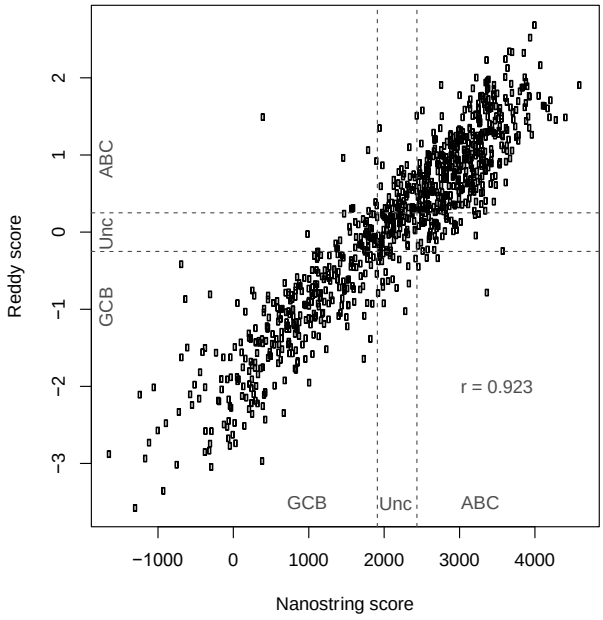
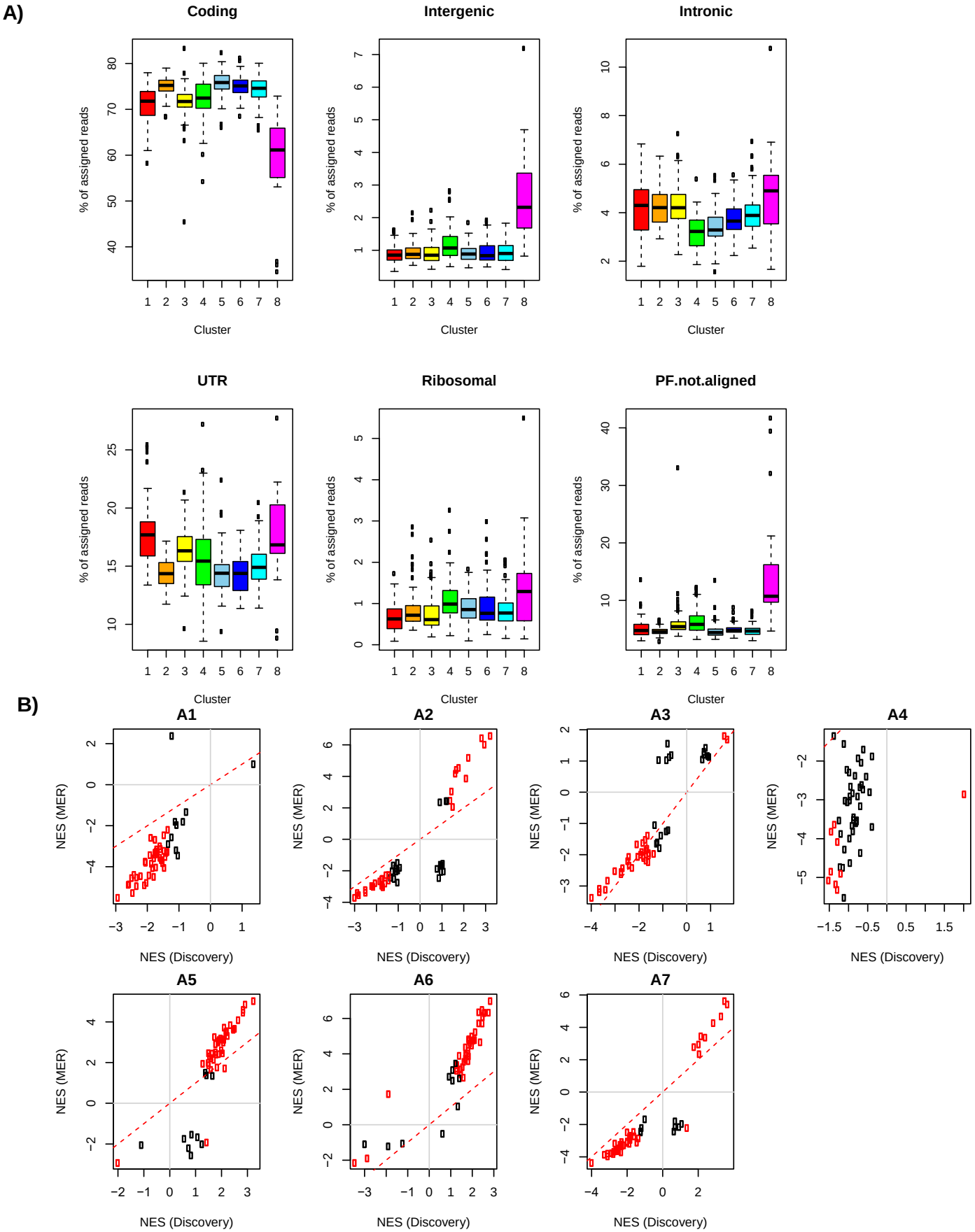
Transcriptomic Classification of Diffuse Large B-Cell Lymphoma Identifies a High-Risk Activated B-cell-like Subpopulation with Targetable MYC Dysregulation

Supplemental Figures

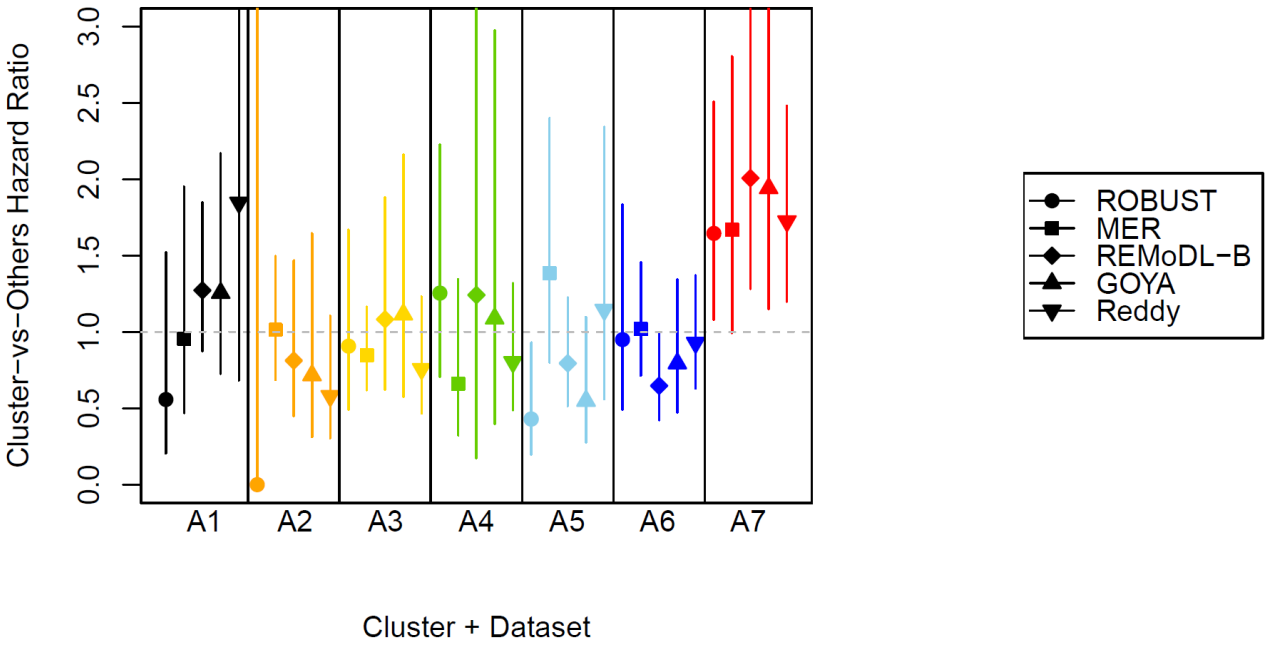
A)



