Title:

Deep immune profiling pinpoints cellular and molecular drivers of lupus immunopathology

Authors:

Masahiro Nakano[1]†*, Michihiro Kono[1,2]†, Kenichiro Asahara[1,3], Takayuki Katsuyama[4], Satoshi Kubo[5], Eri Katsuyama[4,6], Yuya Fujita[5], Takahiro Nishino[1,2], Hajime Inokuchi[1,2,7], Takahiro Arakawa[1], Tsugumi Kawashima[1], Shohei Noma[1], Akiko Minowa[2], Reza Bagherzadeh[1,2], Yukiro Matsumoto[1,2,8], Jun Inamo[1,2], Haruka Takahashi[1,2], Bunki Natsumoto[1], Xuejun Zhang[9,10,11], Sang-Cheol Bae[12,13,14], Akari Suzuki[15], Hiroaki Hatano[1,2], Chikashi Terao[16,17,18], Yoshiya Tanaka[5,19], Yoshinori Matsumoto[4], Kazuhiko Yamamoto[15], Kazuyoshi Ishigaki[1,2,20]*

Affiliations:

[1] Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[2] Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo, Japan

[3] Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

18 [4] Department of Nephrology, Rheumatology, Endocrinology and Metabolism, Okayama University
19 Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan
20 [5] The First Department of Internal Medicine, University of Occupational and Environmental Health,
21 Kitakyushu, Japan
22 [6] Faculty of Health Science, Okayama University Medical School, Graduate School of Health Sciences,
23 Okayama University, Okayama, Japan
24 [7] Division of Rheumatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo,
25 Japan
26 [8] Department of Dermatology, Keio University School of Medicine, Tokyo, Japan
27 [9] Department of Dermatology, the First Affiliated Hospital of Anhui Medical University, Hefei, Anhui,
28 China
29 [10] Institute of Dermatology, Anhui Medical University, Hefei, Anhui, China
30 [11] Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, Hefei, Anhui, China
31 [12] Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea
32 [13] Hanyang Institute of Bioscience and Biotechnology, Seoul, Korea
33 [14] Hanyang University Institute for Rheumatology Research, Seoul, Korea
34 [15] Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama,
35 Japan

[16] Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[17] Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan

[18] The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan

[19] Department of Molecular Targeted Therapeutics, University of Occupational and Environmental Health, Kitakyushu, Japan

[20] Keio University Human Biology-Microbiome-Quantum Research Center (WPI-Bio2Q)

† These authors contributed equally to this work.

*Correspondence:

Masahiro Nakano, M.D., Ph.D.

masahiro.nakano.fc@riken.jp

Kazuyoshi Ishigaki, M.D., Ph.D.

kazuyoshi.ishigaki@keio.jp

Abstract:

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with an unknown etiology. To pinpoint new disease-relevant cell states and their molecular profiles, we performed an in-depth investigation of multimodal single-cell datasets comprising ~2.1 million peripheral blood mononuclear cells from 346 donors. By resolving 123 fine-grained cell states across 27 cell types, we identified previously uncharacterized populations distinctively associated with clinical severity and treatment status, including *GZMK*$^{+}$*GZMH*$^{+}$*HLA-DR*$^{+}$ effector memory CD8^{+} T cells (double-positive [DP] EMCD8) and *FOXO1*$^{+}$*ARHGAP15*$^{+}$ T cells. Through extensive statistical frameworks and multimodal approaches, we delineated their aberrant immune signaling networks, transcriptional regulators, key surface proteins, T cell receptor repertoires, and genetic/epigenetic landscapes, underscoring them as candidate drivers of SLE immunopathology. These findings provide new insights into therapeutic target discovery in SLE.

65 **Main Text:**

66 **Introduction**

67 Systemic lupus erythematosus (SLE) is a prototypic autoimmune disease with a broad spectrum of
68 clinical manifestations, triggered by genetic and environmental factors (*1*). Since the pathophysiology of
69 SLE involves various immune cell types, many bulk and single-cell transcriptome studies on peripheral
70 blood and affected organs have been conducted (*2*–*6*). While these efforts have begun to reveal key
71 gene signatures and core cell populations of SLE, the fine-resolution pathophysiology remains
72 unresolved, and the treatment options are still limited; to date, only two molecular-targeted drugs have
73 been approved for this disease (*7*, *8*).

74 To understand the complex biology of SLE and identify therapeutic targets, it is crucial to
75 characterize disease-relevant cell states and their molecular signatures at high resolution. We previously
76 conducted a large-scale bulk transcriptome study across 27 immune cell types, identifying cell-type-
77 specific signatures linked to SLE development and exacerbation (*6*). However, as an inherent limitation
78 of bulk studies, we failed to identify more granular cell states within each cell type. While single-cell RNA
79 sequencing (scRNA-seq) can theoretically resolve hundreds of cell states, previous SLE studies have
80 not fully leveraged this potential, profiling only ~20 major cell types (*5*) or ~40 cell clusters (*3*, *4*), thereby
81 lacking deeper mechanistic insights. Given that low-frequency cell populations often play pivotal roles in
82 various diseases (*9*–*13*)─such as peripheral helper T cells (Tph) and autoimmune-associated B cells

83 (ABCs) in autoimmunity─a comprehensive identification of disease-relevant cell states and their
84 orchestrated networks is essential to elucidating the complex pathophysiology and clinical heterogeneity
85 of SLE. Moreover, most single-cell studies have not formally distinguished between quantitative cellular
86 abundance shifts and qualitative transcriptional changes within specific cell types. Although recent tools
87 such as milo (*14*) and CNA (*15*) enable the detection of continuous quantitative shifts, they are not
88 designed to directly evaluate qualitative changes. Thus, it is important to extend these concepts and
89 construct a statistical framework capable of jointly dissecting these two scenarios. Finally, as previous
90 SLE studies have focused primarily on gene expression (*3*–*5*), integrating multi-layered information is
91 paramount to uncovering novel therapeutic targets within the complex molecular architecture of the
92 disease.

93 To address these gaps, we extensively investigated multimodal single-cell data of SLE with a
94 total of ~2.1 million cells from 346 donors (Lupus-derived Peripheral immune cell Integrated panels
95 [LuPIN]) to pinpoint granular disease-relevant cell states (**Fig. 1**). Using fine-grained clustering and
96 rigorous validation, we defined 123 cell states across 27 immune cell types. Notably, we identified
97 several previously uncharacterized cell states specifically tracking with severe clinical phenotypes and
98 treatment response. Our statistical framework, dissecting both quantitative and qualitative gene
99 signatures, indicated that key immunological pathways and SLE risk variants were closely linked to cell-
100 state compositional shifts within each cell type. Furthermore, these newly identified cell populations were

101 predicted to engage in intricate immune crosstalk with established pathogenic cell states, potentially
102 driving SLE pathogenesis. Finally, in-depth characterization delineated the multifaceted molecular
103 landscapes of these populations, including transcriptional regulators, epigenetic features, surface
104 proteins, and T-cell receptor (TCR) repertoires. Overall, our intensive approach elucidates the complex
105 immunopathology of SLE, providing a foundation for novel therapeutic strategies.

106

107 **Results**

108 **Overview of this study**

109 LuPIN comprises three datasets, each with a different purpose (**Fig. 1, left**). To robustly identify disease-
110 relevant cell states, we first used the largest, publicly available 3'-end scRNA-seq dataset for SLE (*5*)
111 (dataset 1; **table S1**). Our stringent multi-step quality control (QC) process yielded the final dataset
112 (1,456,041 peripheral blood mononuclear cells [PBMCs] from 97 healthy and 172 SLE samples [159
113 unique SLE donors: 11 donors have 2~3 timepoints]; **fig. S1**). Among 172 SLE samples, 21 exhibited
114 high disease activity (HDA), and 40 were clinically inactive (**Methods**). Following our previous study
115 scheme (*6*), we established two comparison axes: (i) disease-state comparison between healthy controls
116 and inactive SLE patients, reflecting the biology of disease development, and (ii) disease-activity
117 comparison between inactive vs. HDA SLE, reflecting the biology of disease exacerbation. For

118 replication analysis, we used another scRNA-seq dataset (*4*) (dataset 2, 344,235 PBMCs from 17
119 healthy and 41 SLE samples).

120 To overlay in-depth molecular information onto the fine-grained cell state map, we additionally
121 prepared a new multimodal 5'-end dataset using cellular indexing of transcriptomes and epitopes by
122 sequencing (CITE-seq)─a technology profiling gene expression and surface marker proteins from single
123 cells (*16*)─alongside TCR-seq and enhancer RNA profiling (*17*, *18*) (dataset 3, 268,783 PBMCs from 32
124 SLE samples). Through the efficient integration with dataset 1, we searched for key surface proteins,
125 TCR repertoires, and cis-regulatory elements specific to disease-relevant cell states.

126

127 **Definition of 123 robust cell states using fine-grained clustering**

128 To identify granular cell states of SLE, we employed a three-step intensive clustering approach in
129 dataset 1: lineage (layer 1), cell type (layer 2), and cell state (layer 3) (**Fig. 2**). First, Louvain clustering
130 categorized 1,456,041 PBMCs into seven cell lineages (layer 1: $CD4^+$ T cells [CD4 T], $CD8^+$ and
131 unconventional T cells [CD8 and other T], natural killer cells [NK], B cells, monocytes [Mono], dendritic
132 cells [DC], and hematopoietic stem and progenitor cells [HSPC]; **fig. S1D−F**). These lineages were
133 further subdivided into 27 coarse cell types (layer 2: **Fig. 2A−F**, **fig. S2A−F**, and **table S2**). Our analytical
134 pipelines (**table S3** and **Methods**) achieved higher annotation accuracy than the original study(*5*) (**fig.
135 S2G**).

136 To determine if these coarse cell-type-level annotations sufficiently explain disease-relevant
137 compositional changes, we performed co-varying neighborhood analysis (CNA) to identify single-cell-
138 resolution transcriptional neighborhoods associated with clinical phenotypes (*15*). We applied CNA to six
139 cell lineages (excluding HSPC) and examined the abundance of 26 coarse cell types (lineage-level CNA;
140 **fig. S2H**). The extent of expansion or depletion for each cell type was quantified at the single-cell level
141 as "neighborhood correlations". Using two comparison axes─disease-state (inactive SLE vs. healthy,
142 reflecting disease development) and disease-activity CNA (HDA vs. inactive SLE, reflecting disease
143 exacerbation) (*6*)─we recapitulated the depletion of NaiveCD4 and mucosal-associated invariant T cells
144 (MAIT) in disease-state comparison (*5*), while revealing that these patterns were remarkable in disease-
145 activity comparison. Similarly, regulatory T cells (Treg), proliferating (Prolif) CD4/CD8/NK cells, and C1Q^{+}
146 monocytes (Mono) expanded in disease-state and/or -activity CNA. Although neighborhoods in these cell
147 types exhibited universal expansion or depletion patterns, many other cell types showed heterogeneous
148 patterns in lineage-level CNA (**fig. S2H**), underscoring the need for a more fine-grained investigation.
149 To efficiently capture subtle gene-program heterogeneities within each cell type, we further
150 performed clustering on each cell type and defined 123 cell states with unique gene expression profiles
151 (layer 3: **Fig. 2A−F**, **fig. S3**, and **table S4**). As these cell-state definitions form the basis for downstream
152 analyses, we employed multiple rigorous approaches to ensure their robustness. First, a leave-one-
153 batch-out analysis, systematically excluding each of the 23 experimental batches from dataset 1,

154 confirmed the consistent preservation of all 123 cell states, albeit occasionally requiring slightly higher
155 resolution parameters (**Fig. 2G, left**, **fig. S4A**, and **table S5**). Moreover, 122 cell states (except for
156 gamma-delta T cells [gdT]-5: *ISG15*$^{+}$*MX1*$^{+}$) were replicated through independent clustering of datasets 2
157 and/or 3, demonstrating the cross-dataset reproducibility (**Fig. 2G, right**, and **fig. S4B−D**). Second, the
158 identity of each cluster was corroborated by an extensive literature review (*9–13*, *19–39*) (**table S6**).
159 Third, marker genes upregulated in each state exhibited minimal overlap within the same cell type (**fig.**
160 **S4E**), and pathway analysis using these markers revealed enrichment of cell-state-specific pathways
161 (**table S7**), supporting their non-redundancy and biological relevance. Finally, multimodal investigations
162 using dataset 3 identified cell-state-specific surface proteins and TCR features (**Fig. 6−7**), providing
163 orthogonal validation at the immune phenotypic level.

164 Our fine-grained annotations successfully captured multiple disease-relevant cell state
165 candidates (**Fig. 2**), most of which remained unidentified in the original study due to insufficient
166 resolution. We first recapitulated established pathogenic states including: (i) follicular and peripheral
167 helper T cells (Tfh and Tph) within central and effector memory CD4 T cells (CMCD4-3: *ICOS*$^{+}$*PASK*$^{+}$
168 and EMCD4-7: *PDCD1*$^{+}$*TIGIT*$^{+}$) and (ii) autoimmune-associated B cells (ABCs) in memory B cells
169 (MemB-6: *ITGAX*$^{+}$*FCRL5*$^{+}$), which are central to autoantibody production (*9*, *10*, *27*). We also confirmed
170 the presence of subpopulations highly activated by interferon (IFN) signaling (*ISG15*$^{+}$*MX1*$^{+}$ cell states)─a
171 hallmark signature of SLE (*1*, *26*)─across most cell types. The shared gene program across cell types

172 was not limited to IFN signaling; *JUN*$^{+}$*CD69*$^{+}$ cell states with upregulated AP-1 transcription factors (TFs;

173 e.g., *JUN, FOS*) and early activation markers (e.g., *CD69* and *NFKBIA*) (*21*) were identified in most

174 lymphoid cell types (**fig. S3A−D**). Importantly, we identified multiple previously uncharacterized cell

175 states with distinct gene programs, some of which were later found to expand especially in severe SLE

176 (e.g., EMCD8-5: *GZMK*$^{+}$*GZMH*$^{+}$*HLA-DRB1*$^{+}$, **Fig. 3**). Moreover, our approach captured very rare

177 subpopulations such as *ANXA2*$^{+}$*CTLA4*$^{+}$ double-negative T cells (dnT-1) (*32*), *RORA*$^{+}$*GATA3*$^{+}$ innate

178 lymphoid cells (ILC-1) (*33*), *MZB1*$^{+}$*PCNA*$^{+}$ plasmablasts (PlasmaB-1) (*35*), and *LAMP3*$^{+}$*CCR7*$^{+}$ DCs

179 (cDC-5) (*39*). Overall, our intensive clustering approach provided high-resolution cell state information for

180 SLE, serving as the foundation for subsequent analyses.

181

182 **Mapping cell state compositional changes in SLE**

183 To assess the cell-state compositional changes within each cell type, we next applied CNA to 18 major

184 cell types (>5,000 cells; cell-type-level CNA) in disease-state (inactive SLE [n = 40] vs. healthy [n = 97])

185 and disease-activity (HDA [n = 21] vs. inactive SLE [n = 40]) comparisons, accounting for age, sex, and

186 ancestry. Strikingly, we observed skewed neighborhood abundance in 34 out of 36 tests (18 cell types ×

187 2 comparisons, global $P < 0.05/36$), and 29 cell states showed evidence of remarkable expansion (**Fig.

188 3A−B**, **fig. S5A**, and **table S8**). These expanded cell states tended to be rare within each cell type,

189 supporting the importance of the fine-grained strategy (**Fig. 3C** and **fig. S5B**).

190 Among the 29 cell states exhibiting remarkable expansion,18 recapitulated known signals. All
191 IFN-activated (*ISG15*$^{+}$*MX1*$^{+}$) cell states within each cell type expanded in disease-state and -activity CNA
192 (**Fig. 3D**). While Tfh (CMCD4-3) expanded primarily in disease-state, Tph (EMCD4-7) expanded
193 specifically in disease-activity CNA. Notably, in memory B cells (MemB)─which harbor the highest
194 enrichment of SLE risk variants (*40*)─ABC (MemB-6) showed remarkable expansion in disease-activity
195 CNA but not in disease-state CNA (**Fig. 3A**), supporting their relevance in pathogenic autoantibody
196 production, particularly during disease activation (*10*). Furthermore, our ancestry-specific analysis
197 revealed the expansion of this population in Asian SLE compared to European (EUR) SLE (**fig. S5C**),
198 suggesting a potential need for stratified medicine using B-cell targeted therapies (*41*, *42*).
199 The remaining 11 cell states represented previously unreported expansions in SLE (*3*–*5*),
200 nominating them as new disease-relevant candidates. An example is *GZMK*$^{+}$*GZMH*$^{+}$*HLA-DRB1*$^{+}$ cells
201 (EMCD8-5), which showed expansion in both disease-state and -activity CNA (**Fig. 3A−B**). While other
202 EMCD8 cell states express either *GZMK* or *GZMH*, this population expresses both, along with *GZMB* to
203 a lesser extent (**fig. S3B**; hereafter “double-positive [DP] EMCD8”). Notably, this population exhibited
204 high expression of multiple T cell activation markers, including *HLA-DR* genes, *PDCD1*, and *TIGIT* (**fig.
205 S3B**). Another example, *FAM13A*$^{+}$*CDK6*$^{+}$ cells (NaiveCD4-4), further highlighted the necessity of fine-
206 grained analysis: although NaiveCD4 cells were, on average, depleted within the CD4 T lineage (**fig.
207 S2H**), this depletion pattern was not uniform across NaiveCD4 cell states. Lastly, *FOXO1*$^{+}$*ARHGAP15*$^{+}$

208 cell states expanded across most T-cell cell types in disease-state and -activity CNA (**Fig. 3A−B, D**).
209 These states were broadly shared across the majority of SLE samples, rather than being derived from a
210 limited number of donors (**fig. S5D**). Some cell states showed different patterns of abundance changes
211 between disease-state and activity CNA. While *JUN*$^{+}$*CD69*$^{+}$ cell states expanded in disease-state CNA,
212 they tended to deplete in disease-activity CNA across most cell types (**Fig. 3D**), suggesting distinct roles
213 of these states in SLE depending on the disease phase (*6*).

214 Our CNA results demonstrated high reproducibility in multiple approaches. First, cluster-based
215 abundance testing (MASC) (*11*) yielded concordant results with CNA (**fig. S5E**). Second, most CNA
216 signals were globally consistent between Asian and EUR ancestries (**fig. S5F**). Third, by independently
217 applying CNA to low (LDA) and moderate disease activity (MDA) cases─both of which were excluded
218 from the discovery set─we observed concordant signals with a gradual increase in disease-activity
219 effects, reflecting the continuous spectrum of SLE activity (**Fig. 3E**). Lastly, disease-state and activity
220 CNA using independent dataset 2 well replicated the signals detected in dataset 1 (**Fig. 3F**). Collectively,
221 these results robustly captured cell-state compositional changes linked to disease development and
222 exacerbation of SLE.

223

224 **Statistical framework to dissect quantitative and qualitative changes within cell types**

225 Having identified SLE-associated cellular compositional shifts within each cell type, we next introduced a
226 statistical framework to address the limitations of conventional cell-type-level analyses. Cell-type-level
227 analyses define genes driven by either of the following two distinct scenarios as disease-associated
228 differentially expressed genes (DEGs, **Fig. 4A**): (i) a gene is specifically expressed in a particular cell
229 state whose abundance shifts with phenotype (quantitative change) or (ii) a gene's per-cell expression is
230 directly dysregulated by the phenotype, regardless of cell-state abundance (qualitative change). To
231 dissect their respective pathogenic effects, our generalized linear mixed-effect model jointly regresses
232 each single-cell expression by (i) a phenotype-driven cell-state abundance term and (ii) a direct
233 phenotype term (e.g., inactive SLE vs. healthy) capturing per-cell gene dysregulation, while adjusting for
234 donor/batch effects and other covariates (**Fig. 4B**; **Methods**). Specifically, we incorporated single-cell-
235 level neighborhood correlations from cell-type-level CNA (**Fig. 3**) into the model as metrics for cell-state
236 abundance changes. By applying this model across 18 major cell types, we identified (i) "abundance
237 signature genes" (quantitative changes) and (ii) "dysregulated signature genes" (qualitative changes,
238 **table S9**). Notably, quantitative changes drove approximately 50% of the variation in the top 1,000
239 pseudobulk DEGs for each cell type (**fig. S6A**).

240 We tested the robustness of this model through multiple approaches. First, permutation tests
241 showed that this model is well calibrated (**fig. S6B**). Second, a leave-one-batch-out analysis,
242 systematically excluding each batch from dataset 1, confirmed the stability of model statistics across

genes (**fig. S6C**), demonstrating the robustness to technical confounders and donor-specific effects. Finally, application of this model to dataset 2 showed good reproducibility of model statistics (**fig. S6D**). For simplicity, we hereafter present disease-state comparison results in **Fig. 4C−F** and disease-activity results in **fig. S6E−F**.

IFIH1 and *OAS1* are IFN response genes (IRGs) linked to the genetic loci in genome-wide association studies of SLE (*43*) (SLE-GWAS; **table S10**). While cell-type-level pseudobulk analysis identified both as DEGs in NaiveCD4 (false discovery rate [FDR] < 0.05), our model revealed different behaviors of qualitative and quantitative changes for these genes (**Fig. 4C**). While *OAS1* expression localized to NaiveCD4-7 ($ISG15^{+}MX1^{+}$), which expanded in SLE (abundance signature gene), *IFIH1* did not show such localization and was globally overexpressed in SLE compared with healthy cells (dysregulated signature gene). Hierarchical clustering of 100 IRGs (**table S11**) based on model statistics across 18 cell types identified distinct modules representing abundance and dysregulated signature genes, respectively, suggesting heterogeneous roles for IRGs in SLE pathophysiology (**Fig. 4D** and **fig. S6E**).

Next, we performed pathway enrichment analyses on these signatures for each cell type (**Fig. 4E**). IFN signaling was enriched in both dysregulated and abundance signatures across almost all cell types (FDR < 0.05; one-sided Fisher's exact test), demonstrating that IFN signaling exerts pathogenic effects through both quantitative and qualitative changes. In contrast, other classical immunological

261 pathways were primarily enriched in abundance signature genes of relevant cell types (e.g., complement
262 signaling in myeloid cells; **Fig. 4E**, **fig. S6F**, and **table S12**), suggesting that key biological processes in
263 SLE are more closely linked to abundance signatures than dysregulated signatures.

264 Finally, we integrated these model statistics with the largest-scale SLE-GWASs in East Asian
265 (EAS) (*44*) and EUR (*45*). We evaluated the polygenic enrichment of SLE risk variants around signature
266 genes using stratified linkage disequilibrium score regression (S-LDSC) (*40*), followed by a meta-
267 analysis of the EAS and EUR results **(fig. S6G**). While SLE risk variants were enriched around both
268 dysregulated and abundance signature genes across multiple cell types (FDR < 0.05), the enrichment
269 was more pronounced for abundance signatures ($P = 5.3 \times 10^{-4}$; paired Wilcoxon test; **Fig. 4F**). Notably,
270 this enrichment was specifically observed in abundance signature genes with positive statistics (**fig.**
271 **S6H**). Consistent with previous studies (*40*), strongest enrichment was identified in B cells. Although
272 these findings do not imply that SLE risk variants directly regulate cell-state compositional changes, our
273 framework suggests that SLE risk genes exert their effects primarily within these expanded cell states.
274

275 **Characterization of aberrant immune networks between disease-relevant cell states**

276 Motivated by our finding that key immunological pathways and SLE risk variants were predominantly
277 enriched in cell-state compositional changes (**Fig. 4**), we further investigated their clinical relevance.
278 Hierarchical clustering of all samples and cell states based on their proportions revealed that samples

279 broadly partitioned into SLE (sample group [SG]1) and healthy controls (SG2) (**Fig. 5A**). While a subset
280 of SLE patients clustered within SG2, SG1 was further stratified into four subgroups (SG1-1 to SG1-4).
281 Similarly, clustering of cell states identified six cell-state groups (CGs). Intriguingly, while all IFN-
282 activated (*ISG15*$^{+}$*MX1*$^{+}$) cell states correlated within CG3, other SLE-expanded states─including well-
283 known pathogenic (Tfh, Tph, and ABCs) and novel populations (e.g., DP EMCD8 and
284 *FOXO1*$^{+}$*ARHGAP15*$^{+}$ cells)─formed another distinct group (CG2). This suggests that these populations
285 are immunologically interconnected, partially independent of IFN signaling. Indeed, SG1-1 was
286 characterized by a remarkable expansion of CG3, while SG1-2 (and a subset of SG1-3) exhibited CG2
287 expansion, highlighting the cellular heterogeneity underlying SLE. Both SG1-1 and SG1-2 comprised a
288 higher proportion of HDA patients; conversely, SG1-4 and SG2 were enriched for inactive or LDA
289 patients (**Fig. 5B**). SG1-4 showed *JUN*$^{+}$*CD69*$^{+}$ (CG6) expansion (**Fig. 5A**), potentially marking a clinically
290 quiescent state. We found no significant associations between SGs and organ activity or treatment
291 agents (**fig. S7A**).

