.[14–16] However, precise characterization of immune cell subpopulation heterogeneity across autoimmune conditions requires systematic delineation.

Despite multiple integral pan-cancer single-cell RNA sequencing (scRNA-seq) analysis of immune cells having been performed in diverse cancer types,[17–20] scRNA-seq technology also provides the opportunity to investigate fluctuations in cell

status and the strength of cell plasticity and it is increasingly applied in autoimmune diseases.[21,22] scRNA-seq profiling of cerebrospinal fluid in MS has identified localized T and B cell interactions as critical disease mediators, exemplified by the upregulation of the BAFFR-driven non-canonical NF-κB pathway in memory B cells during active disease, which confirms active *TNFSF13B* (*BAFF*) signaling within the central nervous system.[23,24] In RA, analyses of synovial tissue have delineated distinct cellular and molecular pathways underlying seropositive versus seronegative disease subtypes.[25] Studies on SLE demonstrate three hallmark features: upregulation of type I interferon-stimulated genes (ISGs) in monocytes, a concomitant reduction of naive *CD4*$^+$ T cells correlating with monocytic ISG expression, and the expansion of clonally restricted *CD8*$^+$ *GZMH*$^+$ cytotoxic T cells.[26] What is more, SLE stratification based on *TNFSF13B* and IFN-I bioactivity for biologics and implications of *TNFSF13B* produced by glomeruli in lupus nephritis.[27] Research on UC has identified disease-specific *CD8*$^+$ T cell phenotypes with clonal expansion patterns and characterized terminally differentiated, dysfunctional IL-26-expressing *CD8*$^+$ T cell populations.[28] However, it is important to note that current scRNA-seq investigations remain constrained by limited sample sizes and disease-specific focus, raising concerns about the generalizability of these findings across the broader autoimmune disease spectrum.

Previous single-cell studies in pan-autoimmunity have provided valuable insights by focusing on either circulating cells (peripheral blood mononucleal cells),[29] or specific, isolated immune lineages, such as *CD4*$^+$ T cells[30] or myeloid cells.[31] However, these studies have largely centered on specific aspects of the immune system, a more comprehensive immune landscape that integrates multiple cell lineages and tissues remains to be established.

In this study, we integrated public and in-house scRNA-seq data to construct the immune landscape across 6 distinct autoimmune diseases, characterizing the molecular features of various immune cell types.[25,28,32–38] This provides a more comprehensive atlas than previous works by integrating data from all major immune lineages. Critically, we also included samples from local disease-affected tissues, not just peripheral blood, allowing us to construct a more complete single-cell map of the immune state in patients with autoimmune disease. We developed a practical disease classifier to distinguish autoimmune diseases using random forest algorithm. Our comprehensive immune cell compendium would provide an unprecedented resource for understand the global properties of the immune cells and guide the development of immunotherapies for multiple human autoimmune diseases.

RESULTS

Landscape of immune cells in 6 autoimmune diseases revealed by single cell RNA analysis

To characterize immune cells in autoimmune diseases, we compiled scRNA-seq data from both our newly generated dataset with 12 patients diagnosed with one of the four autoimmune diseases and ten additional published datasets. These data covered six autoimmune diseases, including 239 samples from 123 patients and 54 healthy controls (HCs) (Figure 1A, Tables S1 and S2). In total, our integrated atlas contains 41 samples with scTCR-seq information (Table S3). However, the public data within this set was limited almost exclusively to ulcerative colitis (UC). Our 12 in-house samples, which are included in this total, make a critical contribution by providing high-quality, paired scRNA-seq and scTCR-seq data for pSS, SLE, and RA. This addition was indispensable, as it uniquely allowed us to quantify clonal expansion and directly link transcriptional states to specific clonotypes, enabling a cross-disease T cell analysis that was not previously possible. After strict quality control and filtering, we collected a total of 1,332,505 high-quality cells derived from local tissues and peripheral blood across six autoimmune diseases (Figure 1B).

To unbiasedly reveal the population features of immune cells in autoimmune diseases, we integrated all datasets using the batch effect correction algorithm Harmony and performed unsupervised clustering with Seurat. As excepted, unsupervised clustering on immune cells revealed five distinct populations of T cells, NK cells, B cells, myeloid cells, and platelets (Figure S1A). Our integration validation confirms that technical batch effects (“Dataset”) were successfully removed while true biological structure (“Disease” and “Tissue”) was preserved (Figures 1D, S1B, and S1C). Therefore, our integration is robust, and our downstream analyses reflect genuine biological signals.

Further clustering of the aforementioned groups gave rise to 12 T cell subpopulations, two NK cell subpopulations, 17 myeloid cell subpopulations, six B cell subpopulations, and one platelet cell subpopulation, totaling 38 subpopulations, which were identified by differential expression of canonical cell markers (Figures 1C, 1D, and S1D–S1F). Specifically, T cells were identified by expression of CD3D, NK cells by FCGR3A and XCL1, B cells by CD79A, macrophages by CD68, pDC by CLEC4C, and plasma by MZB1 and XBP1 expression (Figures S1G–S1J). The Jaccard index results between our identified cell annotations and the original article’s clustering annotations demonstrate the stability of the clustering (Figure S2).

To systematically delineate shared and divergent immune dysregulation across autoimmune pathologies, we employed proportionality analysis of our integrated datasets. This approach revealed pan-disease enrichment of CD8T_CT_1, B_Naive cells, and non-classical monocytes compared to HCs, while other myeloid subpopulations exhibited disease-specific distribution patterns with distinct pathological enrichment profiles across different disorders. Specifically, patients with SSc showed an enrichment of the Macro_CXCL2 subpopulation, whereas the CD14Mono_AMBRA1 subpopulation was prominent in MS, pSS, and UC. Although AMBRA1 is not canonical, this finding is biologically robust: AMBRA1 regulates autophagy, a process implicated in autoimmunity,[39] and is linked to myeloid cell infiltration.[40] Meanwhile, the CD14Mono_LYZ subpopulation was distinctly enriched in patients with SLE (Figure 1E). In summary, we constructed a detailed transcriptomic landscape of immune cells at single-cell resolution across multiple autoimmune diseases.

Characterization of disease- and cell type-specific transcriptional features across autoimmune diseases

To identify shared and disease-specific transcriptional programs across diverse autoimmune pathologies, we developed a

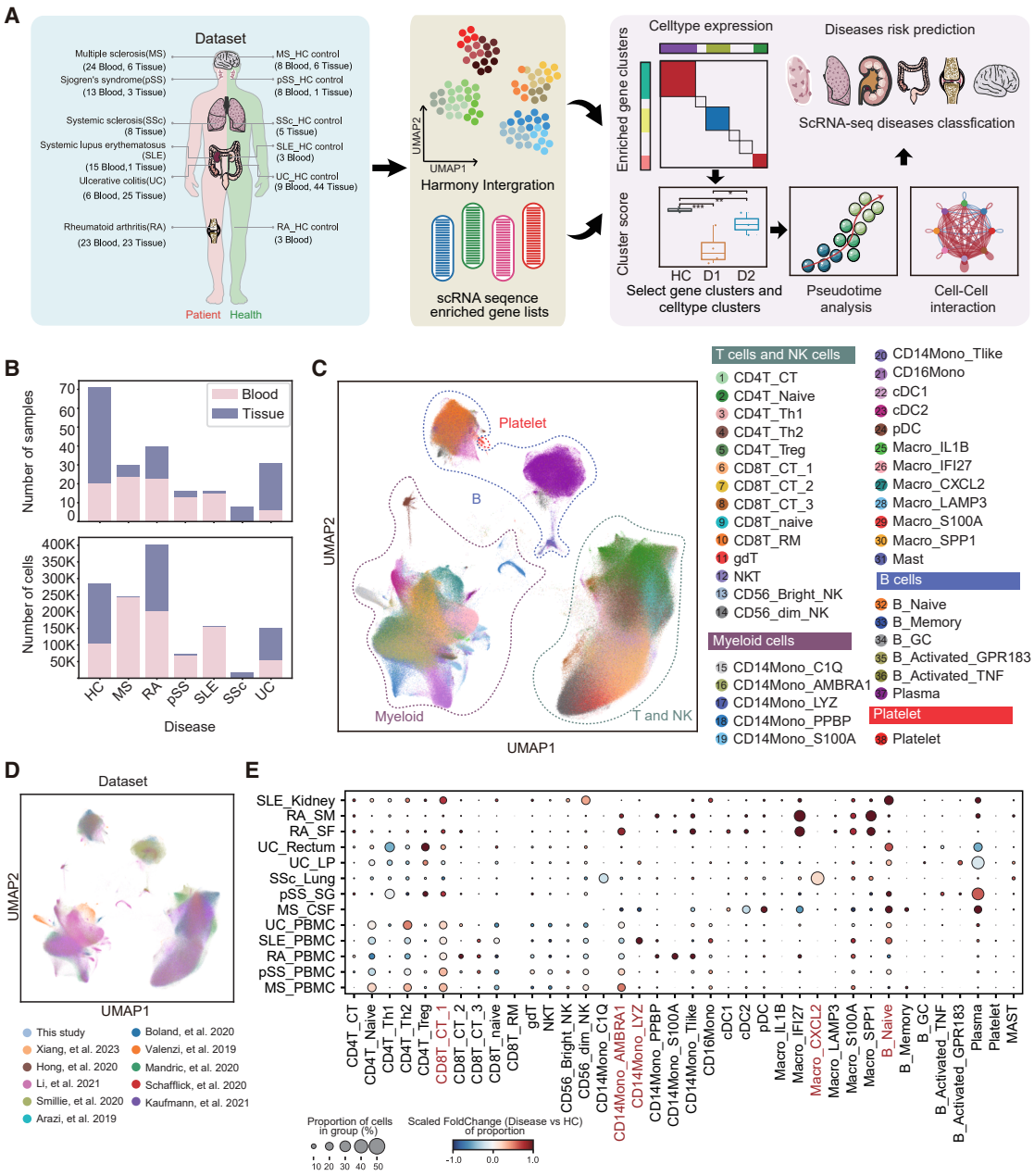


Figure 1. Overview of the pan-autoimmune disease immune cell atlas

(A) Schematic overview of the methodology of construction of the atlas, from data curation, through processing to analysis.

(B) Bar graphs representing the number of samples (top) and the number of cells (bottom) for each disease, stratified by tissue of origin.

(C) UMAP visualization of immune cells in the atlas.

(D) UMAP showing the distribution of data sources.

(E) Bubble plots show the expansion of cell subsets in the disease tissue relative to the healthy state. The bubble size represents the proportion of that cell subset to all cells in the disease of the corresponding tissue, and the bubble color represents the scaled fold change of the disease of the corresponding tissue relative to the change in the proportion of cells in the healthy person of the corresponding tissue.

See also Figures S1 and S2 and Tables S1, S2, S3, and S4.

framework to group co-regulated genes. We first calculated differentially expressed genes within each cell type-disease-tissue combination and then applied hierarchical clustering to group genes with similar expression patterns (see STAR Methods for details). This systematic approach identified 41 distinct gene clusters (GCs), which we define as modules of genes that exhibit coordinated transcriptional behavior across different cell types and pathological conditions (Figures 2A and

