

Supplementary Methods

Library building and data processing

BD Rhapsody system was used following the manufacturer's instructions. Sequencing libraries were generated using a nano-well based system following the AbSeq protocol reported previously¹. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

Raw gene expression matrices were generated for each sample by the Cell Ranger (v.3.0.2) pipeline coupled with human reference version GRCh38. The output filtered gene expression matrices were analyzed by R software (v.4.0.4) with the Seurat package (v.4.0, <https://satijalab.org/seurat>). In brief, gene expressed at a proportion >0.1% of the data and cells with >200 genes detected were selected for further analyses. Low-quality cells were removed if they: 1) <800 unique molecular identifiers (UMIs); 2) <200 genes; or 3) >25% UMIs derived from the mitochondrial genome. After removal of low-quality cells, gene expression matrices were normalized by the `NormalizeData` function and 2000 features with high cell-to-cell variation were calculated using `FindVariableFeatures` function. To reduce dimensionality of the datasets, the `RunPCA` function was conducted with default parameters on linear-transformation scaled data generated by the `ScaleData` function. Next, the `ElbowPlot` functions were used to identify the true dimensionality of each dataset, as recommended by the Seurat developers. Finally, we clustered cells using the `FindNeighbors` and `FindClusters` functions and performed nonlinear dimensional reduction with the `RunUMAP` function with default settings.

After that, we identified “anchors” between individual datasets with the *FindIntegrationAnchors* function and inputted these anchors into the *IntegrateData* function to create a batch-corrected expression matrix of all cells, which allowed cells from different datasets to be integrated and analyzed together. Expression levels of 9 cell surface proteins were also generated for PBMCs from 5 IgG4-RD patients (P1, P2, P3, P4, and P17) and 3 HCs (HC5, HC6, HC7). Seurat v.4.0 was used to integrate the individual mRNA and cell surface protein expression profiles for 8 samples, generating an integrate matrix of transcriptomic and proteomic features.

Cell type annotation and cluster marker identification

After nonlinear dimensional reduction and projection of all cells into two-dimensional space by UMAP, cells were clustered together according to common features. The *FindAllMarkers* function in Seurat was used to find markers for each of the identified clusters. Clusters were then classified and annotated based on expressions of canonical markers of particular cell types. Clusters expressing two or more canonical cell type markers were classified as doublet cells and excluded from further analysis.

DEG identification and functional enrichment

Differentially expressed gene (DEG) testing was performed using the *FindMarkers* function in Seurat with parameter “test.use=wilcox” by default and the Benjamini-Hochberg method was used to estimate the false discovery rate (FDR). DEGs were filtered using a minimum \log_2 (fold change) of 0.25 and a maximum p value of 0.001. Enrichment analysis for the functions of the DEGs was conducted using the Metascape webtool (www.metascape.org). Gene set enrichment analysis was

performed with 1000 permutations with the GSEA software (v4.1.0) on a matrix of all genes.

Defining cell state scores

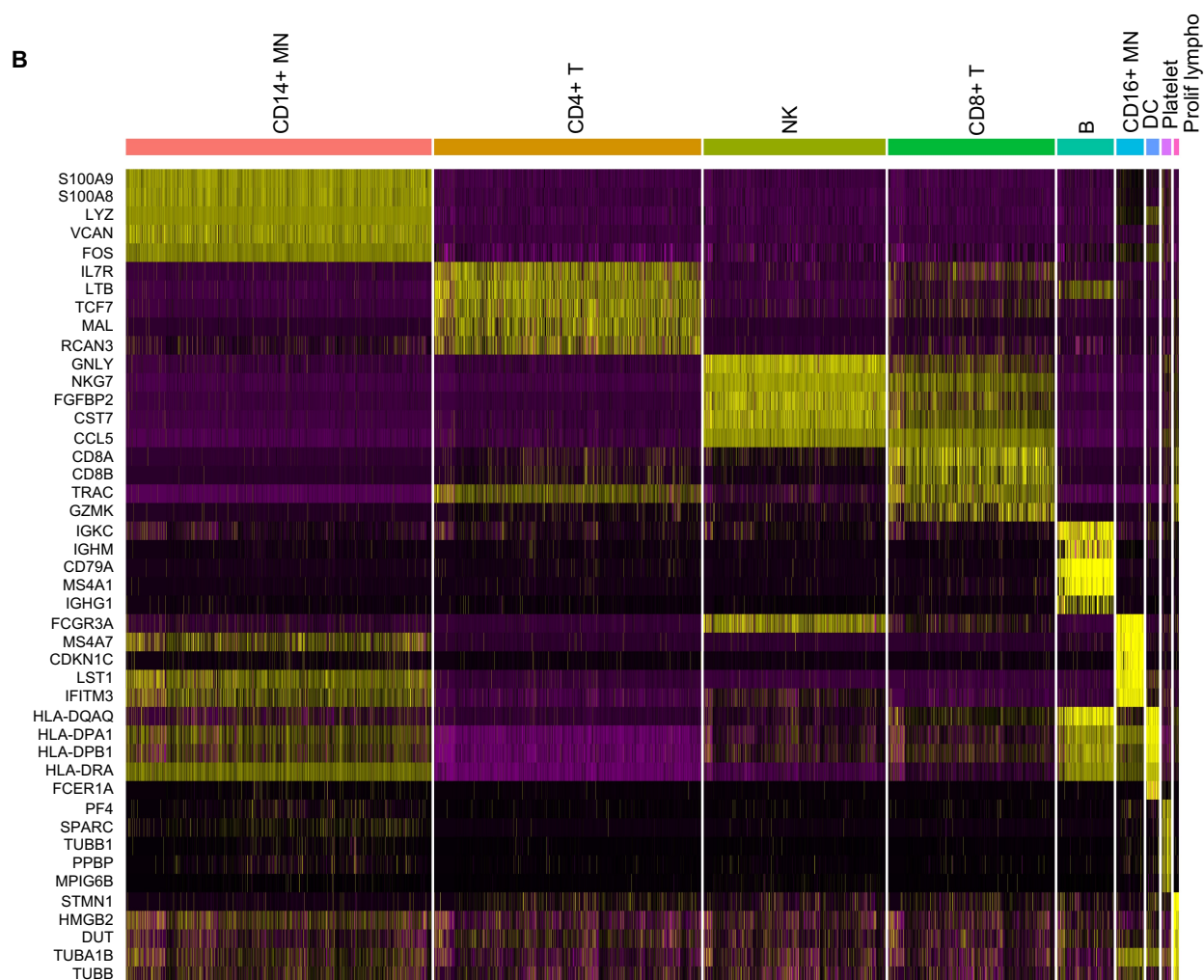
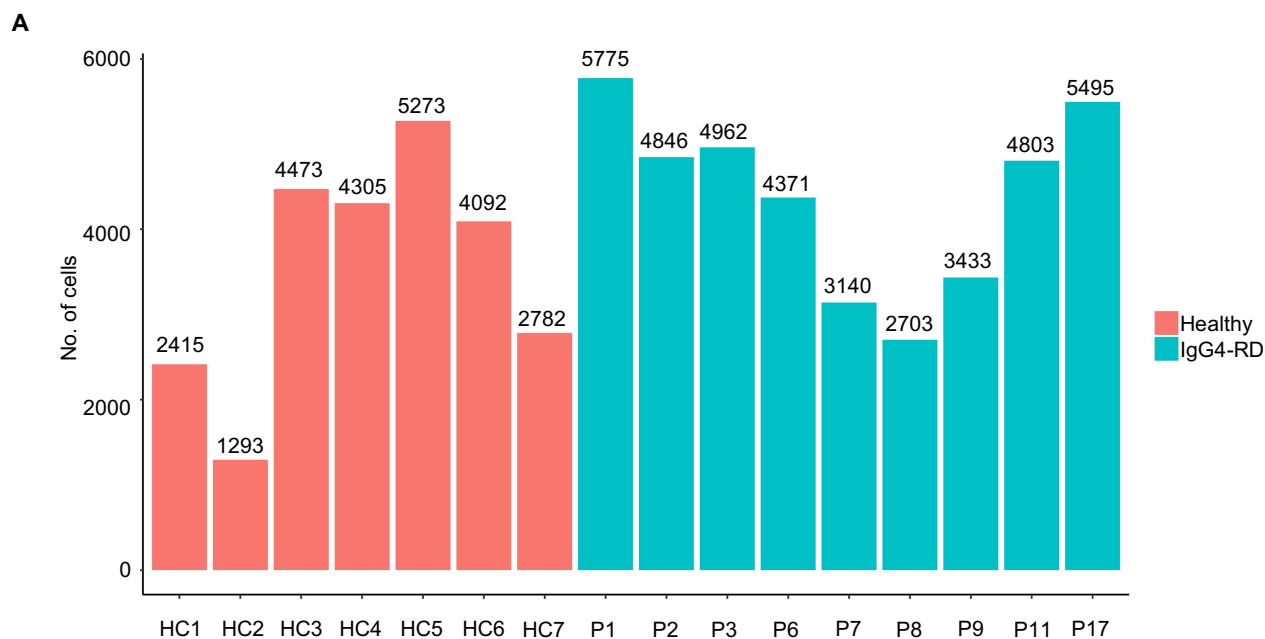
The *AddModuleScore* function in Seurat was used to implement the method with default settings. We used chemotaxis-associated genes (SPN, AKT1, S100A9, RPS19, RAC2, CCL5, CCL4, CX3CR1, ITGB2, CD74, CALR, ADAM8) and cytotoxicity-associated genes (FGFBP2, GNLY, NKG7, PRF1, GZMA, GZMB, GZMH, CTSW, CST7, KLRB1, KLRD1, KLRK1, IFNG) to define the chemotaxis and cytotoxicity of T cell subpopulations. Inhibitory receptors including TIGIT, CTLA4, PDCD1, CD274, BTLA, CD160, LAG3, HAVCR2, CD244, CD200R1, and PECAM1 were used to evaluate the inhibitory signals on cytotoxic T cells.