292 Leveraging longitudinal samples in dataset 1 (from “Flare” to “Treated” samples from 10
293 donors), we compared changes in cell-state proportions. Notably, while many CG2 cell states (e.g., Tph
294 and DP EMCD8) decreased following treatment, CG3 (*ISG15*$^{+}$*MX1*$^{+}$) remained stable or even increased
295 (**Fig. 5C** and **fig. S7B**), suggesting that CG2 might serve as a cellular biomarker reflecting disease
296 activity and treatment response.

297 Next, we inferred the inflammatory networks underlying these correlated cell states. By
298 evaluating 4,701 ligand−receptor (LR) interactions (*46*) across all 123 × 123 cell-state pairs, we screened
299 an extensive signaling space and identified 1,451,770 potentially active signals (LR score > 0.6;
300 **Methods**). To minimize redundancy, we prioritized the most upregulated cell-state pair as the
301 representative for each unique LR interaction within each of the 27 × 27 cell-type pairs. Among these
302 118,629 cell-state-specific interactions (of which 52,599 occurred within 18 × 18 major cell-type pairs),
303 9,428 were predicted to occur between states within the same CG. Intriguingly, within-CG2 interactions
304 accounted for more than one-third (3,386) of these signals, suggesting a prominent immune network
305 among SLE-expanded cell states (**Fig. 5D**). Within CG2, we delineated both established and novel
306 candidate disease-relevant networks (**Fig. 5E** and **table S13**). First, our analysis recapitulated enhanced
307 IL21-IL21R/IL2RG signaling from EMCD4-7 (Tph) to NaiveB-4 (*SOX5*$^{+}$*FCRL5*$^{+}$), a signaling axis
308 essential for their differentiation into MemB-6 (ABC) (*10*). Notably, we identified potential IFNG-IFNGR
309 signaling from EMCD8-5 (DP EMCD8) to NaiveB-4, which is also implicated in ABC differentiation (**Fig.
310 5E−F**), suggesting that Tph and DP EMCD8 may synergistically drive the maturation of pathogenic B
311 cells (*10*, *47*). Furthermore, Tph and DP EMCD8 were predicted to engage in bidirectional costimulatory
312 crosstalk (e.g., CD70-CD27 signaling) (*48*, *49*), mutually enhancing their effector functions (**Fig. 5E, G**).
313 Additionally, other CG2 cell states, including NaiveCD4-4 (*FAM13A*$^{+}$*CDK6*$^{+}$), NaiveCD4-5 and CMCD4-4
314 (*FOXO1*$^{+}$*ARHGAP15*$^{+}$), were predicted to amplify these pro-inflammatory cascades (*50–52*) (**Fig. 5E** and

315 **fig. S7C−E**). Collectively, high-resolution cell–cell interaction analysis provided a putative map of intricate
316 immune networks between established and newly identified pathogenic cell states, potentially driving the
317 complex immunopathology of SLE.
318

319 **Multimodal molecular profiling across fine-grained cell states**

320 To characterize multi-layered molecular profiles across fine-grained cell states, we prepared a new
321 dataset (dataset 3; 32 SLE donors) comprising CITE-seq (130 surface marker proteins), TCR-seq, and 5'
322 transcript-based enhancer RNA (eRNA) profiling. Reference mapping (*53*) successfully integrated
323 268,783 PBMCs from dataset 3 into the latent transcriptional space in dataset 1, extending our fine-
324 grained analysis across multimodal layers (**fig. S8A**). First, we comprehensively interrogated cell-state-
325 specific surface protein expression, identifying an average of nine upregulated proteins per cell state
326 (FDR < 0.05; **fig. S8B** and **table S14**). This analysis recapitulated well-established markers, such as (i)
327 CD278 (ICOS), CD279 (PD1), and TIGIT in Tph (EMCD4-7) (*9*), (ii) CD11c and CD19 in ABC (MemB-6)
328 (*10*), and (iii) CD38 and CD31 in *SOX4*$^{+}$ naive T cell states (NaiveCD4-3 and NaiveCD8-3) (*22*).
329 Furthermore, we identified several key proteins characterizing new disease-relevant cell states, which
330 facilitated our subsequent downstream investigations (**Fig. 6−7**).

331 We next examined TCR repertoire profiles. While we and others have shown that V gene usage
332 differs across cell types (*54*, *55*) and represents a potential therapeutic target for autoimmunity (*56*),

333 systematic analysis at the cell-state resolution has been lacking. Principal component analysis (PCA) of
334 V-gene frequencies confirmed that cell-state embeddings globally clustered by cell type (**fig. S8C**).
335 Notably, among 94 V genes (45 α- and 49 β-chains), ~3.8 V genes per cell state exhibited skewed usage
336 (FDR < 0.05; two-sided Fisher's exact test; **fig. S8D** and **table S15**), highlighting distinct TCR signatures
337 unique to specific cell states. While most of them were observed in EMCD8 cell states, we identified
338 several cell-state-specific V genes in other cell types (e.g., NaiveCD4-5; **Fig. 6**).
339 Leveraging 5' transcript sequencing, we mapped transcription start sites (TSSs) at single-cell
340 resolution. Recent studies have shown that bidirectional transcription around TSSs marks enhancer
341 activity (*17*, *18*). Using this approach, we defined ~1,564 transcribed enhancers per cell type (**Methods**),
342 which were enriched for established PBMC enhancer marks (*57*) (**fig. S8E**). Among them, we identified
343 ~68 cell-state-specific enhancers per cell type (**table S16**). Thus, the efficient integration of multimodal
344 data with large-scale scRNA-seq data provided a comprehensive molecular map of fine-grained cell
345 states.
346

347 ***FOXO1* is a candidate driver of the pathogenic CD4$^+$ T cell state gene program**

348 We next investigated the molecular profiles of disease-relevant cell state candidates that exhibited
349 expansion in both disease-state and disease-activity CNA (**Fig. 3A−B**). Specifically, we interrogated the

350 molecular features of newly identified CD4$^+$ T cell states within CG2 (**Fig. 5**): (1) *FAM13A*$^+$*CDK6*$^+$ cells
351 (NaiveCD4-4), and (2) *FOXO1*$^+$*ARHGAP15*$^+$ cells (NaiveCD4-5 and CMCD4-4).

352 Pathway analysis of upregulated marker genes revealed an enrichment of JAK-STAT signaling
353 in *FAM13A*$^+$*CDK6*$^+$ cells and TCR signaling in *FOXO1*$^+$*ARHGAP15*$^+$ cells, reflecting their distinct
354 immunological roles (**Fig. 6A** and **table S7**). Cell-state-specific enhancers near the marker genes
355 suggested unique epigenetic regulations of these states (**fig. S9A** and **table S16**). Regarding ligand-
356 receptor signaling, both cell states were predicted to engage in pro-inflammatory networks within CG2,
357 including ADA–DPP4, CALM1–PTPRA, MIF−CD74/CD44, and COPA−CD74 axes (*50*–*52*), which
358 interact with Tph and DP EMCD8 cells (**Fig. 5D**, **fig. S7D−F, and table S13**). Beyond the CG2 network,
359 these states exhibited unique ligand-receptor expression profiles, potentially forming extensive crosstalk
360 with multiple cell types (**Fig. 6B**). Consistent with the upregulated TCR signaling pathway, NaiveCD4-5
361 and CMCD4-4 highly expressed *CD28*, *TNFSF8*, *PTPRA*, *CD44*, and *ITGA4*, suggesting an "effector-
362 primed" states triggered by persistent antigen stimulation (*49*, *50*, *58*); indeed, NaiveCD4-5 further
363 exhibited a unique TCR V gene usage pattern (**Fig. 6C** and **table S15**).

364 Surface protein profiling further characterized these states: *FAM13A*$^+$*CDK6*$^+$ cells highly
365 expressed CD127 (IL-7Rα) (FDR < 0.05; **fig. S9B** and **table S14**), reflecting their distinct homeostatic
366 requirements (*59*). In *FOXO1*$^+$*ARHGAP15*$^+$ cells, the upregulation of specific activation markers including
367 CD244, CD32, CD33, and CD38 was consistent with chronic antigen exposure via TCR signaling (*60*,

368 *61*) (**Fig. 6D**). Intriguingly, these upregulations were less pronounced at the mRNA level (**table S4**),
369 highlighting the significance of multi-layered investigation (*13*, *16*, *19*). We also examined the localization
370 of GWAS-prioritized genes: in *FAM13A*$^+$*CDK6*$^+$ cells, the transcription coactivator *ARID5B* was linked to
371 an SLE-GWAS locus (*43*) (**Fig. 6E** and **table S10**). In *FOXO1*$^+$*ARHGAP15*$^+$ cells, multiple genes
372 including *ARHGAP15* and *PRKCB* were associated with SLE-GWAS loci.

373 To identify transcriptional regulators driving the gene program of *FAM13A*$^+$*CDK6*$^+$ and
374 *FOXO1*$^+$*ARHGAP15*$^+$ cells, we integrated computational and experimental approaches, focusing on 22
375 candidate TFs upregulated in these states: (i) pathway and TF enrichment analyses, TF activity
376 inference based on (ii) decoupleR (*62*) and (iii) curated ChIP-seq data (*63*) limited to blood tissue, and
377 (iv) an arrayed CRISPR-KO experiment in primary CD4$^+$ T cells (**fig. S9C** and **table S17−19**). Notably,
378 *FOXO1* consistently showed significant enrichment for NaiveCD4-5 and CMCD4-4 signature genes
379 across all approaches (Bonferroni-corrected $P < 0.05$, **Fig. 6F**). Within EMCD4 cells, *FOXO1* exhibited
380 its highest expression in Tph (EMCD4-7; **Fig. 6G** and **fig. S3A**). Trajectory analysis (*64*) on whole CD4$^+$
381 T cells also suggested a potential transcriptional continuity between *FOXO1*$^+$*ARHGAP15*$^+$ cells and Tph.
382 Recent studies have established that *FOXO1* is essential for maintaining T cell stemness and memory
383 programming (*24*, *25*). While persistent antigen exposure drives the T-cell activation program in
384 *FOXO1*$^+$*ARHGAP15*$^+$ cells, *FOXO1* may act as a differentiation gatekeeper, preventing terminal
385 exhaustion and maintaining a long-lived reservoir of pathogenic effectors. In contrast, *ARID5B* did not

386 robustly account for the *FAM13A*$^{+}$*CDK6*$^{+}$ gene program across our approaches (**fig. S9D**), although the
387 limitations of the in vitro context must be considered. Together, our multi-layered investigation pinpointed
388 key regulatory candidates in pathogenic CD4^{+} T cell states of SLE.
389

390 **DP EMCD8 cells as a candidate pathogenic player in SLE**

391 Among the new cell states within CG2, we next performed in-depth investigation of DP EMCD8 (EMCD8-
392 5; *GZMK*$^{+}$*GZMH*$^{+}$*HLA-DRB1*$^{+}$). Notably, this population shared several transcriptional signatures with
393 Tph beyond cell lineages. Pathway analysis revealed an enrichment of oxidative phosphorylation
394 signaling in both populations, suggesting shared metabolomic reprogramming (**Fig. 7A** and **table S7**).
395 Furthermore, DP EMCD8 exhibited the highest "classical" exhaustion signature (e.g., *PDCD1* and *TIGIT*,
396 **table S20**) within EMCD8, mirroring the pattern of Tph within EMCD4. (**Fig. 7B**). Similarly, *RGS1*—an
397 SLE GWAS-prioritized gene and exhaustion marker—was upregulated in both (**fig. S10A**). While these
398 signatures were traditionally regarded as markers of T cell exhaustion, recent studies in autoimmunity
399 indicate that they reflect a highly activated state induced by chronic antigen exposure, rather than a
400 dysfunctional state (*9*, *65*). Consistently, DP EMCD8 showed evidence of TCR clonal expansion in
401 multiple SLE donors (**Fig. 7C**), with some expanded clonotypes potentially reactive to cytomegalovirus or
402 Epstein-Barr virus epitopes (**table S21**). Although many clones were shared with other EMCD8 states,

403 nine clonotypes and nine V genes were specific to DP EMCD8 (FDR < 0.05; two-sided Fisher's exact
404 test; **fig. S10B−C, table S15** and **21**).

405 As discussed above (**Fig. 5E−G**), DP EMCD8 cells were predicted to engage in pro-
406 inflammatory networks with other pathogenic populations. Specifically, DP EMCD8 and Tph may
407 synergistically drive the maturation of pathogenic B cells (ABCs) via *IFNG* and *IL21* signaling, while
408 potentially maintaining a feed-forward activation loop through bidirectional costimulatory crosstalk (e.g.,
409 CD70−CD27). Longitudinal analysis also showed that both DP EMCD8 and Tph decreased following
410 treatment (**Fig. 5C**), tracking with disease activity. Furthermore, DP EMCD8 exhibited high expression of
411 pro-inflammatory cytokines (*CCL5*, *IL16*, and *IL32*) and functional markers (*CD74*, *CD2*, and *ITGA4*; **Fig.
412 7D**), underscoring their activated effector function rather than exhaustion. Intriguingly, while DP EMCD8
413 cells expressed lower *GZMB* levels than classical cytotoxic states (EMCD8-2; **fig. S3B**), they showed
414 marked *GZMA* upregulation (**Fig. 7D**), suggesting that they might also mediate caspase-independent
415 cytotoxicity (*66*).

416 To identify key surface proteins, we integrated the multimodal dataset 3 into dataset 1,
417 identifying HLA-DR, CD27, CD38, and CD28 as the most upregulated proteins in DP EMCD8 (**Fig. 7E
418 and table S14**). Notably, CD45RA was downregulated, indicating that these cells are distinct from
419 terminally differentiated effector memory (TEMRA) cells (*67*). In-silico gating on HLA-
420 $DR^+CD27^+CD38^+CD28^+$ achieved >90% purity for DP EMCD8 (**fig. S10D**), enabling validation via flow

421 cytometry (FCM) and FACS. We sorted target EMCD8 cells (CCR7$^-$HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$) and
422 performed bulk RNAseq and ATAC-seq (**Fig. 7F**, **fig. S10E**, and **table S22**), confirming a gene
423 expression profile comparable to DP EMCD8 (**fig. S10F**). Motif analysis for differentially accessible
424 regions (DARs) in this population compared with whole EMCD8 (**table S23**) showed minimal TF
425 enrichment (**fig. S10G**). Combined with the few upregulated TFs in this population (**table S4**), this
426 suggests that the primary driver of DP EMCD8 may be extrinsic humoral signals rather than intrinsic
427 transcriptional reprogramming.

428 Since no DARs were initially detected near the *IFNG* locus in standard ATAC-seq, we employed
429 UNIChro-seq (*68*) to enrich tagmented DNA fragments in this region (**table S24**). A ~20-fold enrichment
430 compared with standard ATAC-seq identified a unique DAR (**Fig. 7G**). Consistently, this population
431 exhibited higher IFN-γ protein expression than whole EMCD8 upon in-vitro stimulation, alongside
432 elevated PD-1 and TIGIT levels (**Fig. 7H** and **fig. S10H**), suggesting dysregulation of the IFN-γ axis
433 across multimodal layers. Accumulating evidence suggests that IFN-γ (type II IFN), primarily produced by
434 CD8$^+$ T cells, plays a pivotal role in SLE, alongside type I IFN (*69–71*). Our strategy successfully
435 pinpointed DP EMCD8 as the cellular origin of this key cytokine with a pro-inflammatory phenotype,
436 underscoring its role as a candidate pathogenic player in SLE.

437

438 **Discussion**

439 In this study, we employed an in-depth investigation strategy to pinpoint disease-relevant cell states and
440 molecules from large-scale single-cell datasets of SLE. Fine-grained clustering resolved 123 cell states
441 across 27 cell types, including previously uncharacterized populations distinctively associated with
442 clinical severity and treatment status. Subsequent multimodal characterization delineated cell-state-
443 specific immune signaling networks, transcriptional regulators, epigenetic features, surface proteins, and
444 TCR repertoires.

445 While recent studies suggest that granular cell states within each cell type can exhibit
446 continuous plasticity (*72*), our results demonstrate that a fine-grained cell-state definition is essential for
447 a deeper understanding of disease pathophysiology, facilitating: (i) the interpretation of cluster-free cell
448 abundance testing (**Fig. 3**), (ii) patient stratification with high clinical relevance based on cell-state
449 abundance (**Fig. 5**), (iii) the delineation of aberrant immune crosstalk (**Fig. 5**), and (iv) the functional
450 characterization of disease-relevant populations and the identification of their potential therapeutic
451 targets (**Fig. 6−7**). Since disease-relevant cell states are often low-frequency populations easily obscured
452 by coarse clustering, our approach is consistent with studies that successfully identified rare pathogenic
453 subsets (*9*, *10*, *13*). Leveraging an extensive literature review (*9–13*, *19–39*) and rigorous validation, we
454 reconstructed 123 cell states with distinct immunological features from the largest publicly available
455 single-cell PBMC dataset (*5*) (~1.5 million cells). While recent studies on inflamed tissues in autoimmune
456 diseases have identified critical parenchymal subpopulations (*13*, *73*), the resolution of immune cells in

457 those studies remained limited (~60 immune cell states from 0.2~0.4 million cells). To benefit the
458 research community, we have released LuPIN, a high-resolution immune cell reference panel integrated
459 with a comprehensive atlas of immune signaling networks and multimodal features. Researchers can
460 map their datasets onto LuPIN for precise cell-state annotation and mechanistic inference, providing
461 opportunities to uncover new pathophysiology across multiple diseases.

462 We also proposed a statistical framework to dissect quantitative and qualitative changes linked
463 to disease within each cell type. Although methods like miloDE (*74*) infer differential expression unique to
464 transcriptional neighborhoods, they were not designed to jointly dissect abundance shifts and per-cell
465 gene dysregulation. Distinguishing these scenarios is a critical step in identifying therapeutic targets, as
466 each may play distinct roles in pathophysiology. The significant enrichment of SLE risk variants within
467 expanded cell states further suggests that these populations are pivotal for variant-to-function research
468 (*75*). While our model demonstrated reasonable robustness and reproducibility, its performance relies on
469 the latent space defined by current sparse single-cell gene expression. The distinction between
470 abundance and dysregulated signatures may be further refined as single-cell technologies advance.

471 The identification of cell-state-specific regulators and surface proteins has significant
472 implications for drug discovery. For instance, given that FOXO1 has been implicated in the efficacy of
473 cancer immunotherapy (*24*, *25*), modulating FOXO1 could represent a novel treatment option for
474 autoimmunity by targeting aberrant “effector-primed” $CD4^+$ T cell states in SLE. Regarding surface

475 proteins, current therapeutic agents often target ubiquitous markers (e.g., CD19 and CD20 in B cells)
476 (*41*, *42*), which can deplete normal cells alongside autoreactive ones. Given the recent development of
477 bispecific antibodies (*76*), narrowing the target to disease-relevant cell states based on specific markers
478 could provide more precise and efficient treatment options.

479 This study has some limitations. First, we have yet to identify the optimal combination of surface
480 markers to purify *FOXO1*$^{+}$*ARHGAP15*$^{+}$ T cells; more comprehensive panels may address this in the
481 future. Second, extremely rare cell states within minor cell types could not be included in several
482 downstream analyses due to limited cell numbers (**Fig. 3−4**); improvements in throughput will likely
483 uncover their contribution to pathogenesis. Third, as this study focused on PBMCs, key tissue-resident
484 populations may have been overlooked; thus, our approach should be extended to inflamed tissues
485 across various diseases. Fourth, our study design lacked in-vivo functional validation of the identified cell
486 states. Collectively, our findings contribute to a deeper understanding of SLE pathophysiology and
487 facilitate the discovery of novel therapeutic targets.
488

489 **Acknowledgments**

490 The super-computing resource was provided by Human Genome Center, Institute of Medical Sciences,
491 The University of Tokyo. We appreciate the generosity of the donors for a RIKEN solicited donation
492 project, Support for SLE research and development. We appreciate the RIKEN-IMS Genome Platform
493 for its help in the sequencing experiments. We thank AIDA consortium and RIKEN-IMS Laboratory for
494 Gene Structure and Regulation for helpful feedback.

495

496 **Funding**

497 This study was supported by the Japan Agency for Medical Research and Development (AMED)
498 (JP22tm0424223, JP22ek0410099, JP23tm0524005, JP25ek0410139, JP22ama121015 and
499 JP223fa627010, JP22gm1810002, JP22gm1810005), JSPS Grants-in-Aid for Scientific Research
500 (JP21K20647, JP22H03114, JP23KJ2185, JP23K15361, JP25K19621), the Uehara Memorial
501 Foundation, the Nakajima Foundation, Japan Intractable Diseases Research Foundation, GSK Japan
502 Research Grant, Daiichi Sankyo Foundation of Life Science, Mochida Memorial Foundation for Medical
503 and Pharmaceutical Research, Takeda COCKPI-T Funding, Kanzawa Medical Research Foundation,
504 SENSHIN Medical Research Foundation, Japan Research Foundation for Clinical Pharmacology, the
505 Japan Rheumatism Foundation, the Ichiro Kanehara Foundation, Keio University Academic

Development Funds, and Keio University Program for the Advancement of Next Generation Research Projects.

Authors contributions

M.N. and K.I. designed the study. M.N., K.I., and M.K. wrote the manuscript with critical inputs from T.N., Y.T., and K.Y. M.N. conducted analyses with support from K.I., M.K., K.A., H.I., T.N., J.I., H.T., and H.H. M.K. M.N., and K.A. conducted experiments with support from T. A., T. K., S. N., R. B., Y.M., B.N., A.S., and K.Y. T.K., S.K., E.K., Y.F., Y.T., and Y.M. contributed to sample collection. X.Z., S.B. and C.T. managed EAS SLE-GWAS data. All authors contributed to the final manuscript.

Competing interests

RIKEN has filed a patent application related to UNIChro-seq (WO/2024/242037). K.I, M.K, and T.A are inventors on this patent. S.K. has received speaking fees from Eli Lilly, GlaxoSmithKline, Bristol-Myers, Abbvie, Eisai, Pfizer, Astra-Zeneca, and research grants from Daiichi-Sankyo, Abbvie, Boehringer Ingelheim, and Astellas.

Data, code, and materials availability

Single-cell count matrices of gene expression, surface protein, and TCR features in the CITE-seq dataset will be available at the National Bioscience Database Center (NBDC) Human Database (https://humandbs.dbcls.jp/en/) as of the date of publication. Public data used in this study are available at https://cellxgene.cziscience.com/collections. Reference panels for each lineage and each cell type are available on GitHub (https://github.com/MasahiroNakano-hub/LuPIN). Original codes used in this study are available on GitHub (https://github.com/MasahiroNakano-hub/LuPIN).