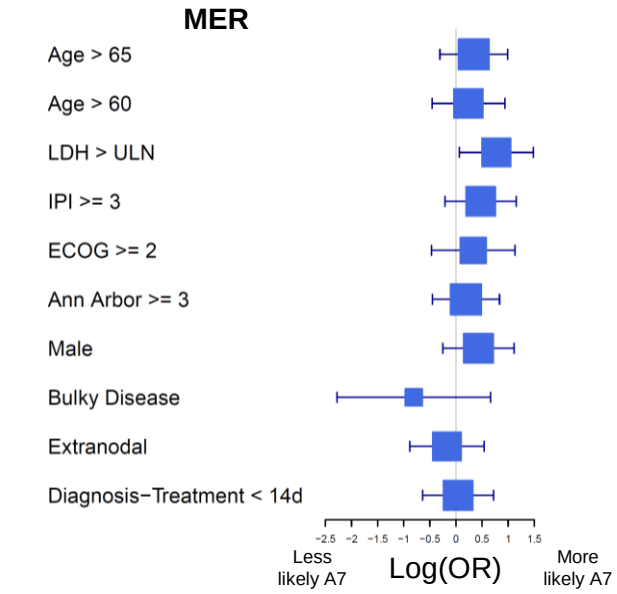
Supplemental Figure S1 – Comparison of COO-scoring methods. A) Correlation between Reddy COO scores and Nanostring LST COO calls (ROBUST). ABC, GCB and unclassified thresholds are shown. Source data are provided as a Source Data file.

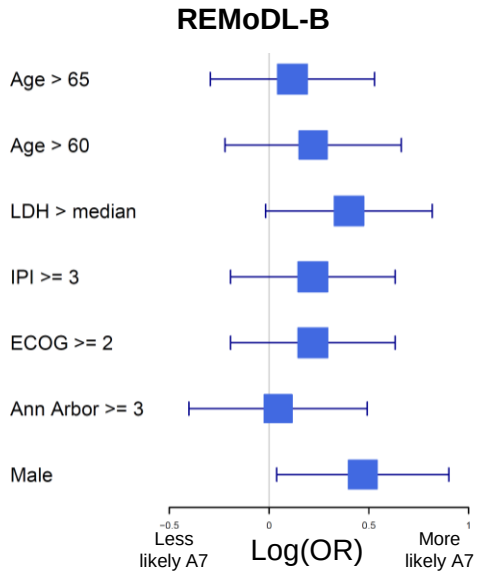


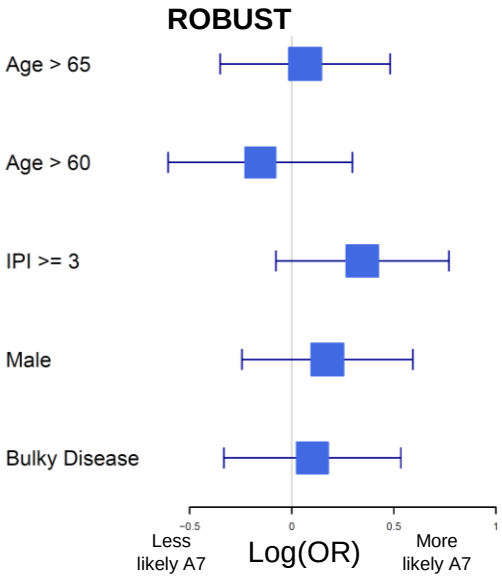
Supplemental Figure S2 – Sample quality measures per cluster. A) Samples clustering into A8 have significantly lower alignment to coding regions and significantly higher rates of intergenic, ribosomal, and unaligned reads than other samples (Discovery). B) Correlation of Hallmark pathway Normalized Enrichment Scores (NES) between Discovery and MER. Significantly dysregulated pathways ($p<0.05$) are highlighted in red. Source data are provided as a Source Data file.

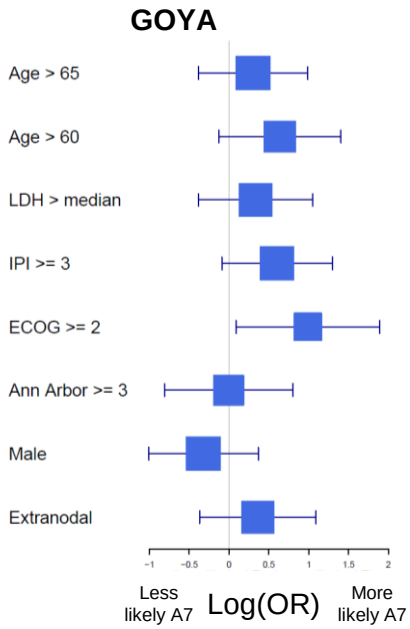


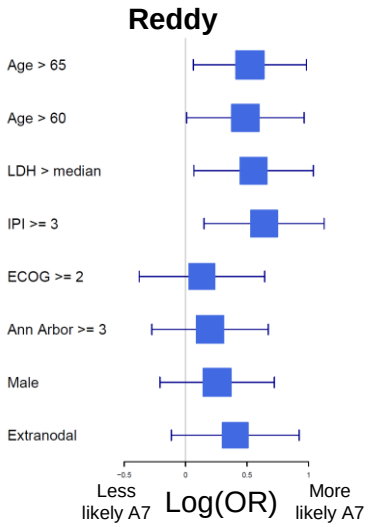
Supplemental Figure S3 – Cluster-specific Hazard Ratios. Hazard ratios comparing binarized models of each cluster versus all others in each cohort. Error bars represent the 95% confidence interval, with statistically significant HRs failing to cross HR=1. Source data are provided as a Source Data file.