Trajectory analysis and SCENIC analysis

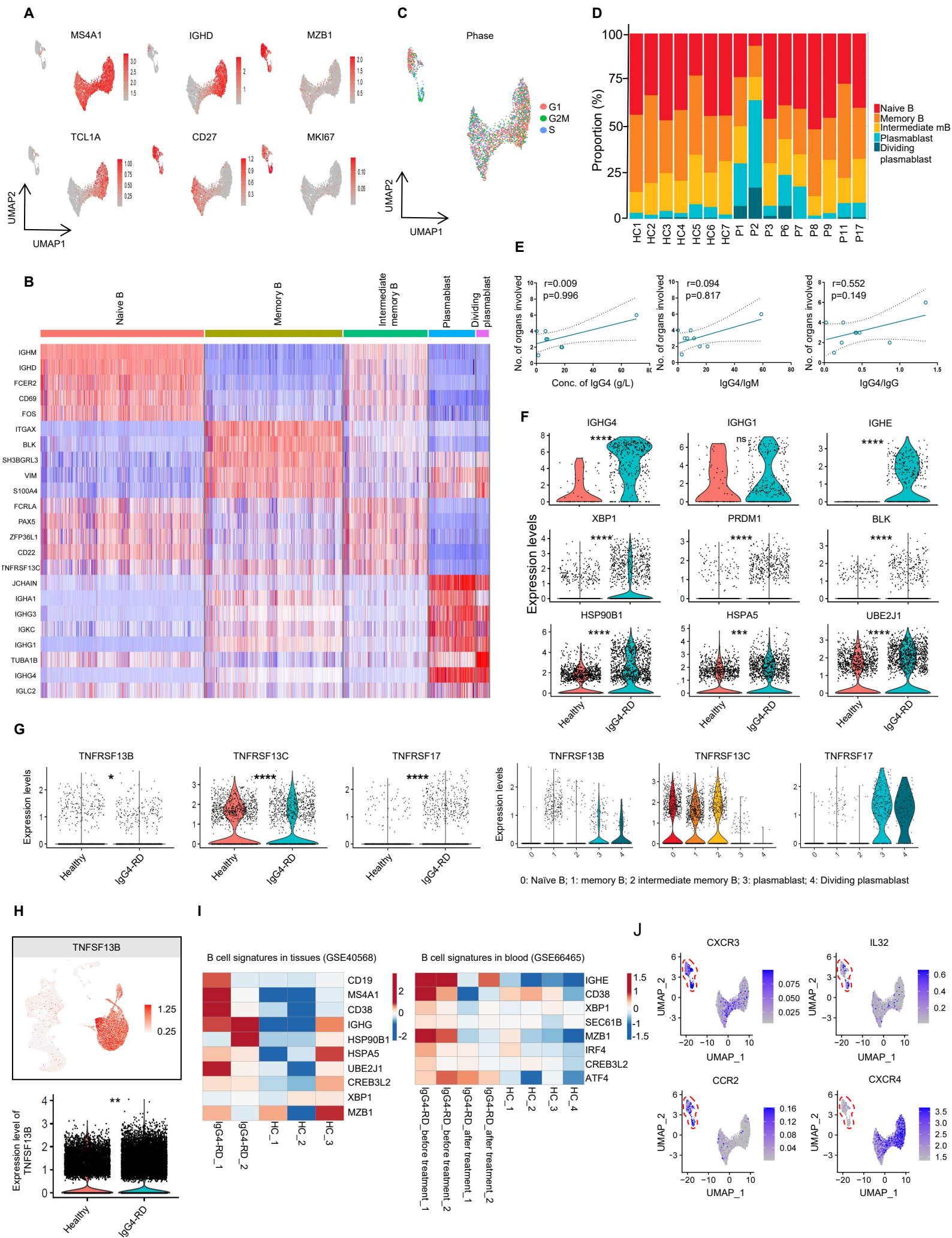
Trajectory analysis was carried out using the *OrderCells* function of Monocle 2², setting the naïve cells, identified by canonical markers, as root node. This was done separately for B cells, CD4⁺ T cells, and CD8⁺ T cells. The SCENIC analysis was run using the pycenic (version 0.9.19) and hg19-500bp-upstream-10species databases for RcisTarget, GRNboost, and AUCell³. The input normalized expression matrix was from Seurat.