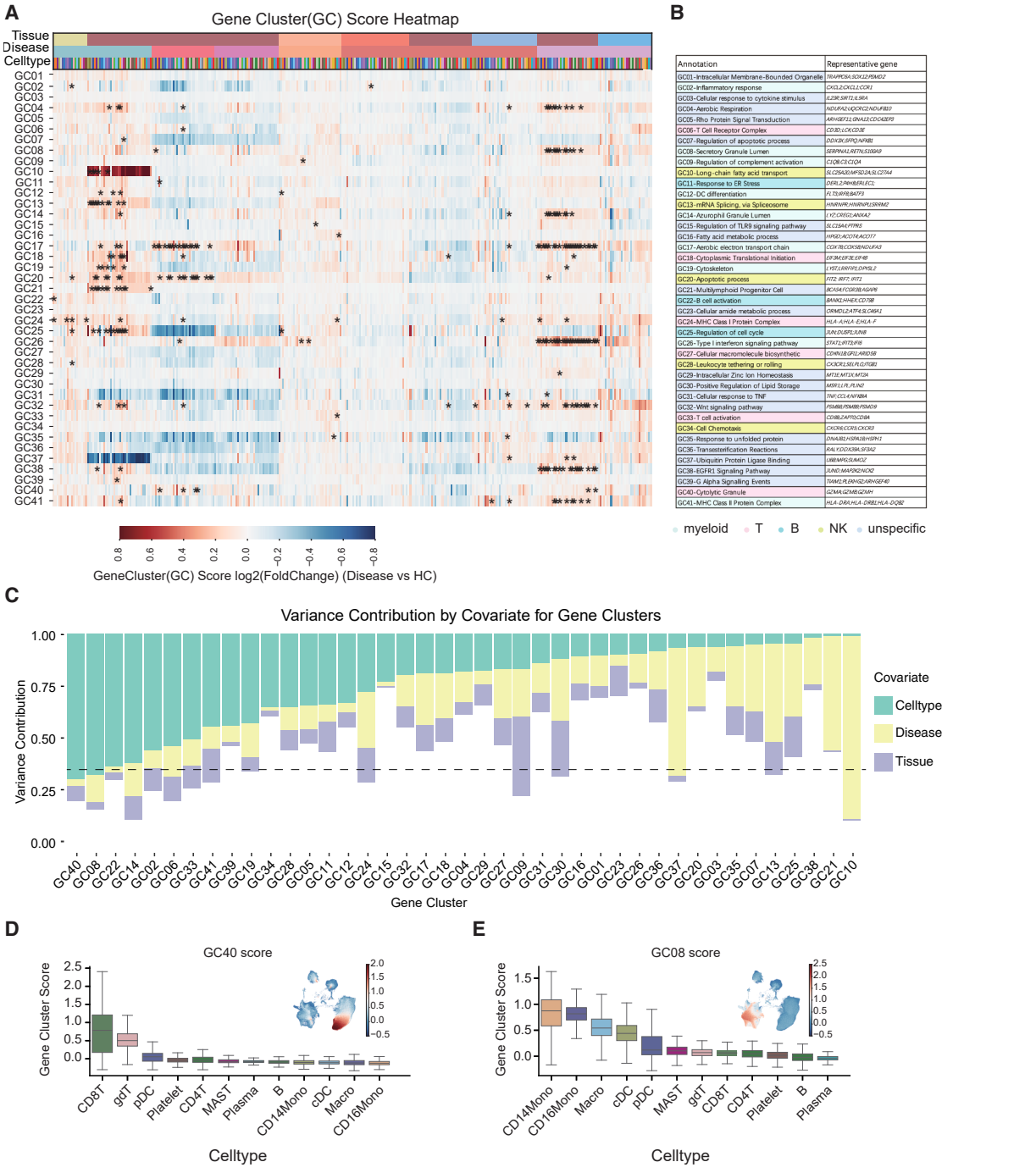


Figure 2. Identification of gene cluster (GC) through cross-comparison of autoimmune disease immune cell atlas

(A) Heatmap of gene cluster scores in different tissue disease cell subsets, asterisk indicates a gene cluster with p value <0.05, Mann-Whitney test, and |log2(fold change) of gene cluster score (disease vs. HC)| > 0.5.

(B) Annotation information of gene cluster and its representative genes.

(C) Bar plot of variance contributions for gene-list covariates, including cell subgroup, disease, and tissue, ranked from largest to smallest contribution by cell subgroup covariate.

(legend continued on next page)

S3A; Table S5). Furthermore, to ensure our GCs were robust to technical artifacts, we performed a variance contribution analysis (see Method details). This analysis (Figure S3B) confirmed that the variance explained by ''Dataset'' (batch) was minimal for all GCs (less than 0.15). Subsequent association analyses of these GCs revealed distinct, cell type-specific patterns. Eight GCs were associated with myeloid cells and were enriched in interferon responses (GC26), secretory granule lumen (GC08), and signal transduction pathways (GC09 and GC41), along with other clusters spanning metabolic, inflammatory, and regulatory processes while five GCs were correlated with NK cells, including cell chemotaxis (GC34) and cell rolling (GC28). Moreover, six GCs were correlated with T cell biology, including cytolytic granule (GC40) and T cell activation (GC33), while three GCs showed B cell linkage, involving B cell activation pathways (GC22). The remaining GCs lacked specific cellular associations, suggesting potential tissue-specific or disease-agnostic functions (Figure 2B).

To quantify the relative contributions of disease status, cellular identity, and tissue origin to transcriptional variation, a systematic variance partitioning analysis was conducted. Three distinct patterns emerged: (1) GC40 and GC08 exhibited predominant associations with cellular identity (*CD8*$^+$ T cells and *CD14*$^+$ monocytes, respectively) alongside additional contributions from disease status; (2) GC10, GC37, GC21, and GC13 demonstrated strong disease-driven regulation, representing conserved transcriptional programs across pathologies; and (3) a substantial proportion of GCs displayed hybrid regulatory influences, indicating complex interplay between biological variables (Figure 2C). To resolve the cell type-specific functional architecture of prioritized GCs, we then scored the GCs with a variance contribution score higher than 0.65 for each cell type and visualized the scores of each GC on the UMAP plot and boxplot. Cytotoxic GC40 demonstrated exclusive enrichment in *CD8*$^+$ T cell compartments, aligning with their effector roles (Figure 2D). Conversely, myeloid-enriched clusters (GC08) showed peak activity in *CD14*$^+$ monocytes, mirroring their functional specialization in secretory granule lumen (Figure 2E). Quantitative validation through boxplot analyses confirmed these distribution patterns with statistical rigor, demonstrating significant differences in GC activity scores across cellular subsets. Collectively, our data delineated disease-specific transcriptional networks from constitutive cellular processes while revealing conserved pathogenic modules across immune disorders, establishing a mechanistic hierarchy of transcriptional regulation through systems-level resolution.

Enhanced cytotoxicity and clonal expansion of CD8T_CT_1 cells in autoimmune diseases

Given the central role of *CD8*$^+$ T cells in adaptive immunity,[41] we conducted quantitative scoring analyses to delineate the association between CD8T-associated GC40 and *CD8*$^+$ T subpopulations. Systematic evaluation revealed CD8T_CT_1 as the dominant subset exhibiting the strongest GC40 correlation (Figures 3A–3C). In CD8T_CT_1 subgroup, we found significant differences in GC40 scores between MS, RA, SLE, pSS, and healthy subjects (Figure S4A). To molecularly validate this association, we performed gene ontology(GO) enrichment analysis,[43] demonstrating concordant enrichment of both CD8T_CT_1 upregulated genes and GC40 components in cytolytic granule pathway (Figure S4B). Expanding this observation to disease contexts, single-cell profiling revealed significant elevation of GC40 scores in autoimmune conditions compared to HCs, particularly across cytotoxicity and TCR signaling pathways (Figure 3D and Table S6), suggesting conserved pathogenic activation patterns. Consistent with this, quantification of cytotoxic molecule expression between morbid and healthy subpopulations confirmed disease-specific functional upregulation (Figure 3E). UMAPs show the expression of representative cytotoxic molecules, *GZMA* and *GZMH*, across disease conditions, further supporting the disease-specific activation of cytotoxic programs (Figure S4C).

To delineate the functional trajectory of autoimmune-associated cytotoxic T cell populations, we conducted longitudinal profiling of clonal dynamics across disease progression stages. Notably, comparative cellular indexing demonstrated preferential expansion of CD8T_CT_1 within the CD8$^+$ T cell compartment (Figure 3F), hinting at clonal proliferation dynamics. This hypothesis was substantiated by TCR repertoire analysis, where CD8T_CT_1 exhibited the most pronounced clonal expansion signature in pathological states (Figure 3G). Collectively, these findings establish enhanced cytotoxic functionality coupled with clonal expansion as hallmark features of CD8$^+$ T cells in autoimmune pathogenesis.[44–47]

To trace the developmental origins of CD8T_CT_1, we integrated STARTRAC lineage tracing with single-cell trajectory modeling (Figure S5A), showing that the transition probability (pIndex.tran) from CD8T_naive to the CD8T_CT_1 state was stronger in disease patients than in HCs. Computational deconvolution revealed naive T cells as the predominant precursors differentiating into cytotoxic effectors under disease conditions. Careful analysis of individual TCR clonal distribution revealed identical clonotypes shared between the CD8T_naive population and its terminally differentiated progeny, CD8T_CT_1 progeny (Figure 3H). Spatiotemporal validation through CellRank pseudotemporal ordering delineated a progressive differentiation trajectory from naive states to CD8T_CT_1 (Figure 3I). When examining each disease individually, *CD8*$^+$ T cells exhibited the same transition trend from CD8T_Naive subset to CD8T_CT_1 subset (Figures S5B and S5C). Furthermore, pseudotemporal analysis along this trajectory (Figure 3J) revealed distinct pathogenic dynamics: ''Disease'' cells exhibited a more rapid upregulation of the cytotoxic marker *GZMA*, concurrent with a faster loss of the naive marker IL7R, compared to HCs.

(D) Boxplot showing the distribution of GC40 score (left). UMAP showing the distribution of GC40 score (right). (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR).

(E) Boxplot showing the distribution of GC08 score (left). UMAP showing the distribution of GC08 score (right). (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR).

See also Figure S3 and Table S5.

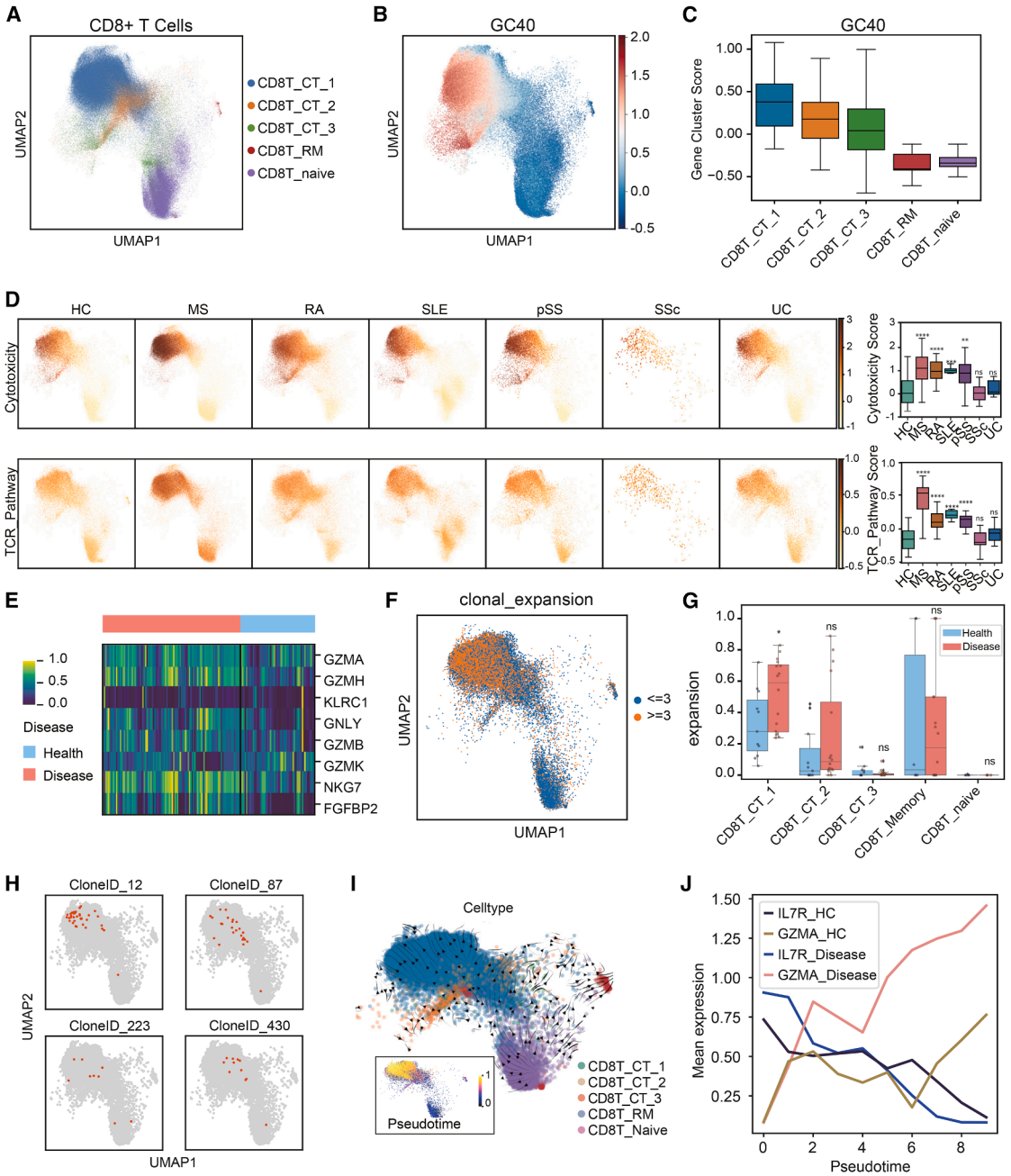


Figure 3. T cells are significantly expanded and their cytotoxicity function is upregulated in patients with autoimmune diseases

(A) UMAP plots showing distribution of *CD8*$^+$ T cell subsets.

(B) UMAP plots as in (A) but showing distribution of GC40 score.

(C) Boxplot showing the GC40 score in different *CD8*$^+$ T subsets. (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR).

