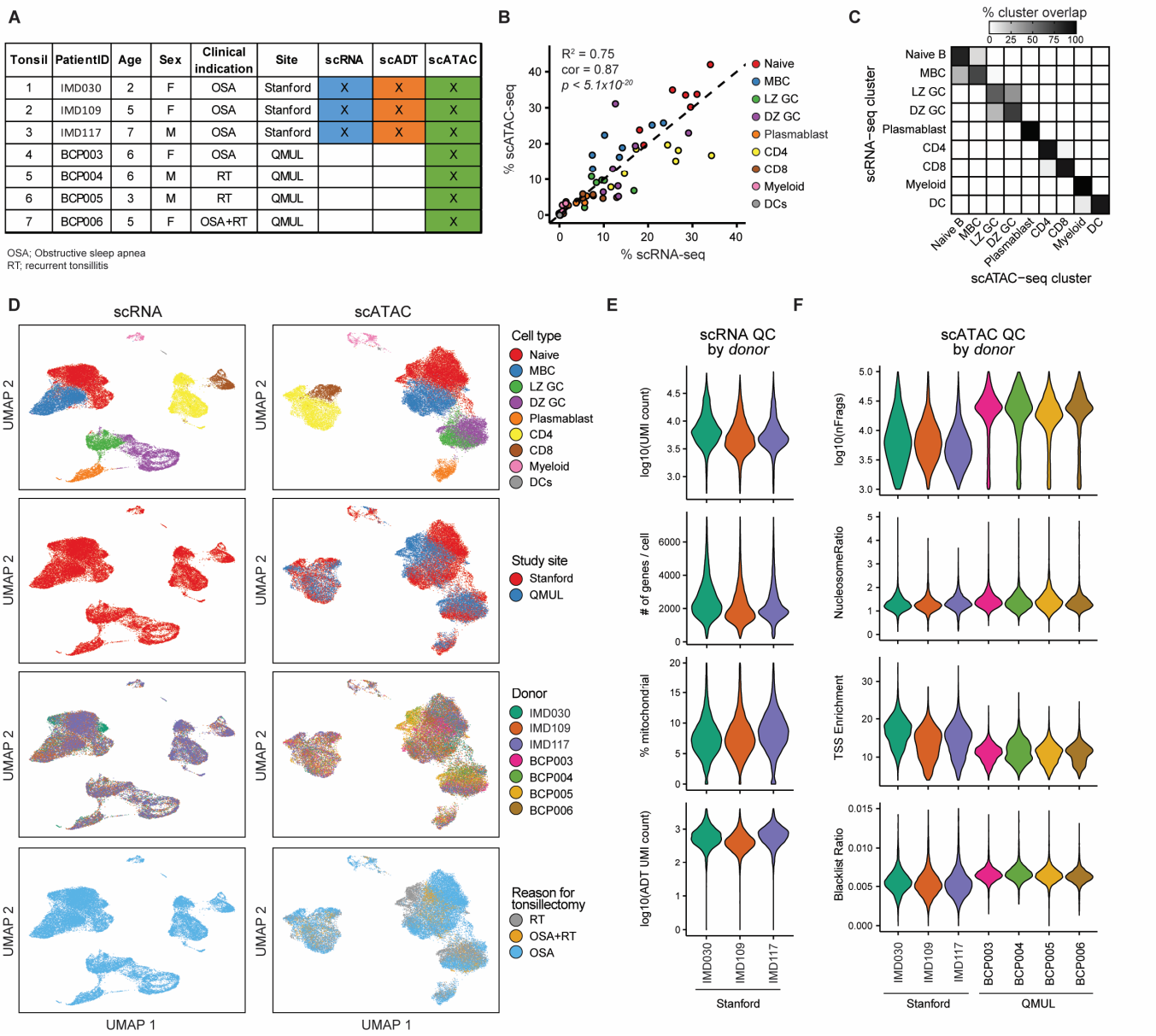

Tonsil	PatientID	Age	Sex	Clinical indication	Site	scRNA	scADT	scATAC
1	IMD030	2	F	OSA	Stanford	X	X	X
2	IMD109	5	F	OSA	Stanford	X	X	X
3	IMD117	7	M	OSA	Stanford	X	X	X
4	BCP003	6	F	OSA	QMUL			X
5	BCP004	6	M	RT	QMUL			X
6	BCP005	3	M	RT	QMUL			X
7	BCP006	5	F	OSA+RT	QMUL			X

Figure S1. Single-cell library metadata, integration, batch correction and quality control.

A) Metadata of tonsillectomy patients analyzed by single-cell genomics in this study, including age, sex, reason for tonsillectomy (RT; recurrent tonsillitis, OSA; obstructive sleep apnea) and site of study.

B) Correlation of relative cluster frequencies for scRNA-seq and scATAC-seq within each donor.

C) Confusion matrix depicting overlap between transferred scRNA-seq cluster identities to scATAC-seq clusters.

D) UMAP visualization of major cell type clusters, site of study, donor and clinical indication for tonsillectomy for scRNA-seq and scATAC-seq datasets

E) Quality control metrics for scRNA-seq datasets by donor, including unique molecular identifier (UMI) counts per cell barcode, number of genes detected per cell barcode, percentage mitochondrial gene expression and cell surface ADT UMI counts.

F) Quality control metrics for scATAC-seq datasets by donor, including number of unique fragments per cell barcode, ratio of nucleosomal to non-nucleosomal fragment sizes, transcription start site enrichment score, and ratio of fragments in genomic blacklist regions (see ArchR for details).

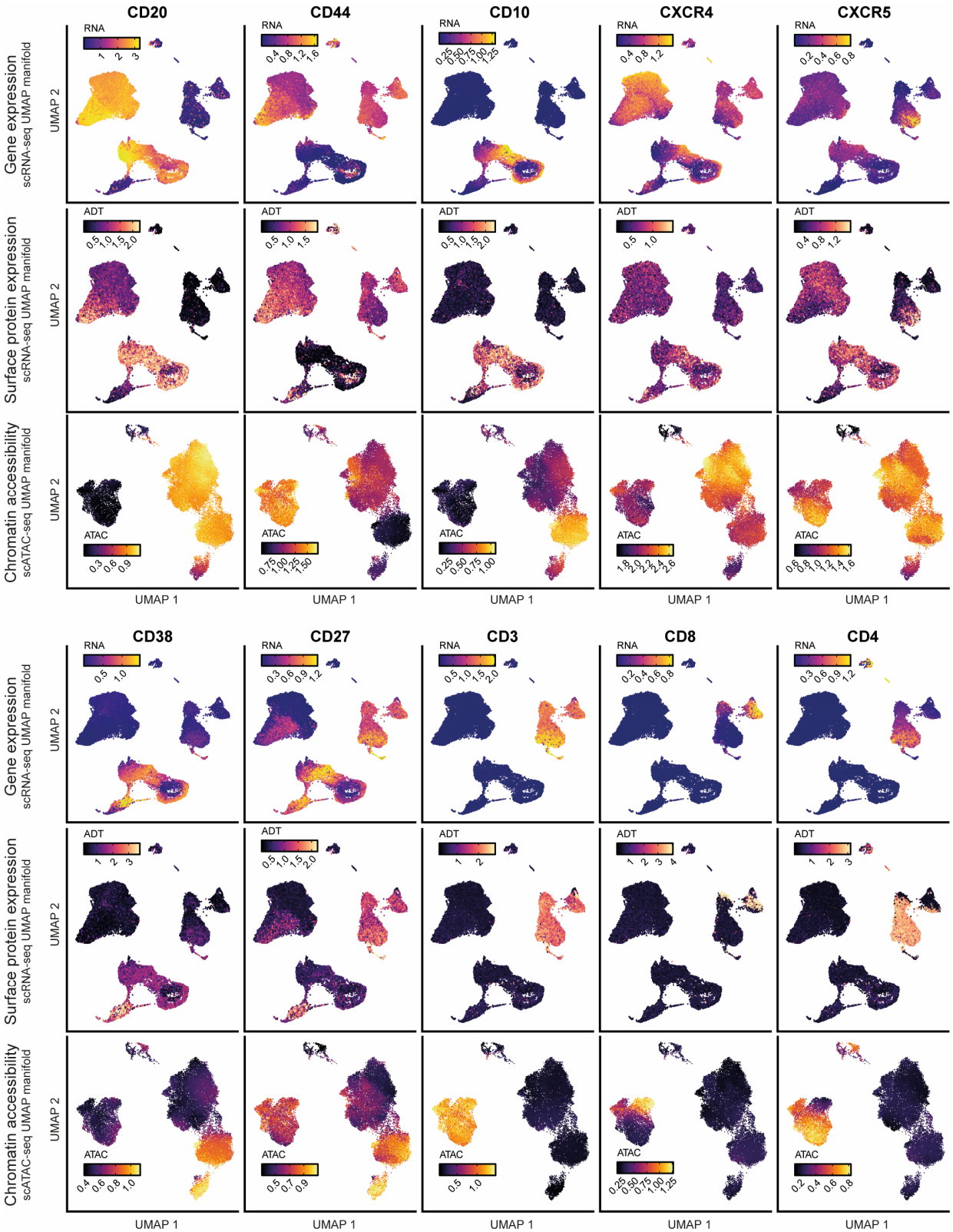

Figure S2. Comparison of RNA expression, cell surface protein expression and chromatin accessibility of key marker genes.

scRNA-seq gene expression (top rows), CITE-seq surface protein expression (middle rows) and chromatin accessibility gene scores (bottom rows) for key marker genes. Gene expression and surface protein expression are visualized on the scRNA-seq UMAP manifold and chromatin accessibility scores are visualized on the scATAC-seq UMAP manifold (see Figure 1b).