Cell-cell interaction

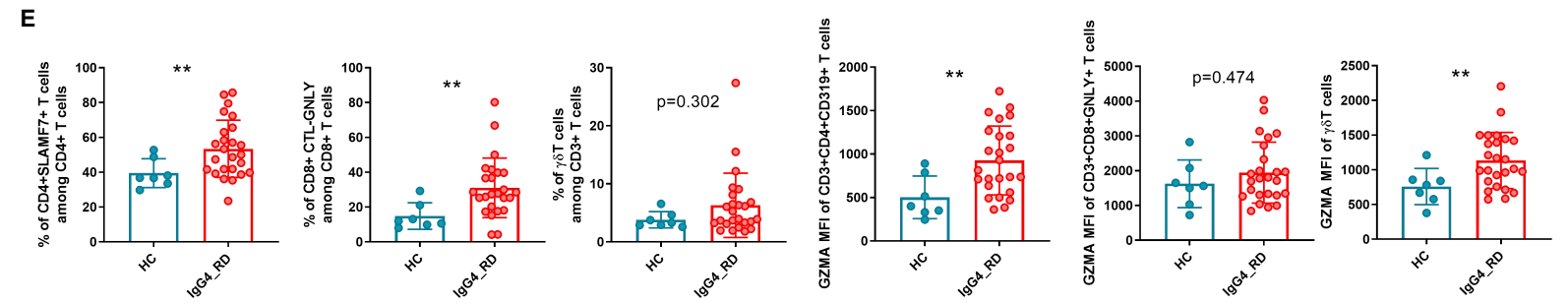
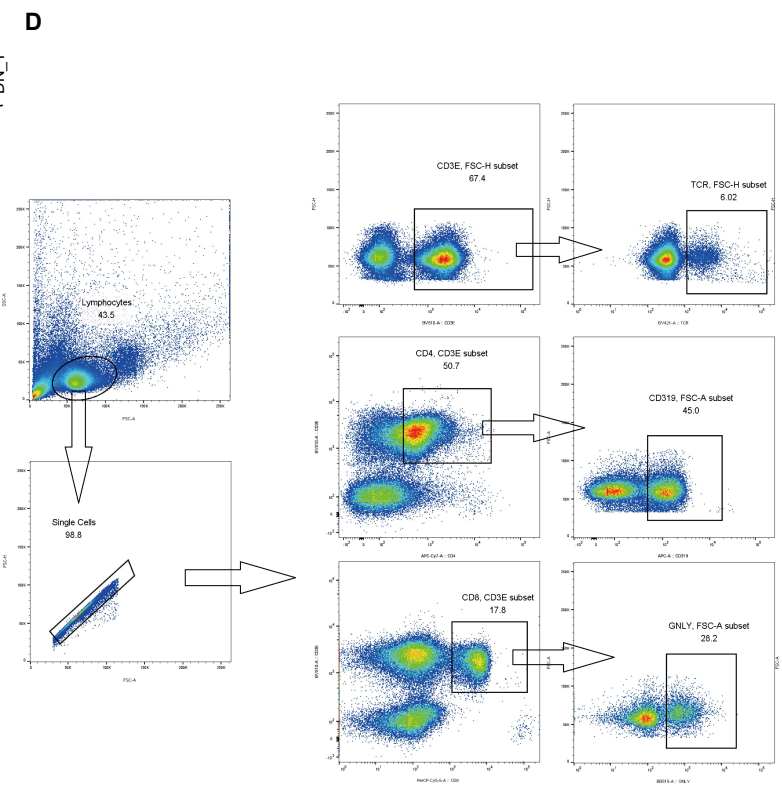
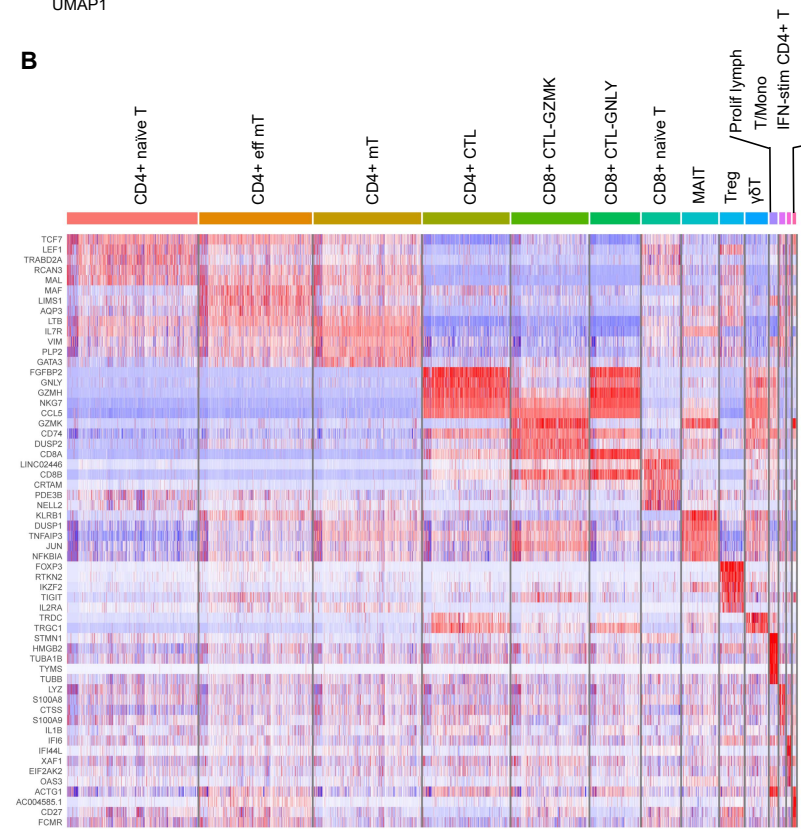
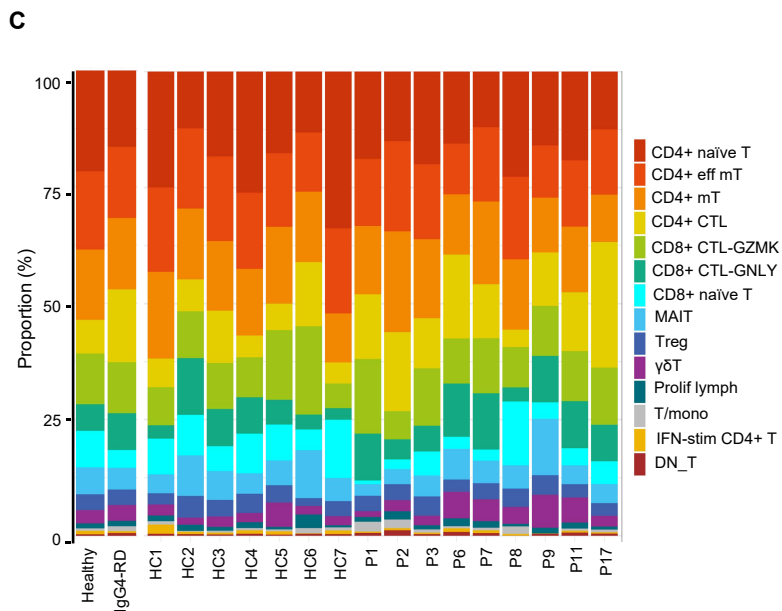
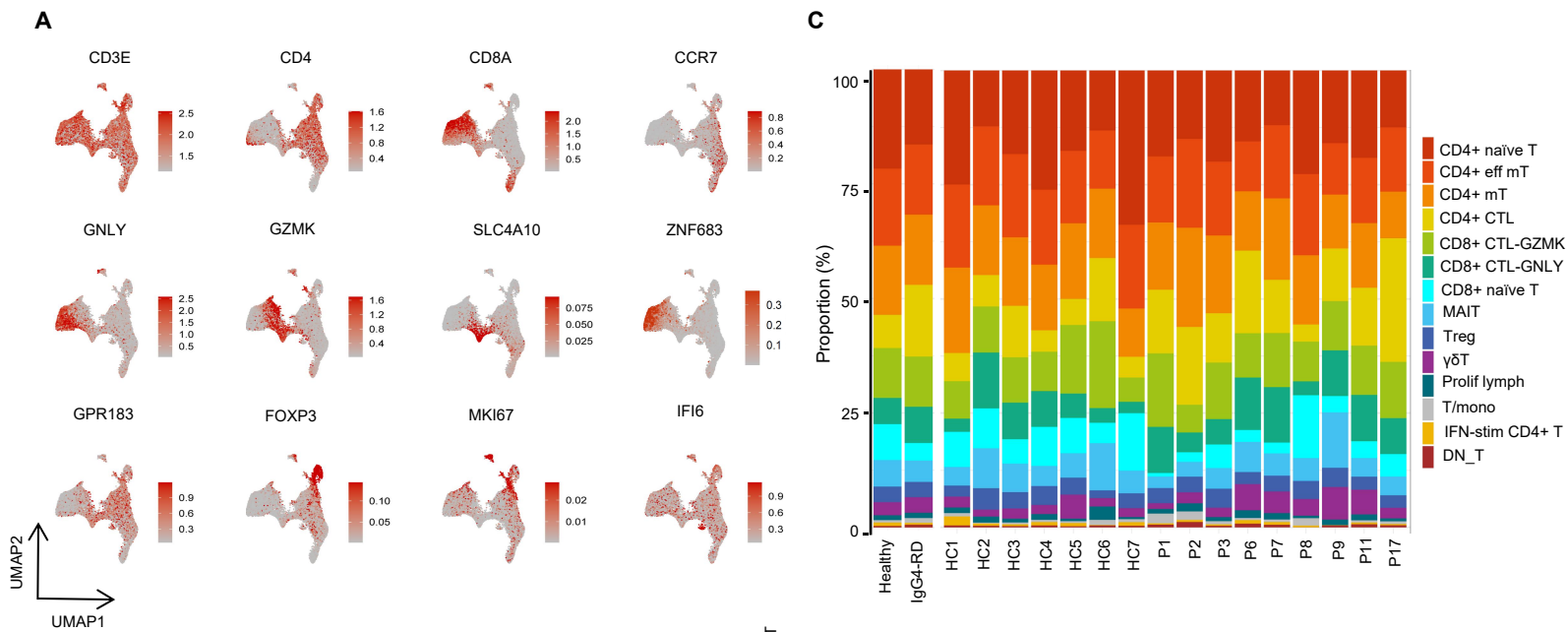
Cell-cell interaction analysis was performed by using the CellPhoneDB package⁴ and Cellchat⁵. The existence of the potential interactions between two cell types was explored by evaluating the expression levels of annotated ligands and receptors between the two cell types.