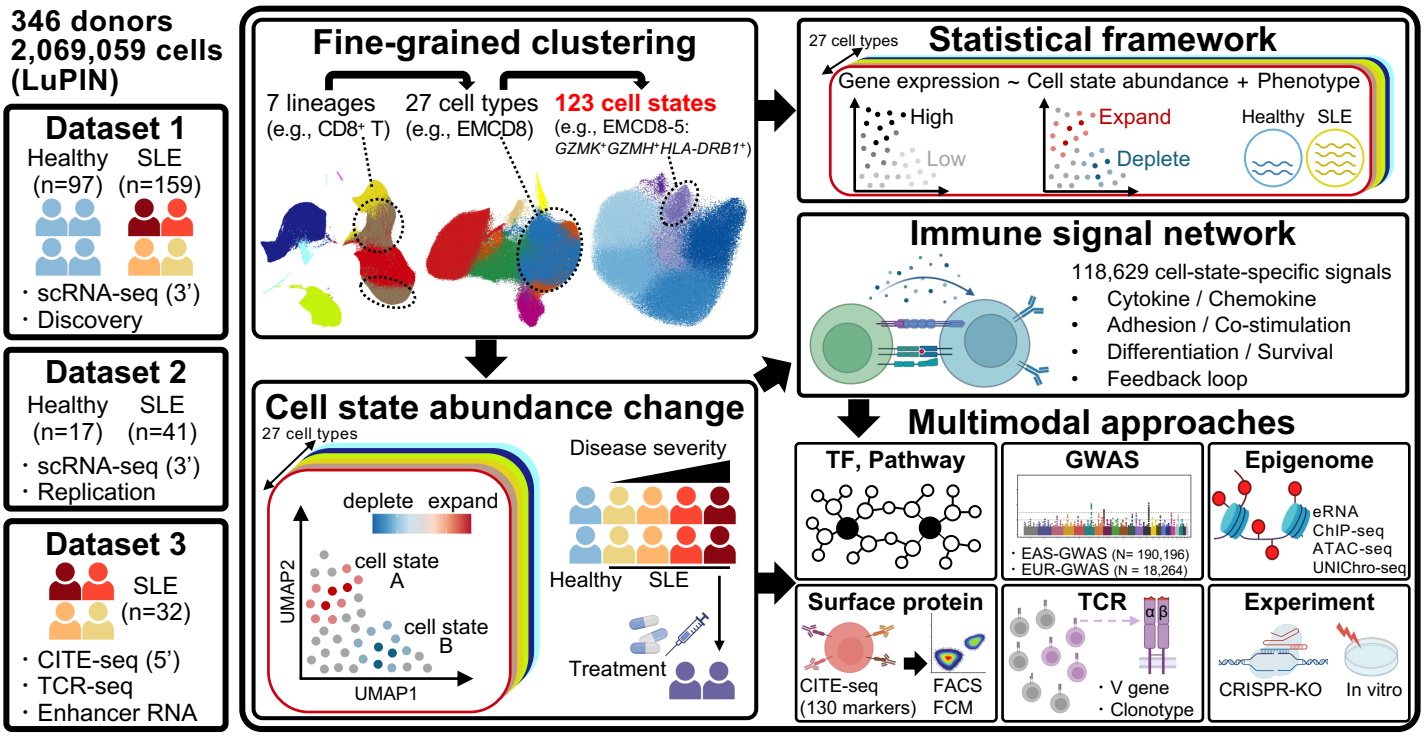

Fig. 1. Overview of this study. (left) We extensively investigated multimodal single-cell data of SLE of ~2.1 million cells from 346 donors to pinpoint granular disease-relevant cell states. **(middle)** Using fine-grained clustering, we defined 123 cell states across 27 immune cell types, identifying previously uncharacterized cell states associated with clinical severity and treatment status. **(right)** Our statistical framework indicated that key immunological pathways and SLE risk variants were linked to cell-state compositional shifts. Furthermore, these newly identified cell populations were predicted to engage in immune crosstalk with established pathogenic cell states, alongside in-depth multimodal characterization of their molecular profiles. CITE-seq, cellular indexing of transcriptomes and epitopes by sequencing; TCR, T cell receptor; EMCD8, effector memory $CD8^+$ T cells; TF, transcription factor; GWAS, genome-wide association studies; eRNA, enhancer RNA; UNIChro-seq, unique molecular identifier counting of regional chromatin accessibility with sequencing; FCM, flowcytometry; FACS, fluorescence-activated cell sorting.

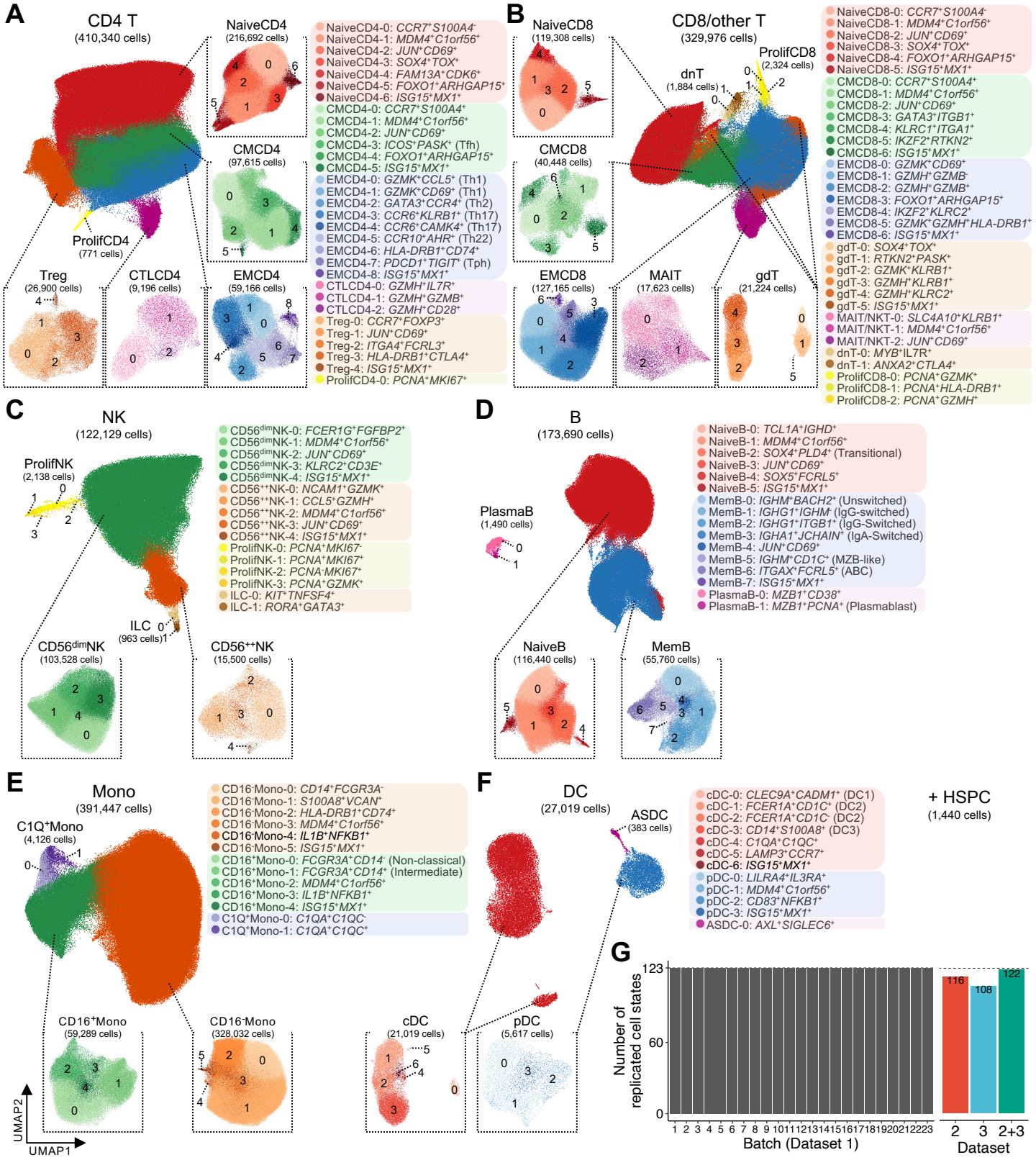

Fig. 2. Definition of 123 robust cell states using fine-grained clustering. (A-F) UMAP plots showing 27 cell types and 123 cell states identified by fine-grained clustering in **(A)** CD4+ T cell (CD4 T), **(B)** CD8+ and unconventional T cell (CD8/other T), **(C)** natural killer cell (NK), **(D)** B cell, **(E)** monocyte (Mono), and **(F)** dendritic cell (DC) lineages (dataset 1; *N* = 269). The hematopoietic stem and progenitor cell (HSPC) lineage is omitted as it comprises a single cell type and cell state. For each panel, large lineage-level UMAPs are colored by cell types, while small cell-type-level UMAPs are colored and numbered by cell states. Cell-state annotations are listed to the right of each panel. See **table S2** for cell type abbreviations. **(G)** Bar plots showing the number of cell states replicated in **(left)** leave-one-batch-out analysis of dataset 1 and **(right)** independent clustering of dataset 2 and 3.

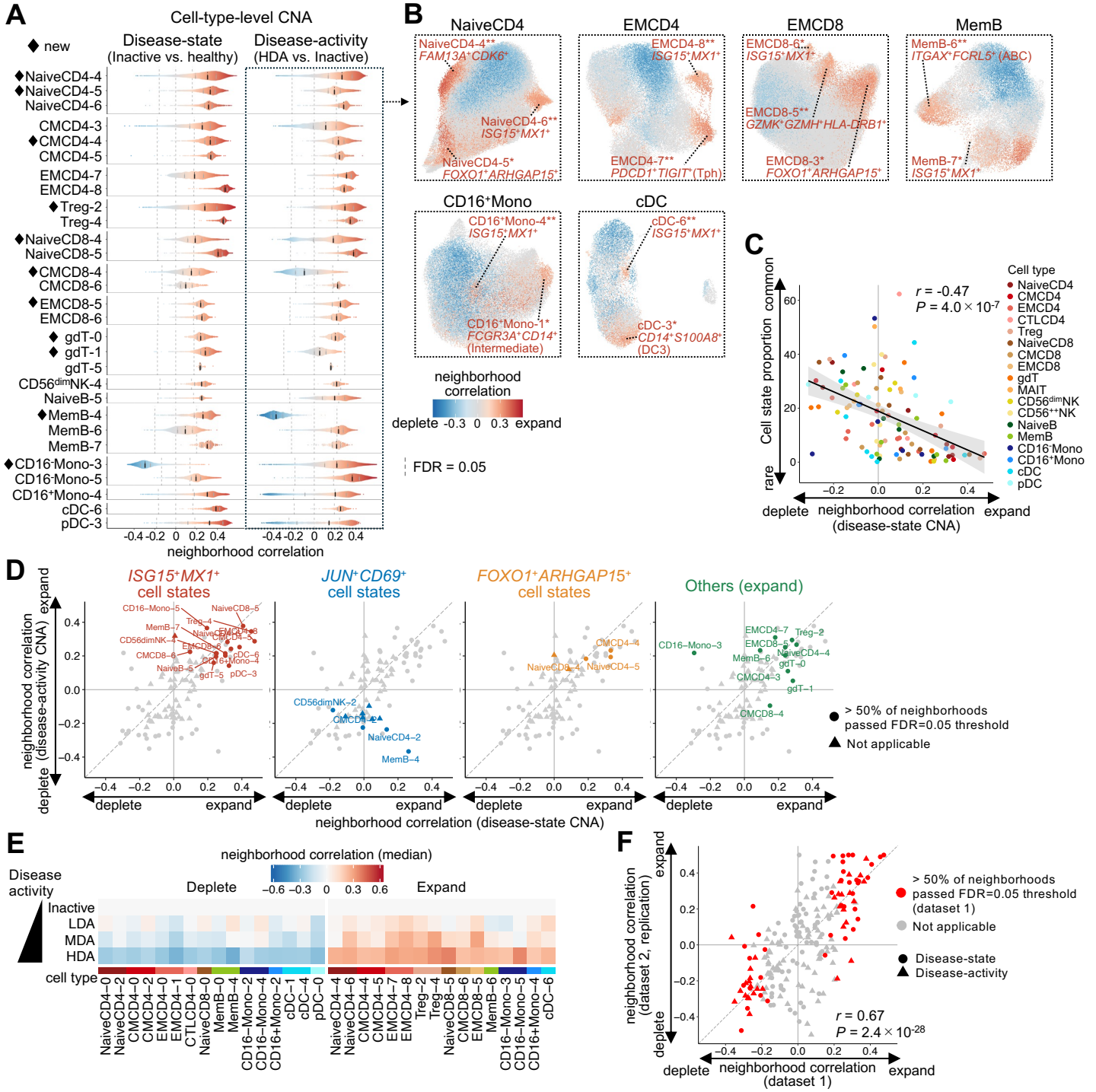

Fig. 3. Mapping cell state compositional changes in SLE. (A) Distributions of neighborhood correlations for each cell state in disease-state (n = 137) or disease-activity (n = 61) comparisons (dataset 1, cell-type-level-CNA). 29 cell states with > 50% of neighborhoods positively associated with phenotypes (FDR < 0.05) are shown (See **fig. S5A** for details). **(B)** UMAP plots of representative cell types, colored by neighborhood correlations. Cell states are highlighted if > 50% (**) or > 25% (*) of their neighborhoods are associated with phenotypes. **(C)** Scatter plot of median neighborhood correlation (disease-state CNA) for each cell state vs. cell-state proportion within cell type. Shaded regions represent 95% confidence intervals. **(D)** Comparison of median neighborhood correlation between disease-state (x-axis) and disease-activity (y-axis) CNA. Cell states relevant to each panel title are highlighted in color. **(E)** Heatmap of median neighborhood correlation for low (LDA) / moderate (MDA) / high disease activity (HDA) vs. Inactive SLE (n = 58/17/21/40, respectively). 32 disease-activity-associated cell states are shown. **(F)** Comparison of median neighborhood correlation for each cell state between dataset 1 (x-axis) and dataset 2 (y-axis; N = 58). In **(C)** and **(F)**, Pearson's r and P values are indicated.

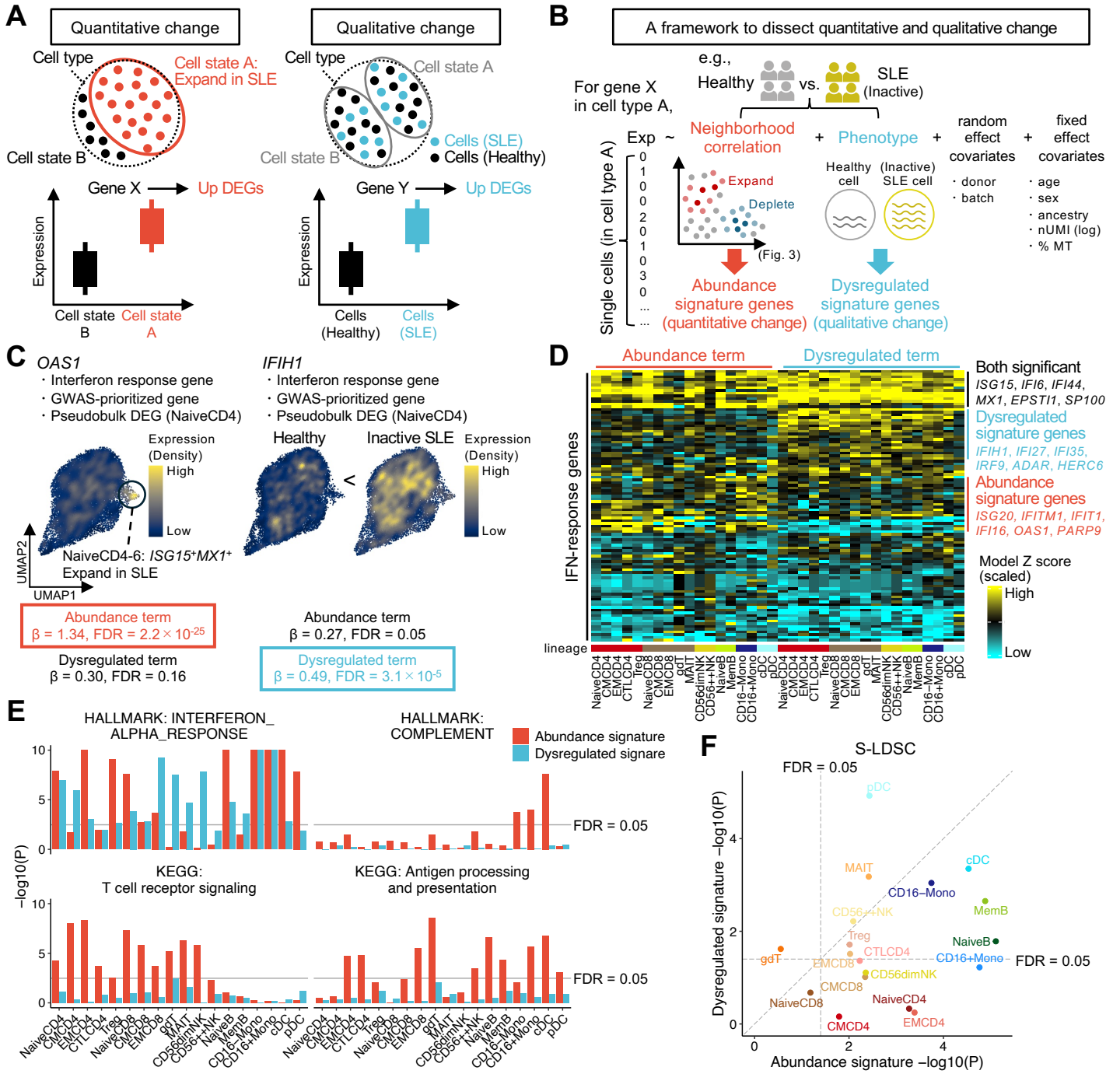

Fig. 4. Statistical framework to dissect quantitative and qualitative changes within cell types. (A) Schematic representation of two distinct scenarios (quantitative and qualitative changes) driving cell-type-level pseudobulk DEGs. **(B)** Schematic of the framework to dissect quantitative and qualitative changes. **(C)** Representative examples of abundance (*OAS1*) and dysregulated (*IFIH1*) signature genes. UMAPs are colored by expression density. Model statistics for both genes are provided below. **(D)** Heatmap of abundance and dysregulated term Z scores across 18 major cell types for 100 IFN-response genes (dataset 1; disease-state comparison; n = 137). Gene order is based on hierarchical clustering of Z scores. **(E)** Bar plots showing the enrichment of representative pathways for each signature across cell types. P, P values from one-sided Fisher's exact tests. **(F)** Comparison of SLE risk variant enrichment around abundance vs. dysregulated signature genes for each cell type using stratified LD score regression (S-LDSC).

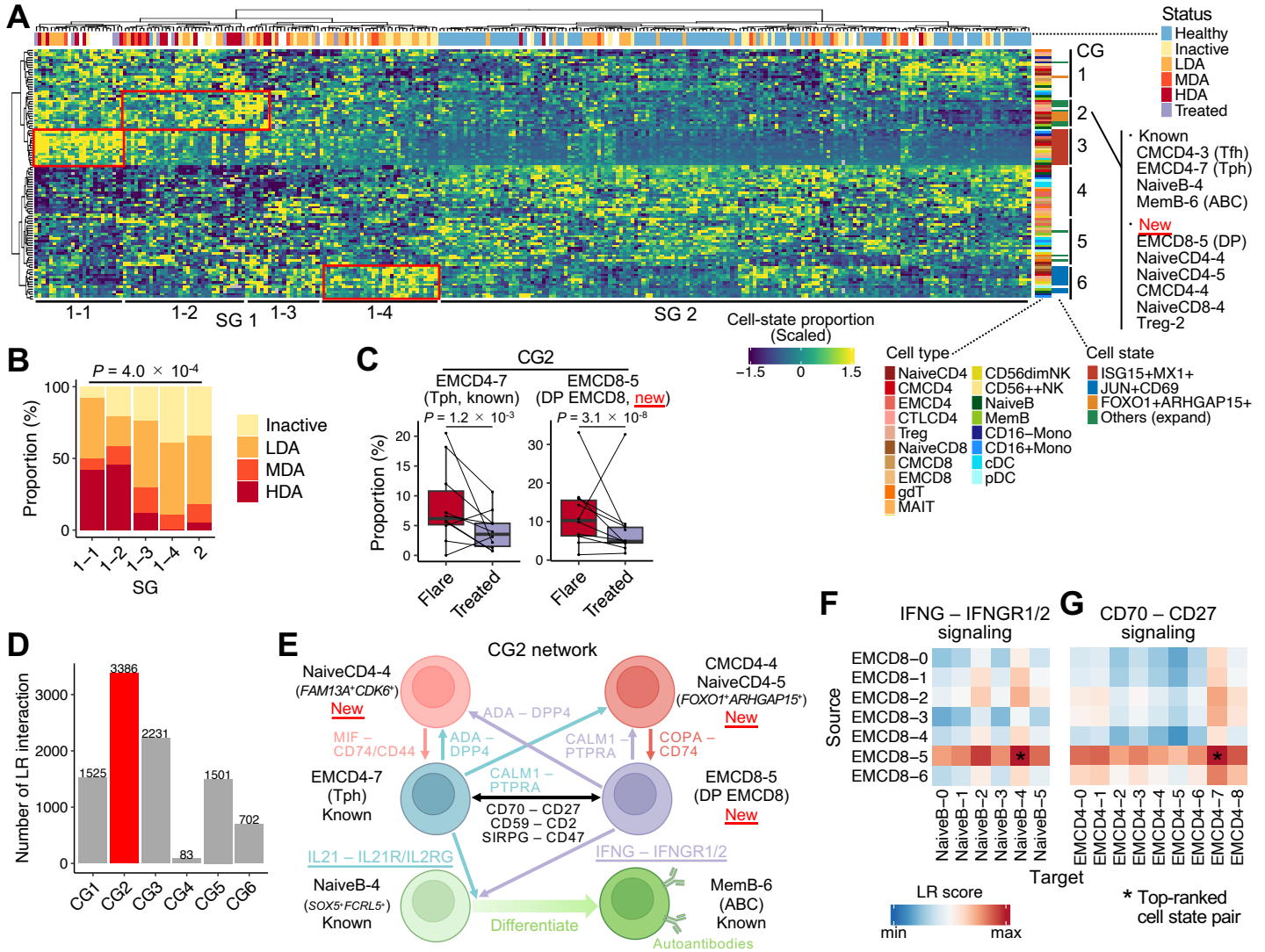

Fig. 5. Characterization of aberrant immune networks between disease-relevant cell states. (A) Heatmap of cell-state proportions across 18 major cell types for all samples (dataset 1, $N = 269$). Sample and cell-state orders are based on hierarchical clustering of cell-state proportions. **(B)** Bar plot showing the frequencies of disease-activity categories for each sample group (SG). *P*, *P* value from two-sided Fisher's exact test. LDA/MDA/HDA; low/moderate/high disease activity. **(C)** Box plots comparing cell-state proportions between "Flare" and "Treated" status in **(left)** Tph and **(right)** DP EMCD8. Samples from the same donor are connected by lines. *P*, *P* values from MASC ($n = 10 \times 2$). **(D)** Bar plot showing the number of potential cell-state-specific ligand-receptor (LR) interactions within each CG. **(E)** Schematic of cell-state-specific LR networks across known pathogenic and newly identified population in cell-state group (CG) 2. **(F-G)** Heatmaps of LR scores for **(F)** IFNG–IFNGR1/2 signaling from EMCD8 to NaiveB cell states and **(G)** CD70–CD27 signaling from EMCD8 to EMCD4 cell states.

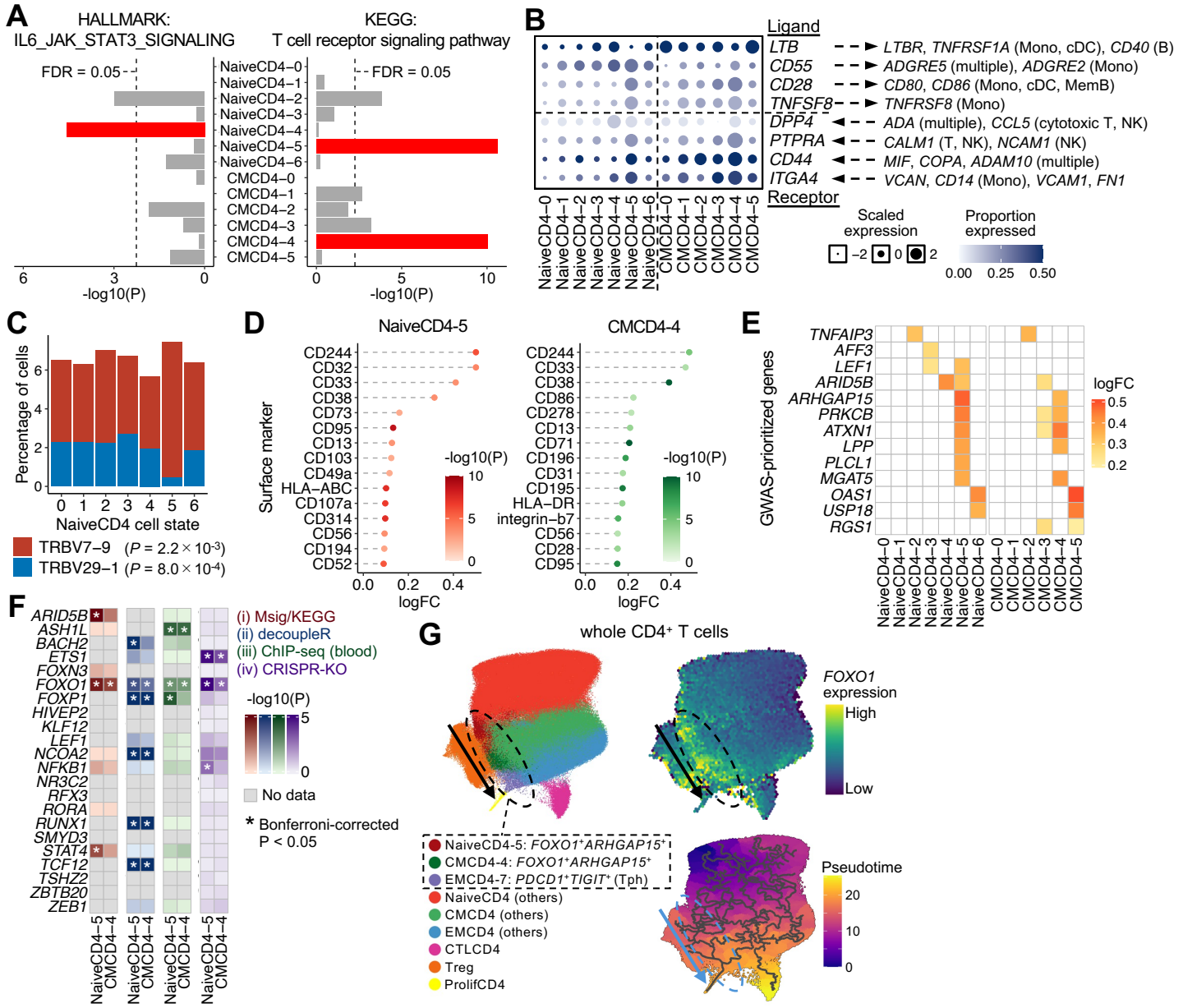

Fig. 6. *FOXO1* is a candidate driver of the pathogenic CD4⁺ T cell state gene program. (A) Bar plots showing the enrichment of representative pathways for upregulated signature genes for NaiveCD4 and CMCD4 cell states. *P*, *P* values from one-sided Fisher's exact tests. **(B)** Dot plot showing scaled expression levels of ligand and receptor genes across NaiveCD4 and CMCD4 cell states. Color intensity represents the proportion of cells expressing each gene. Corresponding interaction partners are indicated on the right. **(C)** Bar plot showing the percentage of two TCR V gene usages skewed in NaiveCD4-5. *P*, *P* values from two-sided Fisher's exact tests. **(D)** Dot plot showing the logFC of top 15 upregulated surface markers for NaiveCD4-5 and CMCD4-4. *P*, *P* values from Seurat FindMarkers. **(E)** Heatmap showing the logFC of GWAS-prioritized genes across NaiveCD4 and CMCD4 cell states. Up to five genes with the highest logFC per cell state are shown. **(F)** Heatmap showing - $\log_{10}$(enrichment *P*) values of each TF for NaiveCD4-5 or CMCD4-4 signature genes based on (i) pathway/TF enrichment analyses, TF activity inference using (ii) decoupleR and (iii) blood-derived ChIP-seq data, and (iv) an arrayed CRISPR-KO experiment. *P*, *P* values from one-sided Fisher's exact tests. **(G)** UMAP plots of whole CD4⁺ T cells, colored by **(left)** focused cell states and other cell types, **(upper right)** *FOXO1* expression, and **(lower right)** pseudotime from trajectory analysis. Edges represent trajectory graph.