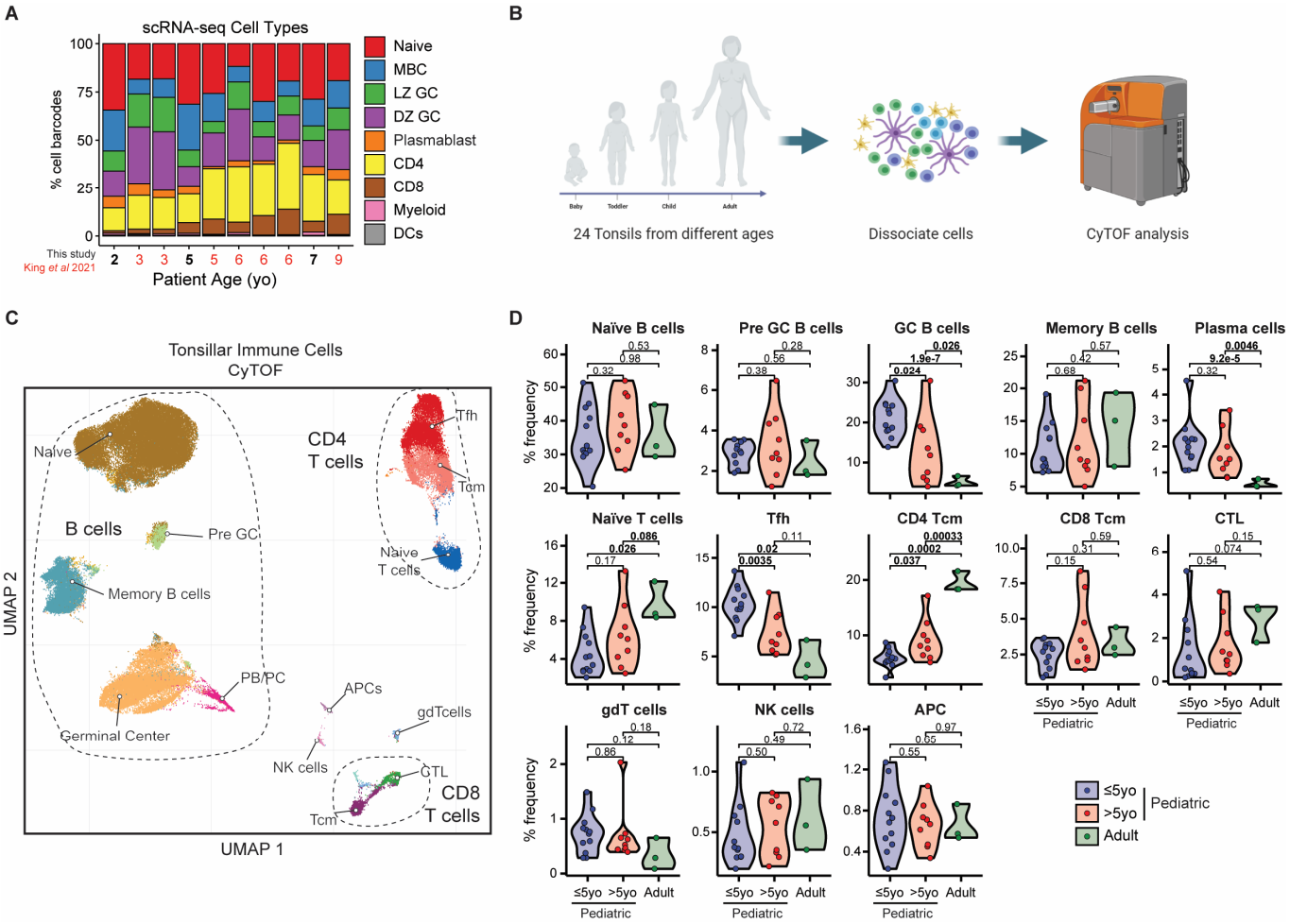

Figure S3. Age-related changes in tonsillar immune cell populations by scRNA-seq and CyTOF.

A) Relative scRNA-seq cluster frequencies of different donors ordered by patient age. Additional tonsillar immune cell samples from King *et al.* (2021) are included (see red labels).

B) Schematic of CyTOF analyses for age-related differences in tonsillar immune cell populations.

C) UMAP visualization of CyTOF analysis of tonsillar immune cells with major immune cell populations (n=24).

D) Quantitation of relative frequencies of immune cell subsets separated by patient age. p values denote results from Student's t test.

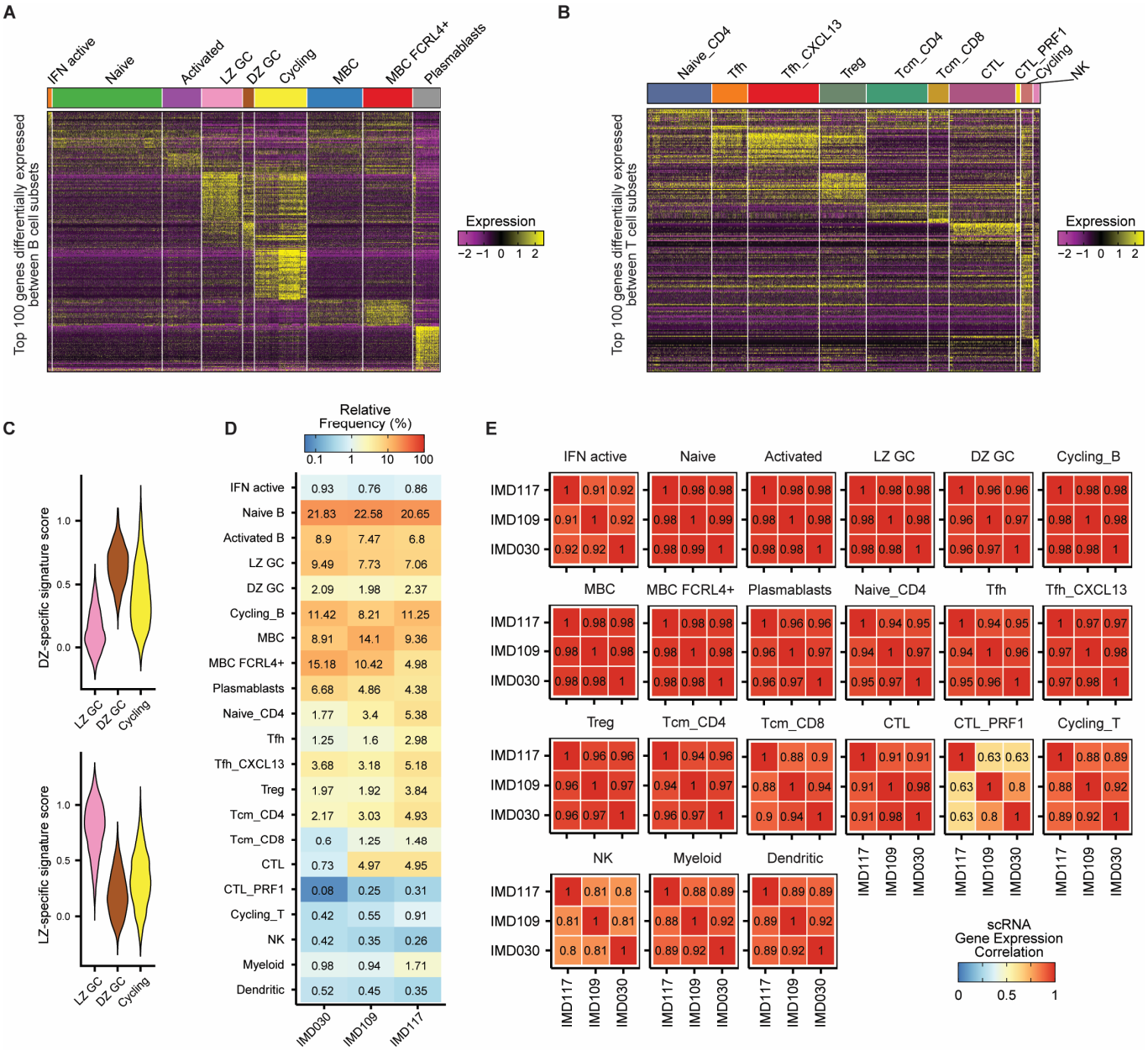

Figure S4. Tonsil scRNA-seq marker gene heatmaps, annotation of GC B cells and reproducibility of cell types across donors.

A) scRNA-seq heatmap of top 100 most differentially expressed genes for high resolution B cell clusters (as in Fig1F-G).
B) scRNA-seq heatmap of top 100 most differentially expressed genes for high resolution T cell clusters (as in Fig1D-E).
C) Gene signature scores for DZ-specific and LZ-specific marker genes in LZ GC, DZ GC and cycling B cell clusters.
D) Relative frequency of high resolution scRNA-seq cell type clusters across the three patient donors.
E) Spearman correlation coefficients of mean gene expression per cell type cluster between patient donors.

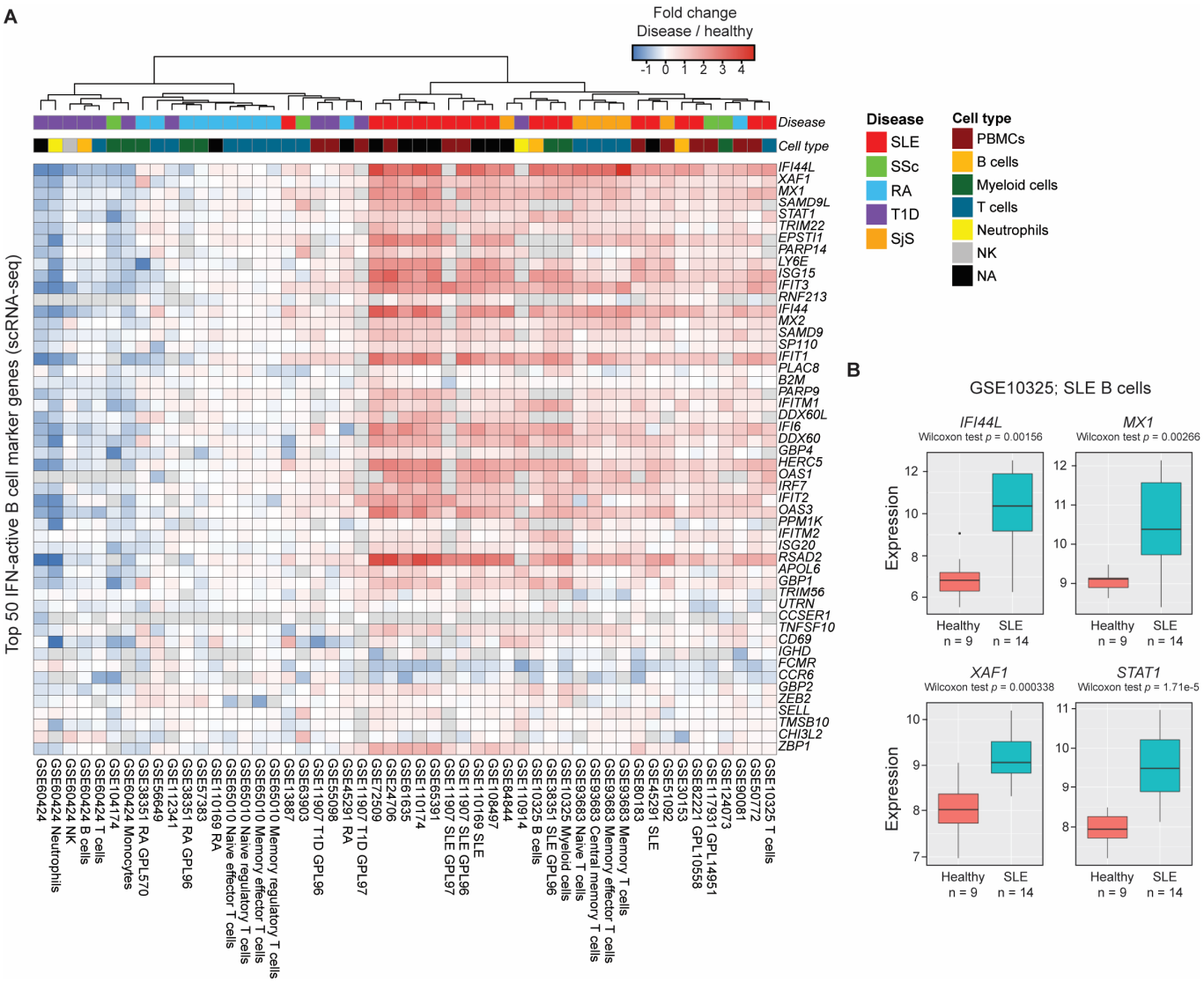

Figure S5. Differential expression in autoimmune disease of top scRNA-seq marker genes for the IFN_active B cell cluster.