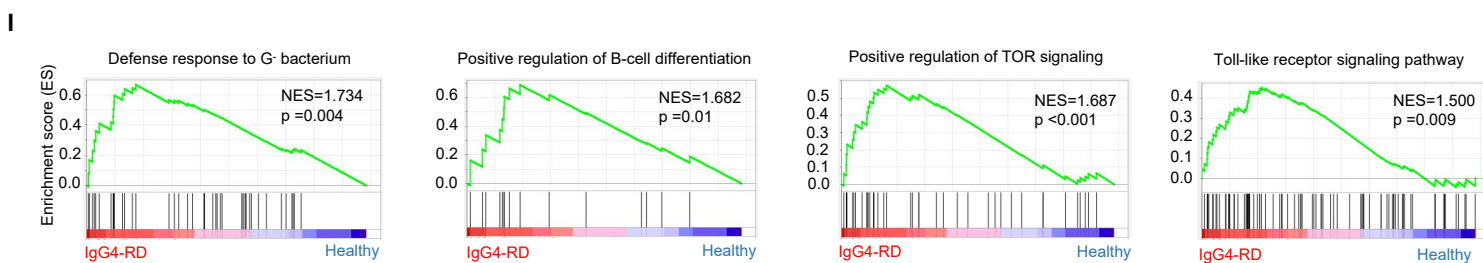
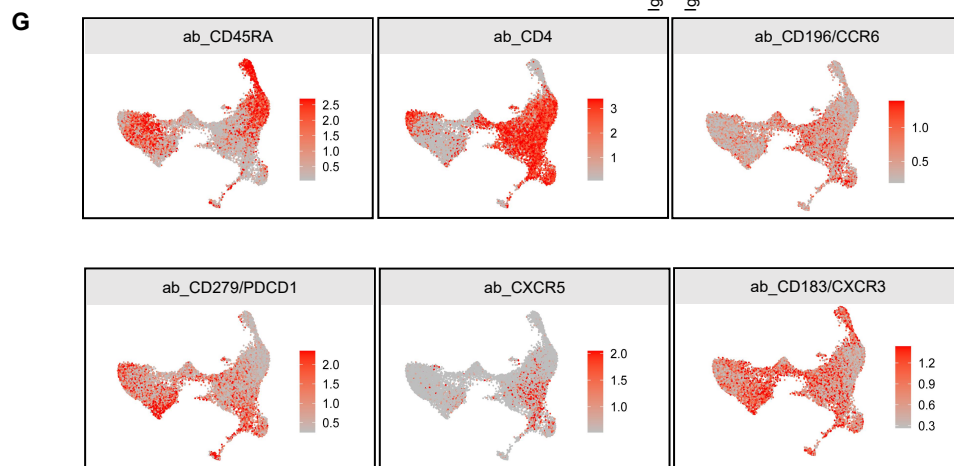
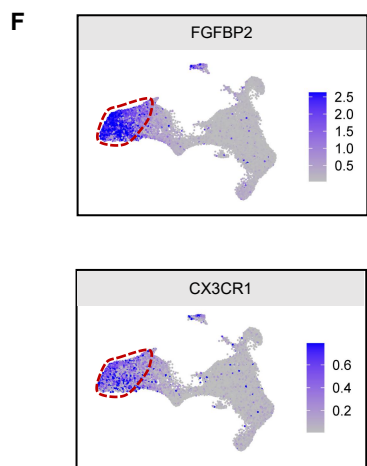
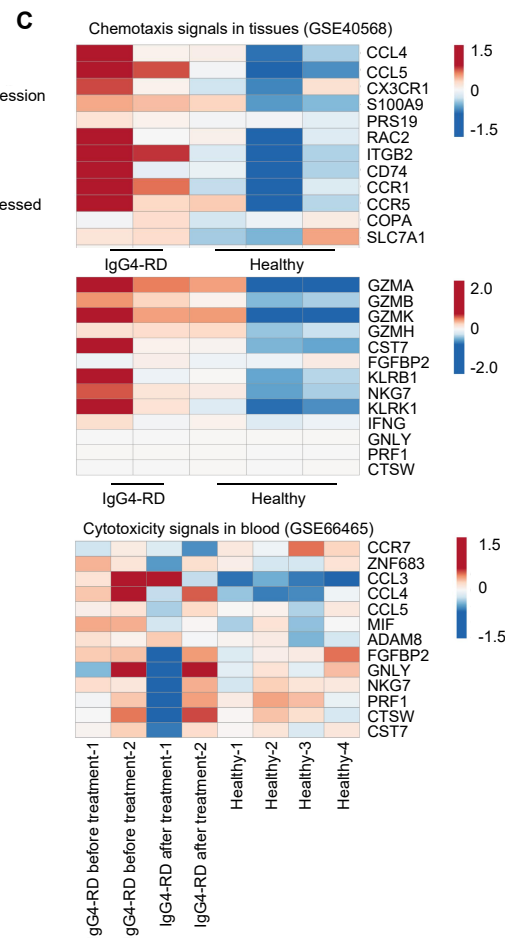
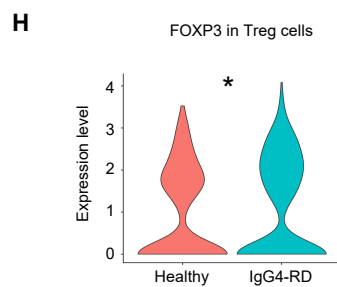
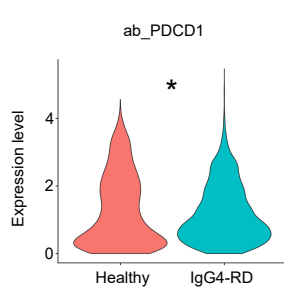
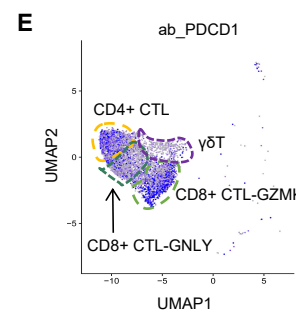
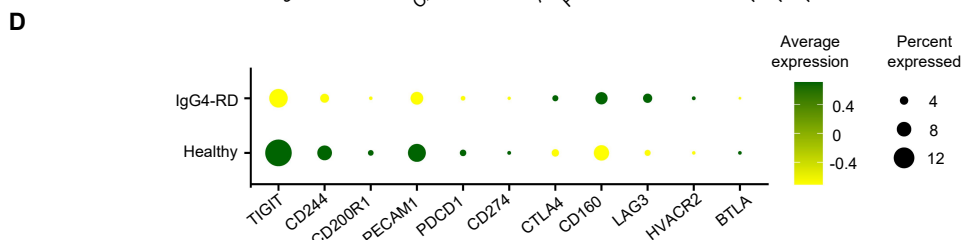
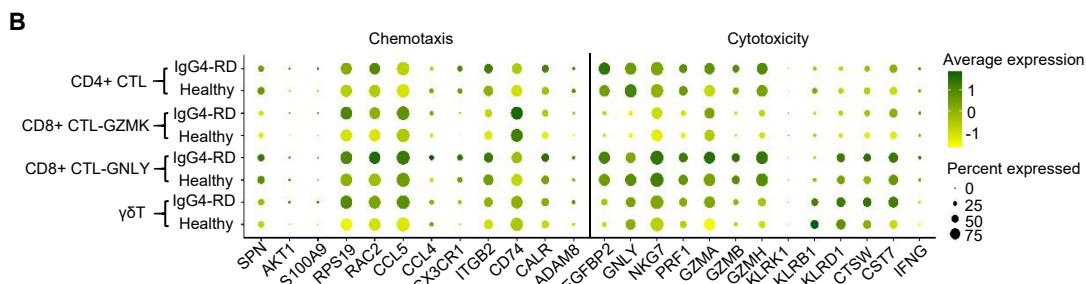
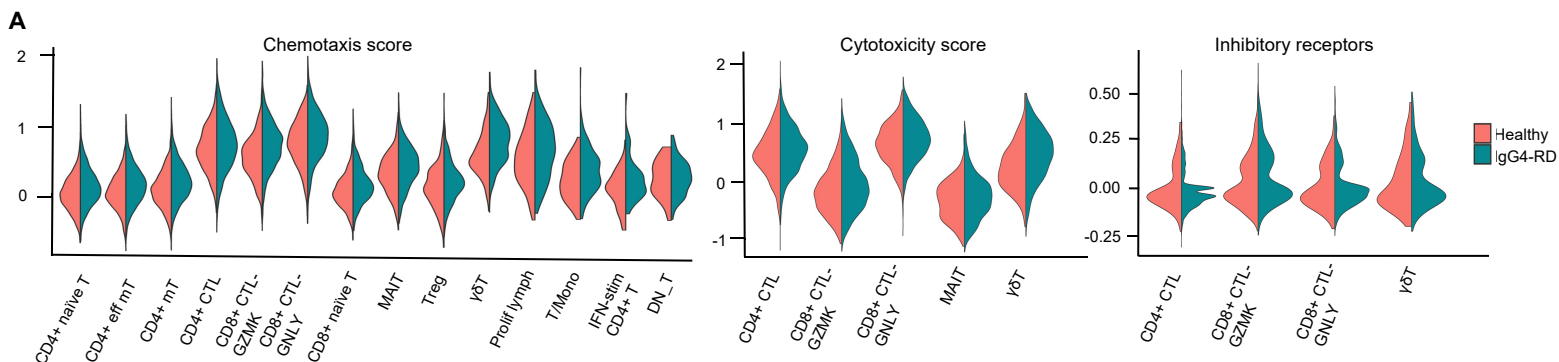
Supplementary Figure 1. A, An overview of number of cells of each sample; **B**, Heatmap showing the expression of top 5 significantly enriched transcripts in each cluster (cluster of interest versus all other cells).