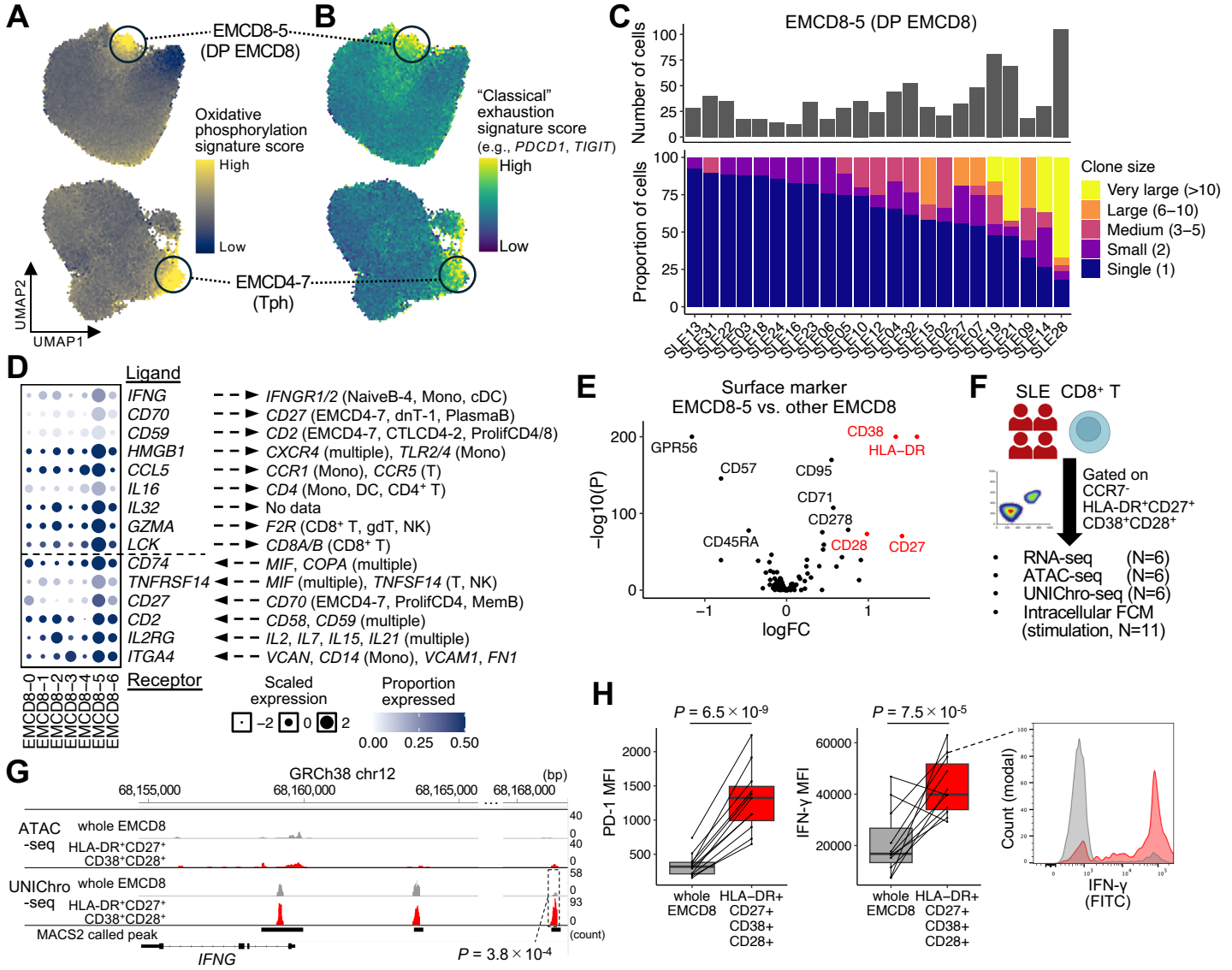

Fig. 7. DP EMCD8 cells as a candidate pathogenic player in SLE. (A-B) UMAP plots of **(top)** EMCD8 and **(bottom)** EMCD4, colored by **(A)** oxidative phosphorylation and **(B)** exhaustion signature scores. **(C)** Bar plots showing **(top)** the absolute number of EMCD8-5 cells per donor and **(bottom)** the proportion of TCR clonotypes categorized into five size groups within EMCD8-5 (dataset 3; *N* = 32; 23 donors with >10 cells are shown). **(D)** Dot plot showing scaled expression levels of ligand and receptor genes across EMCD8 cell states. Color intensity represents the proportion of cells expressing each gene. Corresponding interaction partners are indicated on the right. **(E)** Volcano plot showing differential expression of surface proteins between EMCD8-5 vs. other EMCD8 cells. *P*, *P* values from Seurat FindMarkers. **(F)** Experimental design for flowcytometry (FCM) and FACS of HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ EMCD8 cells. **(G)** ATAC-seq and UNIChro-seq tracks around *IFNG* locus in whole EMCD8 and HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ cells from an SLE donor. *P*, logistic regression *P* values. **(H) (left)** Boxplots comparing the mean fluorescence intensity (MFI) of PD-1 and IFN-γ between HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ and whole EMCD8 cells. Samples from the same donor are connected by lines. *P*, *P* values from linear mixed models. **(right)** Representative histogram comparing IFN-γ fluorescence intensity between HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ and whole EMCD8 cells from an SLE donor.

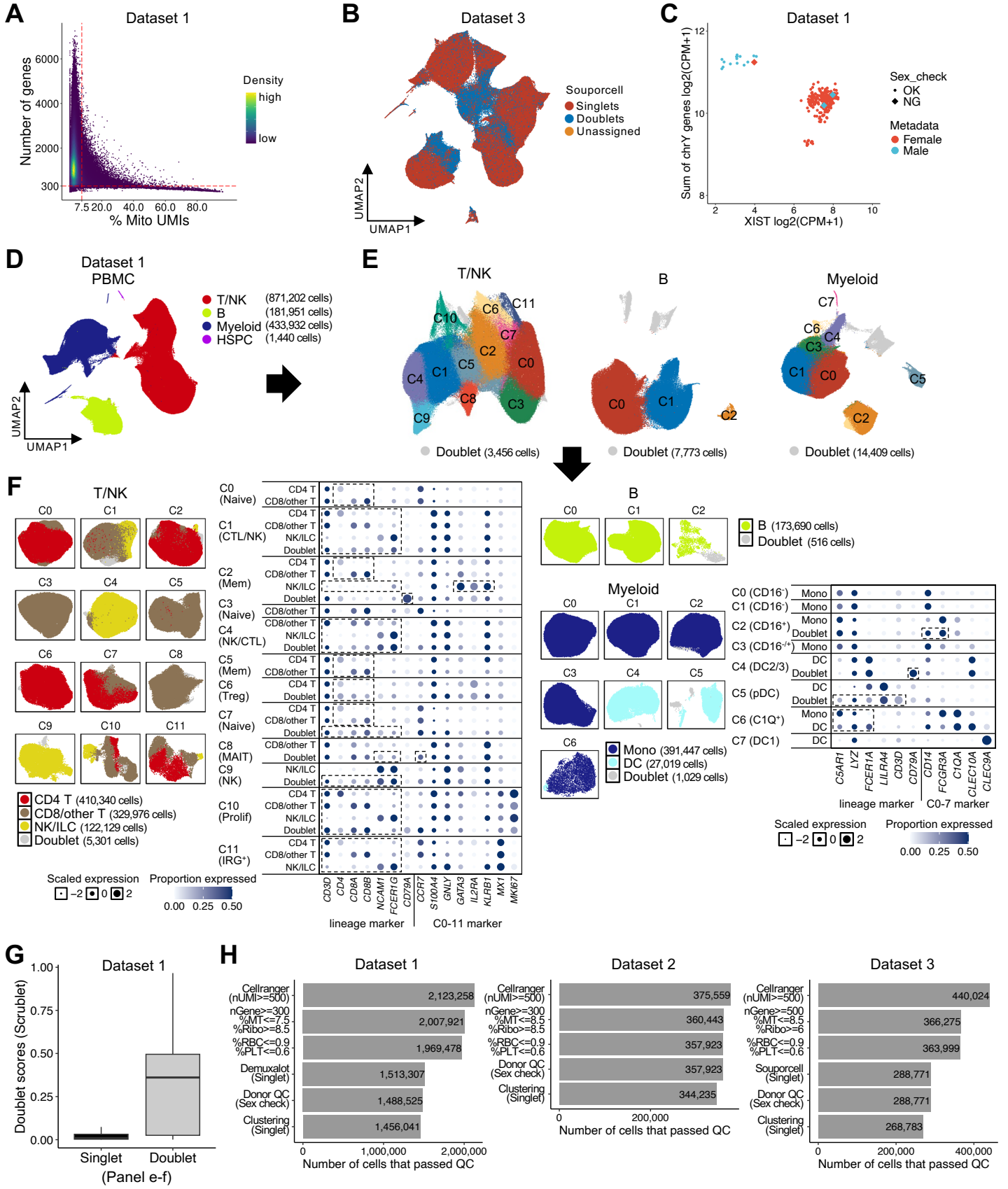

Fig. S1. Quality control of datasets. (A) Density plot showing cell distributions based on the number of detected genes and the percentage of mitochondrial UMIs (%Mito) in dataset 1 (N = 269). **(B)** UMAP plot showing PBMCs in dataset 3 (N = 32), colored by Souporcell-based singlet/doublet/unassigned classification. **(C)** Scatter plot showing donor-level pseudobulk expression of XIST and Y-chromosome-related genes in dataset 1, colored by metadata-derived sex information. **(D)-(F)** Flowchart illustrating the strategy to partition PBMCs into seven lineages in dataset 1. In **(F)**, dot plots display scaled expression levels of lineage or subcluster marker genes, with color intensity indicating the proportion of expressing cells. **(G)** Boxplot comparing Scrublet-derived doublet scores between singlets and doublets identified in **(E)** and **(F)**. **(H)** Bar plots showing the number of cells retained after each QC step in datasets 1, 2, and 3.

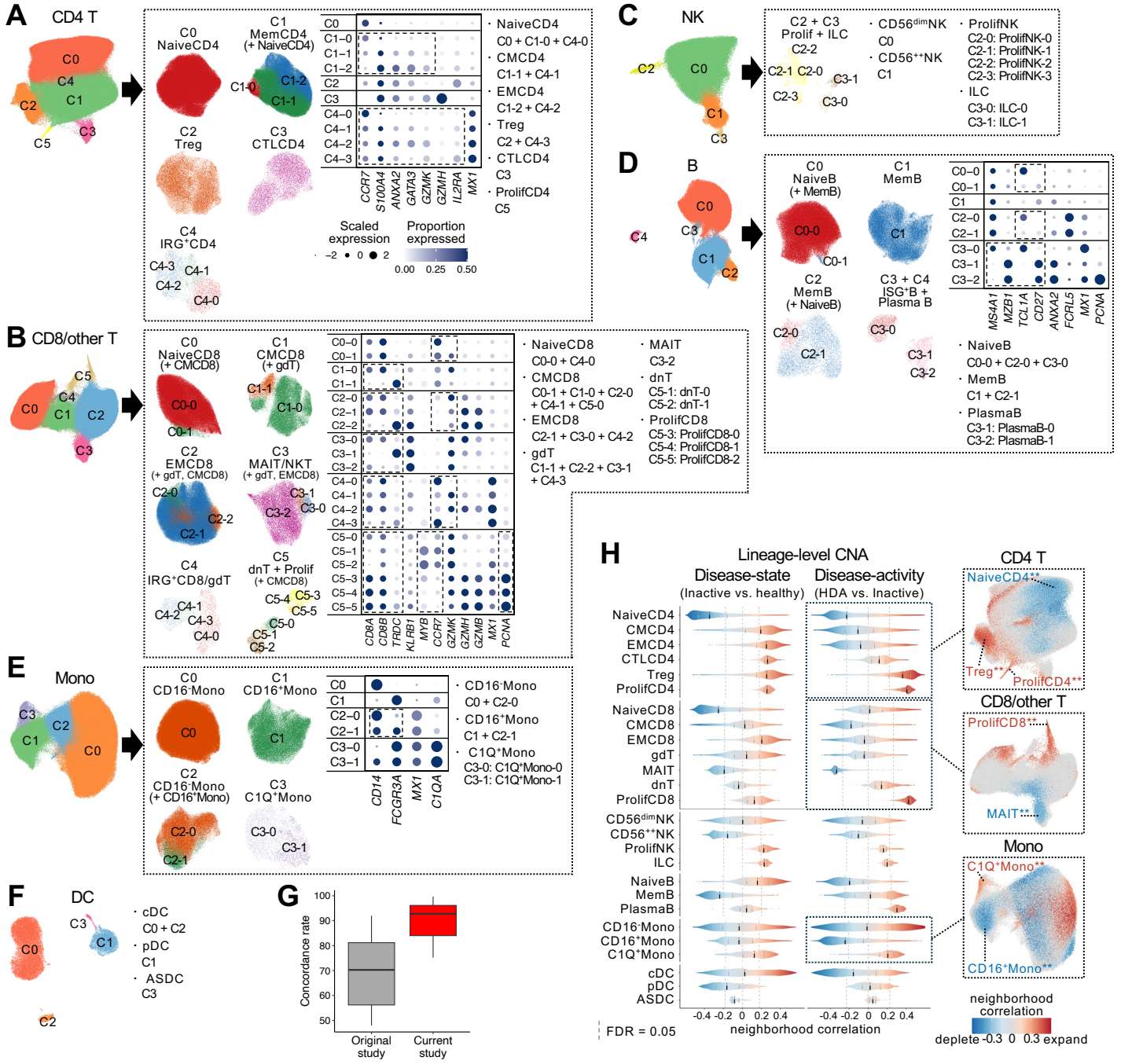

Fig. S2. Definition of 123 robust cell states using fine-grained clustering. (A-F) Flowcharts illustrating the hierarchical separation of lineage cells into cell types in **(A)** CD4$^+$ T cell (CD4 T), **(B)** CD8$^+$ and unconventional T cell (CD8/other T), **(C)** natural killer cell (NK), **(D)** B cell, **(E)** monocyte (Mono), and **(F)** dendritic cell (DC) lineages (dataset 1; N = 269). The hematopoietic stem and progenitor cell (HSPC) lineage is omitted as it contains only a single cell type. Dot plots display scaled expression levels of cell-type marker genes, with color intensity representing the proportion of cells expressing each gene. **(G)** Box plot showing the concordance rates of cell-type-level annotation in the original and the current study with the predicted cell-type labels in the reference mapping onto the independent PBMC dataset. **(H) (left)** Distributions of neighborhood correlations for each cell type in disease-state (n = 137) or disease-activity (n = 61) comparisons (dataset 1, lineage-level-CNA). **(right)** UMAP plots in representative lineages. Cell types are highlighted if > 50% of neighborhoods are associated with phenotypes. See **table S2** for cell type abbreviations.

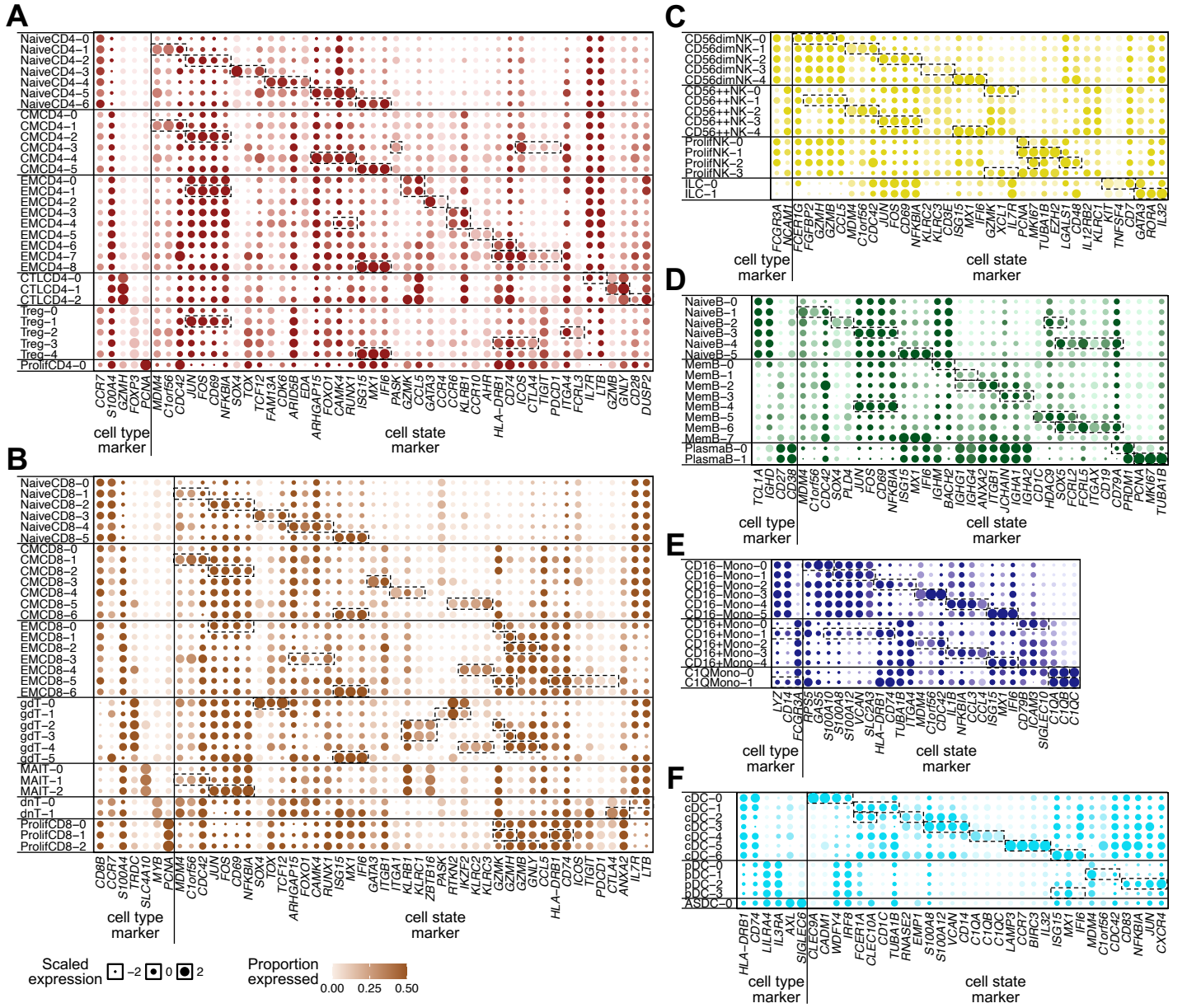

Fig. S3. Marker gene expression across 123 cell states. (A–F) Dot plots showing scaled expression levels of cell-type and cell-state marker genes for each cell state in **(A)** CD4⁺ T cell (CD4 T), **(B)** CD8⁺ and unconventional T cell (CD8/other T), **(C)** natural killer cell (NK), **(D)** B cell, **(E)** monocyte (Mono), and **(F)** dendritic cell (DC) lineages (dataset 1; N = 269). Color intensity represents the proportion of cells expressing each gene.

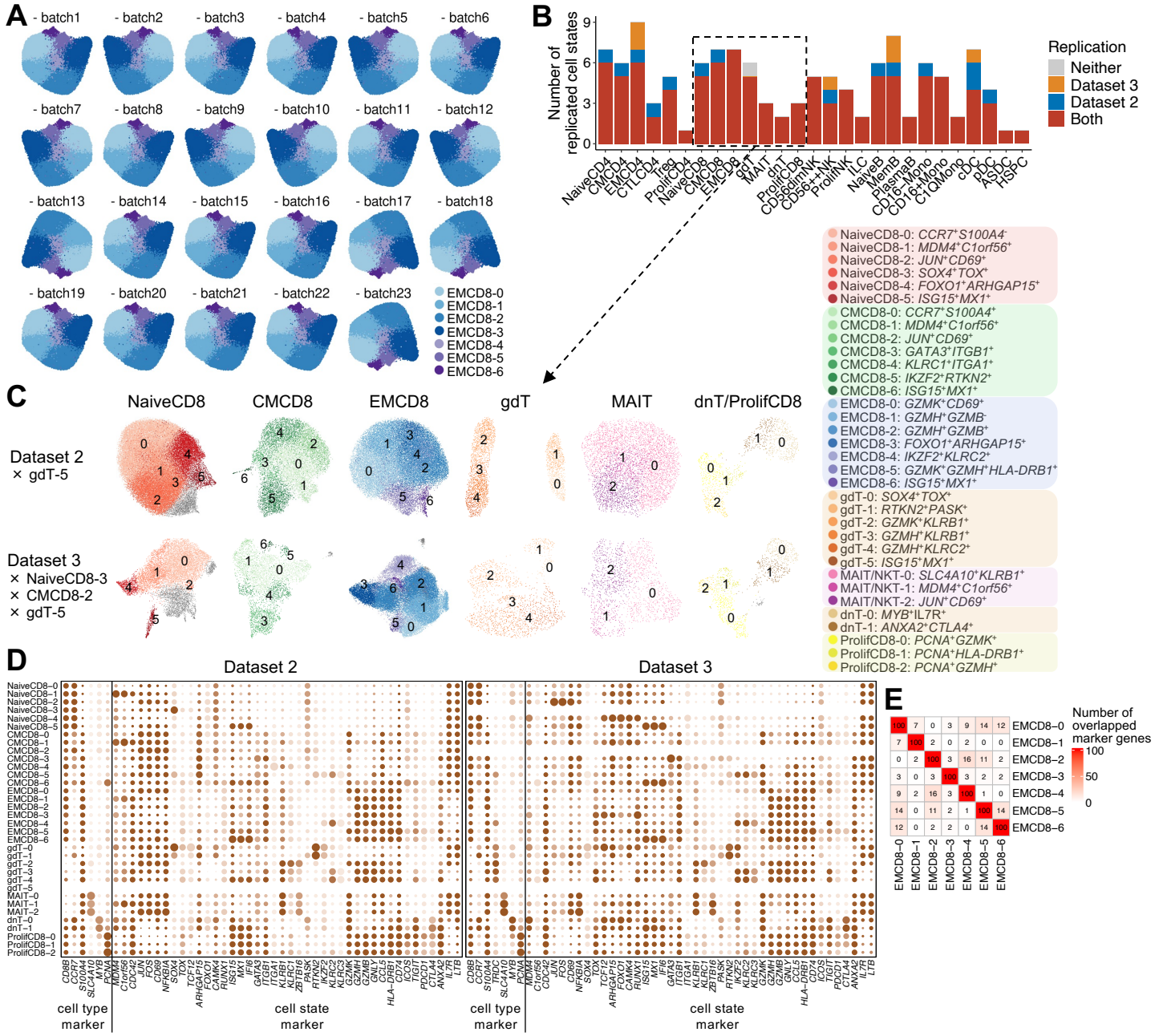

Fig. S4. Robustness and replicability of 123 cell states. (A) UMAP plots of representative cell type (EMCD8), colored by cell states for each leave-one-batch-out condition. **(B)** Bar plots showing the number of cell states replicated in independent clustering of dataset 2 and/or 3. **(C)** UMAP plots of representative cell types in the CD8/other T lineage, colored by cell states in independent clustering of dataset 2 and 3. Cell states not replicated in each condition are listed to the left, and cell-state annotations are listed to the right. **(D)** Dot plots showing scaled expression levels of cell-type and cell-state marker genes for each cell state in the CD8/other T lineage from dataset 2 and 3. Color intensity represents the proportion of cells expressing each gene. **(E)** Heatmap showing the number of overlapping genes among the top 100 upregulated marker genes for each cell state in EMCD8.