(D) UMAP showing cytotoxicity score (up left) and TCR pathway score (bottom left) in different diseases. Boxplot showing the cytotoxicity scores (up right) and TCR pathway scores (bottom right) between autoimmune disease and healthy control T cells. (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR, Mann-Whitney test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$).

(E) Mean expression of cytotoxicity molecules in disease and healthy states. Rows represent individual genes, and columns represent individual samples. The color scale indicates relative gene expression levels, ranging from low (purple) to high (yellow).

(F) UMAP showing the distribution of clonal expansion in T cells. Cells with a clonal expansion number greater than three are shown in orange, and the remaining cells are shown in blue.

(G) Proportion of subsets with clonal expansion greater than 3 within *CD8*$^+$ T cell subsets between autoimmune disease and healthy control. (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR, *t* test, *$p < 0.05$).

(legend continued on next page)

Mechanistically, CD8T_CT_1-mediated immunopathology appears driven by stromal crosstalk. Comparative interactome profiling demonstrated disease-specific intensification of CD8T_CT_1-stromal interactions as assessed by CellChat (Figure S6A). NicheNet network analysis of differentially expressed ligand-receptor pairs identified a conserved upregulation of IFNG signaling in CD8T_CT_1 subsets across multiple pathologies, with parallel activation of cognate receptor pathways in stromal compartments (Figure S6B). Cross-disease validation confirmed IFNG hyperactivation in pSS and SLE microenvironments (Figure S6C), positioning IFNG-mediated T cell-stromal communication as a central axis of cytotoxic effector function in autoimmune pathogenesis. Thus, our findings identified GC40-associated cytotoxic transcriptional programming and CD8T_CT_1 clonal expansion as central drivers of autoimmunity, unified through IFNG-mediated stromal crosstalk.

Secretory *CD14*$^+$ monocytes subpopulations drive enhanced *TNFSF13B* signaling in autoimmune diseases

To visualize the distribution of disease-associated gene programs, we focused on the myeloid populations, which we clustered into distinct subtypes (Figure 4A). Our integration analysis shows that circulating monocyte populations are well-integrated across datasets, while the biological separation of tissue-resident/disease-specific macrophages is correctly preserved (Figures S7A–S7C). This confirmed that these subtypes represent robust biological entities, not technical artifacts. Having validated this UMAP, we then projected the GC08 gene score, revealing its specific enrichment pattern within these myeloid subtypes (Figure 4B). To further elucidate myeloid cell subtypes associated with disease clusters GC08, we performed GC08 gene scores across myeloid subpopulations. CD14Mono populations received the highest enrichment scores (Figure 4C). Comparative assessment across disease cohorts and HCs revealed significant upregulation of GC08 in MS, SLE, and pSS patients (Figure 4D). GO enrichment analysis[43] provided molecular validation for this association, revealing a shared enrichment of upregulated genes in *CD14*$^+$ monocytes and GC08 constituents within the secretory granule lumen (Figure S7D and Table S6). The secretory granule lumen acts as a crucial reservoir within myeloid cells, concentrating and storing pre-synthesized cytokines. This enables the cell to mount a rapid and massive response by releasing these potent signaling molecules immediately upon immune stimulation through a process called degranulation. Building on this principle, we compared the expression of cytokine-related genes across different myeloid cell subsets to identify their specific cytokine profiles. By highlighting genes expressed in over 20% of cells within a given subpopulation (fraction of cells in group >0.2), our analysis revealed that *CD14*$^+$ monocytes predominantly express a distinct set of cytokines, including *TNFSF13B*, *CCL5*, *PPBP*, *IL32*, *TNFSF12*, and *IL1B*. Notably, *TNFSF13B* emerged as a key factor, demonstrating particularly high and widespread expression throughout the *CD14*$^+$ monocyte population (Figure S7E). To investigate the clinical relevance of these cytokines characteristically secreted by *CD14*$^+$ monocytes, we compared their gene expression levels under various disease states. We found that the vast majority of these cytokines were upregulated in MS, RA, SLE, and pSS relative to HCs. Specifically, the expression of *TNFSF13B* was elevated in MS, SLE, and pSS compared to healthy individuals, highlighting its potential role in the pathology of these autoimmune diseases (Figure S7F).

Recognizing the functional convergence of secretory mechanisms in intercellular communication, we performed a systematic interaction analysis using CellChat. This revealed that, compared to HCs, disease states were characterized by a notable increase in both the strength and number of interactions between the *CD14*$^+$ monocyte population and both plasma cells and macrophages (Figure S7G). Delving into the specific outgoing signals driving this change, we identified a notable upregulation in the *TNFSF13B-TNFRSF17* interaction between *CD14*$^+$ monocytes and plasma cells, as well as in the PPIA-BSG interaction between *CD14*$^+$ monocytes and macrophages, in disease states relative to HCs (Figure 4E).

Our interaction analysis identified significant signaling from *CD14*$^+$ Monocytes to Plasma Cells, which was enriched in disease states. The primary drivers of this predicted interaction were the *TNFSF13B* (BAFF) and APRIL signaling pathways. Subsequent NicheNet analysis focusing on the plasma cells (receiver) and *CD14*$^+$ monocytes (sender) populations identified TNFSF13B as the second-ranked ligand, based on area under the precision-recall curve (AUPR) analysis, along with augmented *TNFSF13B-TNFSF17* ligand-receptor interactions (Figure S7H, left). Dissecting these CellChat results at the individual disease level, we found that the interaction between *CD14*$^+$ monocytes and plasma cells was significantly upregulated in MS, pSS, and SSc relative to HCs—a trend consistent with the GC08 score behavior across these conditions (Figure 4F). A more granular analysis of the contributing ligand-receptor pairs confirmed that this enhanced crosstalk was predominantly driven by *TNFSF13B*-mediated signaling, specifically through the *TNFSF13B-TNFRSF17* and *TNFSF13B-TNFRSF13C* interactions. Our study elucidates a key pathogenic mechanism wherein *CD14*$^+$ monocytes, through the upregulation of the secretory ligand *TNFSF13B*, drive enhanced cellular crosstalk with plasma cells across a spectrum of autoimmune diseases. Building on our finding of enriched monocyte-to-plasma cell communication, we propose a key pathogenic mechanism. The over-expression of *TNFSF13B* (*BAFF*) and *APRIL* by activated *CD14*$^+$ monocytes likely creates a pro-survival niche. This excessive signaling can rescue auto-reactive plasma cell clones from apoptosis, allowing them to persist and secrete pathogenic auto-antibodies.[48] This mechanism is consistent with the well-documented role of *BAFF* in autoimmune pathogenesis[49] and the clinical efficacy of BAFF-inhibiting therapies in conditions like SLE.[50]

(H) UMAP showing the distribution of the same clones in T cells.
(I) Developmental trajectory of T subsets inferred by CellRank.[42]
(J) The patterns of gene expression along with the pseudotime. (Data are represented as the mean expression level).
See also Figures S4–S6 and Table S6.

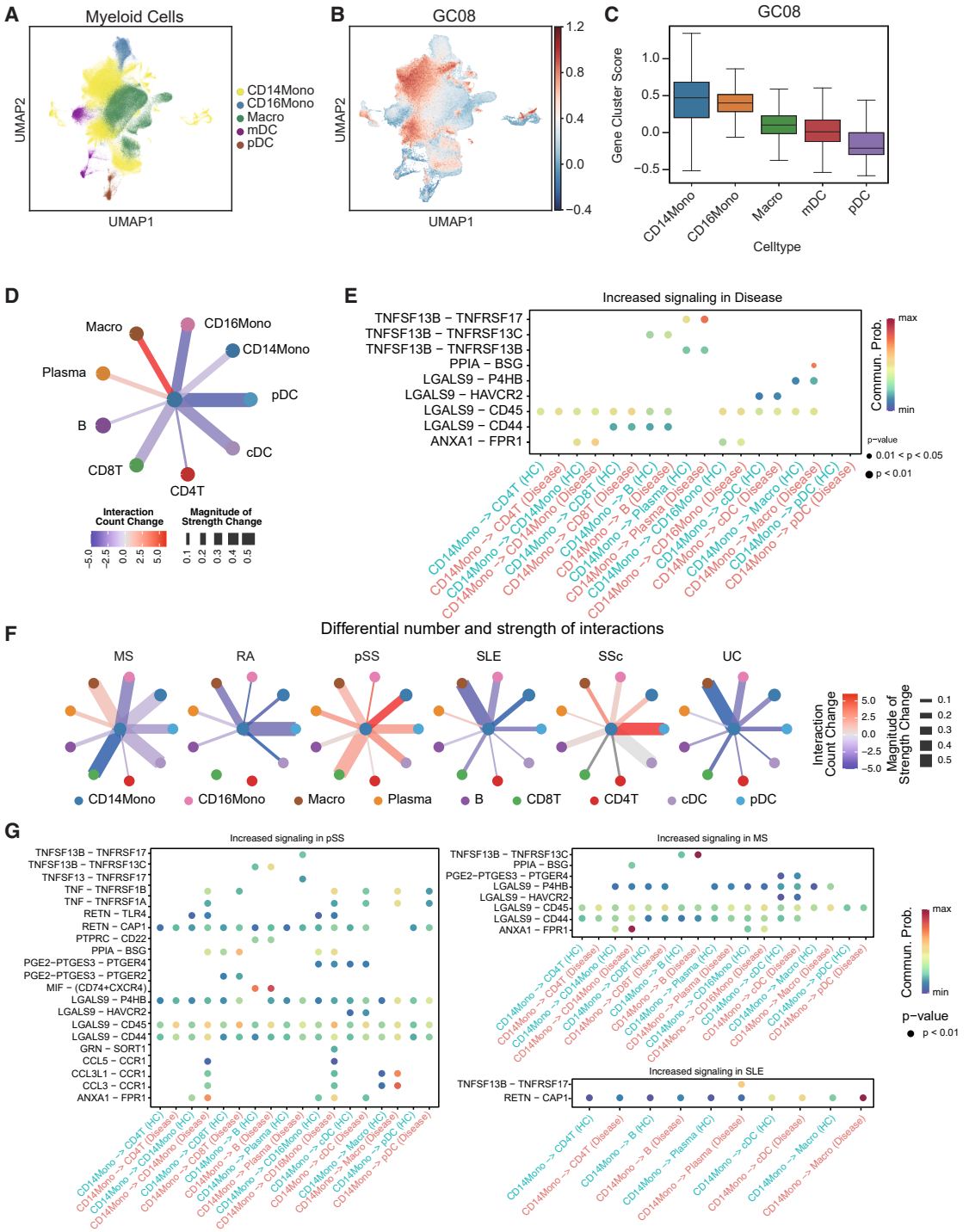


Figure 4. Secretory *CD14*$^{+}$ monocytes drive enhanced *TNFSF13B* signaling in autoimmune diseases

(A) UMAP plots showing distribution of myeloid cell subsets.

(B) UMAP plots as in (A) but showing distribution of GC08 score.

(C) Boxplot showing the GC08 score in different myeloid subsets. (Data are represented as median and interquartile range, with whiskers extending to 1.5×IQR).

(D) Differential number and strength of interactions between autoimmune disease and healthy subpopulations in *CD14*$^{+}$ monocytes. Red lines indicate interactions number upregulated in the disease state, blue lines indicate interactions upregulated in the healthy state, and line width reflects interaction strength.

(E) The bubble plot shows significantly upregulated communications in autoimmune disease samples compared to healthy controls. Each row corresponds to a ligand-receptor pair, and each column represents an interaction from *CD14*$^{+}$ monocytes source cell to other cells. The color indicates the communication

(legend continued on next page)

Application of random forest classifier in predicting autoimmune diseases

Recent advancements in artificial intelligence have revolutionized tumor classification methodologies and deepened our understanding of disease pathogenesis.[51–54] Building upon the newly identified shared and disease-specific gene signatures, we developed a machine learning framework for the automated classification of autoimmune diseases through a two-stage hierarchical approach.[55,56] The primary screening step effectively differentiates between patients and HCs, while the secondary subclassification module enables the precise identification of specific autoimmune conditions, including MS, RA, SLE, SSc, pSS, and UC (Figure 5A). The model demonstrated exceptional performance with near-perfect accuracy in distinguishing autoimmune patients from HCs. Disease-specific classification achieved robust accuracy, with scores exceeding 0.8 for major autoimmune conditions. However, it showed limited efficacy in differentiating UC from HCs (Jaccard index = 0.70), a finding that may be attributed to genetic and phenotypic similarities between UC and non-disease states (Figure 5B). To rule out batch effects, we performed a ''leave-one-study-out'' cross-validation using our pSS cohort. We held out one pSS dataset (''GSE157278'') and re-trained the classifier on the remaining data, including another pSS dataset (''HRA003613''). The model demonstrated strong generalization when predicting the held-out ''GSE157278'' samples (Figure 5C), confirming it learned a robust biological signature for pSS. To test the generalizability of our classifier, we projected an external dataset from an unseen autoimmune disease not included in our training data: psoriasis (PS) (Table S8).[57–66] As shown in Figure 5C (left and middle), cells from this novel disease mapped almost entirely within our ''unlabeled'' (general disease) region. We quantified this projection using a Jaccard index heatmap (Figure 5C, right), which confirmed that 91% of the PS cells were correctly classified as ''unlabeled'' (the general disease state).