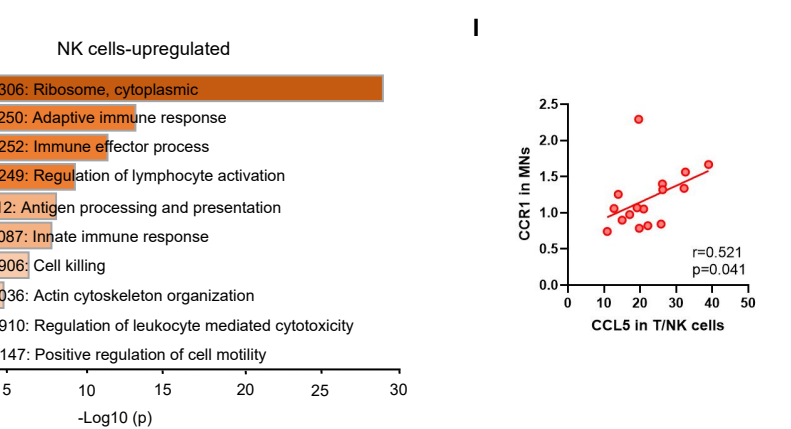
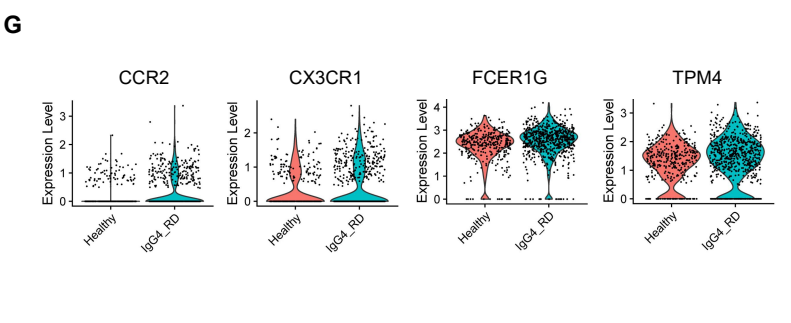
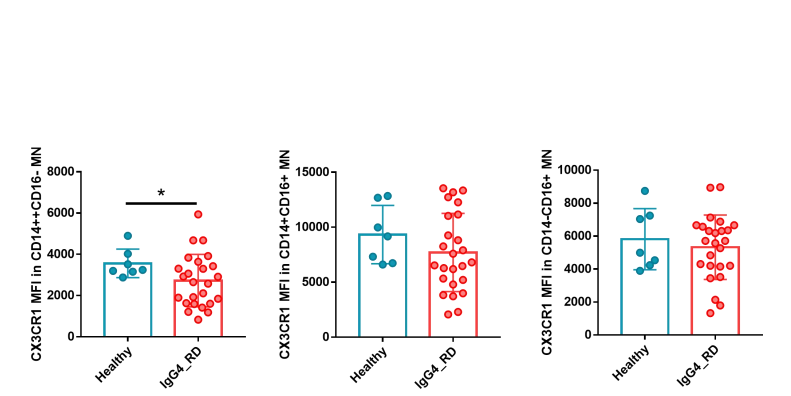
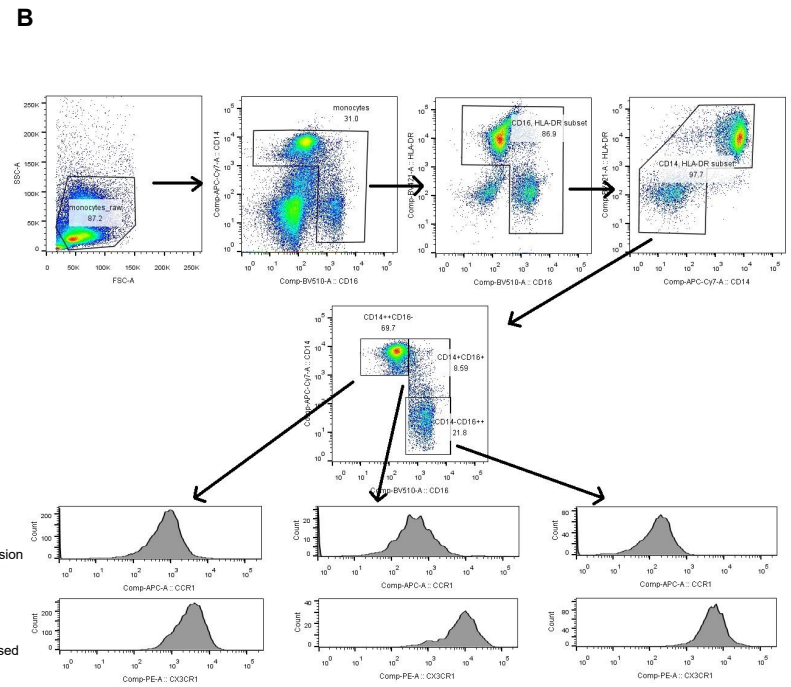
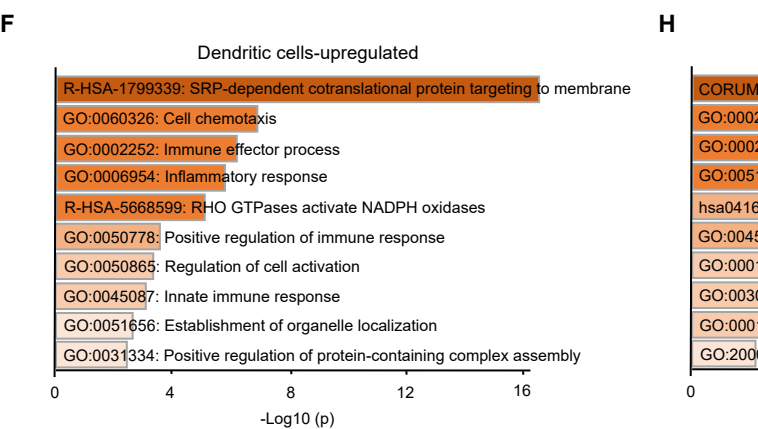
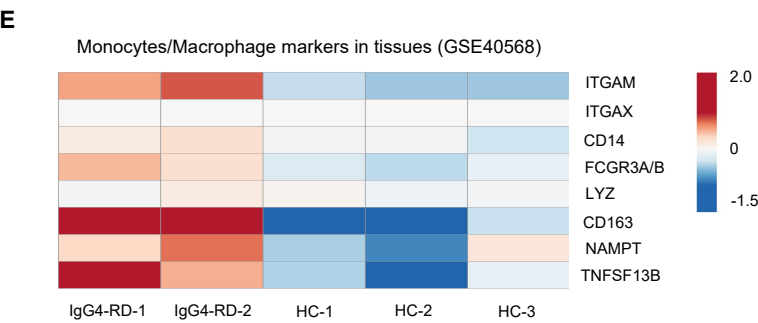
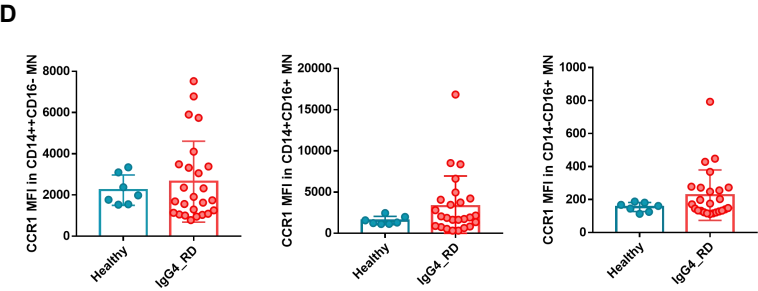
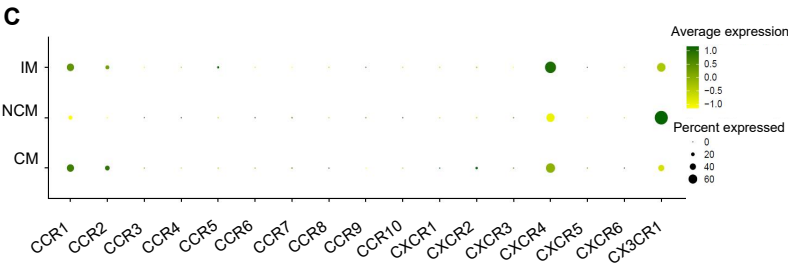
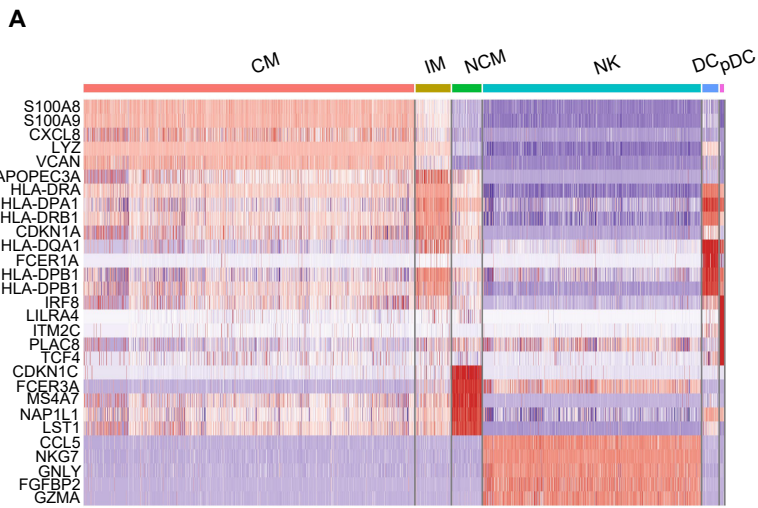
Supplementary Figure 2. **A**, Expression of selective canonical genes for five major subtypes in B cells. **B**, Heatmap showing the expression of top 5 significantly enriched transcripts in each cluster. **C**, Distribution of cells in different cell phase. **D**, Barplot showing proportions of B subpopulations in each sample. HC: health controls; P: patients. **E**, Scatter plots depicting the correlation between serum concentration of IgG4 (g/L, Spearman's r), IgG4/IgM ratio (Spearman's r), or IgG4/IgG ratio (Pearson's r) and numbers of involved organs in IgG4-RD patients (n=9). Exact two-tailed P values and Pearson's/Spearman's r were presented. **F**, Violin plots showing expression level of immunoglobulin-coding genes, B cell activation genes, and ER stress-associated genes in B cells of IgG4-RD and HC. **G**, Violin plots showing expression levels of genes coding receptors for BAFF in B cells of IgG4-RD and HC and in B clusters. **H**, TNFSF13B (encoding BAFF) was specifically expressed in monocytes, and was upregulated in IgG4-RD. **I**, Heatmap showing the expression of B cell maturation-associated genes in tissues (GSE40568) and blood (GSE66465). **J**, Plasmablasts and dividing plasmablasts have upregulated expression of CX3CR1, CCR2, and IL32, but absence of CXCR4 expression. Significance was determined by student's T test or Mann–Whitney test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.