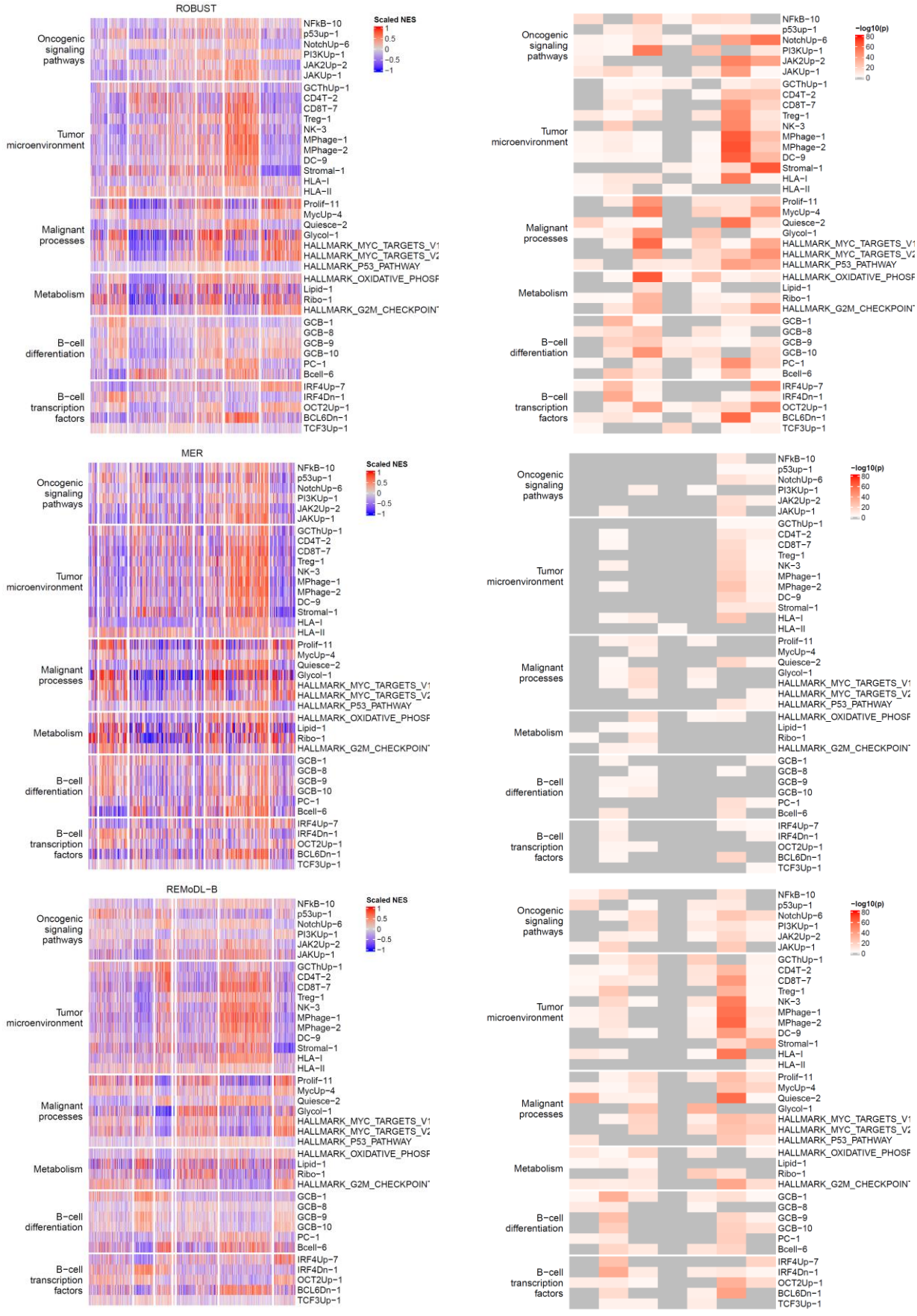




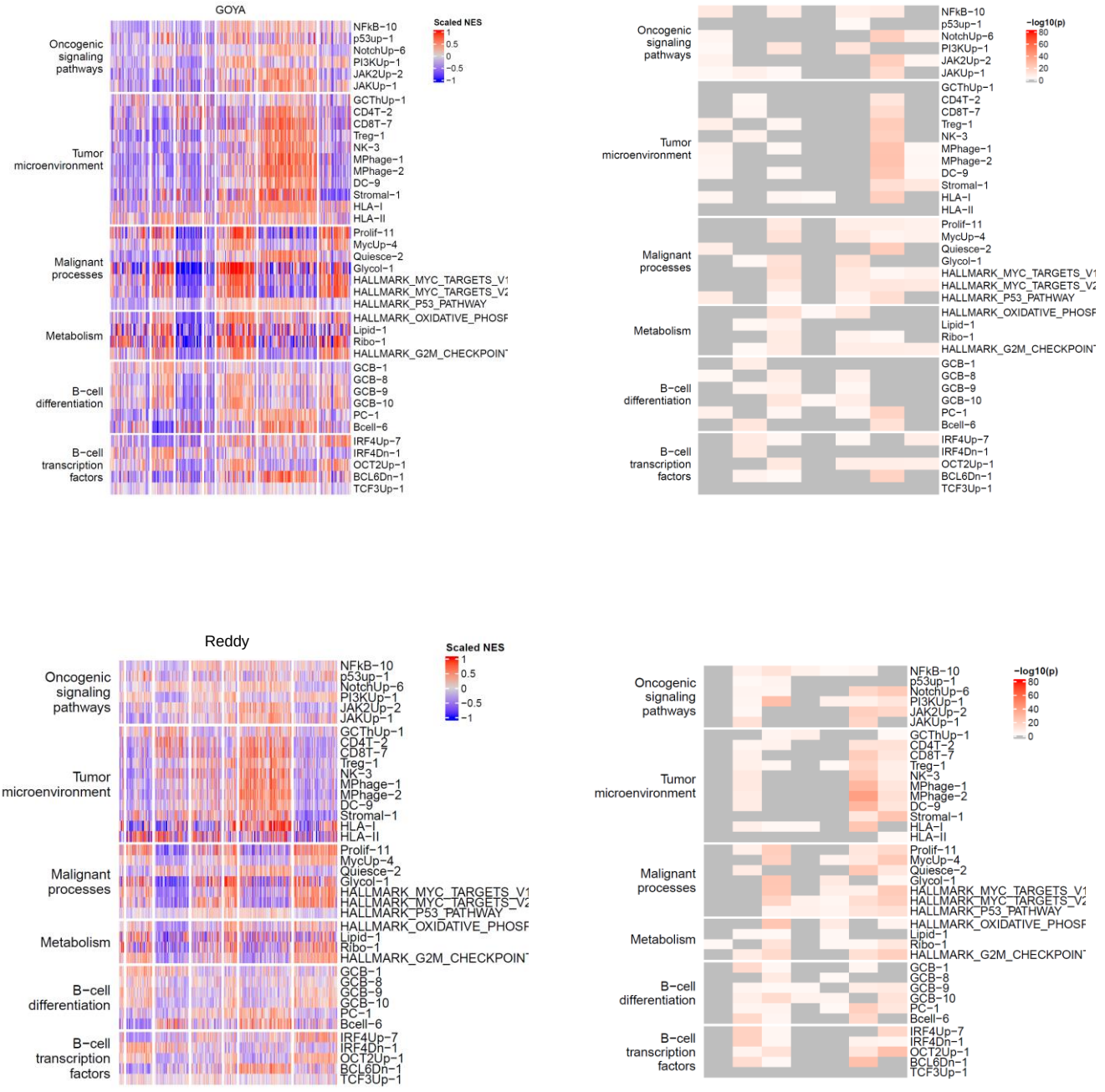




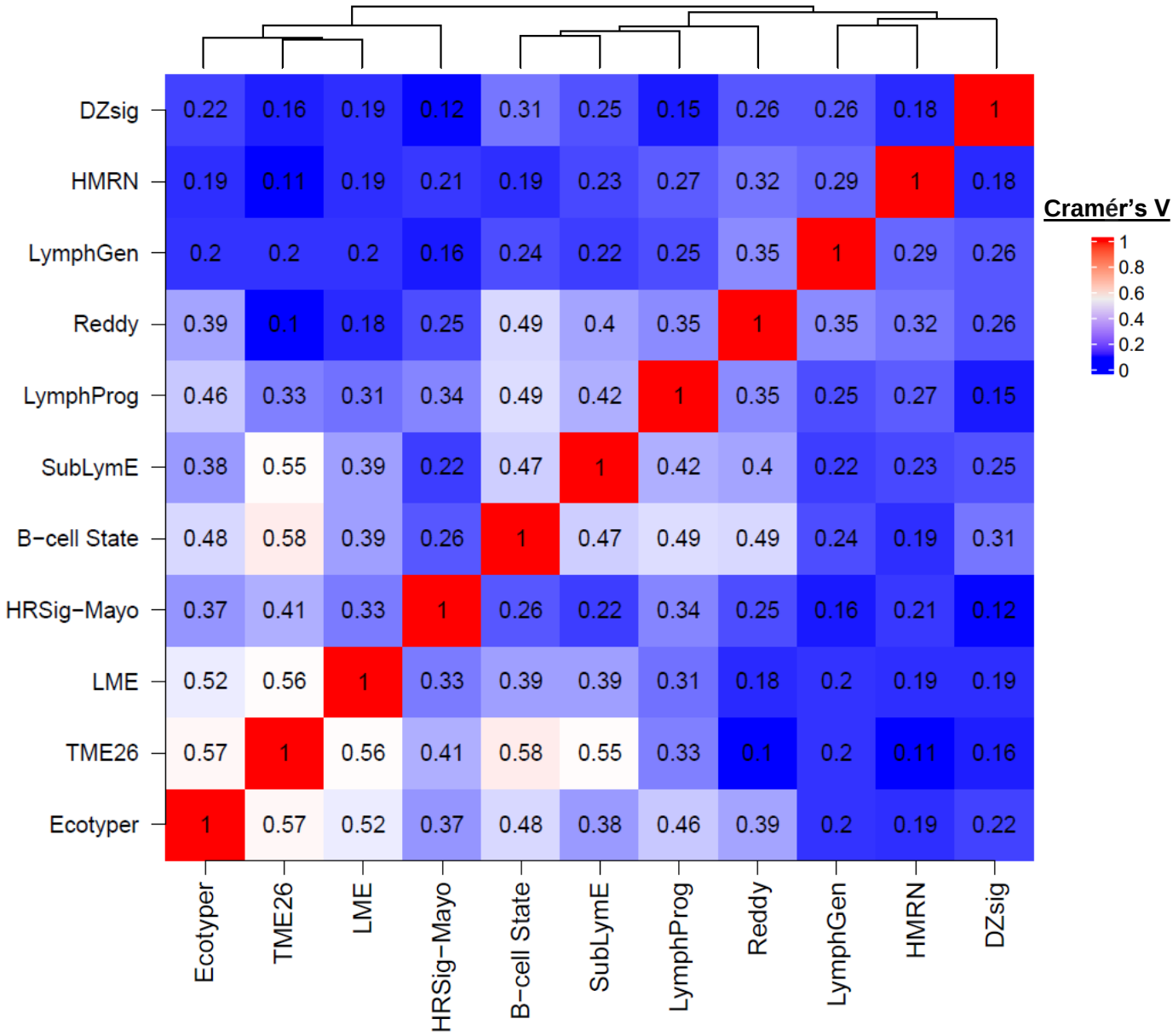
Supplemental Figure S4 – Clinical variables' associations with A7. Forest plots showing the log-odds ratio of being an A7 case given prognostic clinical variable states in ROBUST, MER, REMoDL-B, GOYA, and Reddy. The 95% confidence interval is displayed for each variable. Source data are provided as a Source Data file.