We further evaluated the classifier's performance using ROC curves, accuracy, and precision metrics (Figures 5E–5G). Consistent with the Jaccard index results, the classifier effectively differentiated between disease and non-disease states, achieving an overall prediction accuracy of 95%. Additionally, the model's ROC-AUC for classifying HCs was 0.87, a result likely constrained by the low discriminability between HC and UC, the pair with the poorest classification performance. This robust performance across multiple evaluation parameters confirms that our disease-differentiating gene signatures effectively capture critical pathophysiological features of autoimmune disorders. To ensure methodological rigor, we validated our findings using two distinct random forest implementations (SCVI and Seurat), both of which demonstrated performance comparable to our primary classifier (Figure S8). These consistent results across multiple analytical platforms highlight the reliability of our model. Taken together, we present an efficient random forest classifier for the prediction of autoimmune diseases, showcasing its potential as a valuable tool for disease diagnosis and classification.

DISCUSSION

Our comprehensive analysis of over 1.3 million cells from 239 samples across six autoimmune diseases systematically delineates the previously underappreciated complexity of immune cell heterogeneity, providing new insights into disease-specific and cell type-driven transcriptional reprogramming. By integrating single-cell transcriptomic profiling with functional annotations, we reveal how distinct immune cell subsets contribute to autoimmune pathogenesis through divergent molecular mechanisms and intercellular crosstalk.

The identification of 41 GCs with cell lineage- and disease state-specificity provides a granular view of autoimmune-associated transcriptional dysregulation. Variance decomposition further highlights a critical bifurcation in regulatory control: while clusters such as GC40 and GC08 reflect cell-intrinsic signatures (e.g., T cell cytotoxicity or myeloid secretory responses), suggesting their potential utility as context-dependent biomarkers.

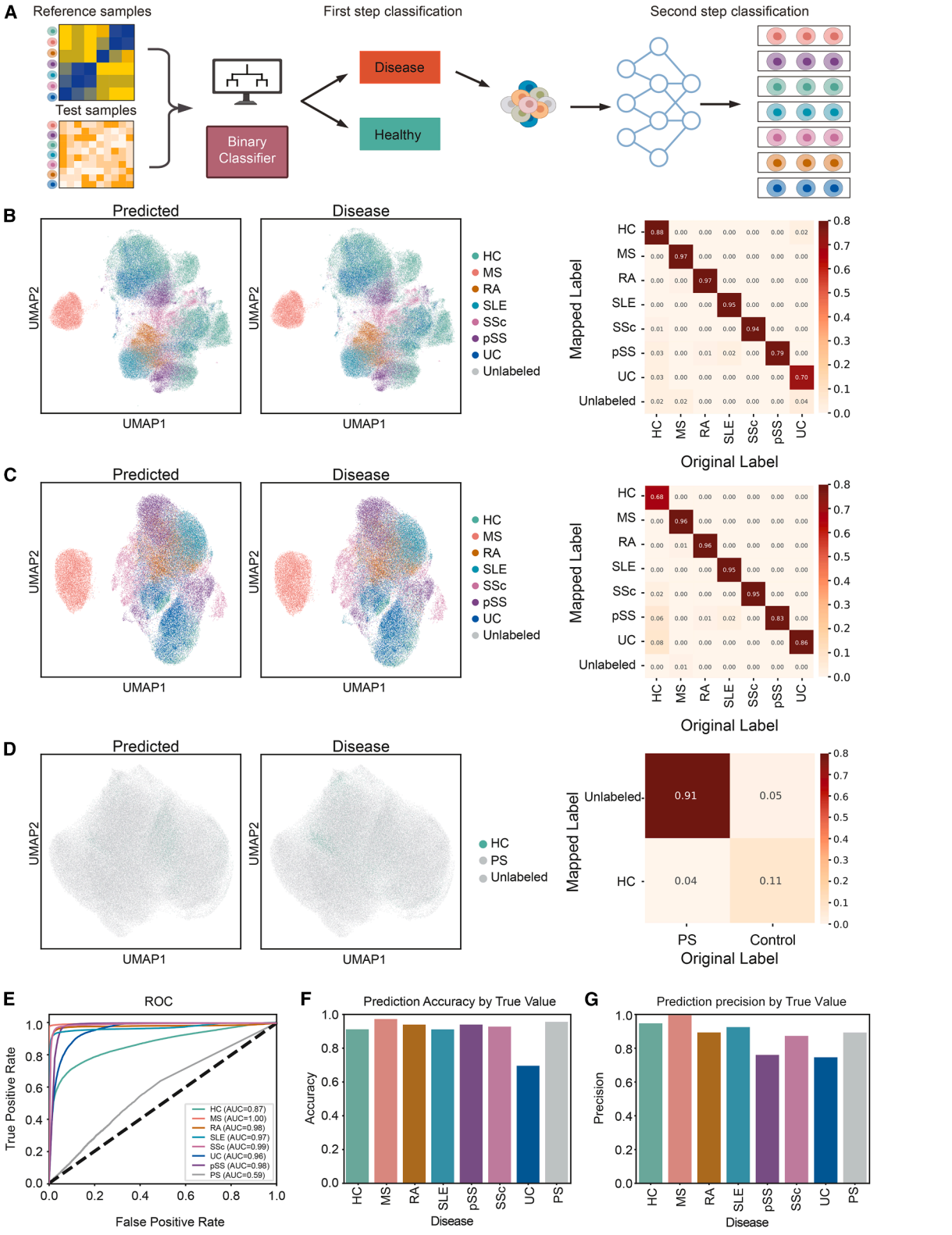
A central finding is the clonal expansion of cytotoxic CD8T_CT_1 cells from naive precursors in autoimmune states, marked by heightened expression of effector molecules (*GZMB* and *GZMK*) and TCR signaling activation. This observation aligns with emerging paradigms of antigen-driven T cell pathology in autoimmunity[67,68] and positions CD8T_CT_1 as a therapeutic target for mitigating cytotoxic tissue damage.[44] In line with previous studies, CD8T-CT_1 was found to interact with stromal cells[47,69,70] through interferon gamma.[71,72] Similarly, interferons have been highlighted as critical mediators in autoimmune disease pathogenesis.[73–76] Complementary myeloid analyses reveal that *CD14*$^+$ monocytes orchestrate inflammatory cascades via GC08-associated secretory programs and *TNFSF13B-TNFRSF17/TNFRSF13C*-mediated crosstalk with plasma cells.[72] Previous studies have shown that upregulation of *TNFRSF17* signaling in autoimmune diseases is associated with disease development and progression.[14–16,24,27,77–80] Our comprehensive single-cell atlas across six autoimmune diseases identifies two complementary arms of immunopathology. We reveal the antigen-driven clonal expansion of cytotoxic *CD8*$^+$ T cells, alongside the orchestration of inflammatory networks by *CD14*$^+$ monocytes through *TNFSF13B*-mediated crosstalk with

probability in autoimmune disease group, with red denoting higher strength. The dot size corresponds to the *p* value from the differential analysis, where larger dots indicate more significant upregulation in autoimmune disease. (One-sided permutation test, *p* values are shown according to the size of the circles in the figure).

(F) Differential number and strength of interactions between six diseases and healthy subpopulations in *CD14*$^+$ monocytes. Red lines indicate interactions number upregulated in the disease state, blue lines indicate interactions upregulated in the healthy state, and line width reflects interaction strength.

(G) The bubble plot shows significantly upregulated communications in pSS, MS, and SLE samples compared to healthy controls. Each row corresponds to a ligand-receptor pair, and each column represents an interaction from *CD14*$^+$ monocytes source cell to other cells. The color indicates the communication probability in autoimmune disease group, with red denoting higher strength. The dot size corresponds to the *p* value from the differential analysis, where larger dots indicate more significant upregulation in autoimmune disease (one-sided permutation test, $p < 0.01$).

See also Figure S7 and Table S6.



(legend on next page)

plasma cells. These bidirectional interactions may create self-sustaining inflammatory circuits, offering mechanistic explanations for disease chronicity.

Our machine learning framework advances autoimmune diagnostics by achieving >90% classification accuracy for MS, RA, SSc, and SLE through large-scale data integration. The model's reduced efficacy in distinguishing UC from HCs (Jaccard index = 70%) is likely due to overlapping transcriptional signatures, echoing recent reports of UC's unique molecular proximity to non-disease states.[81,82] This limitation highlights the necessity of incorporating proteomic, metabolomic, or epigenetic layers to resolve phenotypically ambiguous conditions.

In conclusion, our study advances the current understanding of autoimmune diseases from an integrative perspective, shedding light on the heterogeneity of immune cells and developing a scRNA-sequencing data-based disease classifier. We hope that our comprehensive dataset will serve as a valuable resource for understanding immune cells in various autoimmune diseases and contribute to new immune cell-targeted therapy development and application in the future.

Limitations of the study

Although our study provides proof of principle as the first single-cell analysis of multiple autoimmune diseases, it has several limitations. First, the limited sample size and lack of comprehensive clinical information from public datasets may affect the generalizability of our findings. In particular, the predominant female representation in our cohort, while reflecting the epidemiology of autoimmune diseases, limits our ability to evaluate sex-specific differences or generalize findings to male patients. A larger cohort with matched clinical information is expected to improve the robustness of our results. Second, we only proved the possible alternative origin of CD8T_CT_1 mainly through bioinformatic trajectory analyses. It would be interesting to explore the spatial relationship between CD8T_CT_1 and naive T cells through mIHC with specific markers. Third, our analysis relies exclusively on transcriptomic data. It is well-established that mRNA levels do not always perfectly correlate with protein expression and subsequent cellular function. Therefore, future studies should incorporate proteomics and direct functional assays to validate that the observed transcriptional changes in pathways like cytotoxicity and cytokine secretion translate to functional cellular actions. Finally, distinguishing autoimmune diseases has been achieved through random forest classifier. In future enhancements, the accuracy and efficiency will need to be tested with extended autoimmune disease dataset and to compare with other machine learning approaches.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kun Qu (qukun@ustc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The sequencing data have been deposited in the Genome Sequence Archive (GSA-human: HRA011522) in the National Genomics Data Center, China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences. These data are under controlled access and available via the Data Access Committee (DAC).
- This article analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- The custom scripts for this study have been deposited at GitHub (https://github.com/QuKunLab/pan-autoimmune_disease) are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all members in the Qu laboratory for helpful discussions. This work was supported by the National Natural Science Foundation of China grants (92574202, T2125012, and 31970858 to K.Q., 32471445 to J.L.), the CAS Project for Young Scientists in Basic Research (YSBR-005 to K.Q.), Anhui Province Science and Technology Key Program (202003a07020021 to K.Q.), Anhui Provincial Natural Science Foundation (2508085MH188 to X.L.), USTC Research Funds of the Double First-Class Initiative (YD9100002032 to K.Q.), USTC Original exploration project (YD9100002026 to K.Q.), and the Fundamental Research Funds for the Central Universities (WK9110000141 to K.Q. and WK9100000063 to X.L.). We thank the USTC Supercomputing Center and the School of Life Science Bioinformatics Center for providing computing resources for this project.

AUTHOR CONTRIBUTIONS

K.Q., X. Liu, and J.L. conceived the project. L.Z., W.S., Q.L., and W.Z. designed the framework. L.Z. and C.J. performed all the wet laboratory experiments with helps from J.L.; L.Z. and W.S. performed all the bioinformatics analysis with helps from K.C., Z.L., M.T., C.J., Y.S., Q.L., W.Z., J.W., and A.Z.; L.Z. collected all the samples with helps from X. Li and G.W.; L.Z., K.Q., and X. Liu wrote the manuscript with inputs from all authors. K.Q. supervised the project.

Figure 5. Two-stage random forest classifier in predicting autoimmune diseases

(A) Two-stage random forest classifier workflow.

(B) UMAP showing the distribution of predicted disease (left) and true disease (middle) using autoimmune disease data, the heatmap shows the Jaccard index of the predicted results and reference values (right).

(C) UMAP showing the distribution of predicted disease (left) and true disease (middle) using autoimmune disease data, the heatmap shows the Jaccard index of the predicted results and reference values (right) with leave-one-out way of the SCVI-tools results.

(D) UMAP showing the distribution of predicted disease (left) and true disease (middle) using psoriasis (PS) data, the heatmap shows the Jaccard index of the predicted results and reference values (right).

(E) ROC curves for predicting different autoimmune diseases.

(F) Prediction accuracy by true value of different autoimmune diseases. Data are represented as the classification accuracy score for each group.