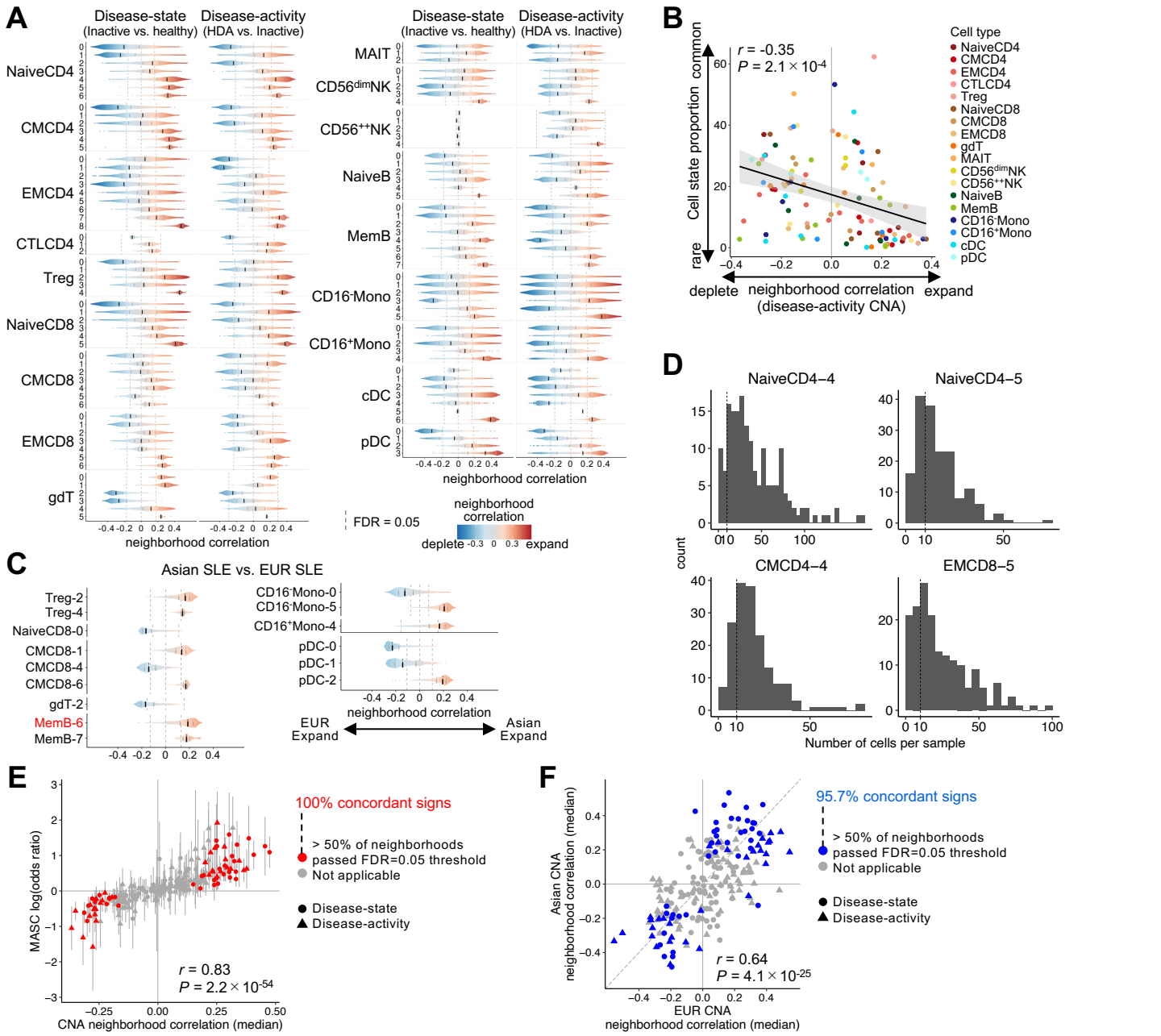

Fig. S5. Mapping cell state compositional changes in SLE. (A) Distributions of neighborhood correlations for each cell state in disease-state (n = 137) or disease-activity (n = 61) comparisons (dataset 1, cell-type-level-CNA). **(B)** Scatter plot of median neighborhood correlation (disease-activity CNA) for each cell state vs. cell-state proportion within cell type. Shaded regions represent 95% confidence intervals. **(C)** Distributions of neighborhood correlations in Asian vs. EUR SLE comparisons. 15 cell states with > 50% of neighborhoods associated with phenotype (FDR < 0.05) are shown. **(D)** Histograms showing the number of cells per SLE sample for representative cell states expanded in SLE. **(E)** Scatter plot of median neighborhood correlation in CNA vs. effect size in MASC for each cell state in disease-state and activity comparisons. **(F)** Scatter plot of median neighborhood correlation of cell-type-level CNA in European vs. in Asian. In **(B)**, **(E)** and **(F)**, Pearson's r and P values are indicated.

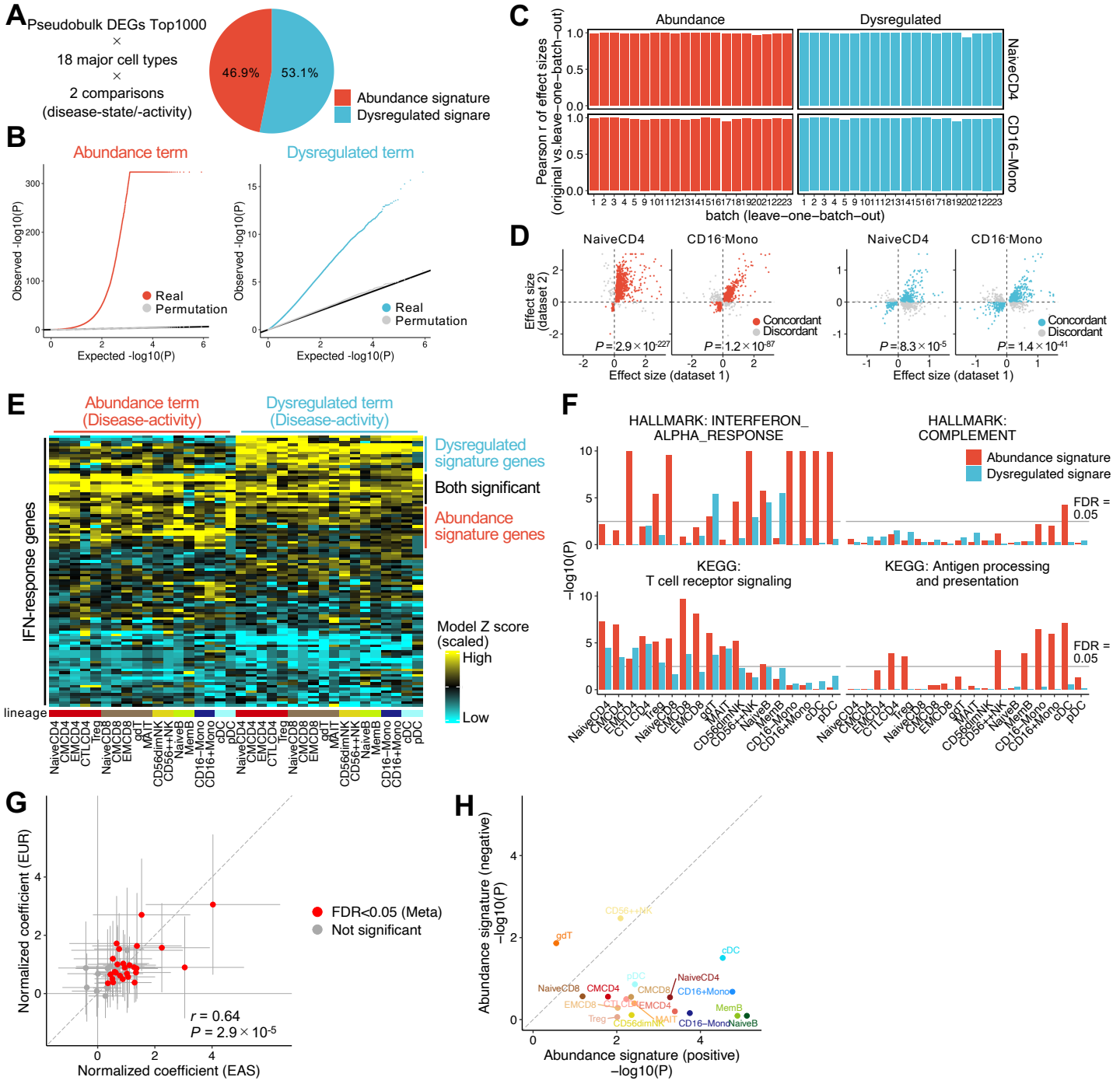

Fig. S6. Statistical framework to dissect quantitative and qualitative changes within cell types. (A) Relative proportion of variance explained by abundance and dysregulated terms for the top 1,000 pseudobulk DEGs across all cell types and comparisons. **(B)** Quantile-quantile plots of observed vs. expected model *P* values for abundance and dysregulated terms across all cell types and comparisons. **(C)** Bar plot showing Pearson's *r* of model effect sizes between the original vs. each leave-one-batch-out analysis for representative cell types. **(D)** Comparison of model effect sizes across genes between dataset 1 and 2. *P*, *P* values from one-sided binomial tests. **(E)** Heatmap of abundance and dysregulated term Z scores across 18 major cell types for 100 IFN-response genes (disease-activity comparison; $n = 61$). Gene order is based on hierarchical clustering of Z scores. **(F)** Bar plots showing the enrichment of representative pathways for each signature across cell types (disease-activity comparison). *P*, *P* values from one-sided Fisher's exact tests. **(G)** Comparison of normalized coefficients in stratified LD score regression (S-LDSC) between EAS vs. EUR SLE-GWAS. Each dot represents a signature for a given cell type. Error bars represent 2 × standard errors. Pearson's *r* and *P* values are indicated. **(H)** Comparison of SLE risk variant enrichment around abundance signature genes with positive vs. negative model statistics for each cell type.

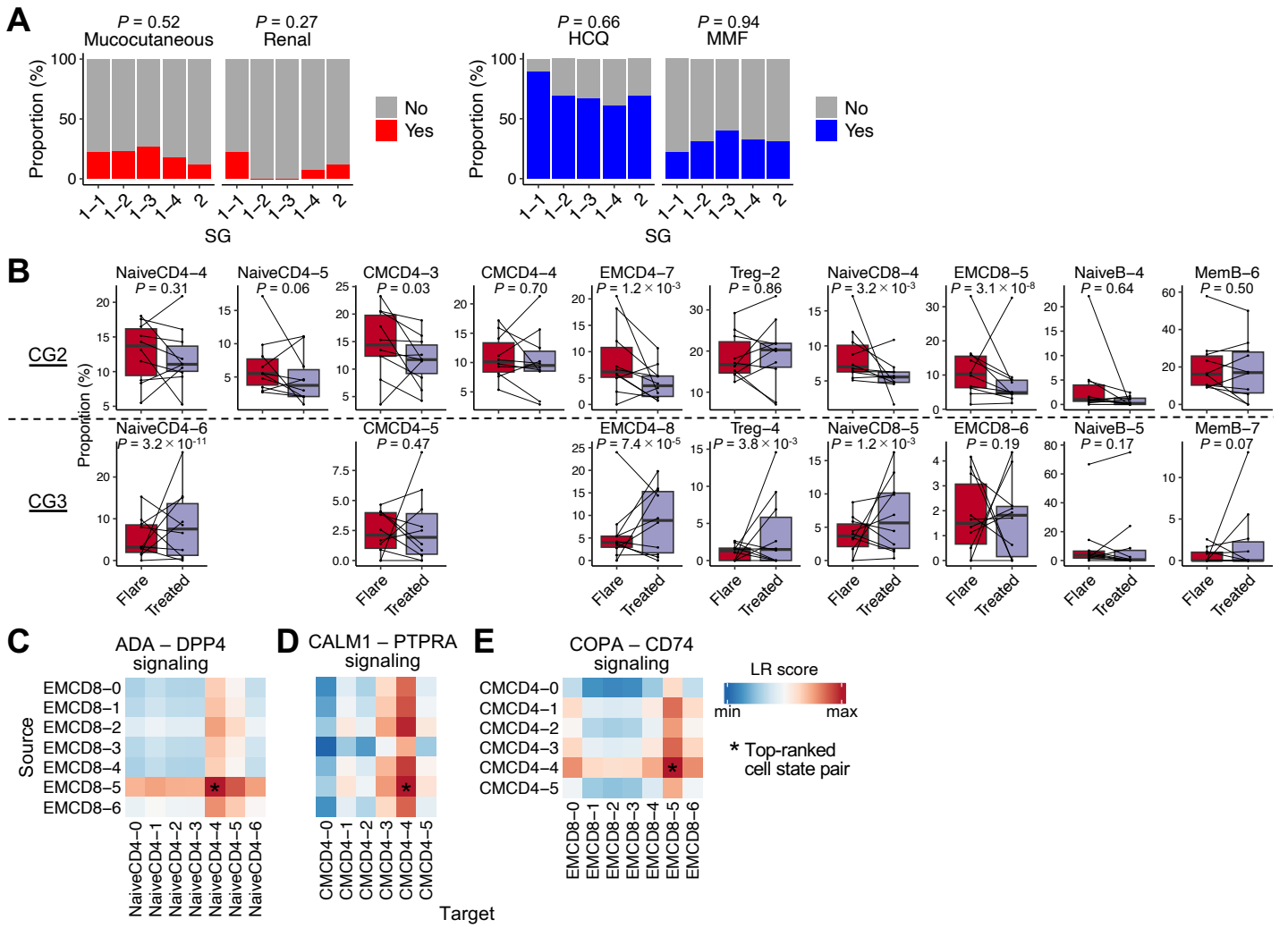

Fig. S7. Characterization of aberrant immune networks between disease-relevant cell states. (A) Bar plot showing the frequencies of SLE samples with **(left)** mucocutaneous or renal activity and **(right)** hydroxychloroquine (HCQ) or mycofenolate mofetil (MMF) treatment for each sample group (SG). *P*, *P* values from two-sided Fisher's exact tests. **(B)** Box plots comparing cell-state proportions between "Flare" and "Treated" status in **(top)** CG2 and **(bottom)** CG3 cell states. Samples from the same donor are connected by lines. *P*, *P* values from MASC ($n = 10 \times 2$). **(C-E)** Heatmaps of LR scores for **(C)** ADA–DPP4 signaling from EMCD8 to NaiveCD4 cell states, **(D)** CALM1–PTPRA signaling from EMCD8 to CMCD4 cell states, and **(E)** COPA–CD74 signaling from CMCD4 to EMCD8 cell states.

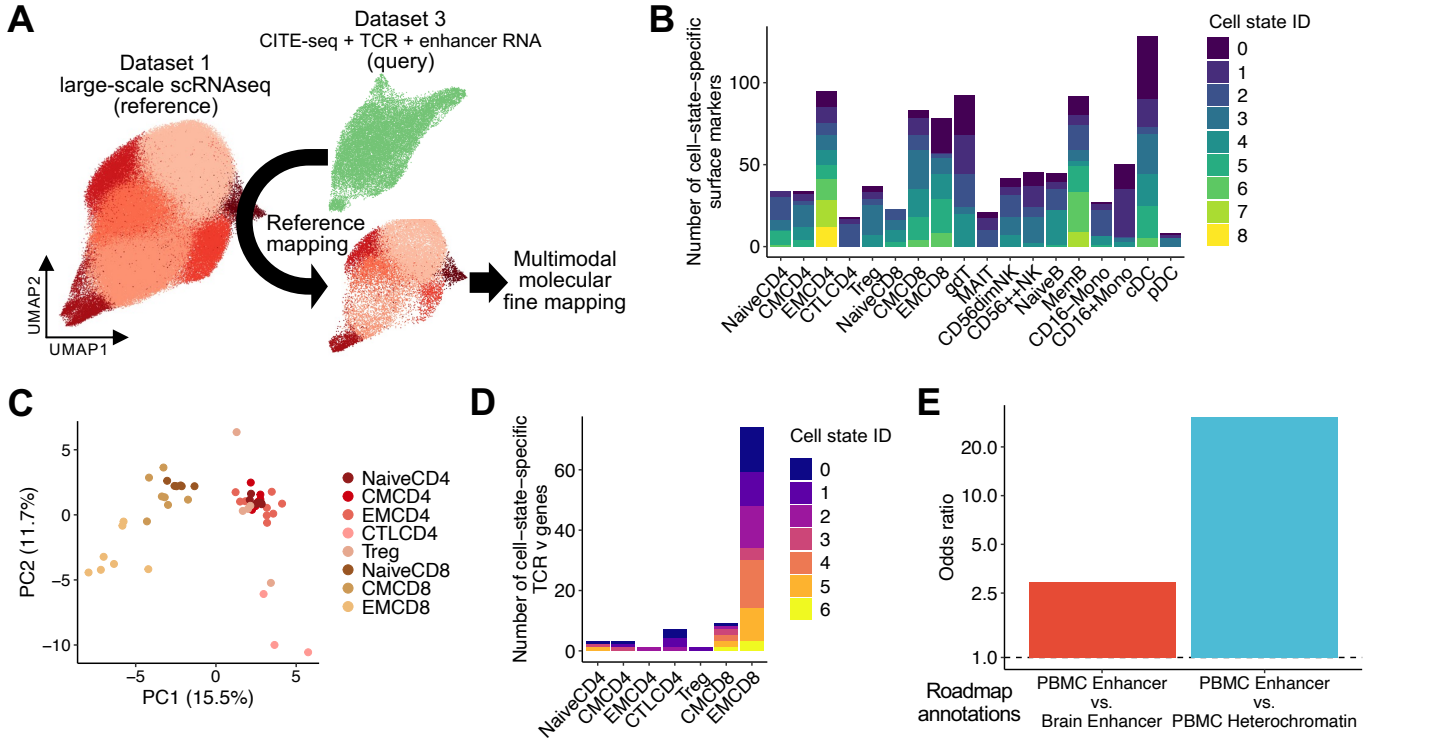

Fig. S8. Multimodal molecular profiling across fine-grained cell states. (A) Schematic of the integration of multimodal dataset 3 (query; *N* = 32) into large scRNA-seq dataset 1 (reference; *N* = 269) through reference mapping. **(B)** Bar plot showing the number of cell-state-specific upregulated surface marker proteins. Colors represent cell-state IDs (e.g., NaiveCD4-0). **(C)** PCA plot of TCR V gene usage frequencies for each cell state, colored by cell type. **(D)** Bar plot showing the number of cell-state-specific (overrepresented) TCR V genes. **(E)** Bar plot showing the enrichment of called transcribed enhancers across all cell types for "PBMC enhancer" compared with **(left)** "Brain enhancer" and **(right)** "PBMC heterochromatin" in Roadmap Epigenomics project annotations.

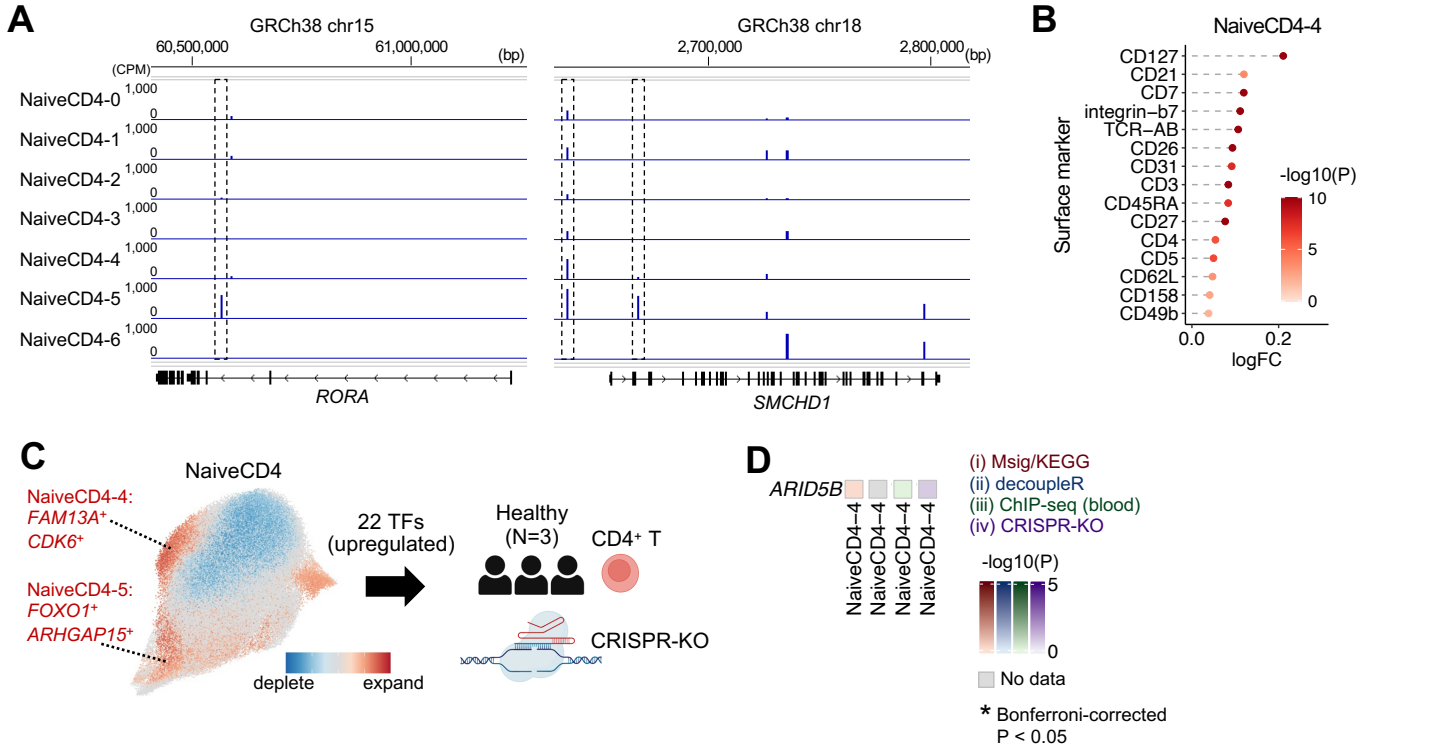

Fig. S9. *FOXO1* is a candidate driver of the pathogenic CD4+ T cell state gene program. (A) Representative examples of cell-state-specific transcribed enhancers near upregulated marker genes in NaiveCD4-5. **(B)** Dot plot showing the logFC of top 15 upregulated surface markers for NaiveCD4-4. *P*, *P* values from Seurat FindMarkers. **(C)** Experiment design of the arrayed CRISPR-KO experiment in primary CD4⁺ T cells. **(D)** Heatmap showing -$\log_{10}$(enrichment *P*) values of *ARID5B* for NaiveCD4-4 signature genes based on (i) pathway/TF enrichment analyses, TF activity inference using (ii) decoupleR and (iii) blood-derived ChIP-seq data, and (iv) a CRISPR-KO experiment. *P*, *P* values from one-sided Fisher's exact tests.