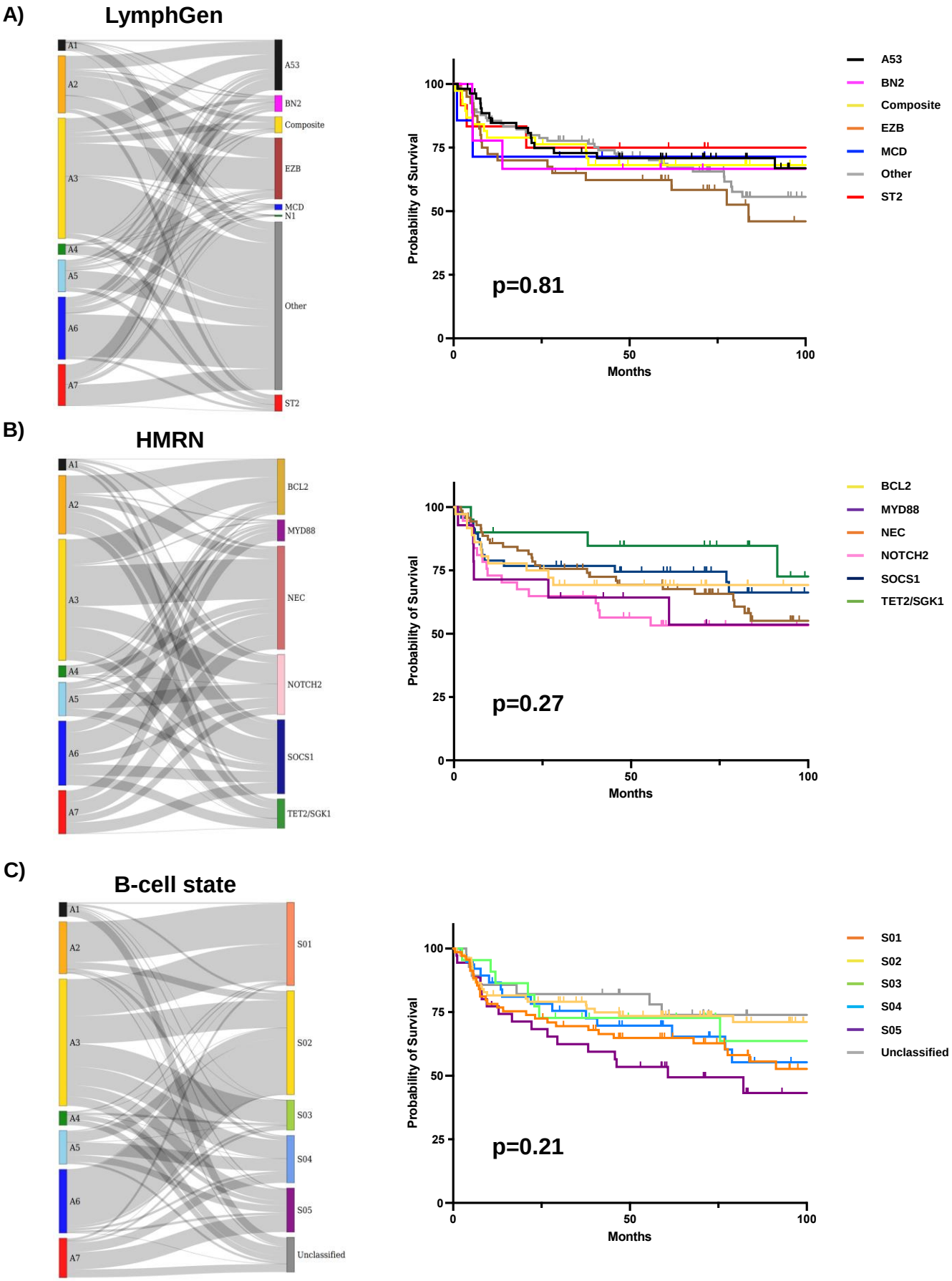
Supplemental Figures S5 – Lymphoma-specific processes across datasets. Shown on the left are the heatmaps of the selected lymphoma-specific processes from Figure 3A, replicated in each dataset. The heatmaps at right highlight statistical significance of pathway dysregulation, as measured by a Wilcoxon test of NES scores between a binarized one-vs-others model. Source data are provided as a Source Data file.



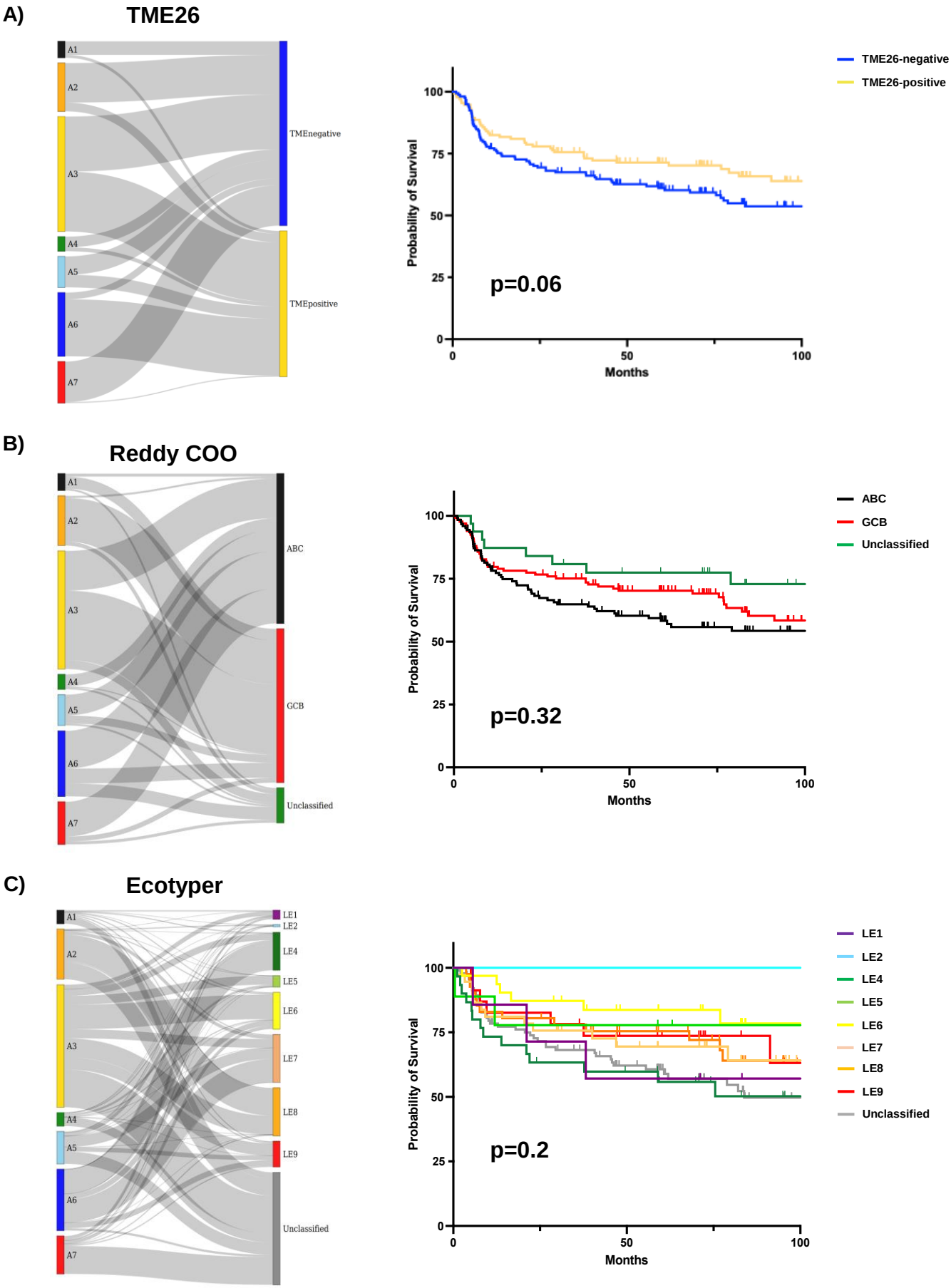
Supplemental Figures S6 – Lymphoma-specific processes across datasets. Shown on the left are the heatmaps of the selected lymphoma-specific processes from Figure 3A, replicated in each dataset. The heatmaps at right highlight statistical significance of pathway dysregulation, as measured by a Wilcoxon test of NES scores between a binarized one-vs-others model. Source data are provided as a Source Data file.



