

Supplemental information

Axes of biological variation in diffuse large B cell lymphoma

Boya Wang, George Wright, Julius C. Enssle, Xin Yu, James D. Phelan, Yandan Yang, Arthur Shaffer III, Arvin E. Ruiz, Da Wei Huang, Zana Coulibaly, Michael Kelly, Bao Tran, Stefania Pittaluga, Mark Roschewski, Wyndham H. Wilson, Thomas Oellerich, Giorgio Inghirami, and Louis M. Staudt

Figure S1

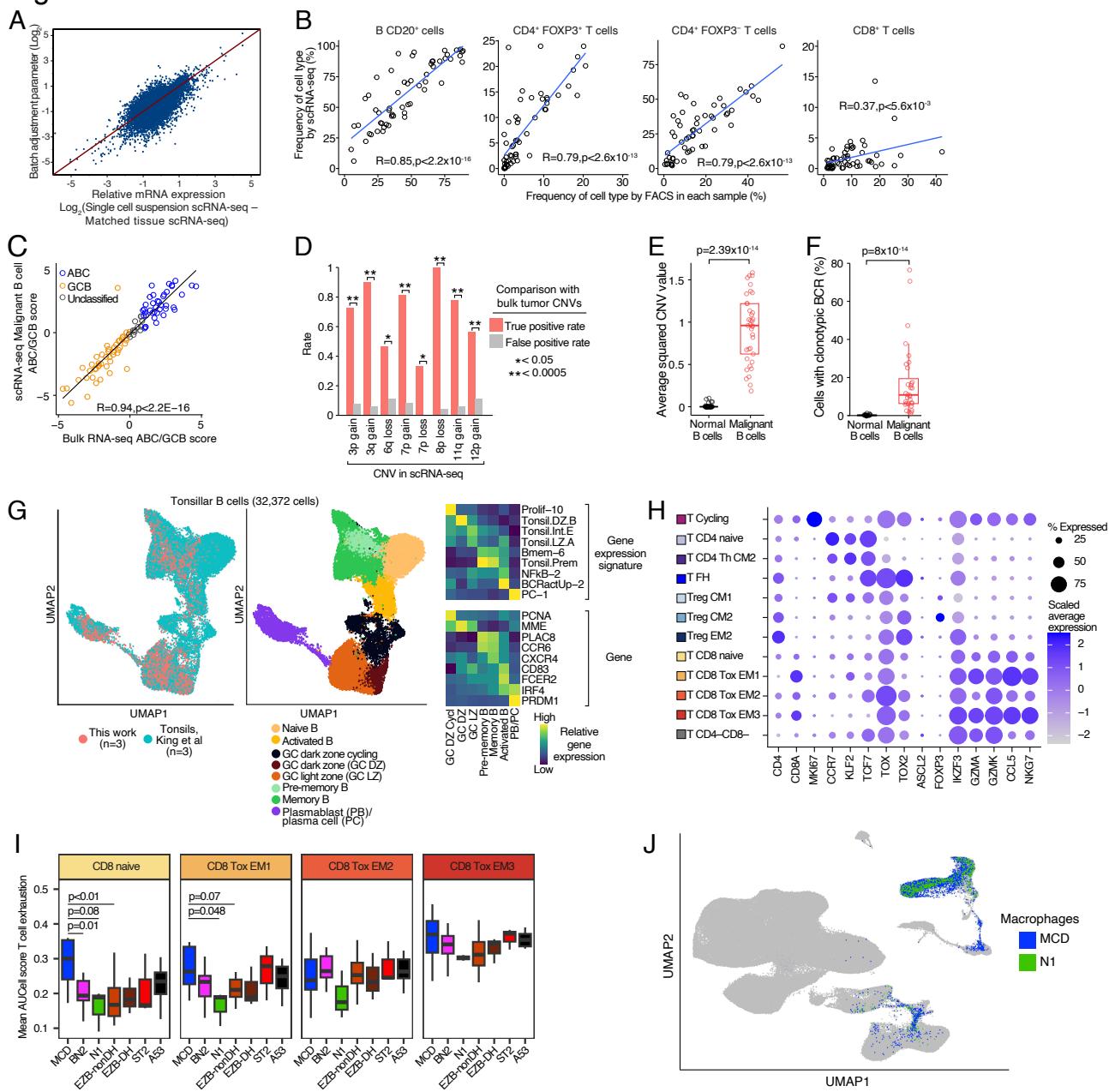


Fig. S1. Confirmation of cell types determined by scRNA-seq clustering, related to Figures 1 and 4.