A) Differential expression of top 50 marker genes for IFN-active B cell cluster (as in Fig1F-G) between control and autoimmune patient gene expression studies, generated from ADEx: Autoimmune Diseases Explorer (https://adex.genyo.es/; (*11*)). SLE; systemic lupus erythematosus. SSc; systemic sclerosis. RA; rheumatoid arthritis. T1D; type I diabetes. SjS; Sjögren's syndrome.

B) Expression of *IFI44L*, *MX1*, *XAF1* and *STAT1* in peripheral blood-derived B cells healthy and systemic lupus erythematosus (SLE) patients (*12*).

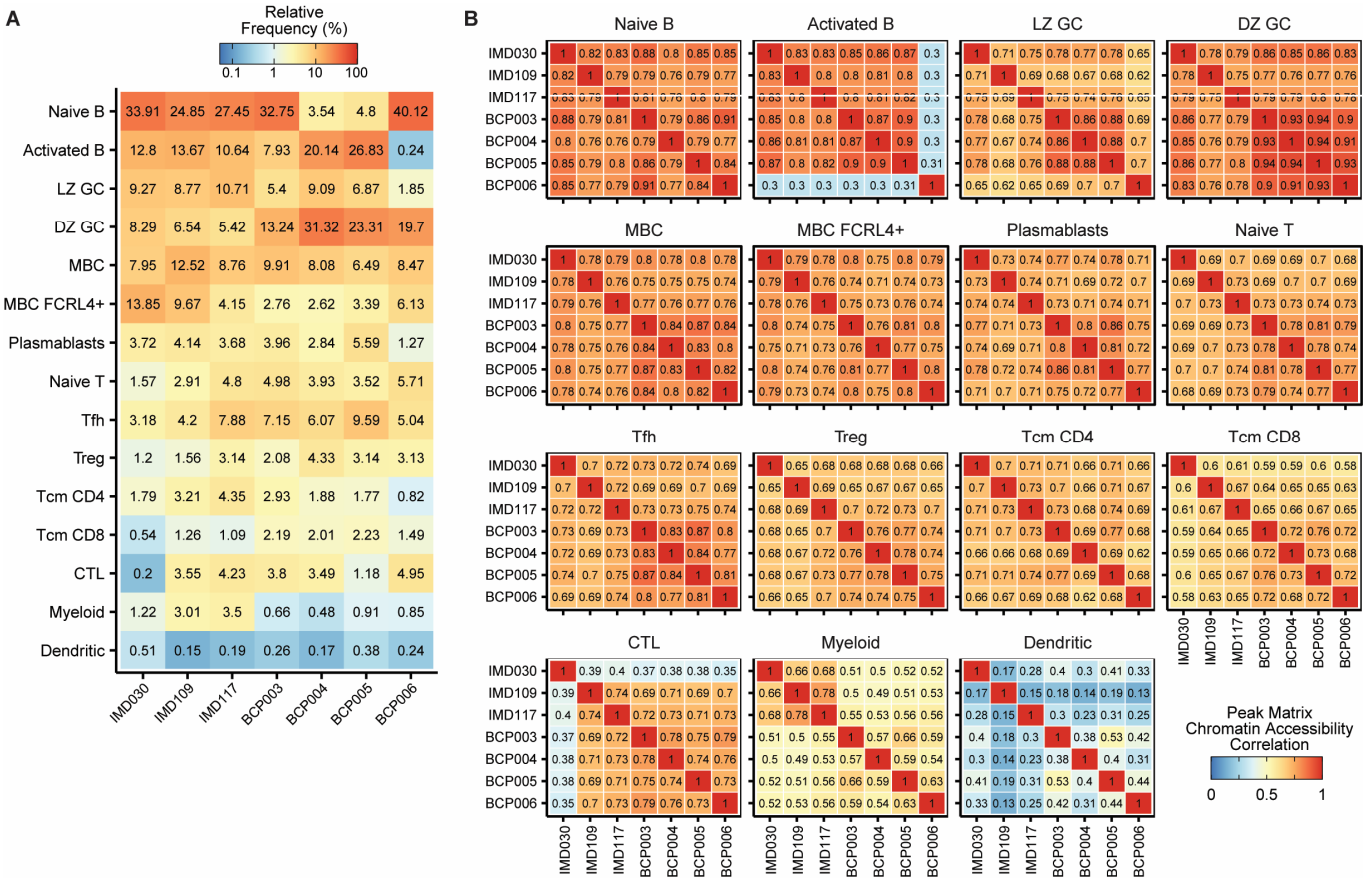

Figure S6. Reproducibility of scATAC-seq cluster frequencies and correlation of peak accessibilities between donors.

A) Relative frequency of high resolution scATAC-seq cell type clusters across patient donors.

B) Spearman correlation coefficients of mean peak accessibility per cell type cluster between patient donors.

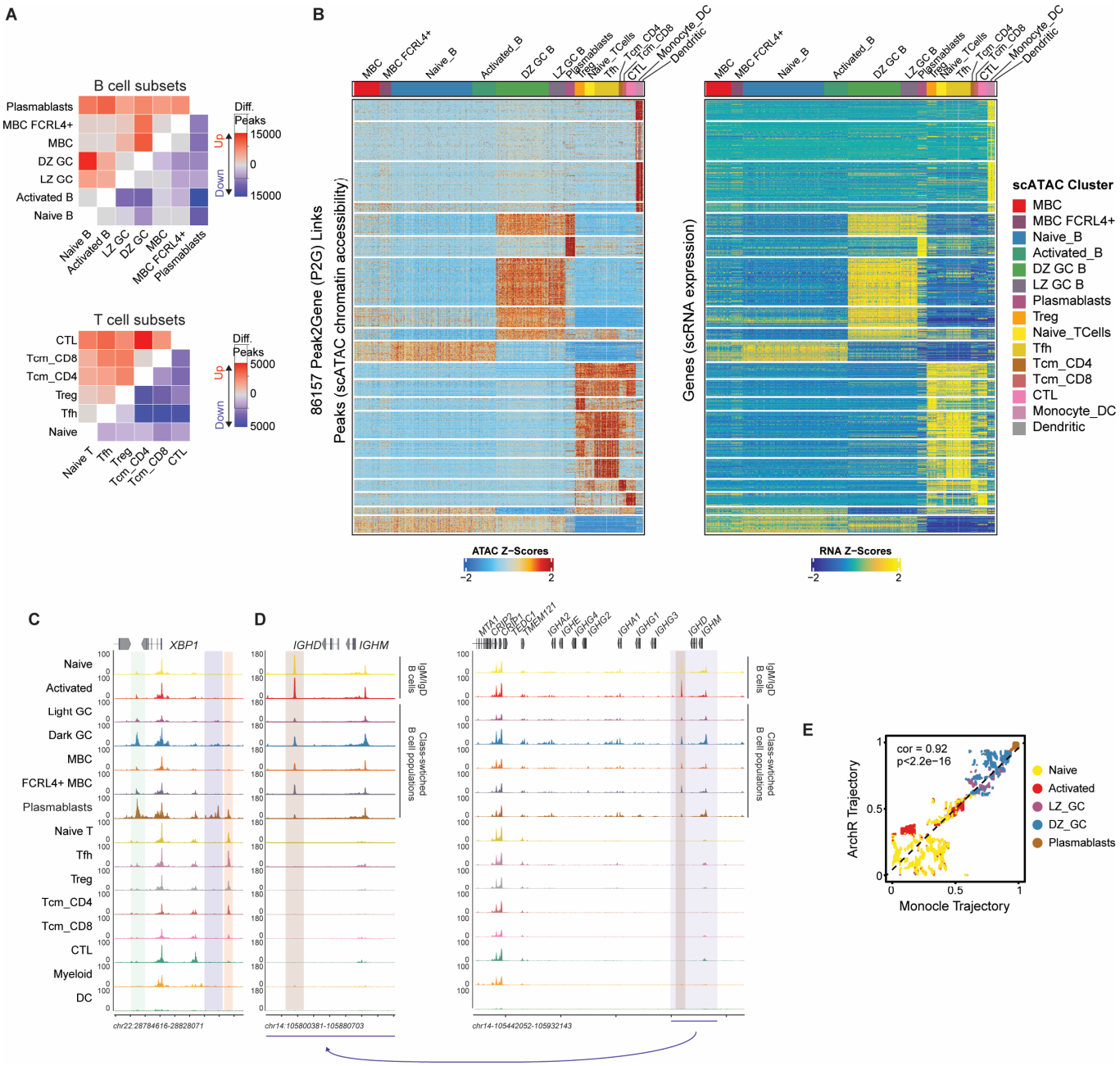

Figure S7. Differential peak analysis of scATAC-seq clusters, peak-to-gene predictions and alternative pseudotime analysis.

A) Differential peak analysis of B (top) and T (bottom) cell subsets comparing number of up-regulated and down-regulated chromatin accessibility regions.

B) Single-cell heatmaps of marker peak accessibility (scATAC) and gene expression (scRNA) for integrative peak2gene predictions in tonsil immune cell type clusters (as in Fig2A).

C) Example genome snapshot of *XBP1* regulatory landscape. B cell-, plasmablast- and T cell-specific regulatory elements are highlighted.

D) Genome snapshot of immunoglobulin heavy chain locus, including closer resolution of regulatory element downstream of *IGHD* and *IGHM* that is lost during class switch recombination (i.e. through deletional recombination).

E) Correlation between two independent pseudotemporal ordering methods (ArchR and Monocle) for naïve, activated, GC and plasmablast B cell lineage. Correlation coefficient and *p* value denotes result from Pearson correlation.