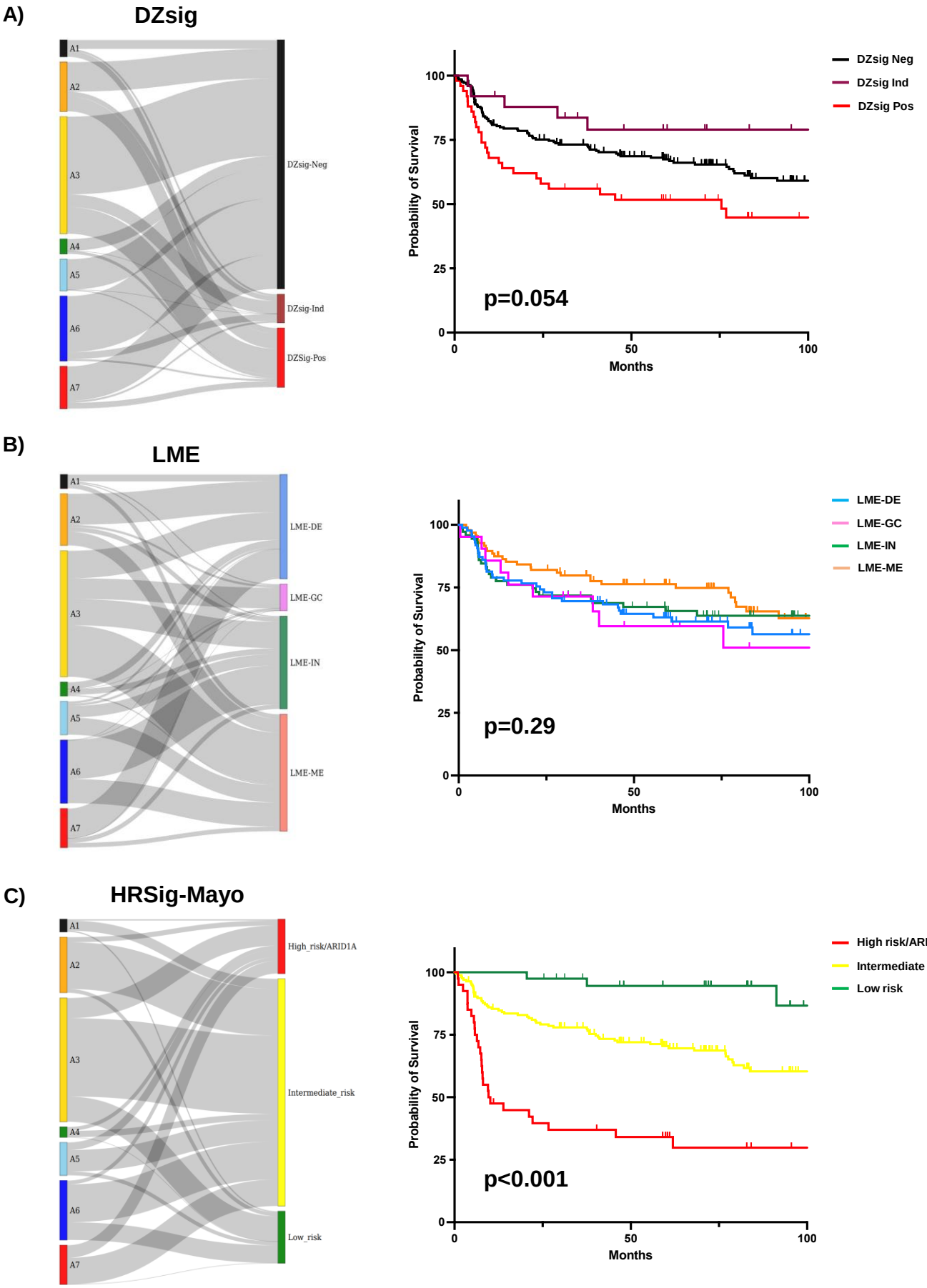
Supplemental Figure S7 – Clustering of classification methods. Heatmap of Cramer's V values measuring global strength of association between different sets of classifier labels (MER). The hierarchical clustering of association values shows groups of methods with similar behavior. The MHG calls were evaluated in the REMoDL-B cohort and are omitted here, but showed high association with the Reddy COO calls in that data (V=0.56). Source data are provided as a Source Data file.



Supplemental Figures S8 – Comparison of cluster memberships and survival outcomes. Sankey plots showing overlap co-incidence of SubLymE classifier labels and each other classification method (MER). Survival stratification (EFS) for each classifier is shown in the accompanying KM plot with log-rank p-value.

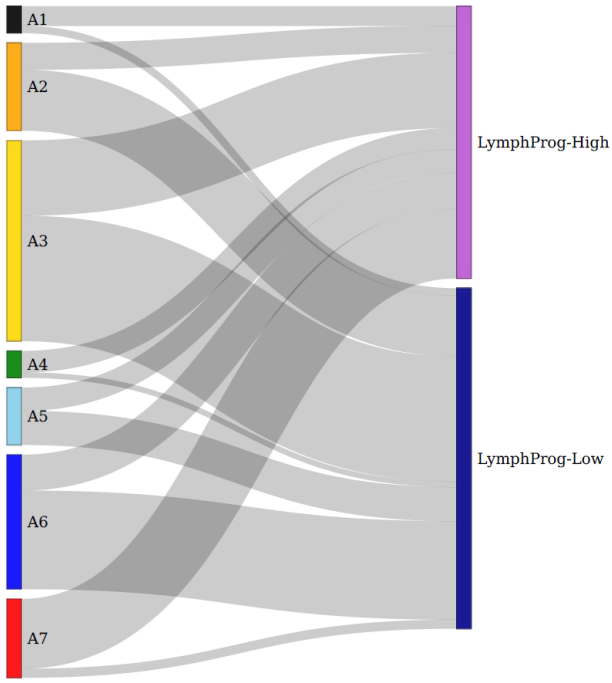


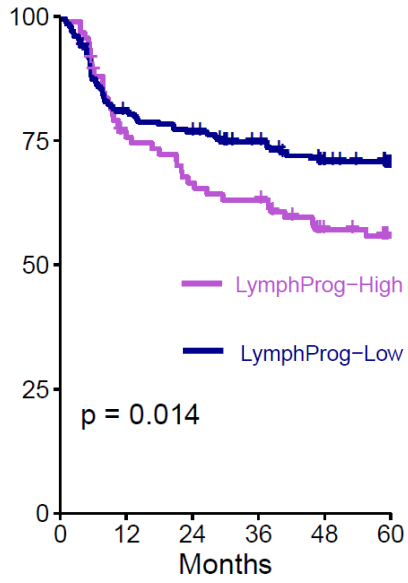
Supplemental Figures S9 – Comparison of cluster memberships and survival outcomes. Sankey plots showing overlap co-incidence of SubLymE classifier labels and each other classification method (MER). Survival stratification (EFS) for each classifier is shown in the accompanying KM plot with log-rank p-value.