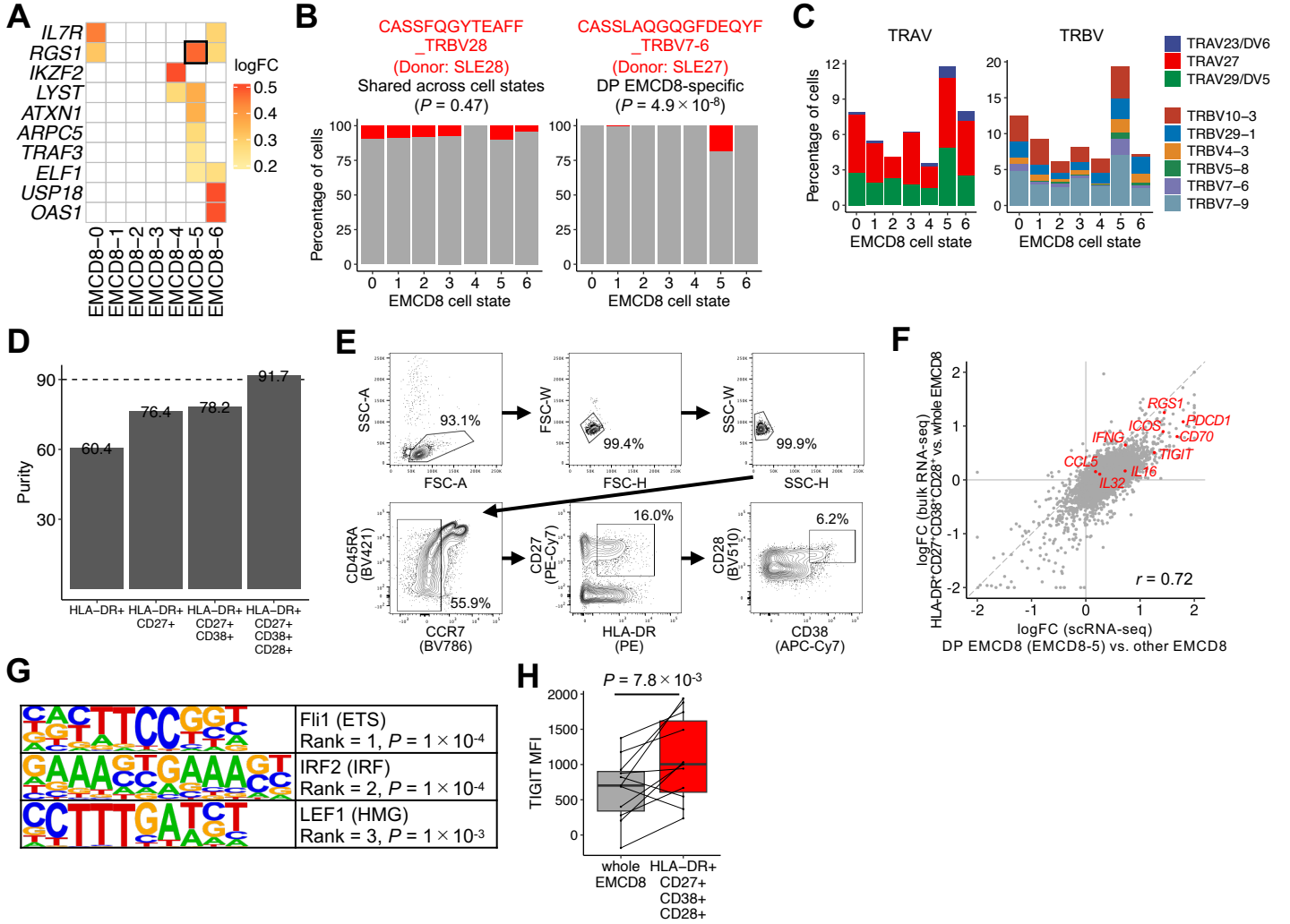

Fig. S10. DP EMCD8 cells as a candidate pathogenic player in SLE. (A) Heatmap showing the logFC of GWAS-prioritized genes across EMCD8 cell states. Up to five genes with the highest logFC per cell state are shown. **(B)** Bar plots showing the percentage of **(left)** a representative clonotype shared across EMCD8 cell states and **(right)** a clonotype specific to EMCD8-5. *P*, *P* values from two-sided Fisher's exact tests. **(C)** Bar plots showing the percentage of **(left)** three TRAV and **(right)** six TRBV gene usages positively skewed in EMCD8-5. **(D)** Bar plot showing the percentage of EMCD8-5 cells within each in-silico gate using dataset 3. **(E)** Gating strategy for defining HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ EMCD8 cells in flowcytometry (FCM) and FACS. **(F)** Scatter plot comparing the scRNA-seq dataset 1 logFC (EMCD8-5 vs. other EMCD8 cells) and bulk RNA-seq logFC (HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ vs. whole EMCD8 cells). **(G)** TF motif logos enriched for differentially accessible peaks in HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ cells compared with whole EMCD8 cells. *P*, enrichment *P* values from HOMER. **(H) (left)** Boxplots comparing the mean fluorescence intensities (MFI) of TIGIT between HLA-DR$^+$CD27$^+$CD38$^+$CD28$^+$ and whole EMCD8 cells. Samples from the same donor are connected by lines. *P*, *P* values from linear mixed models.

530 **Methods**

531 **Subjects**

532 In this study, we utilized two publicly available 3'-end scRNA-seq datasets (datasets 1 and 2). For
533 dataset 1, FASTQ files were downloaded from CELLxGENE (*5*) and raw genotype data were obtained
534 from dbGaP (phs002812.v1.p1) with authorized access. Samples were recruited from the California
535 Lupus Epidemiology Study (CLUES) cohort and the ImmVar Consortium. For dataset 2, count data were
536 downloaded from dbGaP (phs002048.v1.p1) (*4*). For our newly generated 5'-end multimodal dataset
537 (CITE-seq, TCR-seq, and enhancer RNA profiling; dataset 3) and additional FCM/FACS experiments,
538 patients were recruited at Okayama University Hospital and Hospital of the University of the
539 Occupational and Environmental Health under approvals by the Ethics Committees of RIKEN Center for
540 Integrative Medical Sciences (RIKEN−Y-2025-006, 008). Written informed consent was obtained from all
541 participants.

542 Across all datasets, SLE disease activity was categorized into four levels based on the SLE
543 Disease Activity Index (SLEDAI) (*77*): i) inactive (SLEDAI = 0), ii) low (LDA; 1 ≤ SLEDAI ≤ 4), iii)
544 moderate (MDA; 5 ≤ SLEDAI ≤ 9), and iv) high disease activity (HDA; SLEDAI ≥ 10) (**table S1**). In
545 dataset 1, "flare" status was categorized into HDA.

546

547 **Sample processing and quality control**

548 For dataset 3, single-cell suspensions of PBMCs from 32 SLE donors were processed using the 10x
549 Genomics Chromium system. Gene expression (GEX), surface protein (ADT), and TCR (VDJ) libraries
550 were prepared using the Chromium Next GEM Single Cell 5' Kit v2 and TotalSeq™-C Human Universal
551 Cocktail v1.0 (Biolegend, 399905). Eight libraries (16 samples pooled per library; four libraries per batch)
552 were sequenced on an Illumina NovaSeq 6000 (150-bp paired-end).

553 FASTQ files for datasets 1 and 3 were processed using Cell Ranger v7.0.1. We excluded low-
554 quality cells based on their distributions: (i) total unique molecular identifiers (UMIs) (< 500), (ii) number
555 of genes (nGene), (iii) percentage of mitochondrial genes (% MT), or (iv) percentage of ribosomal protein
556 genes (% Ribo) (dataset 1: nGene < 300, %MT > 7.5%, %Ribo < 8.5%; dataset 2: nGene < 300, %MT >
557 8.5%, %Ribo < 8.5%; dataset 3: nGene < 500, %MT > 8.5%, %Ribo < 6%; **fig. S1A**). Cells with high
558 percentage of hemoglobin genes (> 0.9%; *HBB*, *HBA1*, *HBA2*, *HBD*) or platelet (> 0.6%; *PF4*, *PPBP*,
559 *CAVIN2*, *GNG11*) signatures were also removed.

560 For dataset 1, samples were demultiplexed using Demuxalot (*78*) based on imputed genotypes
561 (SHAPEIT2, Minimac3 r^2 > 0.95) or Souporcell (*79*) for samples lacking reference genotypes. For dataset
562 3, Souporcell was used with 1000 Genomes Project common variants (minor allele frequency > 1%),
563 followed by multiplex PCR and MiSeq sequencing of 20 target variants to link genotype clusters to
564 specific individuals. Putative doublets from different genotypes were excluded (**fig. S1B**). In dataset 1,

565 we excluded two samples with < 200 cells recovered and three samples showing sex discordance
566 between metadata and pseudobulk expressions of *XIST* and Y chromosome genes (**fig. S1C**).
567 PBMCs were partitioned into seven lineages (CD4 T, CD8/other T, NK, B, Mono, DC, and
568 HSPC) through iterative Louvain clustering (**fig. S1D−F**). Briefly, gene expression was normalized via
569 SCTransform (Seurat v5.1.0) (*80*) and integrated using 3,000 highly variable genes (HVGs), followed by
570 batch correction via Harmony (group.by.vars = "library", "sample") (*81*). Mitochondrial, Y chromosome,
571 hemoglobin, platelet, and cell-cycle genes were excluded from HVGs. We used 30 Harmony-corrected
572 PCs for UMAP visualization and shared nearest neighbor (SNN) graph construction. After initial
573 separation into four major populations (T/NK, B, Myeloid, and HSPC), we performed recursive sub-
574 clustering to maximize lineage purity based on marker gene expression (**fig. S1D−F**). Hyperparameters
575 for each step are summarized in **table S3**. Sub-clusters representing remaining doublets with multiple
576 lineage markers (**fig. S1F**) and higher Scrublet doublet scores (*82*) (**fig. S1G**) were excluded. After multi-
577 step QC, we retained 1,456,041 (dataset 1), 344,235 (dataset 2), and 268,783 (dataset 3) high-quality
578 PBMCs (**fig. S1H**).
579

580 **Definition of 123 fine-grained cell states**

581 To identify granular cell states, we employed a three-layered iterative clustering strategy in dataset 1:
582 lineage (layer 1), cell type (layer 2), and cell state (layer 3). Following initial lineage separation, we

583 performed SCTransform normalization, HVG selection, PCA, Harmony batch correction, and Louvain
584 clustering for each lineage using optimized parameters (**fig. S2A−F** and **table S3**). Sub-clustering was
585 performed iteratively to ensure cell-type purity based on canonical marker expression. In a few instances
586 (nine clusters), biologically redundant clusters were merged to define a final set of 123 cell states. To
587 validate annotation accuracy, we performed reference mapping of dataset 1 onto an independent PBMC
588 reference (Azimuth) (*19*) and compared concordance rates between our fine-grained annotations and the
589 original labels (**fig. S2G**).

590 Cell-state-specific marker genes were identified using Seurat's FindMarkers (v4.3.0) (min.pct =
591 0.05), comparing each state against all others within the same cell type (**fig. S3** and **table S4**).
592 Upregulated markers (FDR < 0.05 and logFC > 0.1) were used for downstream pathway/TF enrichment
593 analyses. To provide more intuitive differential expression values, we also reported logFC calculated with
594 pseudocount.use = 0 in addition to the default Seurat values (**table S4**).

595 The robustness of the 123 cell states was assessed via leave-one-batch-out analysis, where
596 each of the 23 experimental batches was systematically excluded, and the entire pipeline—from
597 normalization to clustering—was re-executed. We titrated the resolution parameter (in increments of
598 0.05) to confirm that the original clusters were consistently replicated (**fig. S4A** and **table S5**).
599 Furthermore, independent replication analyses were performed using dataset 2 and 3, where the entire

pipeline was re-run from scratch for each dataset (**fig. S4B−C**). We confirmed that the resulting clusters exhibited marker gene profiles comparable to those in dataset 1 (**fig. S4D** and **table S5**).

Co-varying neighborhood analysis

To identify single-cell-resolution transcriptional neighborhoods associated with clinical phenotypes, we applied co-varying neighborhood analysis (CNA v0.1.6) (*15*), to the 18 major cell types (> 5,000 cells each; **Fig. 3**) and six cell lineages (**fig. S2H**) in dataset 1. Following our established framework (*6*), we defined two comparison axes: (i) disease-state (97 healthy controls vs. 40 inactive SLE) and (ii) disease-activity (40 inactive vs. 21 high-activity [HDA] SLE).

Harmony-corrected PCs were used as inputs to infer the neighborhood abundance matrix (NAM), followed by PCA to obtain NAM-PCs. We tested the associations between phenotype labels and NAM-PCs using linear models, adjusting for age, sex, ancestry, and the number of cells per cell type. Given the substantial correlation between disease activity and age (r = -0.60) in dataset 1, age was excluded from covariates in disease-activity comparisons. For cell-type-level CNAs, significance was defined as a global permutation $P < 0.05/36$ (Bonferroni correction). CNA-derived neighborhood correlations (positive for expansion, negative for depletion) were aggregated for each cell state (**Fig. 3** and **table S8**). States were categorized as showing "remarkable expansion/depletion" if > 50% of constituent neighborhoods were associated with the phenotype at an empirical FDR < 0.05, and

618 "moderate" if > 25%. We additionally performed CNA for (1) inactive vs. low-activity (LDA), (2) inactive
619 vs. moderate-activity (MDA), and (3) Asian vs. European SLE (**Fig. 3E** and **fig. S5C**). MASC (*11*) was
620 also applied to test cell-state abundance associations, adjusting for age, sex, and ancestry (**fig. S5E**).

621 For replication, CNA was applied to dataset 2 (**Fig. 3F**). Cells were projected into the dataset 1
622 latent space using Symphony (*53*), and Symphony-corrected PCs served as inputs. To ensure stable
623 CNA performance (requiring ~30 samples) (*15*), we compared 14 SLE (SLEDAI <= 2) vs. 17 healthy
624 donors for disease-state, and 15 SLE (SLEDAI >= 6) vs. 14 SLE (SLEDAI <= 2) for disease-activity
625 comparison.

626

627 **Statistical framework to dissect quantitative and qualitative changes**

628 To decompose cell-type-level gene expression into quantitative and qualitative changes, we introduced a
629 single-cell-level generalized linear mixed effect model (GLMM, **Fig. 4B** and **table S9**). For each gene,
630 the model was fitted using a negative binomial distribution:

631 $$\log(E_i) = \beta_{abund}X_{i,abund} + \beta_{dysreg}X_{d,dysreg} + \beta_{age}X_{d,age} + \beta_{sex}X_{d,sex}$$

632 $$+\beta_{ancestry}X_{d,ancestry} + \beta_{nUMI}X_{i,nUMI} + \beta_{MT}X_{i,MT} + (\phi_d|d) + (\kappa_b|b) + \theta$$

633 where *E* is the expression of the gene in cell *i*, *θ* is an intercept, and *β* coefficients represent fixed effects
634 for cell *i*, donor *d*, or batch *b*. Donor and batch were modeled as random intercepts. The abundance
635 signature term ($X_{i,abund}$) represents neighborhood correlation values from cell-type-level CNA, while the

636 dysregulated signature term ($X_{d,dysreg}$) represents phenotype information (e.g., healthy vs. inactive SLE).
637 The model was applied to genes passing low-expression filters (>=3 counts in >= 10% of samples)
638 across 18 major cell types (> 5,000 cells) for disease-state and activity comparisons in dataset 1, using
639 lme4 (*83*). Significance was determined via likelihood ratio tests (LRT) comparing models with and
640 without the respective terms. ImmVar samples were excluded from disease-state comparisons. To
641 quantify the contribution of each term to pseudobulk expression, we calculated Nagelkerke R^2 values for
642 the top 1,000 DEGs (identified via edgeR (*84*)) per cell type (**fig. S6A**).
643 Type I error calibration was assessed via permutation tests (randomly shuffling neighborhood
644 correlations or phenotype labels; **fig. S6B**). Robustness was further validated through leave-one-batch-
645 out analysis (**fig. S6C**) and independent replication in dataset 2, with effect-size replicability tested via
646 one-sided binomial sign tests (**fig. S6D**). For 100 IFN response genes (IRGs) (*4*) (**table S11**), abundance
647 and dysregulated Z-scores were used for hierarchical clustering (Ward's method, Euclidean distance) to
648 visualize patterns across cell types (**Fig. 4C−D** and **fig. S6E**).
649

650 **Pathway enrichment analysis and gene signature scores**

651 To evaluate the enrichment of pathway annotations for abundance and dysregulated signature genes,
652 we performed over-representation analysis (ORA) using the top 1,000 genes (by LRT P-values) for each
653 signature term in each cell type. ORA was conducted using a one-sided Fisher's exact test via

654 clusterProfiler (*85*) (**Fig. 4E**, **fig. S6F**, and **table S12**), utilizing MSigDB Hallmark and KEGG pathway
655 collections. To capture cell-type-specific biology, the union of the top 1,000 LRT P-value genes across all
656 cell types was used as the background gene set.

657 For pathway analysis of cell-state-specific upregulated marker genes (FDR < 0.05, logFC > 0.1,
658 Seurat FindMarkers), ORA was performed similarly (**Fig. 6A**, **Fig. 7A**, and **table S7**). The background for
659 each cell type consisted of all genes passing the low-expression filter (>=3 counts in >= 10% of
660 samples). To calculate classical exhaustion signature scores in EMCD8 and EMCD4 populations, an
661 exhaustion-related gene set was curated from previous studies (*3*, *5*) (**table S20**). Gene expression was
662 log-normalized and scaled, and the signature score for each cell was defined as the sum of the scaled
663 expression values across the gene set (**Fig. 7B**).

664

665 **Stratified LD score regression analysis**

666 To evaluate the enrichment of SLE heritability (all risk variants irrespective of effect size) around
667 abundance and dysregulated signature genes, we performed stratified linkage disequilibrium score
668 regression (S-LDSC) (*40*) using the sc-linker pipeline (*86*) (**Fig. 4F** and **fig. S6H**). We first calculated
669 min-max-scaled gene scores based on the Z-scores of abundance and dysregulated terms for all tested
670 genes in each cell type. LD scores were then estimated for East Asian (EAS) and European (EUR)
671 populations using 1000 Genomes Project reference panels, with the scaled gene scores serving as

672 annotation weights. In S-LDSC, we examined the enrichment of SLE heritability for common variants
673 within 100-kb windows flanking the transcription start sites (TSS) of genes with positive Z-scores for both
674 signatures, adjusting for the baseline model annotations v1.2 (*40*). Two large-scale SLE-GWAS
675 summary statistics from EAS and EUR ancestries (*44*, *45*) were utilized. To account for differences in
676 GWAS heritability, regression coefficients were normalized by the mean per-SNP heritability as
677 previously described (*40*). Finally, a fixed-effect meta-analysis of the EAS and EUR results was
678 performed using the inverse-variance weighting method based on the normalized coefficients and their
679 standard errors (**fig. S6G**). We reported one-sided *P* values to test whether the regression coefficients
680 were significantly positive.
681

682 **Localization of GWAS prioritized genes**

683 To identify candidate causal genes for SLE-associated genetic variants ($P < 5 \times 10^{-8}$), we used the L2G
684 (locus-to-gene) scores from the Open Targets Platform (*43*). We defined 124 genes with an L2G score >
685 0.5 as SLE-GWAS prioritized genes (**table S10**). For each cell state, we identified the overlap between
686 these 124 prioritized genes and the cell-state-specific upregulated markers (FDR < 0.05, logFC > 0.2;
687 Seurat FindMarkers) (**Fig. 6E** and **fig. S10A**).
688

689 **Sample group (SG) and cell-state group (CG) analysis**

690 To investigate the clinical relevance of fine-grained cell states, we performed bi-hierarchical clustering
691 (Ward's method, Euclidean distance) of all samples in dataset 1 and cell states within 18 major cell types
692 based on their proportions (**Fig. 5A**). Sample groups (SGs) and cell-state groups (CGs) were defined
693 based on the resulting dendrograms. The frequencies of disease-activity categories, organ-specific
694 activity, and treatment agents (**table S1**) were compared across SGs (**Fig. 5B** and **fig. S7A**). For
695 longitudinal analysis in dataset 1, changes in cell-state proportions between paired "Flare" and "Treated"
696 samples (*n* = 10 donors) were tested using MASC (*11*) (**Fig. 5C** and **fig. S7B**).
697

698 **Cell-cell interaction analysis**

699 To infer the inflammatory networks across all cell states, we performed cell-cell interaction analysis using
700 LIANA (46) with patemeters method = c("logfc", "sca") and min_cells = 10. From the consensus set of
701 4,701 ligand–receptor (LR) interactions across all 123 × 123 cell-state pairs, we identified 1,451,770
702 potentially active signals (defined by sca.LRscore > 0.6, logfc.logfc_comb > -0.1, and expression
703 proportions of >=1% for ligands and >=5% for receptors). To minimize redundancy, for each unique LR
704 interaction within each of the 27 × 27 cell-type pairs, we prioritized the cell-state pair with the highest LR
705 score as the representative cell-state-specific interaction (118,629 signals; **Fig. 4D−G**, **Fig. 6B**, **Fig. 7D**,
706 and **fig. S7C−E**).
707

708 **Reference mapping and integration in LuPIN**

709 To integrate the three single-cell datasets into the LuPIN framework, we performed reference mapping
710 using Symphony (*53*), designating the large-scale dataset 1 as the reference and datasets 2 and 3 as
711 queries (**fig. S8A**). For dataset 1, lineage- and cell-type-level reference objects were constructed based
712 on the Harmony-integrated embeddings used in the fine-grained clustering. For the query datasets (2
713 and 3), PBMCs were first partitioned into seven lineages using the same pipeline as for dataset 1.
714 Following normalization and scaling, cells from each lineage were projected into the corresponding
715 lineage-level latent transcriptional space of dataset 1 (lineage-level reference mapping). Cell-type labels
716 were then assigned to query cells using k-nearest neighbor classifiers (k = 5) based on the dataset 1
717 reference. This procedure was repeated at the cell-type level (cell-type-level reference mapping) to
718 annotate each cell in datasets 2 and 3 with the best-predicted cell-state labels.

719

720 **Surface protein analysis**

721 Following reference mapping of the multimodal dataset 3, we investigated cell-state-specific surface
722 protein profiles. For dataset 3, 130 surface proteins and seven isotype controls were profiled using the
723 TotalSeq™-C Human Universal Cocktail v1.0 (Biolegend, 399905). Protein expression for each cell was
724 normalized using a centered-log ratio (CLR) transformation across all 130 proteins. Cell-state-specific
725 surface markers were identified using Seurat FindMarkers, comparing each cell state against all others

726 within the same cell type. Proteins with FDR <0.05 and logFC >0.2 were defined as cell-state-specific
727 upregulated surface markers (**fig. S8B** and **table S14**).
728

729 **TCR repertoire analysis**

730 In dataset 3, we utilized TCR UMI counts for cells meeting the following criteria: (i) passed GEX-based
731 QC; (ii) annotated as NaiveCD4, CMCD4, EMCD4, CTLCD4, Treg, NaiveCD8, CMCD8, or EMCD8 via
732 reference mapping; and (iii) passed Cell Ranger TCR quality filters. To identify cell-state-specific V gene
733 usage, frequencies were compared for each state against all others within the same cell type using two-
734 sided Fisher's exact tests (**Fig. 6C**, **fig. S8D**, **S10C**, and **table S15**). Among all 45 TRAV and 49 TRBV
735 genes, we tested V genes passing a low expression filter (>= 10 total counts per cell type).

736 TCR clonal expansion in EMCD8-5 (DP EMCD8) was assessed by quantifying clonotypes
737 across SLE donors (**Fig. 7C**). Clonotypes were defined based on the combination of the CDR3 amino
738 acid sequence and the V gene of the β-chain. To identify DP EMCD8-specific clonotypes, frequencies
739 were compared between states within each donor using two-sided Fisher's exact tests (**fig. S10B** and
740 **table S21**). Known cognate epitopes for each clonotype were identified using the VDJdb browser (**table**
741 **S21**).
742

743 **Transcribed enhancer analysis**

To detect transcribed enhancers in dataset 3, we employed the ReapTEC pipeline (*18*). 10x Genomics libraries were sequenced on an Illumina NovaSeq 6000 with 150-bp paired-end reads for both Read 1 and Read 2, ensuring the capture of transcription start sites (TSS) within the 5' regions. Reads were mapped using STARsolo to identify those containing unencoded G. After deduplication, reads were assigned to specific cell types based on barcodes. TSS peaks were generated by merging TSSs within 10 bp, and bidirectional transcribed candidate enhancers were identified using the developers' scripts, with promoter regions masked. To validate these enhancers, we tested their enrichment across Roadmap Epigenomics core 15-state reference annotations (*57*) for PBMC-derived enhancers (6_EnhG and 7_Enh in E062), compared with (i) PBMC heterochromatin (9_Het in E062) and (ii) brain hippocampus middle enhancers (6_EnhG and 7_Enh in E071) using two-sided Fisher's exact tests (**fig. S8E**).

Cell-state-specific transcribed enhancers were identified using an entropy-based specificity score (*18*). We first filtered for enhancers with >=5 reads per cell type, aggregated reads across donors, and calculated entropy and specificity scores for each enhancer within each cell type. Enhancers with a specificity score > 0.7 were defined as cell-state-specific (**table S16**).