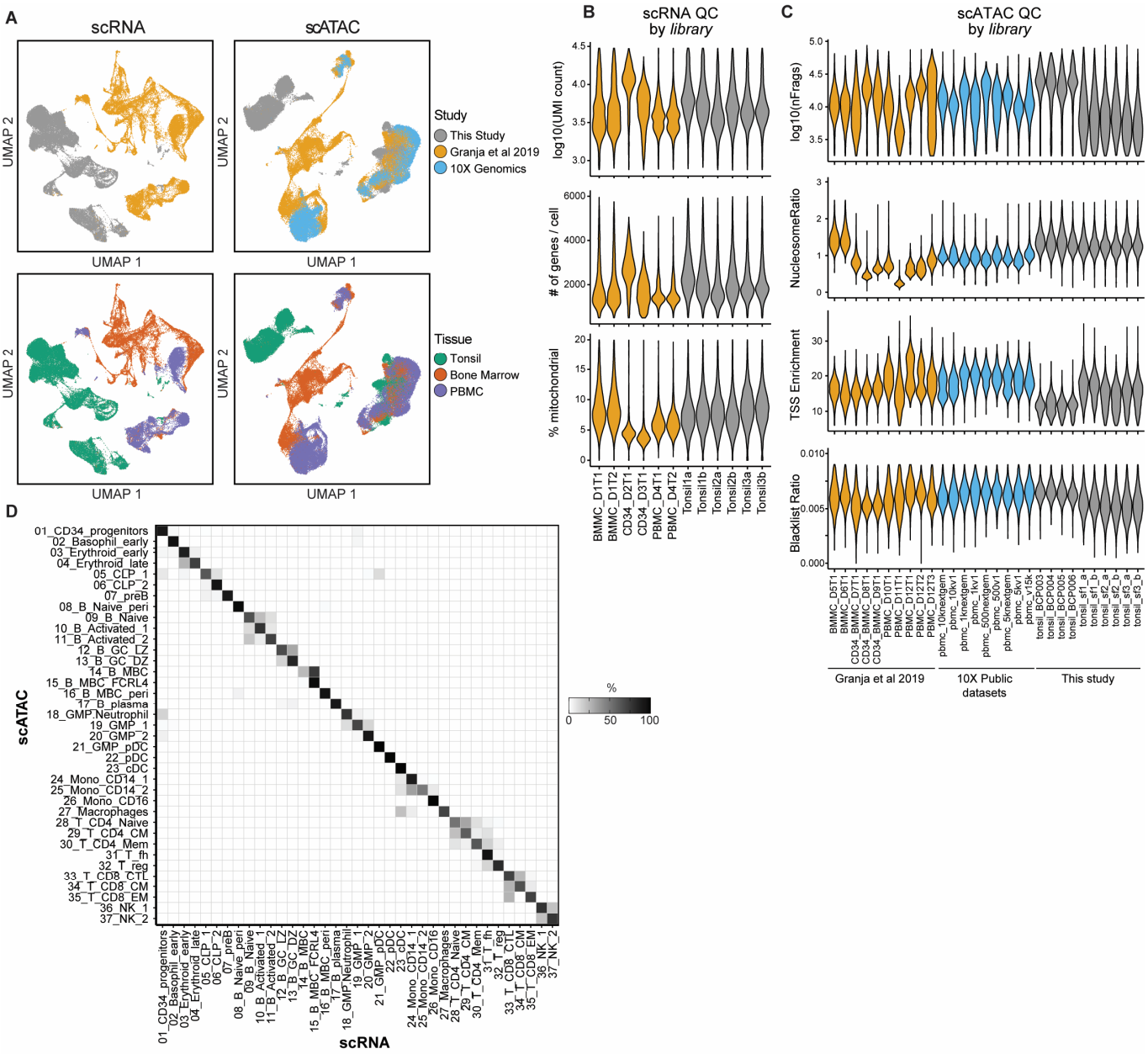

Figure S8. Batch correction and quality control of integrated bone marrow, blood and tonsil immune scRNA-seq and scATAC-seq.

A) UMAP visualization of batch and tissue for tonsil, peripheral blood and bone marrow scRNA-seq and scATAC-seq datasets.
B) Quality control metrics for scRNA-seq datasets by donor, including unique molecular identifier (UMI) counts per cell barcode, number of genes detected per cell barcode, and percentage of mitochondrial gene expression.
C) Quality control metrics for scATAC-seq datasets by donor, including number of unique fragments per cell barcode, ratio of nucleosomal to non-nucleosomal fragment sizes, transcription start site enrichment score, and ratio of fragments in genomic blacklist regions (see ArchR for details).
D) Confusion matrix depicting overlap between transferred scRNA-seq cluster identities to scATAC-seq clusters.

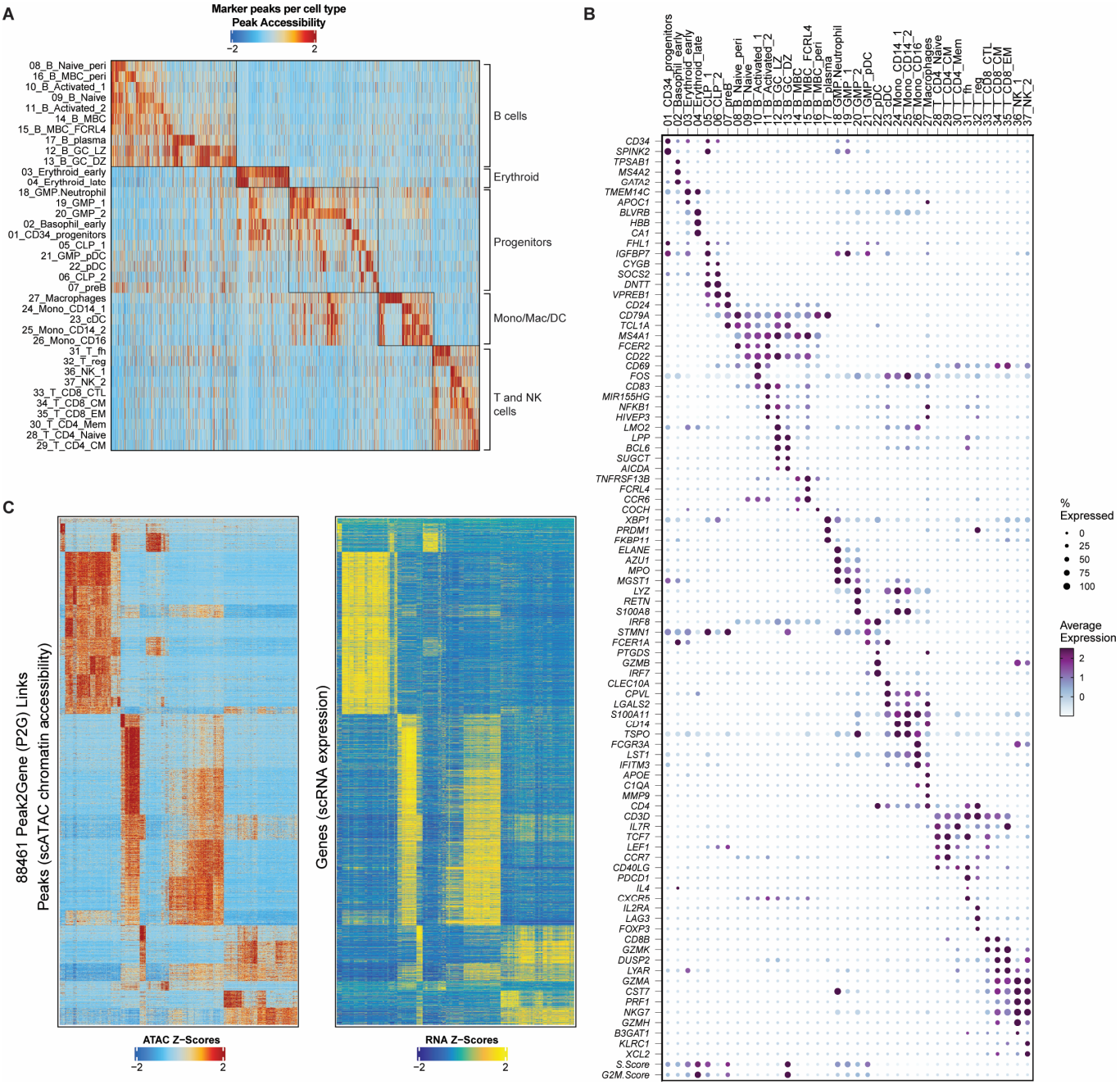

Figure S9. Integrated bone marrow, blood and tonsil scRNA-seq and scATAC-seq markers.

A) Expression of top marker genes for scRNA-seq clusters of integrated bone marrow, blood and tonsil dataset.

B) Chromatin accessibility at cluster-specific peaks for scATAC-seq clusters of integrated bone marrow, blood and tonsil dataset.

C) Peak accessibility (scATAC) and gene expression (scRNA) for integrative peak2gene predictions in tonsil, peripheral blood and bone marrow immune cell type clusters (as in Fig3A).