Supplementary Figure 3. **A**, UMAP plots showing the expression of selected canonical genes. **B**, Heatmap showing the expression of top 5 significantly enriched transcripts in each cluster. **C**, Barplot showing proportions of T subpopulations in each sample. HC: health control; P: patient. **D**, Gating strategy for flow cytometry data of lymphocytes. **E**, Proportions and mean fluorescence intensity of GZMA of each cell type in HCs (n=7) and IgG4-RD patients (n=25) by flow cytometry. Results were shown as mean±SD. Significance was determined by student's T test or Mann–Whitney test. *p<0.05, **p<0.01.



Supplementary Figure 4. **A**, Violin plots showing expression of chemotaxis-associated genes, cytotoxicity-associated genes, and inhibitory receptors shown in Figure 4D. **B**, Dot plot of chemotaxis and cytotoxicity-associated genes in CD4⁺ CTLs, CD8⁺ CTLs-GZMK, CD8⁺ CTLs-GNLY, and $\gamma\delta$ T cells from IgG4-RD and HCs. **C**, Heatmap showing the expression of chemotaxis and cytotoxicity signals in tissues (GSE40568) and blood (GSE66465) of IgG4-RD and HC. **D**, the expression levels of each inhibitory receptors. **E**, the expression level of PD-1 protein in CD4⁺ CTLs, CD8⁺ CTLs and $\gamma\delta$ T cells and comparison between IgG4-RD and HC. **F**, UMAPs depicts the cytotoxic T cells have high expression of FGFBP2 and CX3CR1. **G**, UMAPs showing the expression of CD45RA, CD4, CD279/PDCD1, CXCR5, CCR6, and CXCR3 protein on T cells. **H**, comparison of Foxp3 expression in Treg cells between IgG4-RD and HCs. **I**, GSEA shows top enriched pathways in T cells. NES denotes normalized enrichment score. Significance was determined by student's T test or Mann-Whitney test. *p<0.05.