A. Development of scRNA-seq batch adjustment parameters. Each dot is one gene (n=19,091). X-axis is the estimate from 3 matched DLBCL with single cell data generated from both cell suspension and tissue as starting material. Y-axis is the estimate modeled from all DLBCL scRNA-seq (n=103). **B.** Comparison of cell type proportions as determined by FACS or scRNA-seq clustering in Fig. 1C. Each dot is one sample (n= 57 DLBCL, 3 tonsil). Spearman correlation and p-values are reported. **C.** Correlation of the cell-of-origin (COO) score in paired malignant B cell scRNA-seq and bulk RNA-seq. Each dot is one sample (n=100). **D.** Detection

of copy number alterations in scRNA-seq inferred CNV compared to bulk WES. Significance is by a Fishers exact test. **E.** Classification of B cells in each case (dots) as malignant or non-malignant based on the average squared CNV (n= 37 biopsies). Boxplots show the median, first/third quartiles and 1.5*IQR. P-value is based on a Wilcoxon test. **F.** Frequency of B cells in each case containing the dominant BCR clonotype (n=32 biopsies). p value based on a Wilcox test. **G.** UMAP representation of tonsil B cells (n=3 this work, n=3 from ref¹). Heatmap shows the normalized scRNA-seq expression of published B cell signatures or genes. **H.** Gene expression of marker genes in T cell subsets from Fig. 1D. **I.** Phenotype of CD8 T cells based on AUCell score². Boxplots show the median, first/third quartiles and 1.5*IQR of inferred gene set scores in genetic subtype enriched neighborhoods. P-value based on Wilcoxon test. **J.** UMAP of non-malignant cells from Fig.1D with the color highlighting the macrophages from MCD or N1 samples.