(G) Prediction precision by true value of different autoimmune diseases. Data are represented as the classification precision score for each group.

See also Figure S8 and Tables S7 and S8.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT 3.5 and ChatGPT 4.0 in order to improve the language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human samples and ethical permission
- METHOD DETAILS
  - Single cell collection, sorting, library preparation and sequencing
  - Single-cell RNA-Seq datasets collected in this study
  - Single-cell RNA-seq data processing
  - Integration of multiple scRNA-seq datasets by Harmony
  - Correlation analysis
  - Differential expression analysis
  - Autoimmune disease up-regulated gene set
  - Metacell
  - Gene clusters
  - Scoring celltype by using signature genes for each subset
  - Gene ontology enrichment analysis
  - Gene set variance partition analysis
  - Gene cluster score calculation and disease-wise comparison
  - Monocle
  - CellRank analysis of differentiation potential
  - TCR clone calculation
  - STARTRAC
  - CellphoneDB analysis
  - NicheNet analysis
  - CellChat analysis
  - Two-stage classification
  - Leave-one-study-out cross-validation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical test

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.114515.

Received: August 15, 2025
Revised: November 7, 2025
Accepted: December 17, 2025
Published: December 20, 2025

REFERENCES

1. Davidson, A., and Diamond, B. (2001). Autoimmune diseases. N. Engl. J. Med. *345*, 340–350. https://doi.org/10.1056/nejm200108023450506.
2. Pisetsky, D.S. (2023). Pathogenesis of autoimmune disease. Nat. Rev. Nephrol. *19*, 509–524. https://doi.org/10.1038/s41581-023-00720-1.
3. Abend, A.H., He, I., Bahroos, N., Christianakis, S., Crew, A.B., Wise, L.M., Lipori, G.P., He, X., Murphy, S.N., Herrick, C.D., et al. (2024). Estimation of prevalence of autoimmune diseases in the United States using electronic health record data. J. Clin. Investig. *135*, e178722. https://doi.org/10.1172/JCI178722.
4. Sun, L., Su, Y., Jiao, A., Wang, X., and Zhang, B. (2023). T cells in health and disease. Signal Transduct. Target. Ther. *8*, 235. https://doi.org/10.1038/s41392-023-01471-y.
5. Hampe, C.S. (2012). B Cell in Autoimmune Diseases. Cell in Autoimmune Diseases *2012*, 215308. https://doi.org/10.6064/2012/215308.
6. Morell, M., Varela, N., and Marañón, C. (2017). Myeloid Populations in Systemic Autoimmune Diseases. Clin. Rev. Allergy Immunol. *53*, 198–218. https://doi.org/10.1007/s12016-017-8606-7.
7. Lee, D.S.W., Rojas, O.L., and Gommerman, J.L. (2021). B cell depletion therapies in autoimmune disease: advances and mechanistic insights. Nat. Rev. Drug Discov. *20*, 179–199. https://doi.org/10.1038/s41573-020-00092-2.
8. Bekic, M., and Tomic, S. (2023). Myeloid-derived suppressor cells in the therapy of autoimmune diseases. Eur. J. Immunol. *53*, e2250345. https://doi.org/10.1002/eji.202250345.
9. Chung, J.B., Brudno, J.N., Borie, D., and Kochenderfer, J.N. (2024). Chimeric antigen receptor T cell therapy for autoimmune disease. Nat. Rev. Immunol. *24*, 830–845. https://doi.org/10.1038/s41577-024-01035-3.
10. Skowera, A., Ladell, K., McLaren, J.E., Dolton, G., Matthews, K.K., Gostick, E., Kronenberg-Versteeg, D., Eichmann, M., Knight, R.R., Heck, S., et al. (2015). beta-cell-specific CD8 T cell phenotype in type 1 diabetes reflects chronic autoantigen exposure. Diabetes *64*, 916–925. https://doi.org/10.2337/db14-0332.
11. Shi, Z., Wang, X., Wang, J., Chen, H., Du, Q., Lang, Y., Kong, L., Luo, W., Qiu, Y., Zhang, Y., et al. (2023). Granzyme B + CD8 + T cells with terminal differentiated effector signature determine multiple sclerosis progression. J. Neuroinflammation *20*, 138. https://doi.org/10.1186/s12974-023-02810-0.
12. Friese, M.A., and Fugger, L. (2005). Autoreactive CD8+ T cells in multiple sclerosis: a new target for therapy? Brain *128*, 1747–1763. https://doi.org/10.1093/brain/awh578.
13. Song, Y., Li, J., and Wu, Y. (2024). Evolving understanding of autoimmune mechanisms and new therapeutic strategies of autoimmune disorders. Signal Transduct. Target. Ther. *9*, 263. https://doi.org/10.1038/s41392-024-01952-8.
14. Steri, M., Orrù, V., Idda, M.L., Pitzalis, M., Pala, M., Zara, I., Sidore, C., Faà, V., Floris, M., Deiana, M., et al. (2017). Overexpression of the Cytokine BAFF and Autoimmunity Risk. N. Engl. J. Med. *376*, 1615–1626. https://doi.org/10.1056/NEJMoa1610528.
15. Zhang, Y., Tian, J., Xiao, F., Zheng, L., Zhu, X., Wu, L., Zhao, C., Wang, S., Rui, K., Zou, H., and Lu, L. (2022). B cell-activating factor and its targeted therapy in autoimmune diseases. Cytokine Growth Factor Rev. *64*, 57–70. https://doi.org/10.1016/j.cytogfr.2021.11.004.
16. Xiao, F., Rui, K., Shi, X., Wu, H., Cai, X., Lui, K.O., Lu, Q., Ballestar, E., Tian, J., Zou, H., and Lu, L. (2022). Epigenetic regulation of B cells and its role in autoimmune pathogenesis. Cell. Mol. Immunol. *19*, 1215–1234. https://doi.org/10.1038/s41423-022-00933-7.
17. Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., Ren, X., Wang, L., Wu, X., Zhang, J., et al. (2021). Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science *374*, abe6474. https://doi.org/10.1126/science.abe6474.
18. Tang, F., Li, J., Qi, L., Liu, D., Bo, Y., Qin, S., Miao, Y., Yu, K., Hou, W., Li, J., et al. (2023). A pan-cancer single-cell panorama of human natural killer cells. Cell *186*, 4235–4251.e20. https://doi.org/10.1016/j.cell.2023.07.034.
19. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell *184*, 792–809.e23. https://doi.org/10.1016/j.cell.2021.01.010.

20. Yang, Y., Chen, X., Pan, J., Ning, H., Zhang, Y., Bo, Y., Ren, X., Li, J., Qin, S., Wang, D., et al. (2024). Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. Cell *187*, 4790–4811.e22. https://doi.org/10.1016/j.cell.2024.06.038.

21. Fan, H.C., Fu, G.K., and Fodor, S.P.A. (2015). Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. Science *347*, 1258367. https://doi.org/10.1126/science.1258367.

22. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382. https://doi.org/10.1038/nmeth.1315.

23. Straeten, F., Zhu, J., Börsch, A.L., Zhang, B., Li, K., Lu, I.N., Gross, C., Heming, M., Li, X., Rubin, R., et al. (2022). Integrated single-cell transcriptomics of cerebrospinal fluid cells in treatment-naive multiple sclerosis. J. Neuroinflammation *19*, 306. https://doi.org/10.1186/s12974-022-02667-9.

24. Ramesh, A., Schubert, R.D., Greenfield, A.L., Dandekar, R., Loudermilk, R., Sabatino, J.J., Jr., Koelzer, M.T., Tran, E.B., Koshal, K., Kim, K., et al. (2020). A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. Proc. Natl. Acad. Sci. USA *117*, 22932–22943. https://doi.org/10.1073/pnas.2008523117.

25. Wu, X., Liu, Y., Jin, S., Wang, M., Jiao, Y., Yang, B., Lu, X., Ji, X., Fei, Y., Yang, H., et al. (2021). Single-cell sequencing of immune cells from anticitrullinated peptide antibody positive and negative rheumatoid arthritis. Nat. Commun. *12*, 4977. https://doi.org/10.1038/s41467-021-25246-7.

26. Perez, R.K., Gordon, M.G., Subramaniam, M., Kim, M.C., Hartoularos, G.C., Targ, S., Sun, Y., Ogorodnikov, A., Bueno, R., Lu, A., et al. (2022). Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. Science *376*, eabf1970. https://doi.org/10.1126/science.abf1970.

27. Itotagawa, E., Tomofuji, Y., Kato, Y., Konaka, H., Tsujimoto, K., Park, J., Nagira, D., Hirayama, T., Jo, T., Hirano, T., et al. (2023). SLE stratification based on BAFF and IFN-I bioactivity for biologics and implications of BAFF produced by glomeruli in lupus nephritis. Rheumatology *62*, 1988–1997. https://doi.org/10.1093/rheumatology/keac528.

28. Boland, B.S., He, Z., Tsai, M.S., Olvera, J.G., Omilusik, K.D., Duong, H.G., Kim, E.S., Limary, A.E., Jin, W., Milner, J.J., et al. (2020). Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. Sci. Immunol. *5*, eabb4432. https://doi.org/10.1126/sciimmunol.abb4432.

29. Luo, S., Wang, L., Xiao, Y., Cao, C., Liu, Q., and Zhou, Y. (2023). Single-Cell RNA-Sequencing Integration Analysis Revealed Immune Cell Heterogeneity in Five Human Autoimmune Diseases. BIO Integration *4*, 145–159. https://doi.org/10.15212/bioi-2023-0012.

30. Yasumizu, Y., Takeuchi, D., Morimoto, R., Takeshima, Y., Okuno, T., Kinoshita, M., Morita, T., Kato, Y., Wang, M., Motooka, D., et al. (2024). Single-cell transcriptome landscape of circulating CD4(+) T cell populations in autoimmune diseases. Cell Genom. *4*, 100473. https://doi.org/10.1016/j.xgen.2023.100473.

31. Wang, F.Q., Dang, X., Zhang, C., Su, H., Lei, Y., Chen, X., Yang, J., and Yang, W. (2025). Single-cell transcriptional landscape of peripheral myeloid cells in autoimmune diseases. iScience *28*, 113026. https://doi.org/10.1016/j.isci.2025.113026.

32. Kaufmann, M., Evans, H., Schaupp, A.L., Engler, J.B., Kaur, G., Willing, A., Kursawe, N., Schubert, C., Attfield, K.E., Fugger, L., and Friese, M.A. (2021). Identifying CNS-colonizing T cells as potential therapeutic targets to prevent progression of multiple sclerosis. Med *2*, 296–312.e8. https://doi.org/10.1016/j.medj.2021.01.006.

33. Valenzi, E., Bulik, M., Tabib, T., Morse, C., Sembrat, J., Trejo Bittar, H., Rojas, M., and Lafyatis, R. (2019). Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. Ann. Rheum. Dis. *78*, 1379–1387. https://doi.org/10.1136/annrheumdis-2018-214865.

34. Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell *178*, 714–730.e22. https://doi.org/10.1016/j.cell.2019.06.029.

35. Arazi, A., Rao, D.A., Berthier, C.C., Davidson, A., Liu, Y., Hoover, P.J., Chicoine, A., Eisenhaure, T.M., Jonsson, A.H., Li, S., et al. (2019). The immune cell landscape in kidneys of patients with lupus nephritis. Nat. Immunol. *20*, 902–914. https://doi.org/10.1038/s41590-019-0398-x.

36. Hong, X., Meng, S., Tang, D., Wang, T., Ding, L., Yu, H., Li, H., Liu, D., Dai, Y., and Yang, M. (2020). Single-Cell RNA Sequencing Reveals the Expansion of Cytotoxic CD4(+) T Lymphocytes and a Landscape of Immune Cells in Primary Sjogren's Syndrome. Front. Immunol. *11*, 594658. https://doi.org/10.3389/fimmu.2020.594658.

37. Mandric, I., Schwarz, T., Majumdar, A., Hou, K., Briscoe, L., Perez, R., Subramaniam, M., Hafemeister, C., Satija, R., Ye, C.J., et al. (2020). Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. Nat. Commun. *11*, 5504. https://doi.org/10.1038/s41467-020-19365-w.

38. Schafflick, D., Xu, C.A., Hartlehnert, M., Cole, M., Schulte-Mecklenbeck, A., Lautwein, T., Wolbert, J., Heming, M., Meuth, S.G., Kuhlmann, T., et al. (2020). Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. Nat. Commun. *11*, 247. https://doi.org/10.1038/s41467-019-14118-w.

39. Chaikovsky, A.C., Li, C., Jeng, E.E., Loebell, S., Lee, M.C., Murray, C.W., Cheng, R., Demeter, J., Swaney, D.L., Chen, S.H., et al. (2021). The AMBRA1 E3 ligase adaptor regulates the stability of cyclin D. Nature *592*, 794–798. https://doi.org/10.1038/s41586-021-03474-7.