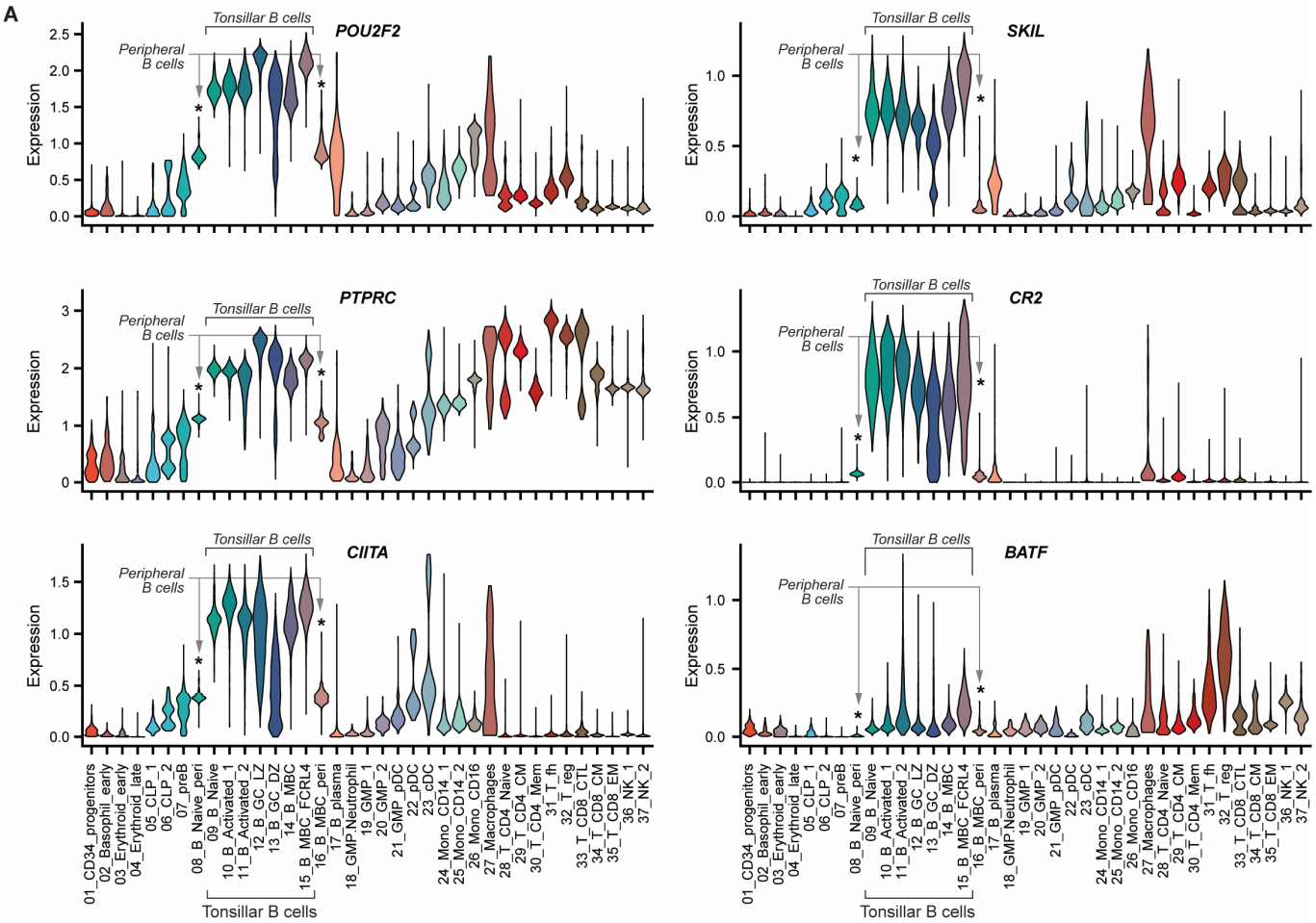

Figure S10. Tonsil B cell-enriched gene expression markers compared to peripheral blood B cells.

A) Expression of genes significantly differentially expressed between tonsil-specific naïve or memory B cell clusters compared to peripheral blood naïve or memory B cell clusters.

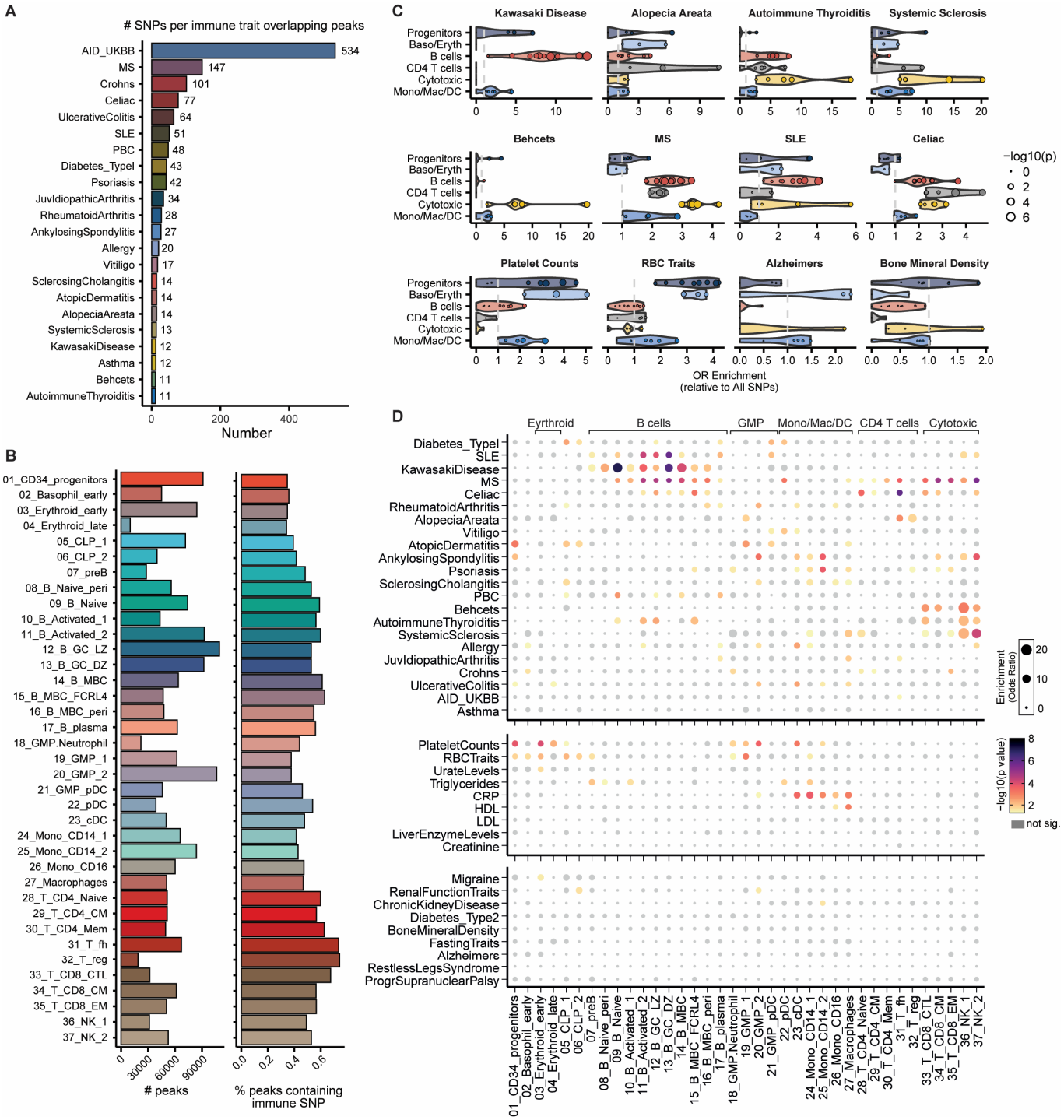

Figure S11. Enrichment of fine-mapped autoimmune variants in immune cell subsets.

A) Number of fine-mapped SNPs per autoimmune trait that overlap with a chromatin accessibility peak in the integrated human bone marrow, peripheral blood and tonsil scATAC-seq atlas. AID_UKBB represents variants identified from fine-mapping a combination of datasets from diverse autoimmune traits. AID; autoimmune disease. MS; multiple sclerosis. SLE; systemic lupus erythematosus. PBC; primary biliary cirrhosis.

B) Number of peaks identified in each scATAC-seq cell type cluster (left) and the percentage of those peaks that overlap with an autoimmune-associated SNP.

C) Fisher enrichment test results for variants specific to selected traits in cell type-specific chromatin. Individual points represent single cell type clusters, separated into five broad lineages. Dot size reflects level of significance of enrichment.

D) Fisher enrichment test results for trait-specific variants in cell type-specific chromatin across all traits in fine-mapped resources analyzed. Dot size conveys enrichment (odds ratio) and color denotes significance of enrichment.

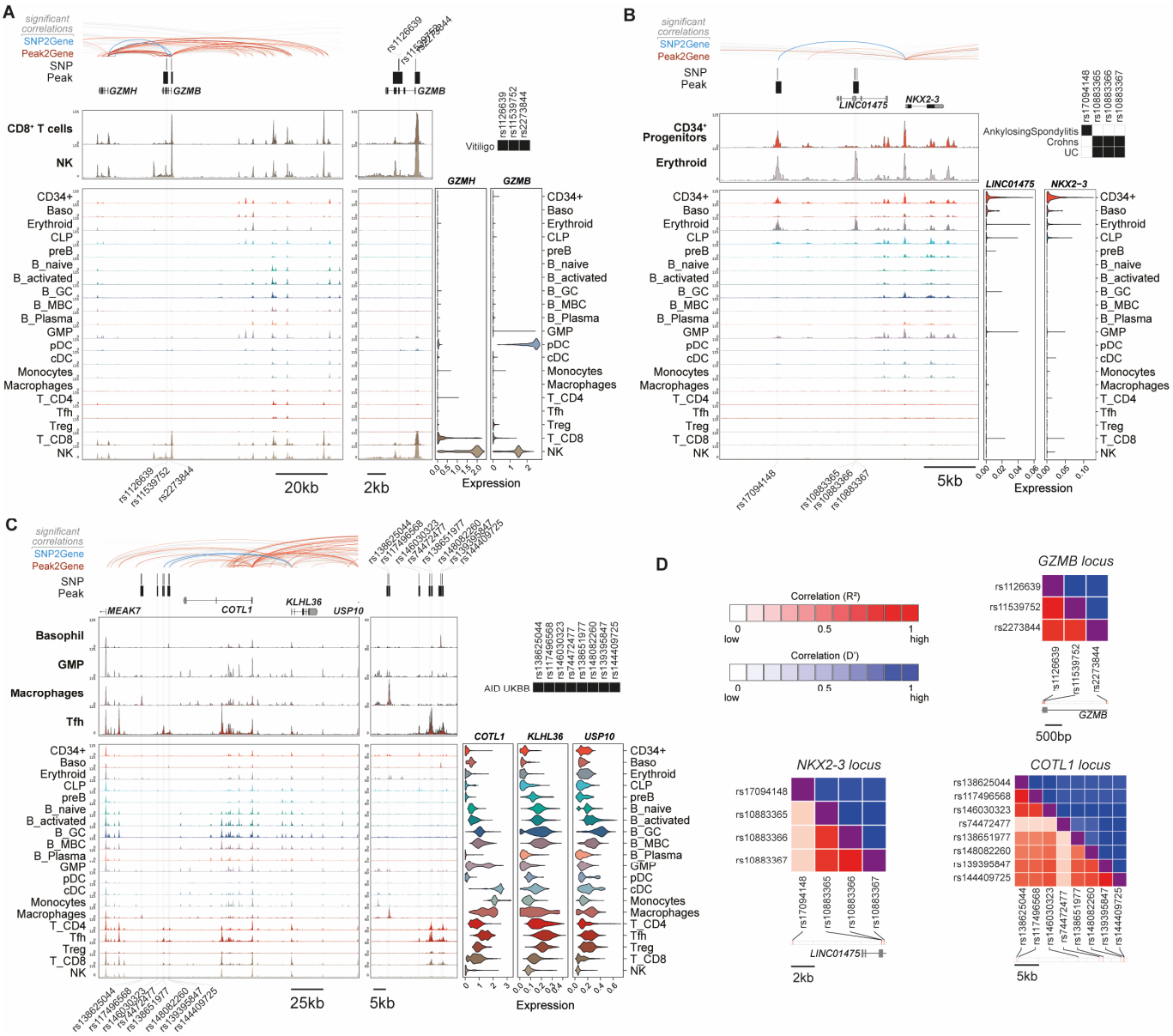

Figure S12. Genome snapshots of fine-mapped autoimmune variants at *GZMB*/*GZMH*, *NKX2-3* and *COTL1*/*KHLH36* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *GZMB*/*GZMH* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks.