Figure S2

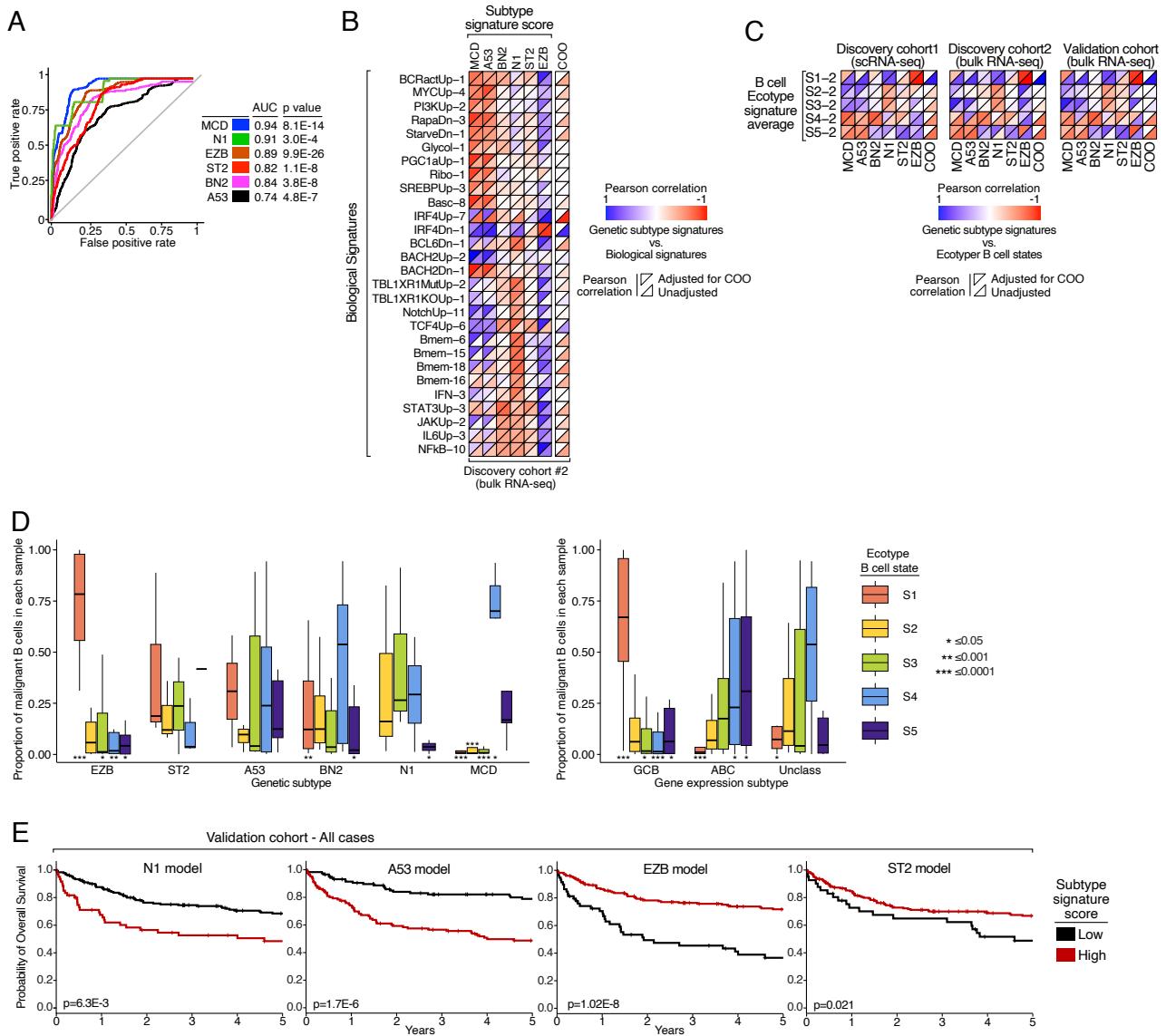


Fig. S2. Subtype signature score, related to figures 2 and 3.

A. Cross validated ROC curve of the subtype signature score in discovery cohort 2. P-value is based on a Wilcox test. **B.** Pearson correlation of gene expression signatures (rows) to subtype signature scores (columns). **C.** Same as Fig. S2B but showing B cell ecotype signature averages from Steen et al. 2021³. **D.** Ecotype B cell state classification of scRNA-seq malignant B cells from Fig. 5A. Boxplots show the median, first/third quartiles and 1.5*IQR, p-value is based on two sided t test. **E.** Kaplan-Meier curve of the validation cohort separating cases based on their subtype signature score. P-value determined by log-rank test.

Fig. S3. Gene expression themes, related to figures 4 and 5.

A. Frequency of subclones in each case (n=103) . Bar plots show mean and SE. **B.** Average pairwise distance between scRNA-seq inferCNV copy number clusters to determine the genetic subclone assignments. Each dot is one genetic subclone. **C.** Generation of gene expression themes. Heatmap shows genetic subclones in cases with >1 subclone (columns, n=221 subclones from 81 cases) and expression of SignatureDB gene expression signatures (rows). Heatmap color indicates the deviation score. Signatures are ordered by centroid clustering used to identify gene expression themes, highlighted in the clustering. **D.** Application of DLBCL signatures to tonsil B cell subsets. **E.** Average Pearson correlation between theme scores and Ecotype signatures across all malignant B cells. **F.** Heatmap of the proportion of cells. UMAP of malignant B cells is from Fig. 5A.

Figure S4

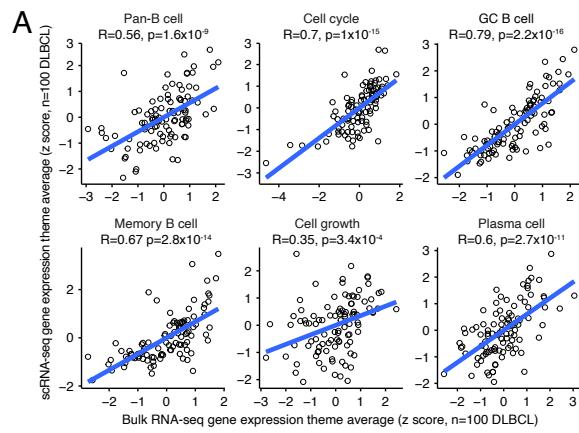


Fig. S4. Intra-tumor heterogeneity is recapitulated by gene expression themes, related to Figure 6.

Pearson correlation and p value of the gene expression theme z score comparing pseudobulk of the malignant B cell scRNA-seq to paired bulk RNA-seq. Each dot is a case (n=100).