40. Akatsuka, H., Kashikawa, T., Masuhara, K., Tokusanai, M., Li, C., Iida, Y., Okada-Yamaguchi, C., Okada, Y., Tanaka, M., Suzuki, T., et al. (2025). Molecular and genetic evidence for the role of AMBRA1 in suppressing S-phase entry and tumorigenesis. iScience *28*, 113054. https://doi.org/10.1016/j.isci.2025.113054.

41. Koh, C.H., Lee, S., Kwak, M., Kim, B.S., and Chung, Y. (2023). CD8 T-cell subsets: heterogeneity, functions, and therapeutic potential. Exp. Mol. Med. *55*, 2287–2299. https://doi.org/10.1038/s12276-023-01105-x.

42. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat. Methods *19*, 159–170. https://doi.org/10.1038/s41592-021-01346-6.

43. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.

44. Li, J., Zaslavsky, M., Su, Y., Guo, J., Sikora, M.J., van Unen, V., Christophersen, A., Chiou, S.H., Chen, L., Li, J., et al. (2022). KIR^{+}CD8^{+} T cells suppress pathogenic T cells and are active in autoimmune diseases and COVID-19. Science *376*, eabi9591. https://doi.org/10.1126/science.abi9591.

45. Dunlap, G., Wagner, A., Meednu, N., Wang, R., Zhang, F., Ekabe, J.C., Jonsson, A.H., Wei, K., Sakaue, S., Nathan, A., et al. (2024). Clonal associations between lymphocyte subsets and functional states in rheumatoid arthritis synovium. Nat. Commun. *15*, 4991. https://doi.org/10.1038/s41467-024-49186-0.

46. Paek, S.-J., Lee, H.S., Lee, Y.J., Bang, S.Y., Kim, D., Kang, B.K., Park, D.J., Joo, Y.B., Kim, M., Kim, H., et al. (2025). Tracking clonal dynamics of CD8 T cells and immune dysregulation in progression of systemic lupus erythematosus with nephritis. Exp. Mol. Med. *57*, 1700–1710. https://doi.org/10.1038/s12276-025-01504-2.

47. Xu, T., Zhu, H.X., You, X., Ma, J.F., Li, X., Luo, P.Y., Li, Y., Lian, Z.X., and Gao, C.Y. (2023). Single-cell profiling reveals pathogenic role and differentiation trajectory of granzyme K+CD8+ T cells in primary Sjögren's syndrome. JCI Insight *8*, e167490. https://doi.org/10.1172/jci.insight.167490.

48. Avery, D.T., Kalled, S.L., Ellyard, J.I., Ambrose, C., Bixler, S.A., Thien, M., Brink, R., Mackay, F., Hodgkin, P.D., and Tangye, S.G. (2003). BAFF

selectively enhances the survival of plasmablasts generated from human memory B cells. J. Clin. Investig. *112*, 286–297. https://doi.org/10.1172/jci18025.

49. Moisini, I., and Davidson, A. (2009). BAFF: a local and systemic target in autoimmune diseases. Clin. Exp. Immunol. *158*, 155–163. https://doi.org/10.1111/j.1365-2249.2009.04007.x.

50. Shin, W., Lee, H.T., Lim, H., Lee, S.H., Son, J.Y., Lee, J.U., Yoo, K.Y., Ryu, S.E., Rhie, J., Lee, J.Y., and Heo, Y.S. (2018). BAFF-neutralizing interaction of belimumab related to its therapeutic efficacy for treating systemic lupus erythematosus. Nat. Commun. *9*, 1200. https://doi.org/10.1038/s41467-018-03620-2.

51. Dohmen, J., Baranovskii, A., Ronen, J., Uyar, B., Franke, V., and Akalin, A. (2022). Identifying tumor cells at the single-cell level using machine learning. Genome Biol. *23*, 123. https://doi.org/10.1186/s13059-022-02683-1.

52. Yates, J., and Van Allen, E.M. (2025). New horizons at the interface of artificial intelligence and translational cancer research. Cancer Cell *43*, 708–727. https://doi.org/10.1016/j.ccell.2025.03.018.

53. Kim, B.H., Yu, K., and Lee, P.C.W. (2020). Cancer classification of single-cell gene expression data by neural network. Bioinformatics *36*, 1360–1366. https://doi.org/10.1093/bioinformatics/btz772.

54. Zhong, Z., Hou, J., Yao, Z., Dong, L., Liu, F., Yue, J., Wu, T., Zheng, J., Ouyang, G., Yang, C., and Song, J. (2024). Domain generalization enables general cancer cell annotation in single-cell and spatial transcriptomics. Nat. Commun. *15*, 1929. https://doi.org/10.1038/s41467-024-46413-6.

55. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

56. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., et al. (2022). A Python library for probabilistic analysis of single-cell omics data. Nat. Biotechnol. *40*, 163–166. https://doi.org/10.1038/s41587-021-01206-w.

57. Reynolds, G., Vegh, P., Fletcher, J., Poyner, E.F.M., Stephenson, E., Goh, I., Botting, R.A., Huang, N., Olabi, B., Dubois, A., et al. (2021). Developmental cell programs are co-opted in inflammatory skin disease. Science *371*, eaba6500. https://doi.org/10.1126/science.aba6500.

58. Gao, Y., Yao, X., Zhai, Y., Li, L., Li, H., Sun, X., Yu, P., Xue, T., Li, Y., and Hu, Y. (2021). Single cell transcriptional zonation of human psoriasis skin identifies an alternative immunoregulatory axis conducted by skin resident cells. Cell Death Dis. *12*, 450. https://doi.org/10.1038/s41419-021-03724-6.

59. Ma, F., Plazyo, O., Billi, A.C., Tsoi, L.C., Xing, X., Wasikowski, R., Gharaee-Kermani, M., Hile, G., Jiang, Y., Harms, P.W., et al. (2023). Single cell and spatial sequencing define processes by which keratinocytes and fibroblasts amplify inflammatory responses in psoriasis. Nat. Commun. *14*, 3455. https://doi.org/10.1038/s41467-023-39020-4.

60. Kim, J., Lee, J., Li, X., Kunjravia, N., Rambhia, D., Cueto, I., Kim, K., Chaparala, V., Ko, Y., Garcet, S., et al. (2023). Multi-omics segregate different transcriptomic impacts of anti-IL-17A blockade on type 17 T-cells and regulatory immune cells in psoriasis skin. Front. Immunol. *14*, 1250504. https://doi.org/10.3389/fimmu.2023.1250504.

61. Kim, S.H., Oh, J., Roh, W.S., Park, J., Chung, K.B., Lee, G.H., Lee, Y.S., Kim, J.H., Lee, H.K., Lee, H., et al. (2023). Pellino-1 promotes intrinsic activation of skin-resident IL-17A producing T cells in psoriasis. J. Allergy Clin. Immunol. *151*, 1317–1328. https://doi.org/10.1016/j.jaci.2022.12.823.

62. Frost, B., Schmidt, M., Klein, B., Loeffler-Wirth, H., Krohn, K., Reidenbach, T., Binder, H., Stubenvoll, A., Simon, J.C., Saalbach, A., and Kunz, M. (2023). Single-cell transcriptomics reveals prominent expression of IL-14, IL-18, and IL-32 in psoriasis. Eur. J. Immunol. *53*, e2250354. https://doi.org/10.1002/eji.202250354.

63. Francis, L., McCluskey, D., Ganier, C., Jiang, T., Du-Harpur, X., Gabriel, J., Dhami, P., Kamra, Y., Visvanathan, S., Barker, J.N., et al. (2024). Single-cell analysis of psoriasis resolution demonstrates an inflammatory fibroblast state targeted by IL-23 blockade. Nat. Commun. *15*, 913. https://doi.org/10.1038/s41467-024-44994-w.

64. Zhu, Y., Zhao, L., Yan, W., Ma, H., Zhao, W., Qu, J., Zheng, W., Zhang, C., Du, H., Yu, M., et al. (2025). Celastrol directly targets LRP1 to inhibit fibroblast-macrophage crosstalk and ameliorates psoriasis progression. Acta Pharm. Sin. B *15*, 876–891. https://doi.org/10.1016/j.apsb.2024.12.041.

65. Peng, L., Liu, W., Cheng, Y., Chen, L., and Shen, Z. (2024). IL-17A/F double producing T cells, unstable Tregs and quiescent TRMs in clinically healed lesions are potential cellular candidates for recurrence of psoriasis. Clin. Immunol. *266*, 110328. https://doi.org/10.1016/j.clim.2024.110328.

66. Kim, J., Lee, J., Lee, J., Kim, K., Li, X., Zhou, W., Cao, J., and Krueger, J.G. (2025). Psoriasis harbors multiple pathogenic type 17 T-cell subsets: Selective modulation by risankizumab. J. Allergy Clin. Immunol. *155*, 1898–1912. https://doi.org/10.1016/j.jaci.2025.02.008.

67. Horai, R., Zárate-Bladés, C.R., Dillenburg-Pilla, P., Chen, J., Kielczewski, J.L., Silver, P.B., Jittayasothorn, Y., Chan, C.C., Yamane, H., Honda, K., and Caspi, R.R. (2015). Microbiota-Dependent Activation of an Autoreactive T Cell Receptor Provokes Autoimmunity in an Immunologically Privileged Site. Immunity *43*, 343–353. https://doi.org/10.1016/j.immuni.2015.07.014.

68. Cho, A., Caldara, A.L., Ran, N.A., Menne, Z., Kauffman, R.C., Affer, M., Llovet, A., Norwood, C., Scanlan, A., Mantus, G., et al. (2019). Single-Cell Analysis Suggests that Ongoing Affinity Maturation Drives the Emergence of Pemphigus Vulgaris Autoimmune Disease. Cell Rep. *28*, 909–922.e6. https://doi.org/10.1016/j.celrep.2019.06.066.

69. Fuschiotti, P. (2018). Current perspectives on the role of CD8+ T cells in systemic sclerosis. Immunol. Lett. *195*, 55–60. https://doi.org/10.1016/j.imlet.2017.10.002.

70. Tasaki, S., Suzuki, K., Nishikawa, A., Kassai, Y., Takiguchi, M., Kurisu, R., Okuzono, Y., Miyazaki, T., Takeshita, M., Yoshimoto, K., et al. (2017). Multiomic disease signatures converge to cytotoxic CD8 T cells in primary Sjogren's syndrome. Ann. Rheum. Dis. *76*, 1458–1466. https://doi.org/10.1136/annrheumdis-2016-210788.

71. Bhat, P., Leggatt, G., Waterhouse, N., and Frazer, I.H. (2017). Interferon-gamma derived from cytotoxic lymphocytes directly enhances their motility and cytotoxicity. Cell Death Dis. *8*, e2836. https://doi.org/10.1038/cddis.2017.67.

72. Uhl, L.F.K., Cai, H., Oram, S.L., Mahale, J.N., MacLean, A.J., Mazet, J.M., Piccirilli, T., He, A.J., Lau, D., Elliott, T., and Gerard, A. (2023). Interferon-gamma couples CD8(+) T cell avidity and differentiation during infection. Nat. Commun. *14*, 6727. https://doi.org/10.1038/s41467-023-42455-4.

73. Tsumiyama, K., Hashiramoto, A., Takimoto, M., Tsuji-Kawahara, S., Miyazawa, M., and Shiozawa, S. (2013). IFN-gamma-producing effector CD8 T lymphocytes cause immune glomerular injury by recognizing antigen presented as immune complex on target tissue. J. Immunol. *191*, 91–96. https://doi.org/10.4049/jimmunol.1203217.

74. Helena Jonsson, A., Z., F., Garrett Dunlap, Gomez-Rivas, E., Watts, G.F.M., Faust, H.J., Vijay Rupani, K., Mears, J.R., Meednu, N., Wang, R., et al. (2023). Granzyme K^+ CD8 T cells form a core population in inflamed human tissue. Sci. Transl. Med. *14*, eabo0686. https://doi.org/10.1126/scitranslmed.abo0686.

75. Boyden, A.W., Brate, A.A., Stephens, L.M., and Karandikar, N.J. (2020). Immune Autoregulatory CD8 T Cells Require IFN-gamma Responsiveness to Optimally Suppress Central Nervous System Autoimmunity. J. Immunol. *205*, 359–368. https://doi.org/10.4049/jimmunol.2000211.

76. Hu, X., and Ivashkiv, L.B. (2009). Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases. Immunity *31*, 539–550. https://doi.org/10.1016/j.immuni.2009.09.002.

77. Daamen, A.R., Alajoleen, R.M., Grammer, A.C., Luo, X.M., and Lipsky, P.E. (2023). Single-cell RNA sequencing analysis reveals the heterogeneity of IL-10 producing regulatory B cells in lupus-prone mice. Front. Immunol. *14*, 1282770. https://doi.org/10.3389/fimmu.2023.1282770.