B) Same as A), at the *NKX2-3* locus. UC; ulcerative colitis.

C) Same as A), at the *COTL1*/*KHLH36* locus. AID; autoimmune disease.

D) Linkage disequilibrium heatmaps for SNPs at loci depicted in A-C. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

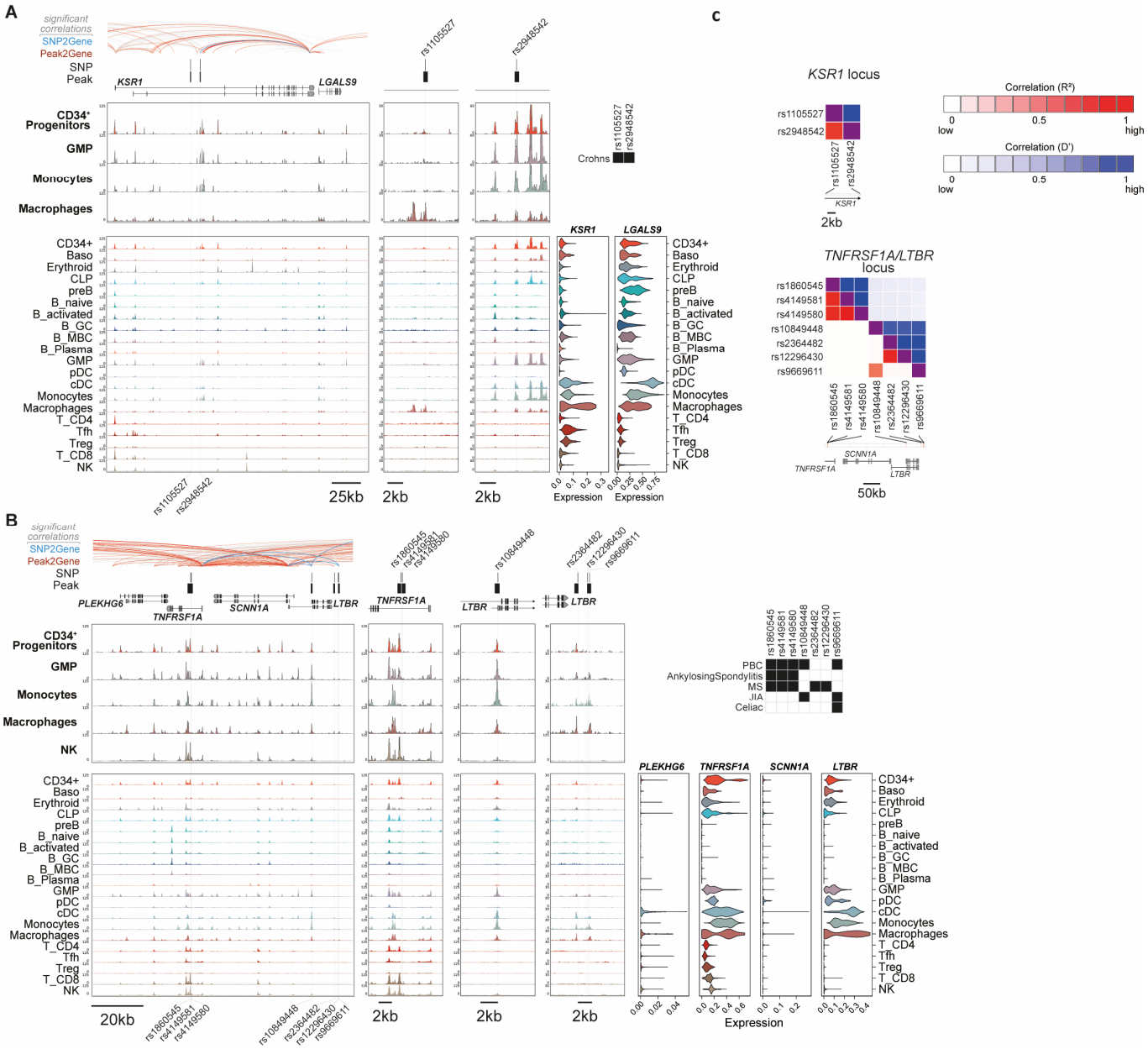

Figure S13. Genome snapshots of fine-mapped autoimmune variants at *KSR1*/*LGALS9* and *TNFRSF1A*/*LTBR* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *KSR1*/*LGALS9* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks.

B) Same as A), at the *TNFRSF1A/LTBR* locus. PBC; primary biliary cirrhosis. MS; multiple sclerosis. JIA; juvenile idiopathic arthritis.

C) Linkage disequilibrium heatmaps for SNPs at loci depicted in A and B. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

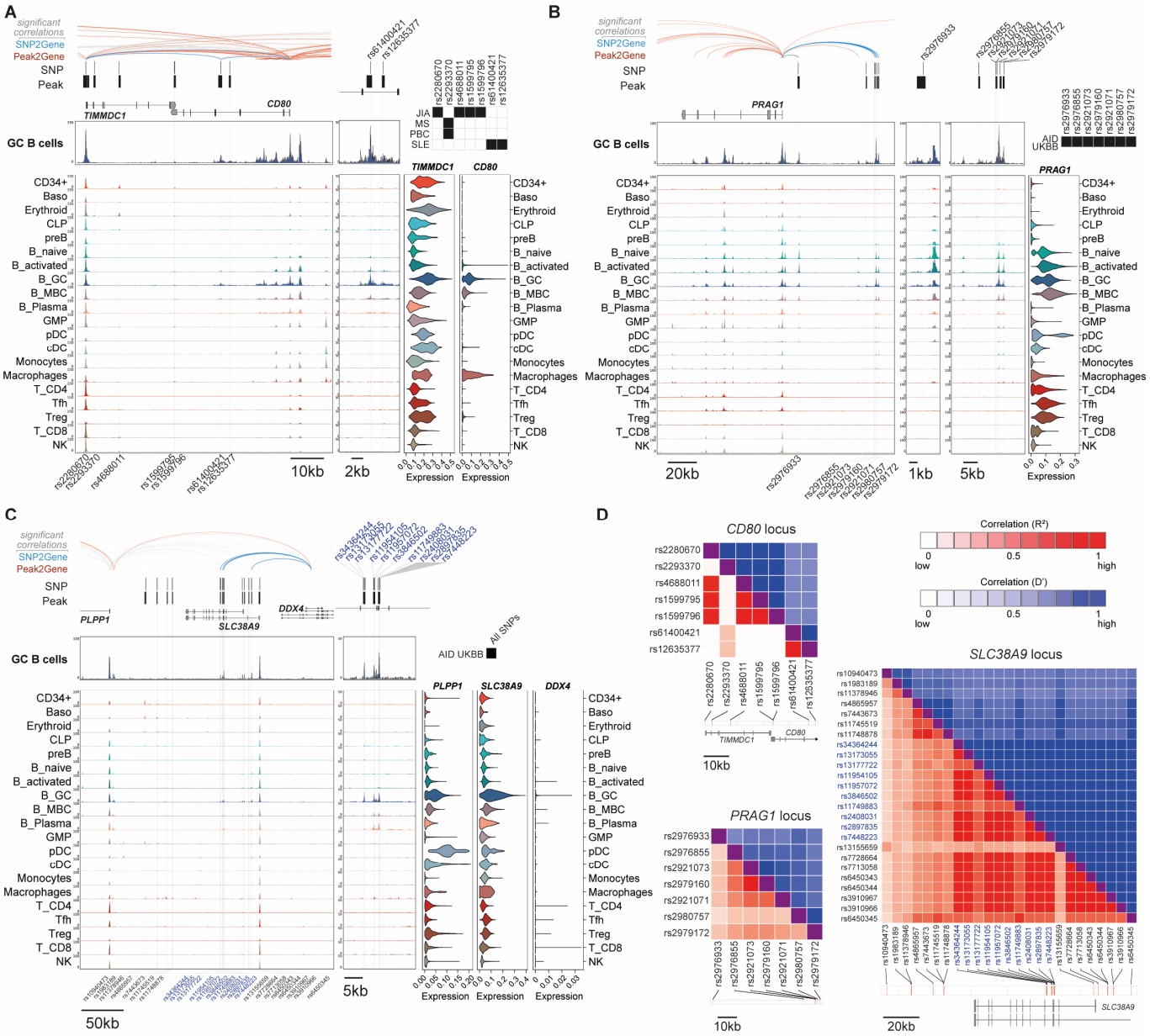

Figure S14. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *CD80*, *PRAG1* and *SLC38A9*/*DDX4* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *CD80* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. PBC; primary biliary cirrhosis. MS; multiple sclerosis. JIA; juvenile idiopathic arthritis. SLE; systemic lupus erythematosus.

B) Same as A), at the *PRAG1* locus. AID; autoimmune disease.

C) Same as A), at the *SLC38A9*/*DDX4* locus.

D) Linkage disequilibrium heatmaps for SNPs at loci depicted in A and B. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