Figure S5

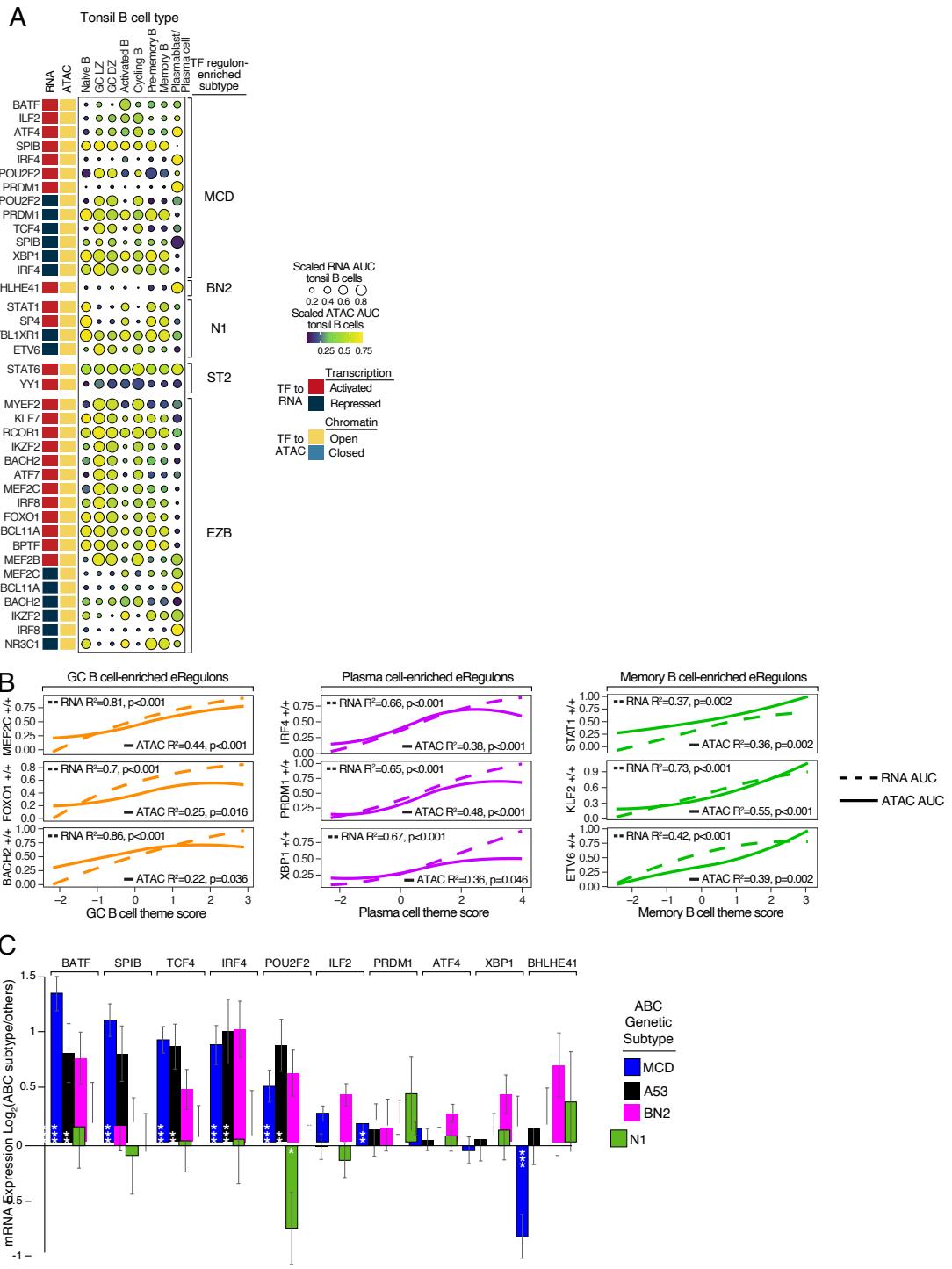


Fig. S5. Analysis of eRegulons, related to Figure 8.

A. RNA AUC and ATAC AUC enrichment scores in the indicated tonsillar B cell subpopulations for TF eRegulons in Fig. 8C. **B.** LOESS correlation for the indicate eRegulon RNA AUC and ATAC AUC enrichment scores with the indicated gene expression themes. **C.** Average fold change in mRNA expression levels comparing the indicated ABC subtypes versus

all other DLBCLs in the validation cohort. Bars depict mean and standard error, *** P<0.001, ** P<0.01, * P<0.05.

References

1. King, H.W., Wells, K.L., Shipony, Z., Kathiria, A.S., Wagar, L.E., Lareau, C., Orban, N., Capasso, R., Davis, M.M., Steinmetz, L.M., et al. (2021). Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci. *Sci Immunol* 6, eabh3768.
2. Roider, T., Baertsch, M.A., Fitzgerald, D., Vohringer, H., Brinkmann, B.J., Czernilofsky, F., Knoll, M., Liao-Cid, L., Mathioudaki, A., Fassbender, B., et al. (2024). Multimodal and spatially resolved profiling identifies distinct patterns of T cell infiltration in nodal B cell lymphoma entities. *Nat Cell Biol* 26, 478-489.
3. Steen, C.B., Luca, B.A., Esfahani, M.S., Azizi, A., Sworder, B.J., Nabat, B.Y., Kurtz, D.M., Liu, C.L., Khameneh, F., Advani, R.H., et al. (2021). The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. *Cancer Cell* 39, 1422-1437 e1410.