Trajectory analysis

761 To infer the developmental trajectory across the CD4$^+$ T cell lineage, we employed Monocle3 (*64*) with
762 default parameters (**Fig. 6G**). For pseudotime estimation, the root was defined as a cell within the
763 SOX4$^+$TOX$^+$ NaiveCD4-3 state. This selection was based on recent evidence that these cells represent
764 recent thymic emigrants (RTEs), the most immature population within the peripheral naive T cell
765 compartment (*22*).
766

767 **Upstream TF activity inference**

768 To identify transcriptional regulators driving disease-relevant CD4$^+$ T cell states, we integrated three
769 complementary approaches (**Fig. 6F** and **fig. S9D**): (i) over-representation analysis (ORA) of TF targets,
770 (ii) TF activity inference via decoupleR (*62*) and (iii) target enrichment analysis using ChIP-Atlas (*63*).
771 Candidates TFs were selected from those upregulated (logFC > 0.3) in NaiveCD4-5 (n = 22) and
772 NaiveCD4-4 (n = 1, *ARID5B*). ORA was conducted using a one-sided Fisher's exact test via
773 clusterProfiler (*85*) for MSigDB C3 and KEGG collections (**table S17**), using signature genes (FDR <
774 0.05, logFC > 0.1; Seurat FindMarkers) from NaiveCD4-4, NaiveCD4-5, and CMCD4-4. For decoupleR,
775 TF activities were inferred using the ulm method based on DoRothEA regulons and FindMarkers logFC
776 values. For the ChIP-Atlas-based analysis, we utilized ChIP-seq datasets restricted to blood tissues
777 (**table S17**). Promoter regions were defined as -5kb to +2kb from the TSS. TF-target gene annotations
778 were constructed by intersecting ChIP-seq peaks with these promoter regions. Enrichment of NaiveCD4-

779 4/5, and CMCD4-4 signature genes within these TF-target sets was evaluated using one-sided Fisher's

780 exact tests.

781

782 **Arrayed CRISPR-KO experiment**

783 To validate the identified transcriptional regulators, we performed an arrayed CRISPR knock-out (KO)

784 experiment targeting 22 candidate TFs (**Upstream TF activity inference**; **fig. S9C**). For each TF, two

785 independent single guide RNAs (sgRNAs) (*87*) were synthesized (IDT; **table S18**) and complexed with

786 recombinant Cas9 (IDT, 1081058) to form ribonucleoprotein (RNP) complexes. Two control samples

787 (four sgRNAs) were prepared for each donor. Editing and frameshift efficiency for each sgRNA was

788 confirmed via Sanger sequencing (on median 79.7% and 68.7%, respectively; **table S18**).

789 Primary $CD4^+$ T cells from three healthy donors were isolated from fresh PBMCs using $CD4^+$ T

790 Cell Isolation Kit (Miltenyi, 130-096-533), expanded for 2 days using Dynabeads Human T-Activator

791 CD3/CD28 (Thermo, DB11132) in complete CTS OpTmizer T-Cell Expansion SFM (Thermo Fisher

792 Scientific, A1048501) supplemented with IL-2 (Peprotech, 200-02), IL-7 (Peprotech, 200-07), and IL-15

793 (Peprotech, 200-15). Seven days after activation, cells were electroporated using the 4D-Nucleofector

794 (Lonza, program EH-140) with 1×10^6 cells per reaction and subsequently cultured in triplicate. Five

795 days post-transfection, total RNA was isolated using MagMAX-96 Total RNA Isolation Kit (Thermo Fisher

Scientific, AM1830) and libraries were prepared using the RamDA-seq method (*88*). Sequencing was performed on an Illumina NovaSeq 6000 (150bp paired-end).

For RNA-seq analysis, adaptors were trimmed using Cutadapt and reads with < 50 bp were removed. Reads were mapped to GRCh38 using STAR (*89*) and quantified via HTSeq in unstranded mode. After excluding three low-quality samples (unique reads < 4×10^6), TMM-normalized counts (*84*) were fitted to a negative binomial GLMM using lme4 (*83*):

$$\log(E_s) = \beta_{KO} X_{s,KO} + (\phi_d | d) + \mathrm{offset}(\log[libsize]) + \theta$$

where E is gene expression in sample s, θ is an intercept, β_{KO} is the fixed effect for KO status, and Donor (d) is a random intercept. Significance was determined via LRT. Enrichment of downregulated genes (FDR < 0.05; **table S19**) within NaiveCD4-4/5 and CMCD-4 signature genes was tested using one-sided Fisher's exact tests (**Fig. 6F**).

FCM and FACS experiments on EMCD8 cells

To characterize EMCD8-5 (DP EMCD8) cells, we performed functional assays using FACSMelody (BD Biosciences) based on the four surface markers with the highest logFC values (HLA-DR, CD27, CD38, and CD28; **Fig. 7E−F** and **table S14**). Two panels were prepared: Panel 1 for surface markers (PD-1, TIGIT) and Panel 2 for intracellular IFN-γ (**table S22**).

For FACS, $CD8^+$ T cells were isolated from six SLE donors using the $CD8^+$ T Cell Isolation Kit (Miltenyi, 130-096-495), blocked with anti-FcγR antibodies (Thermo Fisher Scientific), and stained with Panel 1. Target EMCD8 ($CCR7^-HLA\text{-}DR^+CD27^+CD38^+CD28^+$) and total $CCR7^-$ EMCD8 cells (control) were sorted (5,000 cells per sample; **fig. S10E**) for bulk RNA-seq (SMART-seq, Takara Bio 634772), ATAC-seq, and UNIChro-seq. ATAC-seq tagmentation used TDE1 (Illumina 20034198) and 1% digitonin (Promega G944A). Purified tagmented DNA was amplified for 15–18 cycles using KAPA HiFi ReadyMix (Nippon Genetics KK2602) and Nextera XT Index Kit v2 and size-selected (180–1000 bp) using AMPure XP beads. UNIChro-seq libraries were prepared with custom primers to quantify chromatin accessibility (*68*) at five target *IFNG* and one reference *UBE2D2* loci (**table S24**) and sequenced on Miseq (illumina) using a custom R1 primer (5'- ACACTCTTTCCCTACACGACGCTCTTCAGGTCT).

RNA-seq data were analyzed using the same pipeline as the **Arrayed CRISPR-KO experiment**, with the following GLMM:

$$\log(E_s) = \beta_{sub}X_{s,sub} + \beta_{age}X_{d,age} + \beta_{sex}X_{d,sex} + (\phi_d|d) + \text{offset}(\log[libsize]) + \theta$$

where β_{sub} is the fixed effect for cell subsets. ATAC-seq and UNIChro-seq reads were mapped using Bowtie2 (*90*), with low-quality (q < 30), duplicated (Picard), and ENCODE-blacklisted reads removed. MACS2 (*91*) was used for peak calling. Differentially accessible regions (DARs) were identified using the same GLMM framework as RNA-seq. Motif enrichment for DARs (FDR < 0.1, logFC >0) was performed with HOMER (**fig. S10G**). For UNIChro-seq, we extracted UMI counts at target loci as previously

831 described (*68*). Accessibility at five IFNG loci relative to a UBE2D2 reference was evaluated using a
832 logistic mixed-effect model, regressing cell status (target/control EMCD8) by genomic region
833 (target/reference) with donor as a random effect. Genomic tracks were visualized using Integrative
834 Genomics Viewer (**Fig. 7G**).

835 For FCM, $CD8^+$ T cells from 11 SLE donors were stained with Panel 1 (surface) or Panel 2
836 (intracellular). For Panel 2, cells were stimulated with PMA (50 ng/ml, MP Biomedicals, 16561-29-8) and
837 ionomycin (1 µg/ml) for four hours, with GolgiStop (BD Biosciences, 554724) added after the first hour.
838 Intracellular staining was performed using the Cytofix/Cytoperm kit (BD Biosciences, 554714). Mean
839 fluorescence intensities (MFI) were compared between target and total EMCD8 cells using linear mixed
840 models (fixed effects: age, sex; random effect: donor; **Fig. 7H** and **fig. S10H**). All data were analyzed
841 using FlowJo (BD Biosciences).
842

843 **Reference**

844 1. G. C. Tsokos, M. S. Lo, P. C. Reis, K. E. Sullivan, New insights into the immunopathogenesis of systemic lupus
845 erythematosus. *Nat. Rev. Rheumatol.* **12**, 716–730 (2016).

846 2. R. Banchereau, S. Hong, B. Cantarel, N. Baldwin, J. Baisch, M. Edens, A.-M. Cepika, P. Acs, J. Turner, E. Anguiano, P.
847 Vinod, S. Khan, G. Obermoser, D. Blankenship, E. Wakeland, L. Nassi, A. Gotte, M. Punaro, Y.-J. Liu, J. Banchereau, J.
848 Rossello-Urgell, T. Wright, V. Pascual, Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus
849 Patients. *Cell* **165**, 551–565 (2016).

850 3. the Accelerating Medicines Partnership in SLE network, A. Arazi, D. A. Rao, C. C. Berthier, A. Davidson, Y. Liu, P. J.
851 Hoover, A. Chicoine, T. M. Eisenhaure, A. H. Jonsson, S. Li, D. J. Lieb, F. Zhang, K. Slowikowski, E. P. Browne, A.
852 Noma, D. Sutherby, S. Steelman, D. E. Smilek, P. Tosta, W. Apruzzese, E. Massarotti, M. Dall'Era, M. Park, D. L. Kamen,
853 R. A. Furie, F. Payan-Schober, W. F. Pendergraft, E. A. McInnis, J. P. Buyon, M. A. Petri, C. Putterman, K. C. Kalunian,
854 E. S. Woodle, J. A. Lederer, D. A. Hildeman, C. Nusbaum, S. Raychaudhuri, M. Kretzler, J. H. Anolik, M. B. Brenner, D.
855 Wofsy, N. Hacohen, B. Diamond, The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**,
856 902–914 (2019).

857 4. D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R. J. Rossi, C.-H. Chung,
858 R. P. Huynh, P. Singh, W. F. Flynn, J.-A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M. A. Collet, A. L. Lucido, K. Palucka,
859 P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J. F. Banchereau, Mapping systemic lupus
860 erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020).

861 5. R. K. Perez, M. G. Gordon, M. Subramaniam, M. C. Kim, G. C. Hartoularos, S. Targ, Y. Sun, A. Ogorodnikov, R. Bueno, A.
862 Lu, M. Thompson, N. Rappoport, A. Dahl, C. M. Lanata, M. Matloubian, L. Maliskova, S. S. Kwek, T. Li, M. Slyper, J.
863 Waldman, D. Dionne, O. Rozenblatt-Rosen, L. Fong, M. Dall'Era, B. Balliu, A. Regev, J. Yazdany, L. A. Criswell, N.
864 Zaitlen, C. J. Ye, Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus. *Science* **376**,
865 eabf1970 (2022).

866 6. M. Nakano, M. Ota, Y. Takeshima, Y. Iwasaki, H. Hatano, Y. Nagafuchi, T. Itamiya, J. Maeda, R. Yoshida, S. Yamada, A.
867 Nishiwaki, H. Takahashi, H. Takahashi, Y. Akutsu, T. Kusuda, H. Suetsugu, L. Liu, K. Kim, X. Yin, S.-Y. Bang, Y. Cui, H.-
868 S. Lee, H. Shoda, X. Zhang, S.-C. Bae, C. Terao, K. Yamamoto, T. Okamura, K. Ishigaki, K. Fujio, Distinct transcriptome
869 architectures underlying lupus establishment and exacerbation. *Cell* **185**, 3375-3389.e21 (2022).

870 7. S. V. Navarra, R. M. Guzmán, A. E. Gallacher, S. Hall, R. A. Levy, R. E. Jimenez, E. K.-M. Li, M. Thomas, H.-Y. Kim, M. G.
871 León, C. Tanasescu, E. Nasonov, J.-L. Lan, L. Pineda, Z. J. Zhong, W. Freimuth, M. A. Petri, Efficacy and safety of
872 belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *The*
873 *Lancet* **377**, 721–731 (2011).

874 8. E. F. Morand, R. Furie, Y. Tanaka, I. N. Bruce, A. D. Askanase, C. Richez, S.-C. Bae, P. Z. Brohawn, L. Pineda, A.
875 Berglind, R. Tummala, Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N. Engl. J. Med.* **382**, 211–221
876 (2020).

877 9. D. A. Rao, M. F. Gurish, J. L. Marshall, K. Slowikowski, C. Y. Fonseka, Y. Liu, L. T. Donlin, L. A. Henderson, K. Wei, F.
878 Mizoguchi, N. C. Teslovich, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, Y. C. Lee, D. J. Todd, V. P.
879 Bykerk, S. M. Goodman, A. B. Pernis, L. B. Ivashkiv, E. W. Karlson, P. A. Nigrovic, A. Filer, C. D. Buckley, J. A. Lederer,
880 S. Raychaudhuri, M. B. Brenner, Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid
881 arthritis. *Nature* **542**, 110–114 (2017).

882 10. S. A. Jenks, K. S. Cashman, E. Zumaquero, U. M. Marigorta, A. V. Patel, X. Wang, D. Tomar, M. C. Woodruff, Z. Simon, R.
883 Bugrovsky, E. L. Blalock, C. D. Scharer, C. M. Tipton, C. Wei, S. S. Lim, M. Petri, T. B. Niewold, J. H. Anolik, G. Gibson,
884 F. E.-H. Lee, J. M. Boss, F. E. Lund, I. Sanz, Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7
885 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity* **49**, 725-739.e6 (2018).

886 11. C. Y. Fonseka, D. A. Rao, N. C. Teslovich, I. Korsunsky, S. K. Hannes, K. Slowikowski, M. F. Gurish, L. T. Donlin, J. A.
887 Lederer, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, D. J. Todd, V. P. Bykerk, E. W. Karlson, J.
888 Ermann, Y. C. Lee, M. B. Brenner, S. Raychaudhuri, Mixed-effects association of single cells identifies an expanded
889 effector $CD4^+$ T cell subset in rheumatoid arthritis. *Sci. Transl. Med.* **10**, eaaq0305 (2018).

890 12. J. Li, M. Zaslavsky, Y. Su, J. Guo, M. J. Sikora, V. Van Unen, A. Christophersen, S.-H. Chiou, L. Chen, J. Li, X. Ji, J.
891 Wilhelmy, A. M. McSween, B. A. Palanski, V. V. A. Mallajosyula, N. A. Bracey, G. K. R. Dhondalay, K. Bhamidipati, J. Pai,
892 L. B. Kipp, J. E. Dunn, S. L. Hauser, J. R. Oksenberg, A. T. Satpathy, W. H. Robinson, C. L. Dekker, L. M. Steinmetz, C.
893 Khosla, P. J. Utz, L. M. Sollid, Y.-H. Chien, J. R. Heath, N. Q. Fernandez-Becker, K. C. Nadeau, N. Saligrama, M. M.
894 Davis, KIR^+ $CD8^+$ T cells suppress pathogenic T cells and are active in autoimmune diseases and COVID-19. *Science*
895 **376**, eabi9591 (2022).

896 13. F. Zhang, A. H. Jonsson, A. Nathan, N. Millard, M. Curtis, Q. Xiao, M. Gutierrez-Arcelus, W. Apruzzese, G. F. M. Watts, D.
897 Weisenfeld, S. Nayar, J. Rangel-Moreno, N. Meednu, K. E. Marks, I. Mantel, J. B. Kang, L. Rumker, J. Mears, K.
898 Slowikowski, K. Weinand, D. E. Orange, L. Geraldino-Pardilla, K. D. Deane, D. Tabechian, A. Ceponis, G. S. Firestein, M.
899 Maybury, I. Sahbudin, A. Ben-Artzi, A. M. Mandelin, A. Nerviani, M. J. Lewis, F. Rivellese, C. Pitzalis, L. B. Hughes, D.
900 Horowitz, E. DiCarlo, E. M. Gravallese, B. F. Boyce, Accelerating Medicines Partnership: RA/SLE Network, J. Albrecht, J.
901 L. Barnas, J. M. Bathon, D. L. Boyle, S. L. Bridges, D. Campbell, H. L. Carr, A. Chicoine, A. Cordle, P. Dunn, L. Forbess,
902 P. K. Gregersen, J. M. Guthridge, L. B. Ivashkiv, K. Ishigaki, J. A. James, G. Keras, I. Korsunsky, A. Lakhanpal, J. A.
903 Lederer, Z. J. Li, Y. Li, A. McDavid, M. J. McGeachy, K. Raza, Y. Reshef, C. Ritchlin, W. H. Robinson, S. Sakaue, J. A.
904 Seifert, A. Singaraju, M. H. Smith, D. Scheel-Toellner, P. J. Utz, M. H. Weisman, A. Wyse, Z. Zhu, L. W. Moreland, S. M.
905 Goodman, H. Perlman, V. M. Holers, K. P. Liao, A. Filer, V. P. Bykerk, K. Wei, D. A. Rao, L. T. Donlin, J. H. Anolik, M. B.
906 Brenner, S. Raychaudhuri, Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. *Nature* **623**,
907 616–624 (2023).

908 14. E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, J. C. Marioni, Differential abundance testing on single-cell data
909 using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).

910 15. Y. A. Reshef, L. Rumker, J. B. Kang, A. Nathan, I. Korsunsky, S. Asgari, M. B. Murray, D. B. Moody, S. Raychaudhuri, Co-
911 varying neighborhood analysis identifies cell populations associated with phenotypes of interest from single-cell
912 transcriptomics. *Nat. Biotechnol.* **40**, 355–363 (2022).

913 16. M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert,
914 Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

915 17. J. Moody, T. Kouno, M. Kojima, I. Koya, J. Leon, A. Suzuki, A. Hasegawa, T. Akiyama, N. Akiyama, M. Amagai, J.-C.
916 Chang, A. Fukushima-Nomura, M. Handa, K. Hino, M. Hino, T. Hirata, Y. Imai, K. Inoue, H. Kawasaki, T. Kimura, T.
917 Kinoshita, K. Kubo, Y. Kunii, F. López-Redondo, R. Manabe, T. Miyai, S. Morimoto, A. Nagaoka, J. Nakajima, S. Noma,
918 Y. Okazaki, K. Ozaki, N. Saeki, H. Sakai, K. Seyama, Y. Shibayama, T. Sujino, M. Tagami, H. Takahashi, M. Takao, M.
919 Takeshita, T. Takiuchi, C. Terao, C. W. Yip, S. Yoshinaga, H. Okano, K. Yahamoto, T. Kasukawa, Y. Ando, P. Carninci, J.
920 W. Shin, C.-C. Hon, A single-cell atlas of transcribed *cis* -regulatory elements in the human genome. Genomics [Preprint]
921 (2023). https://doi.org/10.1101/2023.11.13.566791.

922 18. A. Oguchi, A. Suzuki, S. Komatsu, H. Yoshitomi, S. Bhagat, R. Son, R. J. P. Bonnal, S. Kojima, M. Koido, K. Takeuchi, K.
923 Myouzen, G. Inoue, T. Hirai, H. Sano, Y. Takegami, A. Kanemaru, I. Yamaguchi, Y. Ishikawa, N. Tanaka, S. Hirabayashi,
924 R. Konishi, S. Sekito, T. Inoue, J. Kere, S. Takeda, A. Takaori-Kondo, I. Endo, S. Kawaoka, H. Kawaji, K. Ishigaki, H.
925 Ueno, Y. Hayashizaki, M. Pagani, P. Carninci, M. Yanagita, ITEC Consortium‡, N. Parrish, C. Terao, K. Yamamoto, Y.
926 Murakawa, M. Guerrini, H. Hatano, K. Ishigaki, Y. Ishikawa, S. Komatsu, M. Kono, Y. Murakawa, M. Nakano, A. Oguchi,
927 R. Son, A. Suzuki, K. Takeuchi, N. Tanaka, C. Terao, K. Tomizuka, K. Yamamoto, S. Yoshino, An atlas of transcribed
928 enhancers across helper T cell diversity for decoding human diseases. *Science* **385**, eadd8394 (2024).

929 19. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P.
930 Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J.
931 Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data.
932 *Cell* **184**, 3573-3587.e29 (2021).

933 20. M. Watanabe, K. D. Moon, M. S. Vacchio, K. S. Hathcock, R. J. Hodes, Downmodulation of Tumor Suppressor p53 by T
934 Cell Receptor Signaling Is Critical for Antigen-Specific CD4+ T Cell Responses. *Immunity* **40**, 681–691 (2014).

935 21. M. C. Castellanos, C. Muñoz, M. C. Montoya, E. Lara-Pezzi, M. López-Cabrera, M. O. de Landázuri, Expression of the
936 leukocyte early activation antigen CD69 is regulated by the transcription factor AP-1. *J. Immunol. Baltim. Md 1950* **159**,
937 5463–5473 (1997).

938 22. P. Bohacova, M. Terekhova, P. Tsurinov, R. Mullins, K. Husarcikova, I. Shchukina, A. U. Antonova, B. Echalar, J. Kossl, A.
939 Saidu, T. Francis, C. Mannie, L. Arthur, S. D. R. Harridge, D. Kreisel, P. A. Mudd, A. M. Taylor, C. A. McNamara, M.

940 Cella, S. V. Puram, T. Van Den Broek, F. Van Wijk, P. Eghtesady, M. N. Artyomov, Multidimensional profiling of human T
941 cells reveals high CD38 expression, marking recent thymic emigrants and age-related naive T cell remodeling. *Immunity*
942 **57**, 2362-2379.e10 (2024).

943 23. Y. Yasumizu, D. Takeuchi, R. Morimoto, Y. Takeshima, T. Okuno, M. Kinoshita, T. Morita, Y. Kato, M. Wang, D. Motooka,
944 D. Okuzaki, Y. Nakamura, N. Mikami, M. Arai, X. Zhang, A. Kumanogoh, H. Mochizuki, N. Ohkura, S. Sakaguchi, Single-
945 cell transcriptome landscape of circulating CD4+ T cell populations in autoimmune diseases. *Cell Genomics* **4**, 100473
946 (2024).

947 24. J. D. Chan, C. M. Scheffler, I. Munoz, K. Sek, J. N. Lee, Y.-K. Huang, K. M. Yap, N. Y. L. Saw, J. Li, A. X. Y. Chen, C. W.
948 Chan, E. B. Derrick, K. L. Todd, J. Tong, P. A. Dunbar, J. Li, T. X. Hoang, M. N. De Menezes, E. V. Petley, J. S. Kim, D.
949 Nguyen, P. S. K. Leung, J. So, C. Deguit, J. Zhu, I. G. House, L. M. Kats, A. M. Scott, B. J. Solomon, S. J. Harrison, J.
950 Oliaro, I. A. Parish, K. M. Quinn, P. J. Neeson, C. Y. Slaney, J. Lai, P. A. Beavis, P. K. Darcy, FOXO1 enhances CAR T
951 cell stemness, metabolic fitness and efficacy. *Nature* **629**, 201–210 (2024).

952 25. A. E. Doan, K. P. Mueller, A. Y. Chen, G. T. Rouin, Y. Chen, B. Daniel, J. Lattin, M. Markovska, B. Mozarsky, J. Arias-
953 Umana, R. Hapke, I.-Y. Jung, A. Wang, P. Xu, D. Klysz, G. Zuern, M. Bashti, P. J. Quinn, Z. Miao, K. Sandor, W. Zhang,
954 G. M. Chen, F. Ryu, M. Logun, J. Hall, K. Tan, S. A. Grupp, S. E. McClory, C. A. Lareau, J. A. Fraietta, E. Sotillo, A. T.
955 Satpathy, C. L. Mackall, E. W. Weber, FOXO1 is a master regulator of memory programming in CAR T cells. *Nature* **629**,
956 211–218 (2024).

957 26. M. K. Crow, Type I Interferon in the Pathogenesis of Lupus. *J. Immunol.* **192**, 5459–5468 (2014).

958 27. J. He, L. M. Tsai, Y. A. Leong, X. Hu, C. S. Ma, N. Chevalier, X. Sun, K. Vandenberg, S. Rockman, Y. Ding, L. Zhu, W.
959 Wei, C. Wang, A. Karnowski, G. T. Belz, J. R. Ghali, M. C. Cook, D. S. Riminton, A. Veillette, P. L. Schwartzberg, F.
960 Mackay, R. Brink, S. G. Tangye, C. G. Vinuesa, C. R. Mackay, Z. Li, D. Yu, Circulating Precursor CCR7loPD-1hi CXCR5+
961 CD4+ T Cells Indicate Tfh Cell Activity and Promote Antibody Responses upon Antigen Reexposure. *Immunity* **39**, 770–
962 781 (2013).

963 28. T. Koga, C. M. Hedrich, M. Mizui, N. Yoshida, K. Otomo, L. A. Lieberman, T. Rauen, J. C. Crispín, G. C. Tsokos, CaMK4-
964 dependent activation of AKT/mTOR and CREM-α underlies autoimmunity-associated Th17 imbalance. *J. Clin. Invest.*
965 **124**, 2234–2245 (2014).