78. Mockel, T., Basta, F., Weinmann-Menke, J., and Schwarting, A. (2021). B cell activating factor (BAFF): Structure, functions, autoimmunity and clinical implications in Systemic Lupus Erythematosus (SLE). Autoimmun. Rev. *20*, 102736. https://doi.org/10.1016/j.autrev.2020.102736.

79. Kobayashi, G., Okamura, T., Hashimoto, Y., Sakai, K., Sumi, M., Imai, D., Kitagawa, N., Hamaguchi, M., and Fukui, M. (2025). Pathogenesis of Graves' Disease Determined Using Single-Cell Sequencing with Thyroid Autoantigen Peptide Stimulation in B Cells. Cells *14*, 1102. https://doi.org/10.3390/cells14141102.

80. Nobari, S.T., Nojadeh, J.N., and Talebi, M. (2022). B-cell maturation antigen targeting strategies in multiple myeloma treatment, advantages and disadvantages. J. Transl. Med. *20*, 82. https://doi.org/10.1186/s12967-022-03285-y.

81. Ota, M., Nagafuchi, Y., Hatano, H., Ishigaki, K., Terao, C., Takeshima, Y., Yanaoka, H., Kobayashi, S., Okubo, M., Shirai, H., et al. (2021). Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell *184*, 3006–3021.e17. https://doi.org/10.1016/j.cell.2021.03.056.

82. Guthrie, J., Köstel Bal, S., Lombardo, S.D., Müller, F., Sin, C., Hütter, C.V.R., Menche, J., and Boztug, K. (2023). AutoCore: A network-based definition of the core module of human autoimmunity and autoinflammation. Sci. Adv. *9*, eadg6375. https://doi.org/10.1126/sciadv.adg6375.

83. Xiang, N., Xu, H., Zhou, Z., Wang, J., Cai, P., Wang, L., Tan, Z., Zhou, Y., Zhang, T., Zhou, J., et al. (2023). Single-cell transcriptome profiling reveals immune and stromal cell heterogeneity in primary Sjogren's syndrome. iScience *26*, 107943. https://doi.org/10.1016/j.isci.2023.107943.

84. Stanley, S., Vanarsa, K., Soliman, S., Habazi, D., Pedroza, C., Gidley, G., Zhang, T., Mohan, S., Der, E., Suryawanshi, H., et al. (2020). Comprehensive aptamer-based screening identifies a spectrum of urinary biomarkers of lupus nephritis across ethnicities. Nat. Commun. *11*, 2197. https://doi.org/10.1038/s41467-020-15986-3.

85. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

86. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982. https://doi.org/10.1038/nmeth.4402980.

87. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005.

88. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

89. Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A., and Tanay, A. (2019). MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. Genome Biol. *20*, 206. https://doi.org/10.1186/s13059-019-1812-2.

90. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

91. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinf. *17*, 483. https://doi.org/10.1186/s12859-016-1323-z.

92. Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q., Huang, J.Y., et al. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature *564*, 268–272. https://doi.org/10.1038/s41586-018-0694-x.

93. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc. *15*, 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

94. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162. https://doi.org/10.1038/s41592-019-0667-5.

95. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.

96. Prevoo, M.L., van 't Hof, M.A., Kuper, H.H., van Leeuwen, M.A., van de Putte, L.B., and van Riel, P.L. (1995). Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. *38*, 44–48. https://doi.org/10.1002/art.1780380107.

97. Gladman, D.D., Ibañez, D., and Urowitz, M.B. (2002). Systemic lupus erythematosus disease activity index 2000. J. Rheumatol. *29*, 288–291.

98. Bombardier, C., Gladman, D.D., Urowitz, M.B., Caron, D., and Chang, C.H. (1992). Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum. *35*, 630–640. https://doi.org/10.1002/art.1780350606.

99. Seror, R., Ravaud, P., Bowman, S.J., Baron, G., Tzioufas, A., Theander, E., Gottenberg, J.E., Bootsma, H., Mariette, X., and Vitali, C.; EULAR Sjögren's Task Force (2010). EULAR Sjogren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjogren's syndrome. Ann. Rheum. Dis. *69*, 1103–1109. https://doi.org/10.1136/ard.2009.110619.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Peripheral blood mononuclear cells, fully anonymized	The First Affiliated Hospital of University of Science and Technology of China	N/A
Critical commercial assays		
Chromium Single Cell 5′ Library and Gel Bead kit	10× Genomics	Cat #PN-1000006
Chromium Single cell V(D)J Enrichment Kit.Human T cell	10× Genomics	Cat #PN-1000005
Chromium i7 Multiplex kit	10× Genomics	Cat #PN-120262
Dynabeads MyOne SILANE	Thermo Fisher Scientific	Cat #PN-2000048
Chromium Single Cell 5′ Library Construction kit	10× Genomics	Cat #PN-1000020
Nuclease-free Water	Thermo Fisher Scientific	AM9937
Ethanol, Pure (200 Proof, anhydrous)	Millipore Sigma	E7023-500 ML
SPRIselect Reagent Kit	Beckman Coulter	B23318
Qiagen Buffer EB	Qiagen	19086
Deposited data		
Raw and analyzed data	This paper	GSA-human: HRA011522
pSS Analyzed public RNA-seq data	Xiang et al.[83]	HRA003613
pSS Analyzed public RNA-seq data	Hong et al.[36]	GSE157278
RA Analyzed public RNA-seq data	Wu et al.[25]	HRA000155
UC Analyzed public RNA-seq data	Smillie et al.[34]	SCP259
SLE Analyzed public RNA-seq data	Stanley et al.[84]	SDY997
UC Analyzed public RNA-seq data	Boland et al.[28]	GSE125527
SSc Analyzed public RNA-seq data	Valenzi et al.[33]	GSE128169
SLE Analyzed public RNA-seq data	Mandric et al.[37]	GSE137029
MS Analyzed public RNA-seq data	Schafflick et al.[38]	GSE138266
MS Analyzed public RNA-seq data	Kaufmann et al.[32]	GSE144744
PS Analyzed public RNA-seq data	Francis et al.[63]	GSE228421
PS Analyzed public RNA-seq data	Frost et al.[62]	GSE221648
PS Analyzed public RNA-seq data	Gao et al.[58]	GSE162183
PS Analyzed public RNA-seq data	Kim et al.[66]	GSE278330
PS Analyzed public RNA-seq data	Kim et al.[60]	GSE198805
PS Analyzed public RNA-seq data	Kim et al.[61]	GSE183047
PS Analyzed public RNA-seq data	Ma et al.[59]	GSE173706
PS Analyzed public RNA-seq data	Peng et al.[65]	GSE248121
PS Analyzed public RNA-seq data	Reynolds et al.[57]	E-MTAB-8142
PS Analyzed public RNA-seq data	Zhu et al.[64]	GSE230842
Software and algorithms		
Cell Ranger (v5.0.1)(Cell Ranger Single-Cell Software Suite)	10× Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
Scanpy(1.9.3)	Wolf et al.[85]	https://scanpy.readthedocs.io
SCVI(0.20.3)	Gayoso et al.[56]	https://github.com/scverse/scvi-tools
Seurat(4.1.0)	Hao et al.[55]	https://satijalab.org/seurat/
Monocle(2.20.0)	Qiu et al.[86]	http://cole-trapnell-lab.github.io/monocle-release/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Scrublet(0.2.3)	Wolock et al.[87]	https://github.com/swolock/scrublet
Harmony(1.2.1)	Korsunsky et al.[88]	https://github.com/immunogenomics/harmony
Metacell(0.7.0)	Baran et al.[89]	https://tanaylab.github.io/metacell/
Scipy(1.8.0)	Virtanen et al.[90]	https://scipy.org/
GSEApy(0.10.8))	Fang et al.[43]	https://github.com/zqfang/GSEApy
variancePartition(1.22.0)	Hoffman et al.[91]	https://github.com/GabrielHoffman/variancePartition
CellRank(1.5.1)	Lange et al.[42]	https://github.com/theislab/cellrank
STARTRAC(0.1.0)	Zhang et al.[92]	https://github.com/Japrin/STARTRAC
CellphoneDB(2.0.0)	Efremova et al.[93]	https://github.com/Teichlab/cellphonedb
NicheNet(2.2.0)	Browaeys et al.[94]	https://github.com/saeyslab/nichenetr
CellChat(2.1.2)	Jin et al.[95]	https://github.com/sqjin/CellChat
Custom scripts for this study	This paper	https://github.com/QuKunLab/pan-autoimmune_disease

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples and ethical permission

This study was approved by the medical research ethics committee of The First Affiliated Hospital of University of Science and Technology of China (approval No.2025-KY-217), and all patients provided signed informed consent accordingly before the blood sampling.

Patients with autoimmune diseases were eligible for enrollment only if they had not previously received hormonotherapy treatments. Peripheral blood samples (<5 mL) were collected using EDTA anticoagulant tubes.

Clinical data, including disease activity scores (referred to as "severity of illness score" in Table S3), were collected at the time of sampling. Disease activity was assessed using validated, disease-specific clinical indices: the Disease Activity Score 28 (DAS28) for Rheumatoid Arthritis (RA) patients,[96] the SLE Disease Activity Index (SLEDAI) for Systemic Lupus Erythematosus (SLE) patients,[97,98] and the EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI) for Primary Sjögren's Syndrome (pSS) patients.[99]

A total of 12 individuals were recruited for this study, including 3 patients with Systemic Lupus Erythematosus (SLE), 3 with Rheumatoid Arthritis (RA), 3 with Primary Sjögren's Syndrome (pSS), and 3 healthy controls. The cohort consisted of 11 females and 1 males, ranging in age from 17 to 59 years. Detailed demographic and clinical characteristics of the study participants are provided in Table S3. Consequently, statistical analysis of sex-specific differences was not feasible in this specific dataset due to the limited number of male participants. We acknowledge this as a limitation of the study's generalizability regarding sex-based comparisons.

METHOD DETAILS

Single cell collection, sorting, library preparation and sequencing

Peripheral blood was collected using blood collection tubes and carefully layered onto Solarbio Ficoll Plus 1.077 (Catalog No.: P4350-200 mL) for density gradient centrifugation. The sample was centrifuged at 400g (~1500 rpm, 15 cm radius, horizontal rotor) for 20 min, and the whitish lymphocyte ring layer was carefully aspirated. The second layer of cells was collected and transferred into a tube containing 4–5 mL of physiological saline, mixed thoroughly, and centrifuged again at 400g (~1500 rpm) for 20 min. A cryopreservation medium was prepared by mixing serum and DMSO at a 9:1 ratio. After cell counting, the cells were resuspended in the cryopreservation medium at a concentration of 5×10^6 to 1×10^7 cells/mL, and 1–1.5 mL of the suspension was aliquoted into cryovials for freezing.

To thaw cryopreserved cells, remove the cryovial from liquid nitrogen storage and quickly place it in a 37°C water bath. Gently swirl the vial until the contents are almost completely thawed (approximately 1–2 min). Carefully transfer the thawed cell suspension into a 15 mL centrifuge tube and slowly add 8–10 mL of pre-warmed RPMI-1640 medium (GIBCO) dropwise while gently mixing to minimize osmotic shock. Centrifuge the tube at 500g for 5 min to pellet the cells, discard the supernatant, and gently resuspend the cell pellet in 1–2 mL of fresh complete medium.

Single cell suspensions were stained with antibodies against CD45 and DAPI for FACS sorting, performed on a sony SH800S instrument. Based on FACS analysis, single cells were sorted into 1.5 mL tubes (Eppendorf) and counted manually under the microscope. The concentration of single cell suspensions was adjusted to 500–1200 cells/ul. Cells were loaded between 7,000 and 15,000 cells/chip position using the 10× Chromium Single cell 5′ Library, Gel Bead & Multiplex Kit and Chip Kit (10× Genomics, V2 barcoding

chemistry) according to the manufacturer's instructions. All the subsequent steps were performed following the standard manufacturer's protocols. Purified libraries were analyzed by an Illumina Hiseq X Ten sequencer with 150-bp paired-end reads.

Single-cell RNA-Seq datasets collected in this study

We obtained scRNA-seq data on immune cells in 160 samples from 160 patients diagnosed with one of the 6 common autoimmune diseases (Figure 1B and Table S1). 29 of these samples also included TCR information. Data from 177 patients (239 samples) were obtained from published studies, including those previously generated by us. To supplement the publicly available data and cover the cancer types that have not been adequately studied, we collected an additional cohort of 12 treatment-naive patients of 4 autoimmune diseases and 4 health doners, and obtained the scRNA-seq data for 160 samples derived from these patients using the 10× Genomics platform.