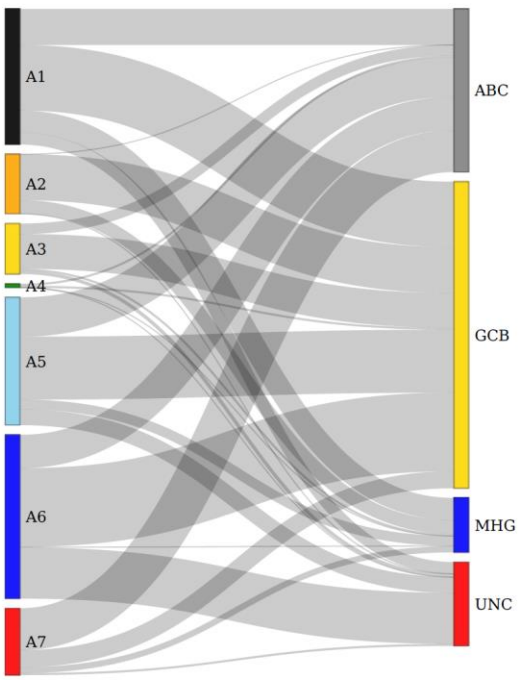
Supplemental Figures S10 – **Comparison of cluster memberships and survival outcomes.** Sankey plots showing overlap co-incidence of SubLymE classifier labels and each other classification method (MER). Survival stratification (EFS) for each classifier is shown in the accompanying KM plot with log-rank p-value.

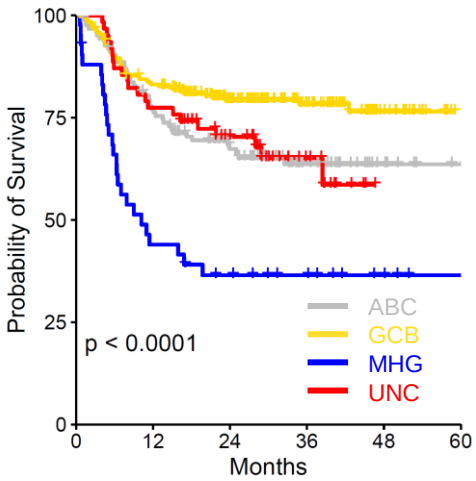
A)



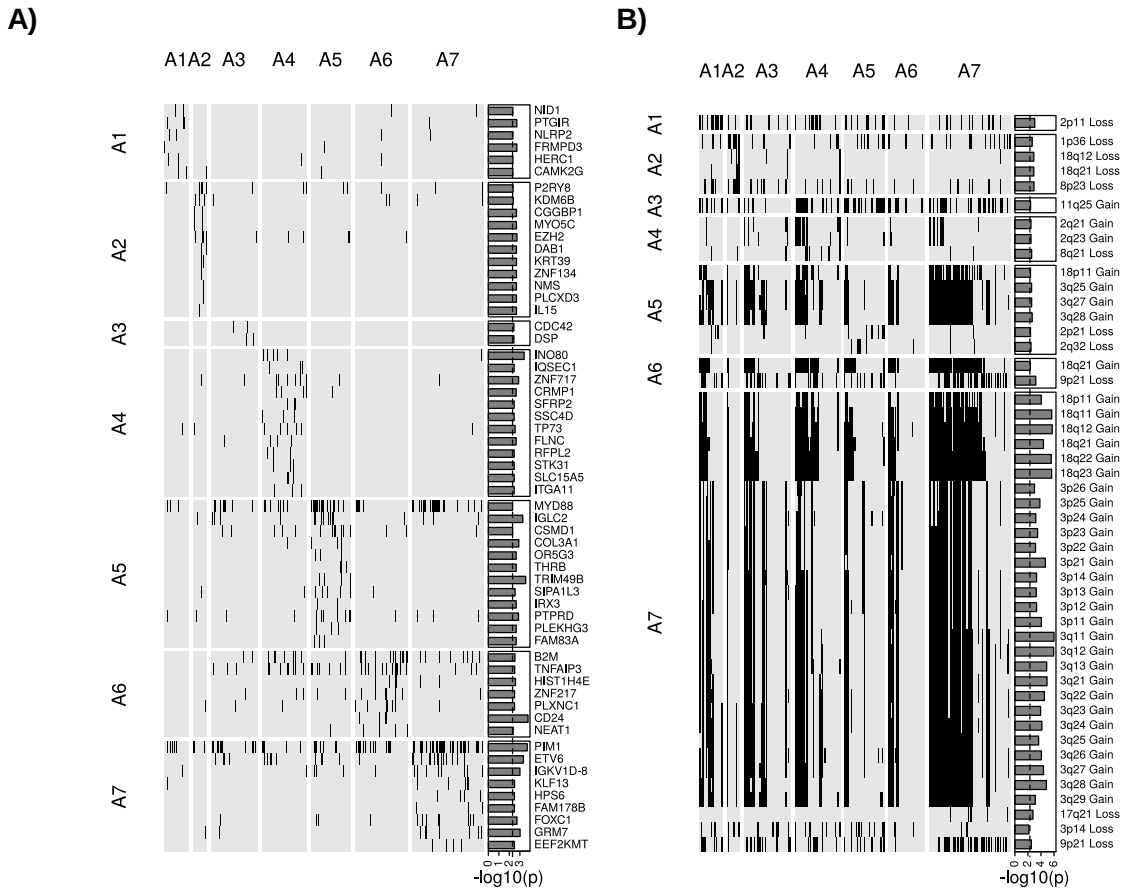


B)