Supplementary Figure 5. A, Heatmap showing the expression of top 5 significantly enriched transcripts in each cluster. **B**, Gating strategy for flow cytometry data of monocytes. **C**, Dot plot of chemokine receptor genes in monocytes. **D**, Flow cytometry showing expression of CCR1 and CX3CR1 on CM, IM, and NCM from HCs (n=7) or IgG4-RD patients (n=25). **E**, Heatmap illustrating the expression of MN/macrophage marker genes in IgG4-RD and HC tissues (GSE40568). **F**, Biological processes enriched for upregulated genes in DCs from IgG4-RD versus HC. **G**, Violin plots showing the comparisons of CCR2, CX3CR1, FER1G, and TPM4 genes in IgG4-RD and HC DCs. **H**, Biological processes enriched for upregulated genes in NK cells from IgG4-RD versus HC. **I**, The relationship between CCR1 expression in MNs and CCL5 expression in T and NK cells (n=16, Pearson's r with 2-tailed p value). CM: classical monocyte, NCM: non-classical monocyte, IM: intermediate monocyte, DC: dendritic cell, pDC: plasmacytoid dendritic cell. Results were shown as mean±SD. Significance was determined by student's T test or Mann–Whitney test. *p<0.05.

Supplementary Table 1 Characteristics of patients with IgG4-RD and healthy controls

No.	sex	age(year)	Allergy	Lymph node	salivary/lachrymal glands	respiratory system	Liver, biliary tract and pancreas	Treatment at the time of enrollment
P-1	F	69	NA	Lymph nodes in the mediastinum, bilateral axillary, retroperitoneal area were enlarged	None	Patient had recurrent cough. Pulmonary CT scan found fibrosis in some areas of both lungs.	CT showed: 1) the pancreas is round and blunt and enveloped by "capsule" like structure. 2) the liver was slightly larger; the wall of intrahepatic and extrahepatic bile ducts and gallbladder was slightly thicker, and the intrahepatic and extrahepatic bile ducts were slightly dilated; 3) the left anterior and lateral pyramidal renal fascia were thickened.	Amlodipine 5mg daily for hypertension; Gliclazide 60mg daily for diabetes
P-2	M	49	none	Lymph nodes in the cervical, bilateral axillary, mediastinum area were enlarged	Bilateral lachrymal glands, parotid glands, and submandibular glands were swollen	The wall of nasopharynx swelled and bilateral pharyngeal recess disappeared. Pulmonary CT	CT scan found several low-density foci in the left lobe of the liver, thickening of extrahepatic bile duct wall and mild dilatation of intrahepatic bile duct. The body and tail of pancreas were slightly dilated and the spleen was enlarged.	NO

						scan found patchy and nodular shadows in both lungs, and two nodules in the upper lobe of the right lung.		
P-3	F	47	amoxicillin, metronidazole, and iodine	Cervical lymph nodes were enlarged	Bilateral lachrymal glands, submandibular, and sublingual glands were enlarged	None	None	NO
P-6	M	69	urticaria history	Cervical lymph nodes were enlarged	Bilateral lachrymal glands, parotid glands, sublingual and submandibular glands were swollen	None	None	NO

P-7	M	74	none	The mediastinum and right axillary lymph nodes were enlarged	Bilateral lachrymal glands were enlarged	Mild obstructive ventilation dysfunction with mild impaired lung function.	None	Acetylcysteine 600mg twice daily and Methoxyphenamine 25mg 3 times/day for cough
P-8	F	37	none	Bilateral inguinal lymph nodes were enlarged	Right lachrymal glands and bilateral submandibular glands were enlarged	New organisms were found in the right nasal cavity.	None	NO
P-9	M	39	none	Bilateral cervical and retroauricular lymph nodes and those near bilateral parotid glands were enlarged	None	Scattered patch and ground glass shadow in both lungs. PET-CT showed abnormal FDG uptake in both lungs.	None	NO

P-11	M	55	bronchial asthma	None	Bilateral lachrymal glands were enlarged	None	None	Amlodipine 5mg daily and Irbesartan 150mg daily for hypertension; Metformin 500mg daily and Acarbose 100mg 3 times/day for diabetes
P-17	M	46	none	Left submandibular lymph nodes were enlarged	Sublingual glands and bilateral lachrymal glands were enlarged	Ground glass shadow on pulmonary CT	None	NO
HC1	F	47	NA	NA	NA	NA	NA	NO
HC2	M	54	NA	NA	NA	NA	NA	NO
HC3	M	49	NA	NA	NA	NA	NA	NO
HC4	F	59	NA	NA	NA	NA	NA	NO
HC5	M	50	NA	NA	NA	NA	NA	NO
HC6	F	45	NA	NA	NA	NA	NA	NO
HC7	M	41	NA	NA	NA	NA	NA	NO

Supplementary Table 2 Laboratory examinations of patients with IgG4-RD

No.	IgG4(g/L)	IgG(g/L)	IgA(mg/L)	IgM(mg/L)	IgE(iu/ml)	kap chain (g/L)	lam chain (g/L)	C3 (g/L)	C4 (g/L)	CRP (mg/L)
P-1	6.37	25.4	6370	702	NA	23.8	9.77	0.437	0.0257	24.4
P-2	70.7	52.6	1200	1200	393	36.7	22.6	0.648	0.139	11.1
P-3	7.7	16.5	2040	1160	158	13.2	7.26	0.875	0.261	3.72
P-6	7.79	18.8	1520	683	893	15.1	8.54	0.462	0.029	1.2
P-7	18.2	21	980	871	NA	15.5	10.2	0.738	0.17	NA
P-8	0.246	10.4	1870	1900	NA	NA	NA	0.759	0.202	NA
P-9	17.6	76.9	6550	1140	1030	53.9	27.7	1.01	0.18	18.9
P-11	1.42	11	2050	513	318	8.67	4.46	0.953	0.278	1.68
P-17	6.93	16.4	2130	1450	<5	12.3	8.69	0.787	0.176	1.6

Red: higher than the normal range; blue: lower than the normal range

Supplementary Table 3 Antibodies conjugated with polyadenylated and antibody-specific barcodes

CD14:MPHIP9-CD14-AHS0037-pAbO	CD45RA:HI100-PTPRC-AHS0009-pAbO
CD16:3G8-FCGR3A-AHS0053-pAbO	CXCR5-CXCR5-AHS0039-pAbO
CD4:SK3-CD4-AHS0032-pAbO	CD279:EH12-1-PDCD1-AHS0014-pAbO
CD8:RPA-T8-CD8A-AHS0027-pAbO	CD196-CCR6-AHS0034-pAbO
CD56:NCAM16.2-NCAM1-AHS0019-pAbO	

References

1. Erickson JR, Mair F, Bugos G, et al. AbSeq Protocol Using the Nano-Well Cartridge-Based Rhapsody Platform to Generate Protein and Transcript Expression Data on the Single-Cell Level. *STAR Protoc* 2020;1:100092.
2. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14:979-82.
3. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083-6.
4. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;15:1484-506.
5. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021;12:1088.