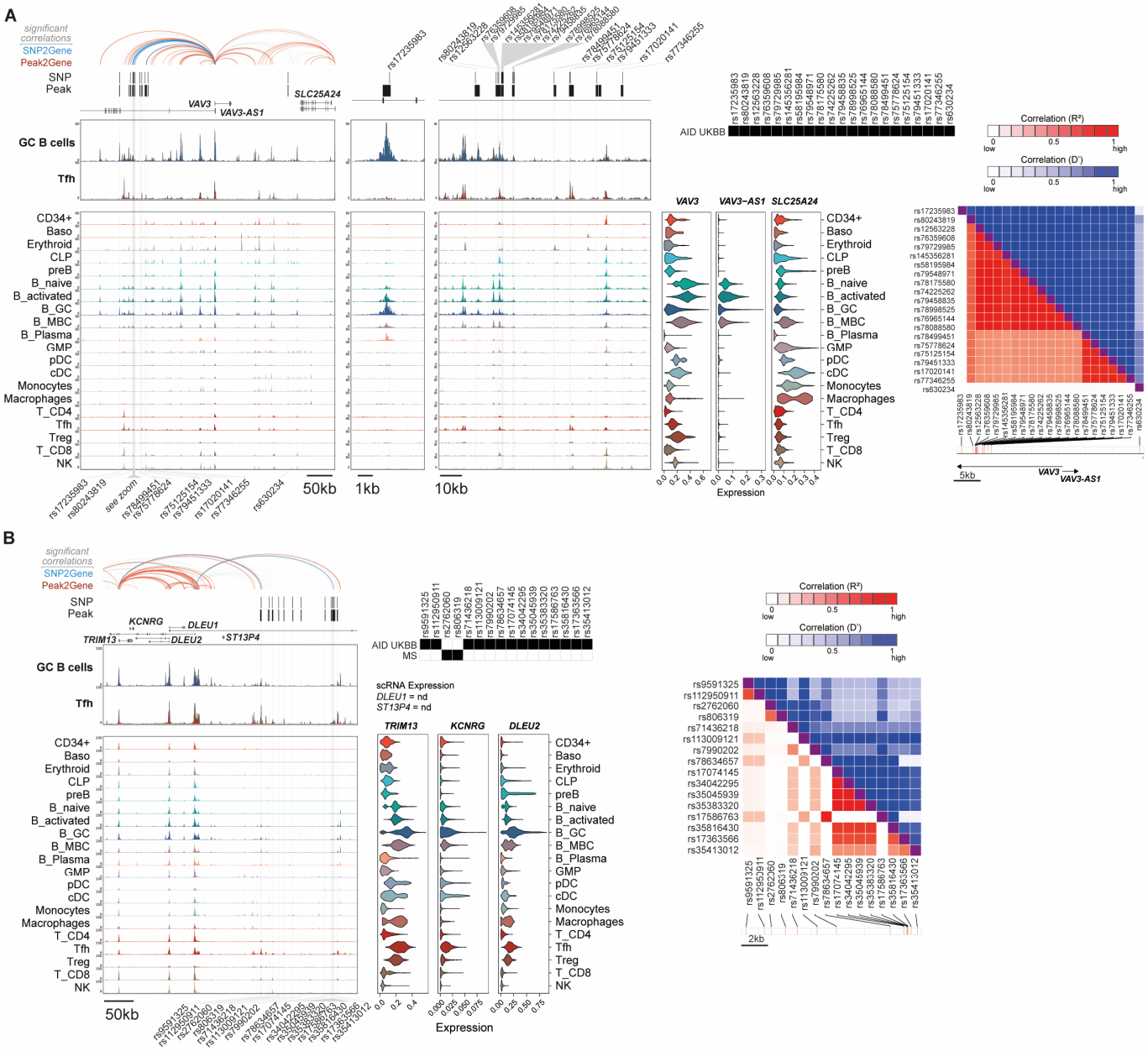

Figure S15. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *VAV3* and *DLEU2* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *VAV3* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmaps are also shown separately. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. AID; autoimmune disease.

B) Same as A), at the *DLEU2* locus. MS; multiple sclerosis.

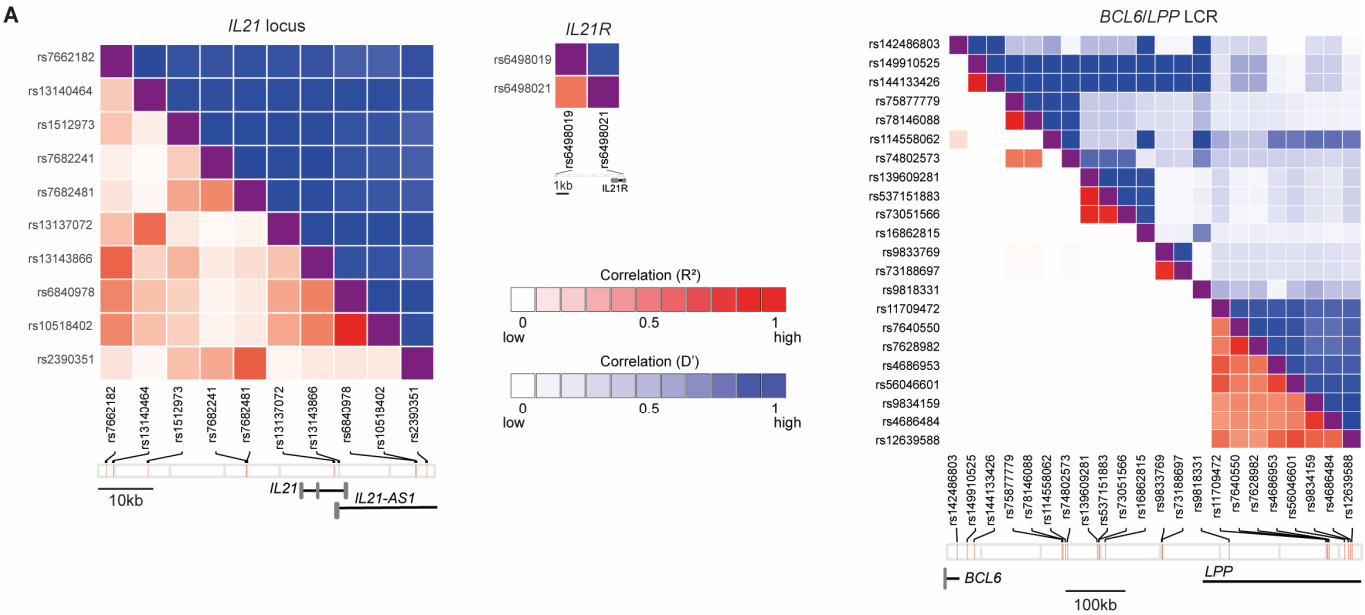

Figure S16. Linkage disequilibrium scores for variants at *IL21*, *IL21R* and *BCL6* loci.

A) Linkage disequilibrium heatmaps for SNPs at *IL21*, *IL21R* and *BCL6*/*LPP* loci depicted in Fig5. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

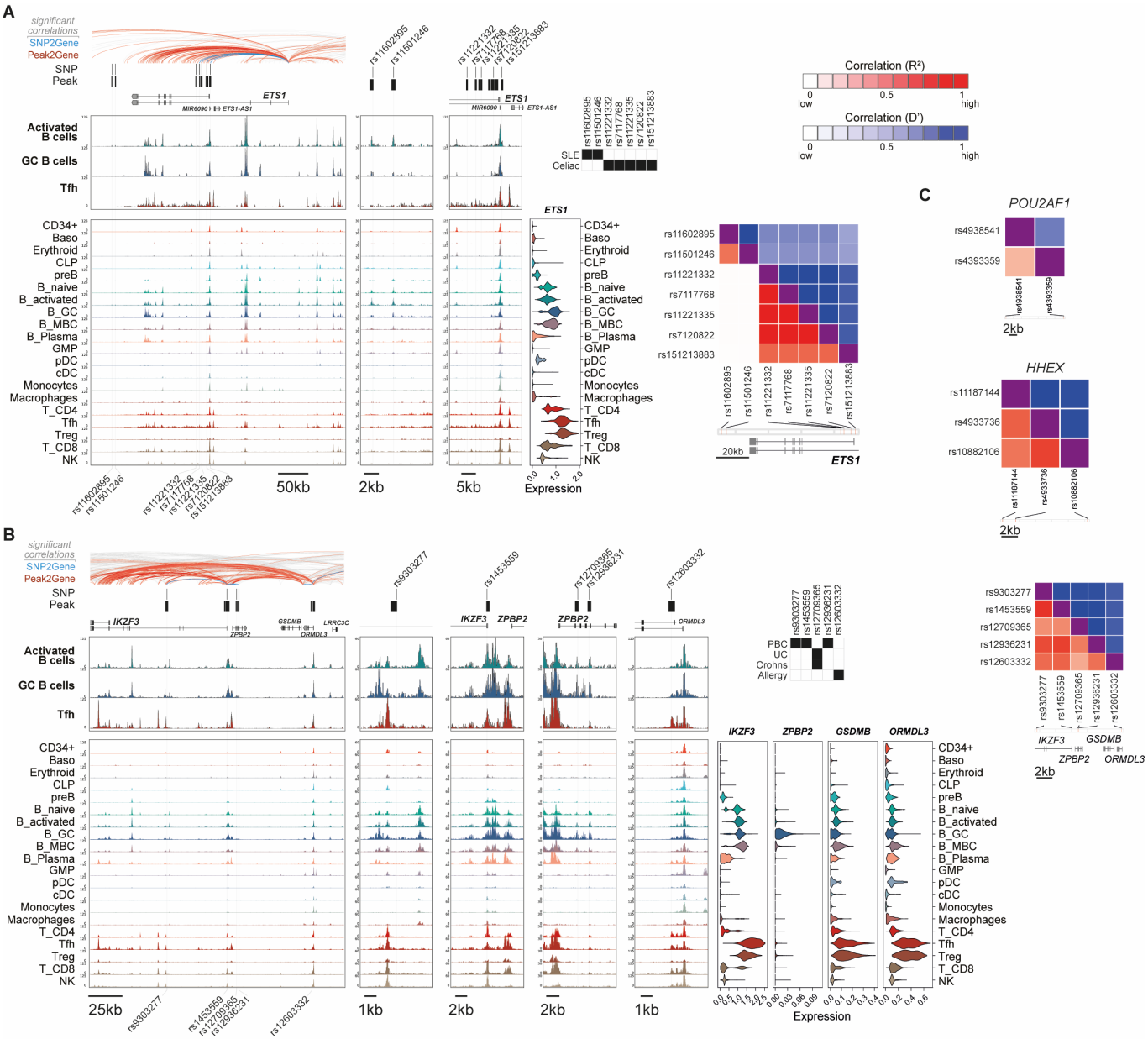

Figure S17. Genome snapshots of fine-mapped autoimmune variants at *ETS1* and *IKZF3* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *ETS1* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmap is also shown. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. SLE; systemic lupus erythematosus.

B) Same as A), at the *IKZF3* locus. PBC; primary biliary cirrhosis. UC; ulcerative colitis.