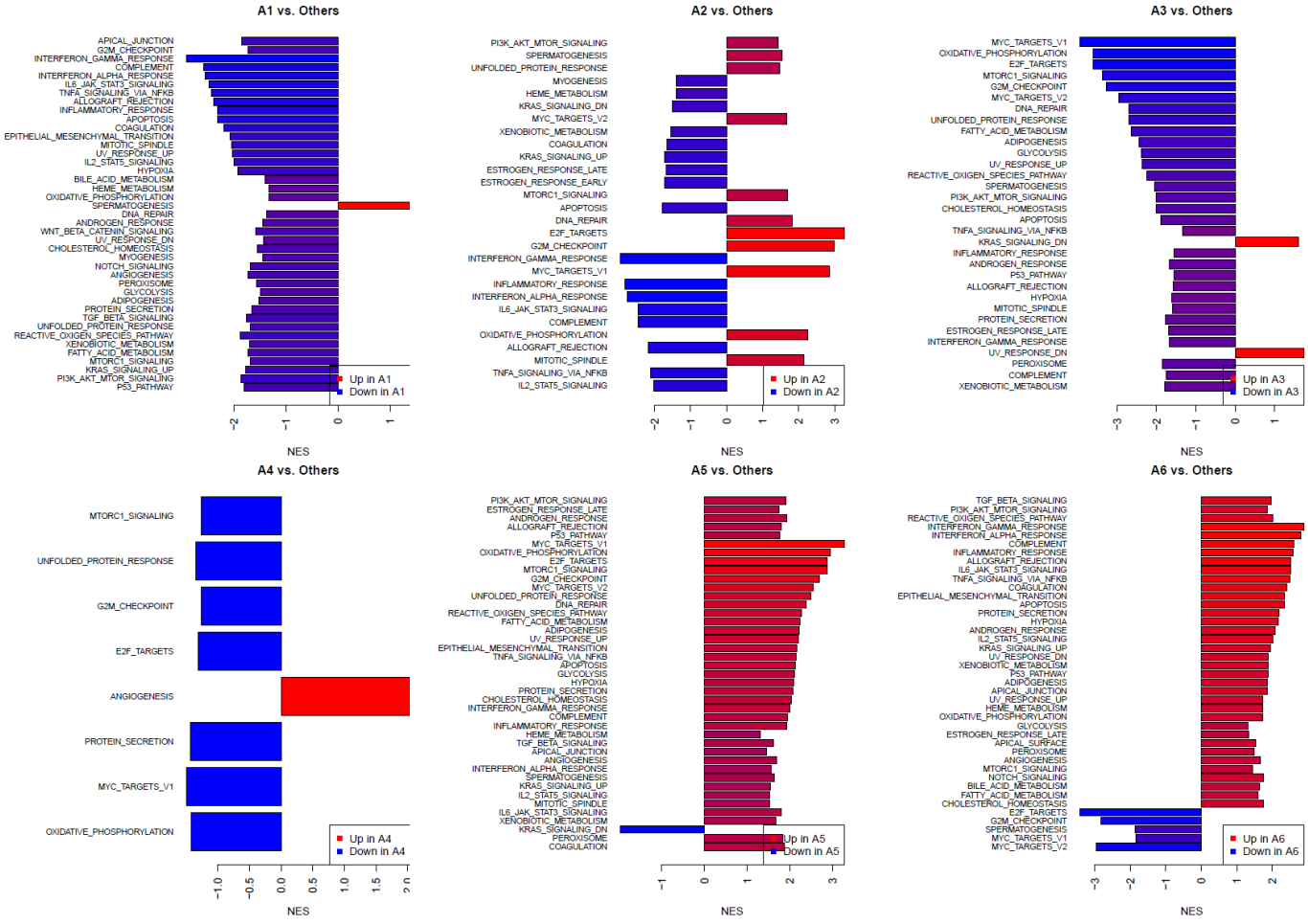




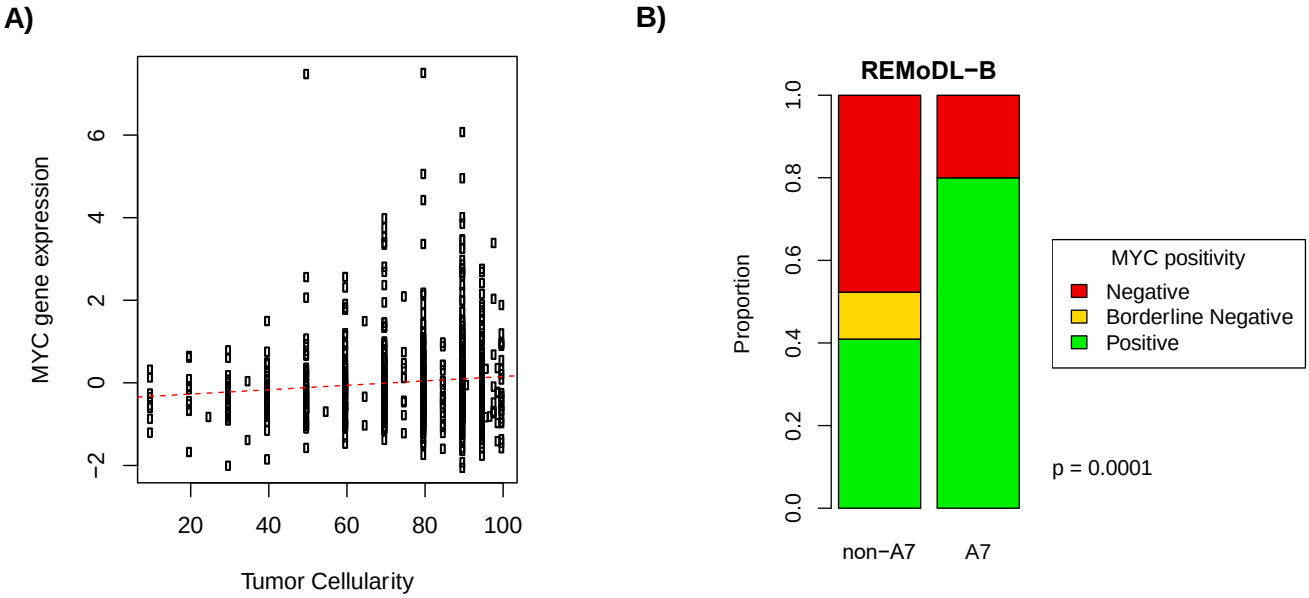
Supplemental Figures S11 – **Comparison of cluster memberships and survival outcomes.** Sankey plots showing overlap co-incidence of SubLymE classifier labels and each other classification method (A, MER; B, REMoDL-B). Survival stratification (EFS) for each classifier is shown in the accompanying KM plot with log-rank p-value.



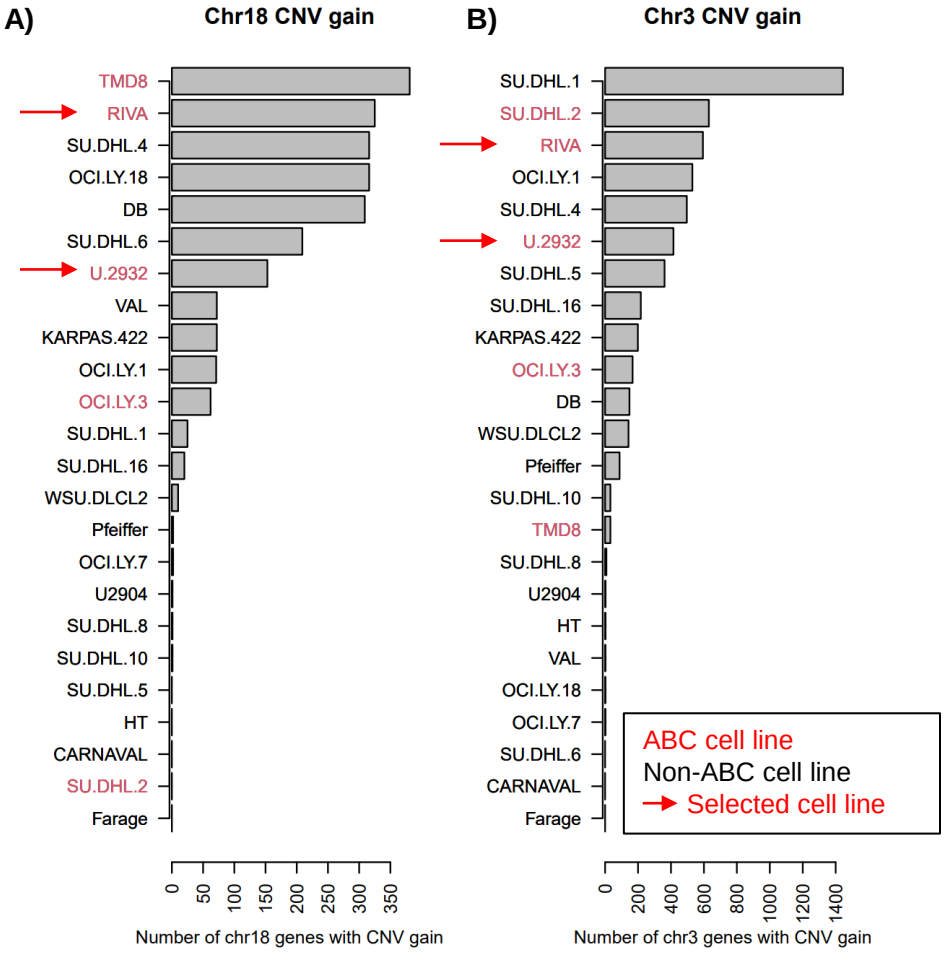
Supplemental Figure S12 – **Genomic features of SubLymE clusters.** A) Nominally significant mutational features associated with the SubLymE clusters. B) FDR-significant CNV features associated with the SubLymE clusters. Repeated chr3 and chr18 CNV features represent arm-level gains. Source data are provided as a Source Data file.