966 29. V. S. Patil, A. Madrigal, B. J. Schmiedel, J. Clarke, P. O'Rourke, A. D. De Silva, E. Harris, B. Peters, G. Seumois, D.
967 Weiskopf, A. Sette, P. Vijayanand, Precursors of human $CD4^+$ cytotoxic T lymphocytes identified by single-cell
968 transcriptome analysis. *Sci. Immunol.* **3**, eaan8664 (2018).

969 30. Y. Luo, C. Xu, B. Wang, Q. Niu, X. Su, Y. Bai, S. Zhu, C. Zhao, Y. Sun, J. Wang, M. Liu, X. Sun, G. Song, H. Cui, X. Chen,
970 H. Huang, H. Wang, M. Han, E. Jiang, L. Shi, X. Feng, Single-cell transcriptomic analysis reveals disparate effector
971 differentiation pathways in human Treg compartment. *Nat. Commun.* **12**, 3913 (2021).

31. A. H. Jonsson, F. Zhang, G. Dunlap, E. Gomez-Rivas, G. F. M. Watts, H. J. Faust, K. V. Rupani, J. R. Mears, N. Meednu, R. Wang, G. Keras, J. S. Coblyn, E. M. Massarotti, D. J. Todd, J. H. Anolik, A. McDavid, Accelerating Medicines Partnership RA/SLE Network, K. Wei, D. A. Rao, S. Raychaudhuri, M. B. Brenner, Granzyme K^+ CD8 T cells form a core population in inflamed human tissue. *Sci. Transl. Med.* **14**, eabo0686 (2022).

32. T. Velikkakam, K. J. Gollob, W. O. Dutra, Double-negative T cells: Setting the stage for disease control or progression. *Immunology* **165**, 371–385 (2022).

33. M. Colonna, Innate Lymphoid Cells: Diversity, Plasticity, and Unique Functions in Immunity. *Immunity* **48**, 1104–1117 (2018).

34. D. Morgan, V. Tergaonkar, Unraveling B cell trajectories at single cell resolution. *Trends Immunol.* **43**, 210–229 (2022).

35. S. L. Nutt, P. D. Hodgkin, D. M. Tarlinton, L. M. Corcoran, The generation of antibody-secreting plasma cells. *Nat. Rev. Immunol.* **15**, 160–171 (2015).

36. K. Mulder, A. A. Patel, W. T. Kong, C. Piot, E. Halitzki, G. Dunsmore, S. Khalilnezhad, S. E. Irac, A. Dubuisson, M. Chevrier, X. M. Zhang, J. K. C. Tam, T. K. H. Lim, R. M. M. Wong, R. Pai, A. I. S. Khalil, P. K. H. Chow, S. Z. Wu, G. Al-Eryani, D. Roden, A. Swarbrick, J. K. Y. Chan, S. Albani, L. Derosa, L. Zitvogel, A. Sharma, J. Chen, A. Silvin, A. Bertoletti, C. Blériot, C.-A. Dutertre, F. Ginhoux, Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883-1900.e5 (2021).

37. J. Leffler, A. A. Bengtsson, A. M. Blom, The complement system in systemic lupus erythematosus: an update. *Ann. Rheum. Dis.* **73**, 1601–1606 (2014).

38. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).

39. Q. Zhang, Y. He, N. Luo, S. J. Patel, Y. Han, R. Gao, M. Modak, S. Carotta, C. Haslinger, D. Kind, G. W. Peet, G. Zhong, S. Lu, W. Zhu, Y. Mao, M. Xiao, M. Bergmann, X. Hu, S. P. Kerkar, A. B. Vogt, S. Pflanz, K. Liu, J. Peng, X. Ren, Z. Zhang, Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* **179**, 829-845.e20 (2019).

40. H. K. Finucane, Y. A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, S. Gazal, P.-R. Loh, C. Lareau, N. Shoresh, G. Genovese, A. Saunders, E. Macosko, S. Pollack, The Brainstorm Consortium, J. R. B. Perry, J. D. Buenrostro, B. E. Bernstein, S. Raychaudhuri, S. McCarroll, B. M. Neale, A. L. Price, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

1001 41. R. P. Taylor, M. A. Lindorfer, Drug Insight: the mechanism of action of rituximab in autoimmune disease—the immune
1002 complex decoy hypothesis. *Nat. Clin. Pract. Rheumatol.* **3**, 86–95 (2007).

1003 42. F. Müller, J. Taubmann, L. Bucci, A. Wilhelm, C. Bergmann, S. Völkl, M. Aigner, T. Rothe, I. Minopoulou, C. Tur, J. Knitza,
1004 S. Kharboutli, S. Kretschmann, I. Vasova, S. Spoerl, H. Reimann, L. Munoz, R. G. Gerlach, S. Schäfer, R. Grieshaber-
1005 Bouyer, A.-S. Korganow, D. Farge-Bancel, D. Mougiakakos, A. Bozec, T. Winkler, G. Krönke, A. Mackensen, G. Schett,
1006 CD19 CAR T-Cell Therapy in Autoimmune Disease — A Case Series with Follow-up. *N. Engl. J. Med.* **390**, 687–700
1007 (2024).

1008 43. E. Mountjoy, E. M. Schmidt, M. Carmona, J. Schwartzentruber, G. Peat, A. Miranda, L. Fumis, J. Hayhurst, A. Buniello, M.
1009 A. Karim, D. Wright, A. Hercules, E. Papa, E. B. Fauman, J. C. Barrett, J. A. Todd, D. Ochoa, I. Dunham, M. Ghoussaini,
1010 An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated
1011 loci. *Nat. Genet.* **53**, 1527–1533 (2021).

1012 44. X. Yin, K. Kim, H. Suetsugu, S.-Y. Bang, L. Wen, M. Koido, E. Ha, L. Liu, Y. Sakamoto, S. Jo, R.-X. Leng, N. Otomo, V.
1013 Laurynenka, Y.-C. Kwon, Y. Sheng, N. Sugano, M. Y. Hwang, W. Li, M. Mukai, K. Yoon, M. Cai, K. Ishigaki, W. T. Chung,
1014 H. Huang, D. Takahashi, S.-S. Lee, M. Wang, K. Karino, S.-C. Shim, X. Zheng, T. Miyamura, Y. M. Kang, D. Ye, J.
1015 Nakamura, C.-H. Suh, Y. Tang, G. Motomura, Y.-B. Park, H. Ding, T. Kuroda, J.-Y. Choe, C. Li, H. Niiro, Y. Park, C.
1016 Shen, T. Miyamoto, G.-Y. Ahn, W. Fei, T. Takeuchi, J.-M. Shin, K. Li, Y. Kawaguchi, Y.-K. Lee, Y. Wang, K. Amano, D. J.
1017 Park, W. Yang, Y. Tada, K. Yamaji, M. Shimizu, T. Atsumi, A. Suzuki, T. Sumida, Y. Okada, K. Matsuda, K. Matsuo, Y.
1018 Kochi, L. C. Kottyan, M. T. Weirauch, S. Parameswaran, S. Eswar, H. Salim, X. Chen, K. Yamamoto, J. B. Harley, K.
1019 Ohmura, T.-H. Kim, S. Yang, T. Yamamoto, B.-J. Kim, N. Shen, S. Ikegawa, H.-S. Lee, X. Zhang, C. Terao, Y. Cui, S.-C.
1020 Bae, Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann.*
1021 *Rheum. Dis.* **80**, 632–640 (2021).

1022 45. C. D. Langefeld, H. C. Ainsworth, D. S. C. Graham, J. A. Kelly, M. E. Comeau, M. C. Marion, T. D. Howard, P. S. Ramos, J.
1023 A. Croker, D. L. Morris, J. K. Sandling, J. C. Almlöf, E. M. Acevedo-Vásquez, G. S. Alarcón, A. M. Babini, V. Baca, A. A.
1024 Bengtsson, G. A. Berbotto, M. Bijl, E. E. Brown, H. I. Brunner, M. H. Cardiel, L. Catoggio, R. Cervera, J. M. Cucho-
1025 Venegas, S. R. Dahlqvist, S. D'Alfonso, B. M. Da Silva, I. De La Rúa Figueroa, A. Doria, J. C. Edberg, E. Endreffy, J. A.
1026 Esquivel-Valerio, P. R. Fortin, B. I. Freedman, J. Frostegård, M. A. García, I. G. De La Torre, G. S. Gilkeson, D. D.
1027 Gladman, I. Gunnarsson, J. M. Guthridge, J. L. Huggins, J. A. James, C. G. M. Kallenberg, D. L. Kamen, D. R. Karp, K.
1028 M. Kaufman, L. C. Kottyan, L. Kovács, H. Laustrup, B. R. Lauwerys, Q.-Z. Li, M. A. Maradiaga-Ceceña, J. Martín, J. M.
1029 McCune, D. R. McWilliams, J. T. Merrill, P. Miranda, J. F. Moctezuma, S. K. Nath, T. B. Niewold, L. Orozco, N. Ortego-
1030 Centeno, M. Petri, C. A. Pineau, B. A. Pons-Estel, J. Pope, P. Raj, R. Ramsey-Goldman, J. D. Reveille, L. P. Russell, J.
1031 M. Sabio, C. A. Aguilar-Salinas, H. R. Scherbarth, R. Scorza, M. F. Seldin, C. Sjöwall, E. Svenungsson, S. D. Thompson,
1032 S. M. A. Toloza, L. Truedsson, T. Tusié-Luna, C. Vasconcelos, L. M. Vilá, D. J. Wallace, M. H. Weisman, J. E. Wither, T.
1033 Bhangale, J. R. Oksenberg, J. D. Rioux, P. K. Gregersen, A.-C. Syvänen, L. Rönnblom, L. A. Criswell, C. O. Jacob, K. L.
1034 Sivils, B. P. Tsao, L. E. Schanberg, T. W. Behrens, E. D. Silverman, M. E. Alarcón-Riquelme, R. P. Kimberly, J. B. Harley,

1035 E. K. Wakeland, R. R. Graham, P. M. Gaffney, T. J. Vyse, Transancestral mapping and genetic load in systemic lupus
1036 erythematosus. *Nat. Commun.* **8**, 16021 (2017).

1037 46. D. Dimitrov, D. Türei, M. Garrido-Rodriguez, P. L. Burmedi, J. S. Nagai, C. Boys, R. O. Ramirez Flores, H. Kim, B. Szalai, I.
1038 G. Costa, A. Valdeolivas, A. Dugourd, J. Saez-Rodriguez, Comparison of methods and resources for cell-cell
1039 communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).

1040 47. E. Zumaquero, S. L. Stone, C. D. Scharer, S. A. Jenks, A. Nellore, B. Mousseau, A. Rosal-Vela, D. Botta, J. E. Bradley, W.
1041 Wojciechowski, T. Ptacek, M. I. Danila, J. C. Edberg, S. L. Bridges, R. P. Kimberly, W. W. Chatham, T. R. Schoeb, A. F.
1042 Rosenberg, J. M. Boss, I. Sanz, F. E. Lund, IFNγ induces epigenetic programming of human T-bethi B cells and promotes
1043 TLR7/8 and IL-21 induced differentiation. *eLife* **8**, e41641 (2019).

1044 48. B. K. Han, N. J. Olsen, A. Bottaro, The CD27–CD70 pathway and pathogenesis of autoimmune disease. *Semin. Arthritis*
1045 *Rheum.* **45**, 496–501 (2016).

1046 49. Q. Zhang, D. A. A. Vignali, Co-stimulatory and Co-inhibitory Pathways in Autoimmunity. *Immunity* **44**, 1034–1051 (2016).

1047 50. S. M. Stanford, M. N. D. Svensson, C. Sacchetti, C. A. Pilo, D. J. Wu, W. B. Kiosses, A. Hellvard, B. Bergum, G. R. A.
1048 Muench, C. Elly, Y.-C. Liu, J. den Hertog, A. Elson, J. Sap, P. Mydel, D. L. Boyle, M. Corr, G. S. Firestein, N. Bottini,
1049 Receptor Protein Tyrosine Phosphatase α–Mediated Enhancement of Rheumatoid Synovial Fibroblast Signaling and
1050 Promotion of Arthritis in Mice. *Arthritis Rheumatol.* **68**, 359–369 (2016).

1051 51. J. Huang, X. Liu, Y. Wei, X. Li, S. Gao, L. Dong, X. Rao, J. Zhong, Emerging Role of Dipeptidyl Peptidase-4 in Autoimmune
1052 Disease. *Front. Immunol.* **13**, 830863 (2022).

1053 52. E. Doherty, L. Osmani, J. Bilsborrow, J. Par-Young, S. Choi, P. Tilstam, M. Shin, M. Piecychna, H. Cai, L. Leng, W.-U. Kim,
1054 I. Kang, R. Bucala, Pathogenic role of MIF receptor (CD74) expressing T cells in inflammatory arthritis. *Proc. Natl. Acad.*
1055 *Sci.* **123**, e2509156123 (2026).

1056 53. J. B. Kang, A. Nathan, K. Weinand, F. Zhang, N. Millard, L. Rumker, D. B. Moody, I. Korsunsky, S. Raychaudhuri, Efficient
1057 and precise single-cell reference atlas mapping with Symphony. *Nat. Commun.* **12**, 5890 (2021).

1058 54. K. A. Lagattuta, J. B. Kang, A. Nathan, K. E. Pauken, A. H. Jonsson, D. A. Rao, A. H. Sharpe, K. Ishigaki, S. Raychaudhuri,
1059 Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat. Immunol.* **23**, 446–
1060 457 (2022).

1061 55. Y. Nagafuchi, M. Nakano, K. A. Lagattuta, M. Ota, H. Hatano, H. Takahashi, T. Itamiya, H. Inokuchi, S. Raychaudhuri, T.
1062 Okamura, K. Fujio, K. Ishigaki, T cell plasticity in systemic lupus erythematosus revealed by large-scale T cell receptor
1063 repertoire and transcriptome studies. Rheumatology [Preprint] (2025). https://doi.org/10.1101/2025.01.06.24319648.

1064 56. O. V. Britanova, K. R. Lupyr, D. B. Staroverov, I. A. Shagina, A. A. Aleksandrov, Y. Y. Ustyugov, D. V. Somov, A. Klimenko,
1065 N. A. Shostak, I. V. Zvyagin, A. V. Stepanov, E. M. Merzlyak, A. N. Davydov, M. Izraelson, E. S. Egorov, E. A.
1066 Bogdanova, A. K. Vladimirova, P. A. Iakovlev, D. A. Fedorenko, R. A. Ivanov, V. I. Skvortsova, S. Lukyanov, D. M.
1067 Chudakov, Targeted depletion of TRBV9+ T cells as immunotherapy in a patient with ankylosing spondylitis. *Nat. Med.*
1068 **29**, 2731–2736 (2023).

1069 57. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P.
1070 Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S.
1071 Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh,
1072 C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E.
1073 Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S.
1074 Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N.
1075 Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang,
1076 X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A.
1077 Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H.
1078 Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A.
1079 Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330
1080 (2015).

1081 58. V. R. Moulton, G. C. Tsokos, Abnormalities of T cell signaling in systemic lupus erythematosus. *Arthritis Res. Ther.* **13**, 207
1082 (2011).

1083 59. A. Meyer, P. J. Parmar, S. Shahrara, Significance of IL-7 and IL-7R in RA and autoimmunity. *Autoimmun. Rev.* **21**, 103120
1084 (2022).

1085 60. L. Sun, X. Gang, Z. Li, X. Zhao, T. Zhou, S. Zhang, G. Wang, Advances in Understanding the Roles of CD244 (SLAMF4) in
1086 Immune Regulation and Associated Diseases. *Front. Immunol.* **12**, 648182 (2021).

1087 61. E. Katsuyama, M. Humbel, A. Suarez-Fueyo, A. Satyam, N. Yoshida, V. C. Kyttaris, M. G. Tsokos, G. C. Tsokos, CD38 in
1088 SLE CD4 T cells promotes Ca2+ flux and suppresses interleukin-2 production by enhancing the expression of GM2 on the
1089 surface membrane. *Nat. Commun.* **15**, 8304 (2024).

1090 62. P. Badia-i-Mompel, J. Vélez Santiago, J. Braunger, C. Geiss, D. Dimitrov, S. Müller-Dott, P. Taus, A. Dugourd, C. H.
1091 Holland, R. O. Ramirez Flores, J. Saez-Rodriguez, decoupleR: ensemble of computational methods to infer biological
1092 activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).

1093 63. Z. Zou, T. Ohta, S. Oki, ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale
1094 regulome data. *Nucleic Acids Res.* **52**, W45–W53 (2024).

1095 64. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C.
1096 Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502
1097 (2019).

1098 65. C. Saggau, P. Bacher, D. Esser, M. Rasa, S. Meise, N. Mohr, N. Kohlstedt, A. Hutloff, S.-S. Schacht, J. Dargvainiene, G. R.
1099 Martini, K. H. Stürner, I. Schröder, R. Markewitz, J. Hartl, M. Hastermann, A. Duchow, P. Schindler, M. Becker, C.
1100 Bautista, J. Gottfreund, J. Walter, J. K. Polansky, M. Yang, R. Naghavian, M. Wendorff, E.-M. Schuster, A. Dahl, A.
1101 Petzold, S. Reinhardt, A. Franke, M. Wieczorek, L. Henschel, D. Berger, G. Heine, M. Holtsche, V. Häußler, C. Peters, E.
1102 Schmidt, S. Fillatreau, D. H. Busch, K.-P. Wandinger, K. Schober, R. Martin, F. Paul, F. Leypoldt, A. Scheffold,
1103 Autoantigen-specific CD4+ T cells acquire an exhausted phenotype and persist in human antigen-specific autoimmune
1104 diseases. *Immunity* **57**, 2416-2432.e8 (2024).

1105 66. Z. Zhou, H. He, K. Wang, X. Shi, Y. Wang, Y. Su, Y. Wang, D. Li, W. Liu, Y. Zhang, L. Shen, W. Han, L. Shen, J. Ding, F.
1106 Shao, Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells. *Science* **368**,
1107 eaaz7548 (2020).

1108 67. A. Perl, L. Morel, Expanding scope of TEMRA in autoimmunity. *eBioMedicine* **90**, 104520 (2023).

1109 68. M. Kono, H. Hatano, K. Asahara, M. Nakano, R. Bagherzadeh, T. Kawashima, T. Arakawa, M. Sato, H. Inokuchi, T.
1110 Nishino, T. Itamiya, H. Takahashi, B. Natsumoto, A. Suzuki, K. Yamamoto, K. Ishigaki, Accurate, sensitive, and efficient
1111 chromatin accessibility quantification at target loci using UNIChro-seq. *medRxiv*, 2025.07.29.25332340 (2025).

1112 69. L. Chiche, N. Jourde-Chiche, E. Whalen, S. Presnell, V. Gersuk, K. Dang, E. Anguiano, C. Quinn, S. Burtey, Y. Berland, G.
1113 Kaplanski, J.-R. Harle, V. Pascual, D. Chaussabel, Modular Transcriptional Repertoire Analyses of Adults With Systemic
1114 Lupus Erythematosus Reveal Distinct Type I and Type II Interferon Signatures. *Arthritis Rheumatol.* **66**, 1583–1595
1115 (2014).

1116 70. M. E. Munroe, R. Lu, Y. D. Zhao, D. A. Fife, J. M. Robertson, J. M. Guthridge, T. B. Niewold, G. C. Tsokos, M. P. Keith, J.
1117 B. Harley, J. A. James, Altered type II interferon precedes autoantibody accrual and elevated type I interferon activity prior
1118 to systemic lupus erythematosus classification. *Ann. Rheum. Dis.* **75**, 2014–2021 (2016).

1119 71. A. Fava, J. Buyon, C. Mohan, T. Zhang, H. M. Belmont, P. Izmirly, R. Clancy, J. M. Trujillo, D. Fine, Y. Zhang, L. Magder,
1120 D. A. Rao, A. Arazi, C. C. Berthier, A. Davidson, B. Diamond, N. Hacohen, D. Wofsy, W. Apruzzese, S. Raychaudhuri, M.
1121 Petri, Integrated urine proteomics and renal single-cell genomics identify an IFN-γ response gradient in lupus nephritis.
1122 *JCI Insight* **5**, e138345 (2020).

1123 72. A. Wagner, A. Regev, N. Yosef, Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**,
1124 1145–1160 (2016).

1125 73. S. Gurajala, N. Sugiarto, M. Curtis, T. Eisenhaure, A. Arazi, A. Fava, Q. Xiao, J. Mears, B. Rovin, C. C. Berthier, P. Izmirly,
1126 J. Barnas, P. J. Hoover, M. Peters, R. Raychowdhury, A. Horsberger, S. Sakaue, Y. Zhao, R. A. Furie, M. Belmont, D.
1127 Hildeman, E. S. Woodle, M. Dall'Era, C. Putterman, D. L. Kamen, M. A. McMahon, J. Grossman, K. Kalunian, J. Hodgin,
1128 F. Payan-Schober, W. Apruzzese, H. Perlman, C. M. Cuda, D. Wofsy, J. M. Guthridge, J. H. Anolik, J. A. James,
1129 Accelerating Medicines Partnerships Rheumatoid Arthritis/Systemic Lupus Erythematosus, D. A. Rao, A. Davidson, M. A.
1130 Petri, J. P. Buyon, N. Hacohen, B. Diamond, S. Raychaudhuri, A population-scale atlas of blood and tissue in lupus
1131 nephritis. *bioRxiv*, 2025.08.11.669754 (2025).

1132 74. A. Missarova, E. Dann, L. Rosen, R. Satija, J. Marioni, Leveraging neighborhood representations of single-cell data to
1133 achieve sensitive DE testing with miloDE. *Genome Biol.* **25**, 189 (2024).

1134 75. T. Lappalainen, D. G. MacArthur, From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).

1135 76. A. F. Labrijn, M. L. Janmaat, J. M. Reichert, P. W. H. I. Parren, Bispecific antibodies: a mechanistic review of the pipeline.
1136 *Nat. Rev. Drug Discov.* **18**, 585–608 (2019).

1137 77. C. Bombardier, D. D. Gladman, M. B. Urowitz, D. Caron, C. H. Chang, A. Austin, A. Bell, D. A. Bloch, P. N. Corey, J. L.
1138 Decker, J. Esdaile, J. F. Fries, E. M. Ginzler, C. H. Goldsmith, M. C. Hochberg, J. V. Jones, N. G. H. L. Riche, M. H.
1139 Liang, M. D. Lockshin, L. R. Muenz, D. L. Sackett, P. H. Schur, Derivation of the sledai. A disease activity index for lupus
1140 patients. *Arthritis Rheum.* **35**, 630–640 (1992).

1141 78. A. Rogozhnikov, P. Ramkumar, K. Shah, R. Bedi, S. Kato, G. S. Escola, Demuxalot: scaled up genetic demultiplexing for
1142 single-cell sequencing. Bioinformatics [Preprint] (2021). https://doi.org/10.1101/2021.05.22.443646.

1143 79. H. Heaton, A. M. Talman, A. Knights, M. Imaz, D. J. Gaffney, R. Durbin, M. Hemberg, M. K. N. Lawniczak, Souporcell:
1144 robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620
1145 (2020).

1146 80. Y. Hao, T. Stuart, M. H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, C.
1147 Fernandez-Granda, R. Satija, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat.*
1148 *Biotechnol.* **42**, 293–304 (2024).

1149 81. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. Loh, S. Raychaudhuri,
1150 Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

1151 82. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic
1152 Data. *Cell Syst.* **8**, 281-291.e9 (2019).

1153 83. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using **lme4**. *J. Stat. Softw.* **67** (2015).

1154 84. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR : a Bioconductor package for differential expression analysis of digital
1155 gene expression data. *Bioinformatics* **26**, 139–140 (2010).

1156 85. S. Xu, E. Hu, Y. Cai, Z. Xie, X. Luo, L. Zhan, W. Tang, Q. Wang, B. Liu, R. Wang, W. Xie, T. Wu, L. Xie, G. Yu, Using
1157 clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).

1158 86. K. A. Jagadeesh, K. K. Dey, D. T. Montoro, R. Mohan, S. Gazal, J. M. Engreitz, R. J. Xavier, A. L. Price, A. Regev,
1159 Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human
1160 genetics. *Nat. Genet.* **54**, 1479–1492 (2022).

1161 87. J. W. Freimer, O. Shaked, S. Naqvi, N. Sinnott-Armstrong, A. Kathiria, C. M. Garrido, A. F. Chen, J. T. Cortez, W. J.
1162 Greenleaf, J. K. Pritchard, A. Marson, Systematic discovery and perturbation of regulatory genes in human T cells reveals
1163 the architecture of immune networks. *Nat. Genet.* **54**, 1133–1144 (2022).

1164 88. T. Hayashi, H. Ozaki, Y. Sasagawa, M. Umeda, H. Danno, I. Nikaido, Single-cell full-length total RNA sequencing uncovers
1165 dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* **9**, 619 (2018).

1166 89. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR:
1167 ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

1168 90. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

1169 91. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X.
1170 S. Liu, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

1171