Single-cell RNA-seq data processing

The newly generated scRNA-seq data from 10× Genomics were aligned and quantified using the Cell Ranger Single-Cell Software Suite against the GRCh38 human reference genome. The preliminary filtered data generated from Cell Ranger were used for downstream analysis. Further quality control was applied to cells based on three metrics step by step, including the total UMI count, number of detected genes and proportion of mitochondrial gene count per cell. Specifically, cells with less than 2000 UMI count and 500 detected genes were filtered, as well as cells with more than 10% mitochondrial gene count. To remove potential doublets, cells with UMI count above 40,000 and detected genes above 5,000 are also filtered out. Notably, we removed potential doublets predicted by Scrublet. For other collected scRNA-seq datasets, we applied the same filtering steps to all datasets. After quality control, we applied the library-size correction method to normalize the raw count by using normalize_total function in Scanpy.[85] Then the logarithmized normalized count matrix was used for the downstream analysis.

Integration of multiple scRNA-seq datasets by Harmony

We run Harmony, an algorithm that could identify and merge shared cell types among multiple datasets. To remove the batch effects within scRNA-seq datasets of 6 autoimmune diseases, group.by.vars were set to Dataset\Tissue1\Sample, and lambda = c(1.2,0.8, 0.8). The results from Harmony integration and batch correction were then used as input data for scRNA-seq analysis, including highly-variable gene identification and dimension reduction.

Correlation analysis

To quantify the similarity between cell phase states and cell type annotations, we computed the Jaccard Index (JI) between each pair of Celltype3 and Celltype1 categories. For each phase-cell type combination, the JI was calculated as the size of the intersection divided by the size of the union of the corresponding cell indices. The resulting JI matrix reflects the degree of overlap between transcriptional phase progression and cell identity.

Differential expression analysis

To identify differentially expressed genes between two groups of clusters, we used wilcox.test in R to evaluate the significance of each gene, with multiple hypothesis correction using the Benjamini-Hochberg procedure. Genes with adjusted p-value less than 0.05 and the log2 fold change (log2FC) > 1.0 were considered as differentially expressed genes.

Autoimmune disease up-regulated gene set

Here, according to the disease-tissue-cell type, we compared the differentially up-regulated genes in different tissues and cell types of patients with different diseases compared with healthy people, and added the up-regulated genes that directly distinguished different diseases and healthy people. The obtained differentially up-regulated genes were merged to obtain the autoimmune disease up-regulated differential gene set.

Metacell

To identify MetaCells, the MetaCell R package was used. For cells from different data sources, facilitates analysis of single cell RNA-seq UMI matrices by computing partitions of a cell similarity graph into small (50) homogeneous groups of cells which are defined as metacells (MCs). The derived MCs are then used for building different representations of the data, allowing matrix or 2D graph visualization forming a basis for analysis of cell types, subtypes, transcriptional gradients, cell-cycle variation, gene modules and their regulatory models and more.

Gene clusters

The average expression of celltype differential gene list specific for each disease subgroup was calculated for metacell relative to foldchange for healthy subgroups to obtain gene score X Celltype tissue Disease foldchange matrix. Hierarchical clustering analysis and visualization were performed on gene score X Celltype tissue Disease foldchange matrix, generated using a random number generator to simulate example data. The clustering was conducted using a heatmap approach implemented with the sns.clustermap function from the Seaborn library. Data normalization was applied along columns using Z score normalization (z_score = 1), and

hierarchical clustering was performed using Ward's method. The clustering result was visualized with a heatmap, using the colormap RdBu_r to display normalized values ranging from −1 to 1. The dendrogram associated with the hierarchical clustering was used to extract row clustering order and the corresponding linkage matrix. To define discrete clusters from the hierarchical tree, the fcluster function from the SciPy library was employed, using a distance threshold of 70.0 to segment the dendrogram into distinct clusters. The resulting clusters were analyzed, and row indices corresponding to each cluster were identified and reported. Finally, the heatmap and dendrogram were visualized using Matplotlib for further interpretation of the clustering results.

Scoring celltype by using signature genes for each subset

To perform gene set enrichment analysis, a predefined list of genes (genelist) was compared to differentially expressed gene (DEG) lists (degs_list) for each cluster. For each cluster, the overlap between the predefined gene list and the DEG list was calculated using the intersect function. Hypergeometric testing was then conducted for each cluster using the phyper function to compute the probability (*p*-value) of observing the overlap by chance, considering the total number of genes in the dataset, the size of the DEG list for the cluster, and the size of the predefined gene list. The hypergeometric test was one-tailed (lower.tail = FALSE) to identify clusters with significant enrichment. The resulting *p*-values, along with the overlap count, were compiled into a data frame and sorted by significance to rank the clusters by their level of enrichment. This analysis was implemented in R using the dplyr package for data manipulation. To evaluate the characteristic subgroups of gene clusters, we used two methods, one of which was gene scoring using scanpy.tl. The Score_genes function calculates a gene expression score for each cell based on a predefined list of genes associated with the analysis. This method calculates the average expression of selected genes per cell and then subtract the average expression of the control gene set to reduce noise. Cells were then grouped according to their respective cell type, and the average gene score was calculated for each cell type. The cell type with the highest mean genetic score was identified, and cells in this subgroup were extracted as the target high-score subgroup for further analysis. This approach enabled the identification of cell subtypes with the most significant expression of selected genes for downstream analysis. The second is that differential genes were calculated for large subgroups and cluster genes was compared to differentially expressed gene (DEG) lists (degs_list) for each cluster. For each cluster, the overlap between the predefined gene list and the DEG list was calculated using the intersect function. Hypergeometric testing was then conducted for each cluster using the phyper function to compute the probability (*p*-value) of observing the overlap by chance, considering the total number of genes in the dataset, the size of the DEG list for the cluster, and the size of the predefined gene list. The hypergeometric test was one-tailed (lower.tail = FALSE) to identify clusters with significant enrichment. The resulting *p*-values, along with the overlap count, were compiled into a data frame and sorted by significance to rank the clusters by their level of enrichment. This analysis was implemented in R using the dplyr package for data manipulation.

Gene ontology enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using the GSEApy Python package (version 0.10.8). The input gene sets were analyzed against the latest GO reference gene sets, including GO_Biological_Process_2023, GO_Cellular_Component_2023, and GO_Molecular_Function_2023, obtained from the Enrichr database.

Enrichment was assessed using a Fisher's exact test (overrepresentation analysis), and *p* values were adjusted for multiple testing using the Benjamini–Hochberg method. GO terms with a false discovery rate (FDR) < 0.05 were considered statistically significant.

Gene set variance partition analysis

We performed Gene Set Variance Partition Analysis using the variancePartition R package[91] to quantify the contributions of biological and technical factors to gene expression variability. Specifically, the analysis was conducted to evaluate the relative contributions of three key variables—disease, tissue, and cell type—to the expression of genes in a predefined gene list. The mixed-effects model framework implemented in variancePartition was used, allowing for the decomposition of gene expression variability into proportions attributed to each factor. For each gene in the gene list, a linear mixed-effects model was fitted, where the variables disease, tissue, and cell type were included as fixed effects, and random effects accounted for other sources of variation such as batch effects. The variance explained by each factor was quantified and normalized to obtain variance fractions. Genes with a high variance fraction for a specific factor were identified as strongly associated with that factor. The results were summarized across the gene list to identify dominant contributors to the variability in gene expression.

Gene cluster score calculation and disease-wise comparison

To quantify defense response activity at the single-cell level, we used the score_genes function from Scanpy to compute a module score for the predefined gene sets. The resulting scores were averaged per sample within each annotated cell subset. Each sample's disease status was annotated based on metadata. For each cell subset, we compared the average predefined gene sets scores between disease and healthy samples using the Mann-Whitney U test. Results were visualized as boxplots with overlaid significance levels, and outliers were excluded for clarity. This analysis enabled the identification of cell types exhibiting significantly altered predefined gene sets in disease conditions.

Monocle

To characterize the developmental origins of *CD8*$^+$ T, we applied the Monocle (version 2.14.0) algorithm[86] with the top 400 signature genes calculated by differentialGeneTest function. The *CD8*$^+$ T cell differentiation trajectory was inferred with the default parameters of Monocle after dimension reduction and cell ordering.

CellRank analysis of differentiation potential

To infer cellular differentiation trajectories and fate potentials independent of RNA velocity, we applied the CytoTRACEKernel implemented in the CellRank framework (version 1.5.1).[42] This approach estimates cell plasticity based on transcriptional entropy and gene count complexity. The analysis was performed in the backward mode, allowing the reconstruction of developmental origins by computing transition probabilities toward progenitor-like states. A soft thresholding scheme was used to compute the transition matrix, and the resulting transition probabilities were projected onto the precomputed UMAP embedding to visualize fate dynamics across cell populations. Cell fate directionality was visualized using streamline plots, with cells colored by annotated subtypes (Celltype3), providing insights into pseudotemporal progression within the immune landscape.

TCR clone calculation

TCR clone calculation was performed to assess clonal diversity and expansion in single-cell TCR sequencing data. TCR sequences, including CDR3 regions and V(D)J genes, were extracted and annotated from the dataset. Clones were defined based on identical CDR3 nucleotide sequences and V(D)J gene usage. For each clone, the frequency of occurrence across cells was calculated to quantify clonal expansion. Clonal diversity was further analyzed using metrics such as clonality, Shannon entropy, and Simpson's diversity index to evaluate the extent of TCR repertoire variation. Visualization of clonal distribution was performed using bar plots and scatterplots, highlighting dominant clones and their association with specific cell types or phenotypes. This analysis provides insights into T cell dynamics, immune responses, and potential antigen-specific expansions.

STARTRAC

The STARTRAC Transition Index was computed using the STARTRAC R package with default parameters. Results were visualized using heatmaps and network plots, where nodes represent clusters or tissues, and edge weights represent the Transition Index between clusters. These visualizations highlight clonal migration patterns, providing insights into the mobility and functional relevance of T cell clones across different microenvironments.

CellphoneDB analysis

The statistical_analysis method was run with default parameters. It evaluates ligand-receptor interactions by computing expression scores between interacting cell types. A permutation test (default: 1000 permutations) was performed to calculate *p*-values for each ligand-receptor pair, based on the null hypothesis that ligands and receptors are randomly distributed across cell types.

NicheNet analysis

We also performed NicheNet analysis with default parameters,[94] an algorithm that predicts ligand-receptor-mediated intercellular communication by linking ligands expressed in sender cells to target genes expressed in receiver cells. NicheNet uses a comprehensive ligand-receptor interaction database combined with a downstream signaling and regulatory network to infer the regulatory impact of ligands on target gene expression. This analysis complements other trajectory and communication studies by providing a ligand-target-centric perspective to identify key signaling pathways mediating intercellular communication in the tissue microenvironment.

CellChat analysis

We also performed CellChat analysis with default parameters,[95] a computational framework that infers and analyzes intercellular communication networks by integrating single-cell RNA sequencing data with a curated database of ligand-receptor interactions. CellChat uses a probabilistic model to quantify the likelihood of communication between cell types based on the expression levels of ligands, receptors, and their cofactors. The analysis provides a comprehensive overview of signaling pathways active in the tissue microenvironment, including the identification of major sender and receiver cell populations, dominant signaling pathways, and the strength of intercellular communication networks. This method complements other trajectory and functional analyses, providing insights into the signaling dynamics that regulate tissue organization and cell behavior.

Two-stage classification

The analysis was performed using SCVI or Seurat, with single-cell data serving as the reference and previously enriched differential genes guiding the process. In the first stage, a binary label (''Disease'' vs. ''Healthy'' status) was used as the reference label to distinguish between disease and healthy states with selected genes (Table S7). In the second stage, a multi-class disease label (e.g., RA, SLE, MS) was employed as the reference label to classify different disease states with intersection genes of all gene clusters, allowing for the identification and labeling of specific diseases in newly added single-cell data.

Leave-one-study-out cross-validation

To validate the classifier's robustness against study-specific batch effects, we performed a leave-one-study-out validation using the pSS cohort, which was represented by two independent studies ("HRA003613" and "GSE157278"). A new training dataset was created by excluding all pSS samples from the "GSE157278" study. The classifier was re-trained on this modified dataset using the same parameters as the original model. The model's performance was then evaluated by predicting the disease labels of the held-out "GSE157278" pSS samples.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical test

All statistical tests used to analyze the data are specifically described in the main text and figure legends or Methods section. Statistical tests included two-sided *t* test, two-sided Mann-Whitney U test, two-sided Wilcoxon rank-sum test, two-sided Fisher's exact test, and Pearson or Spearman cor.test. Unless otherwise specified, boxes show the median and the interquartile range (IQR); the lower whisker indicates Q1 – 1.5 $\times$ the IQR; the upper whisker indicates Q3 + 1.5 $\times$ the IQR. Significant *p* values were indicated as follows: $p \leq 0.05$ (*), $p \leq 0.01$(**) and $p \leq 0.001$(***), $p \leq 0.0001$(****).