C) Linkage disequilibrium heatmaps for SNPs at *POU2AF1* and *HHEX* loci depicted in Fig6. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

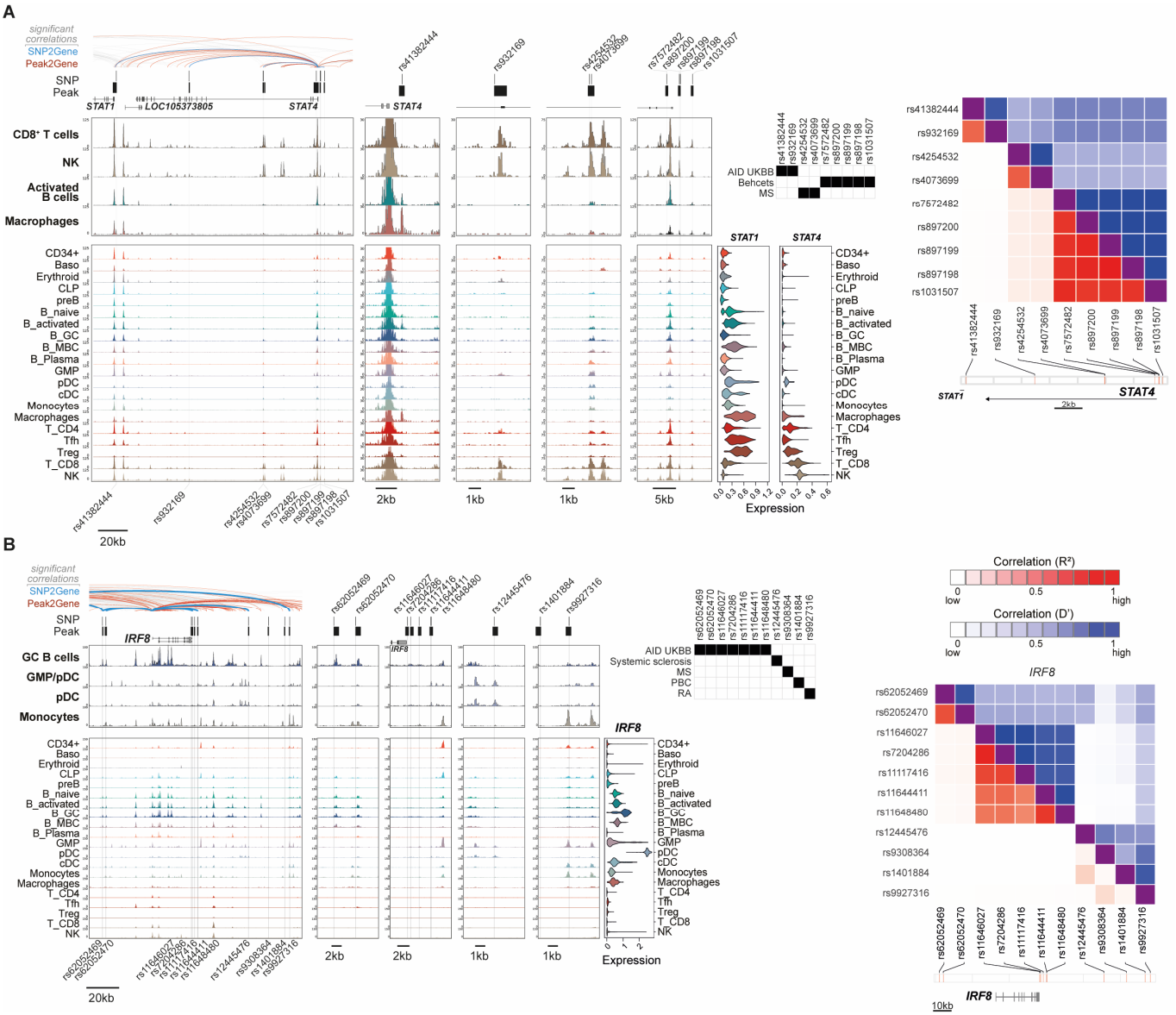

Figure S18. Genome snapshots of fine-mapped autoimmune variants at *STAT4* and *IRF8* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *STAT4* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmap is also shown. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. AID; autoimmune disease. MS; multiple sclerosis.

B) Same as A), at the *IRF8* locus. PBC; primary biliary cirrhosis. RA; rheumatoid arthritis.

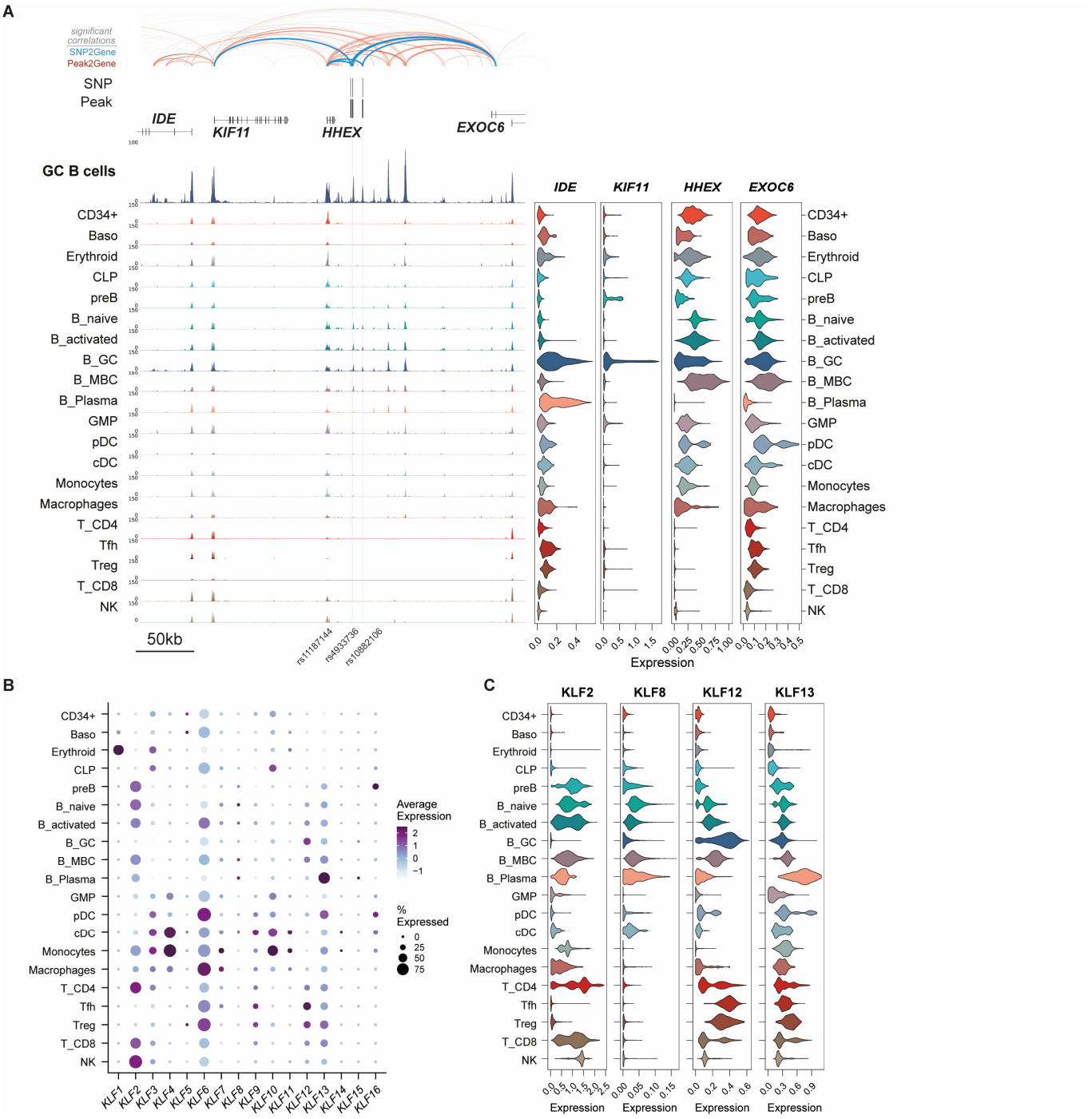

Figure S19. Genomic landscape at *HHEX* and expression of KLF family transcription factors.

A) Broader view of the *HHEX* locus (see Fig6b), showing significantly correlated peak2gene links and gene expression of neighboring genes *KIF11* and *EXOC6*.

B) Mean expression of all KLF family transcription factors detected in scRNA-seq dataset. Dot size denotes percent of cluster in which gene is detected.

C) Single-cell expression of *KLF2*, *KLF8, KLF12* and *KLF13*, with highest expression in B cell subsets.

Table S1. CyTOF phenotyping antibody panel.

metal	marker	clone name	clone source
Y	CD45(1)	HI30	DVS
Pd	CD45(2)	HI30	Biolegend
Pd	CD45(3)	HI30	Biolegend
Pd	CD45(4)	HI30	Biolegend
In	CD45(5)	HI30	Biolegend
In	CD57	HCD57	Biolegend
La	CD16	3G8	Biolegend
Pr			
Nd	CD45RO	UCHL1	Biolegend
Nd			
Nd			
Nd	CD33	WM53	Biolegend
Nd	CD14	M5E2	Biolegend
Nd			
Sm	CD20	2H7	Biolegend
Nd			
Sm	CD25	2A3	Fluidigm
Nd	**CD138**	**DL-101**	**Fluidigm**
Eu			
Sm	CD103	B-Ly7	eBioscience
Eu			
Sm	CD27	LG.7F9	eBioscience
Gd	CD56	NCAM16.2	BD
Gd	CD3	UCHT1	Biolegend
Gd	CD19	HIB19	Biolegend
Gd	CD45RA	HI100	Biolegend
Tb	CXCR5	RF8B2	BD
Gd	CD28	CD28.2	Biolegend
Dy	CD38	HIT2	Biolegend
Dy	**CD8**	**RPA-T8**	**Fluidigm**
Dy			
Dy			
Ho	**CD40**	**5C3**	**Fluidigm**
Er			
Er	int b7	FIB504	Biolegend
Er	CCR7	G043H7	Biolegend
Tm	IgD	IA6-2	Biolegend
Er	IgA	G18-1	BD
Yb	HLA-DR	L243	Biolegend
Yb	**IgM**	**MHM-88**	**Fluidigm**
Yb	TCRgd	5A6.E9	in-house
Yb	**CD4**	**SK3**	**Fluidigm**
Lu	PD-1	EH12.2H7	Fluidigm
Yb	CD127	A019D5	Biolegend
Ir	DNA1	-	Fluidigm
Ir	DNA2	-	Fluidigm
Pt	cisplatin l/d	-	Fluidigm

Table S2. CITE-seq antibody details.

Target	Antibody	Clone	Catalog#
CD3	TotalSeq™-A0034 anti-human CD3 Antibody	UCHT1	300475
CD4	TotalSeq™-A0072 anti-human CD4 Antibody	RPA-T4	300563
CD8a	TotalSeq™-A0080 anti-human CD8a Antibody	RPA-T8	301067
CD20	TotalSeq™-A0100 anti-human CD20 Antibody	2H7	302359
CD27	TotalSeq™-A0154 anti-human CD27 Antibody	O323	302847
CD38	TotalSeq™-A0389 anti-human CD38 Antibody	HIT2	303541
CD10	TotalSeq™-A0062 anti-human CD10 Antibody	HI10a	312231
CXCR4	TotalSeq™-A0366 anti-human CD184 (CXCR4) Antibody	12G5	306531
CXCR5	TotalSeq™-A0144 anti-human CD185 (CXCR5) Antibody	J252D4	356937
CD44	TotalSeq™-A0125 anti-human CD44 Antibody	BJ18	338825
IgD	TotalSeq™-A0384 anti-human IgD Antibody	IA6-2	348243
IgM	TotalSeq™-A0136 anti-human IgM Antibody	MHM-88	314541