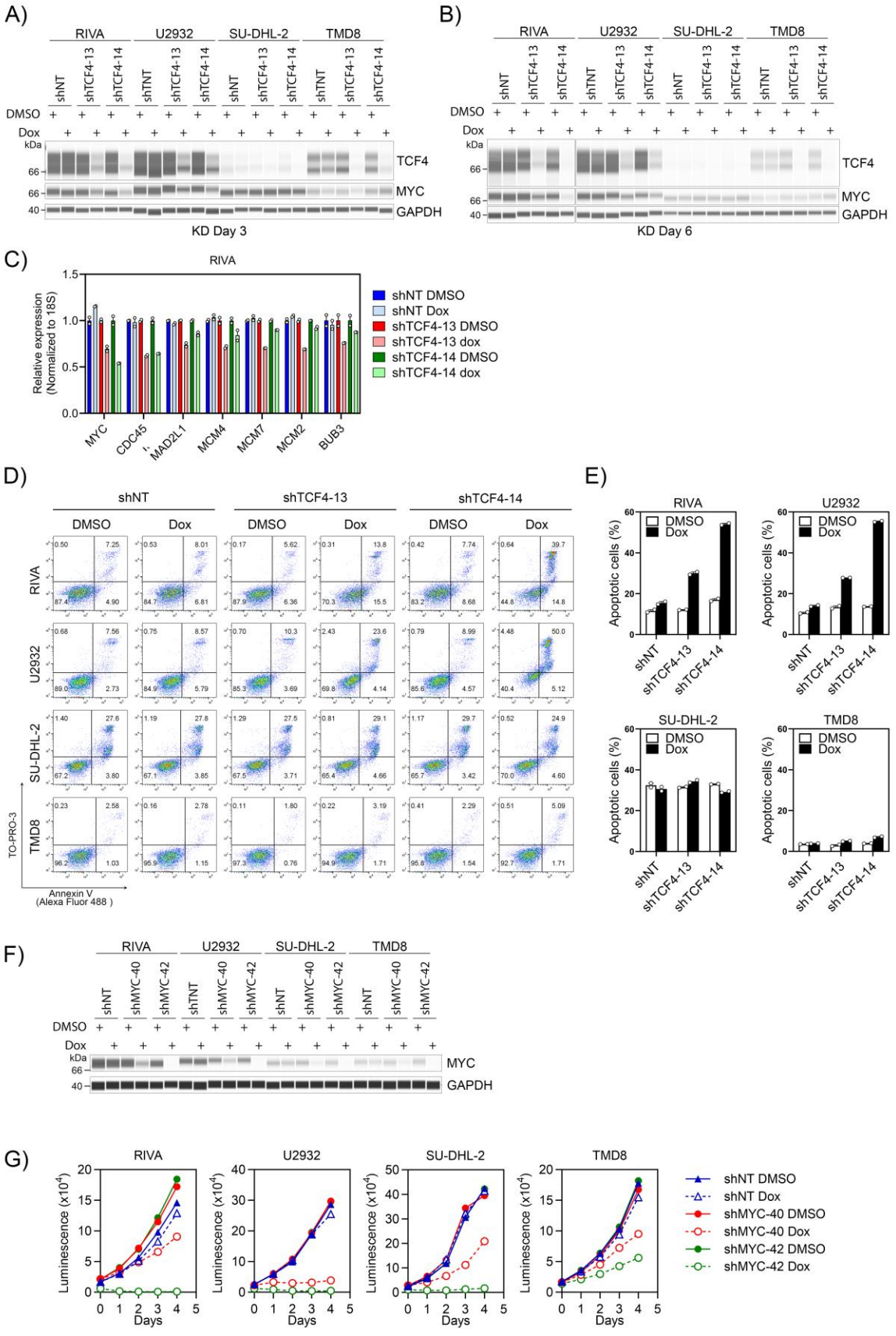
Supplemental Figure S13 – Significantly dysregulated Hallmark pathways per cluster. We calculated the Normalized Enrichment Score (NES) for all Hallmark pathways for each cluster versus all others in the MER cohort. Significantly differential pathways are shown for each cluster, ranked by most significant p-value at the top.



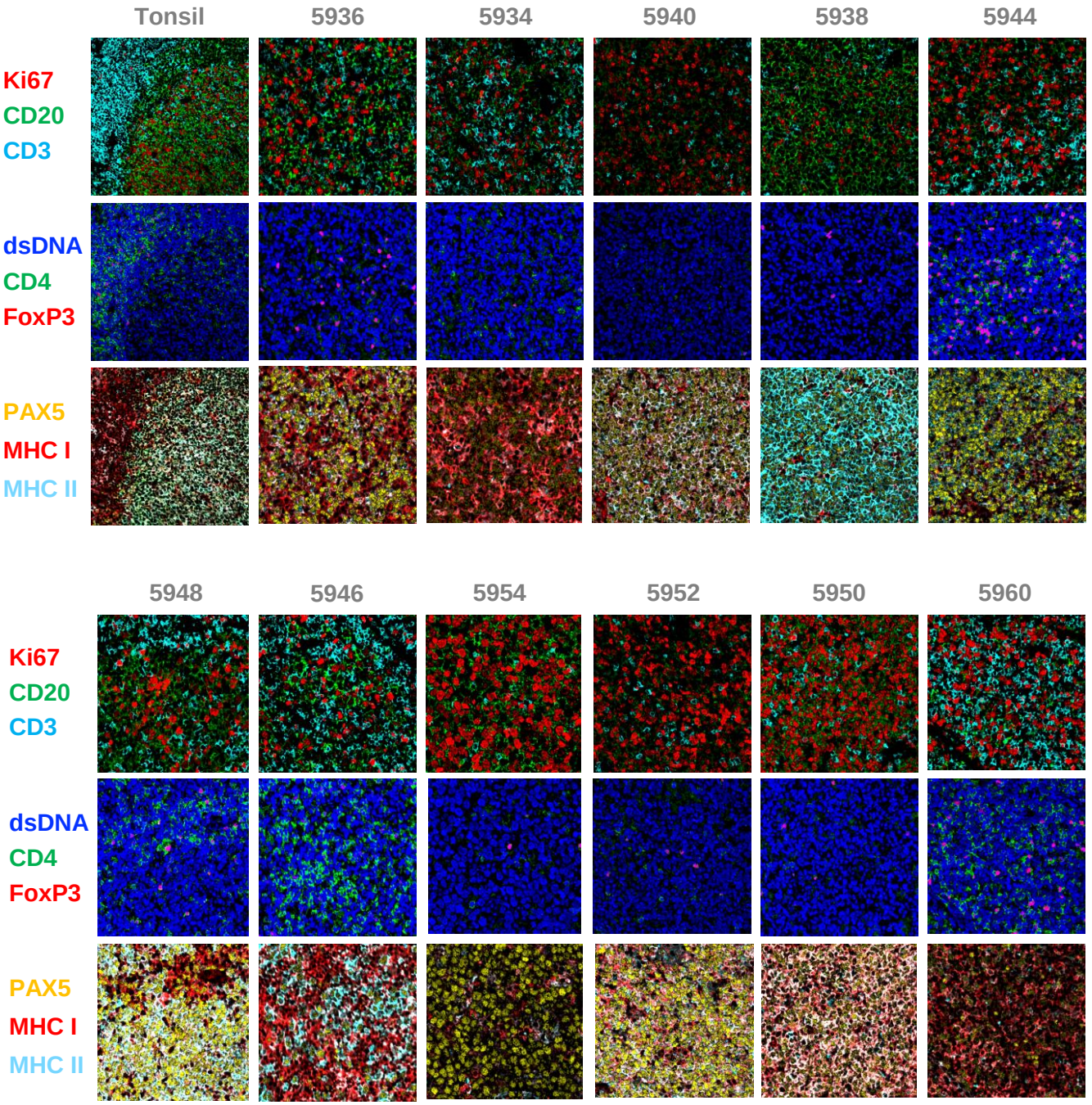
Supplemental Figure S14 – Tumor purity and MYC. A) Tumor purity has near-zero correlation with MYC expression (ROBUST, N=960). B) Percentage of non-A7 and A7 which are MYC-positive by IHC at a threshold of 40% of cells staining positive (REMoDL-B, N=355). Source data are provided as a Source Data file.



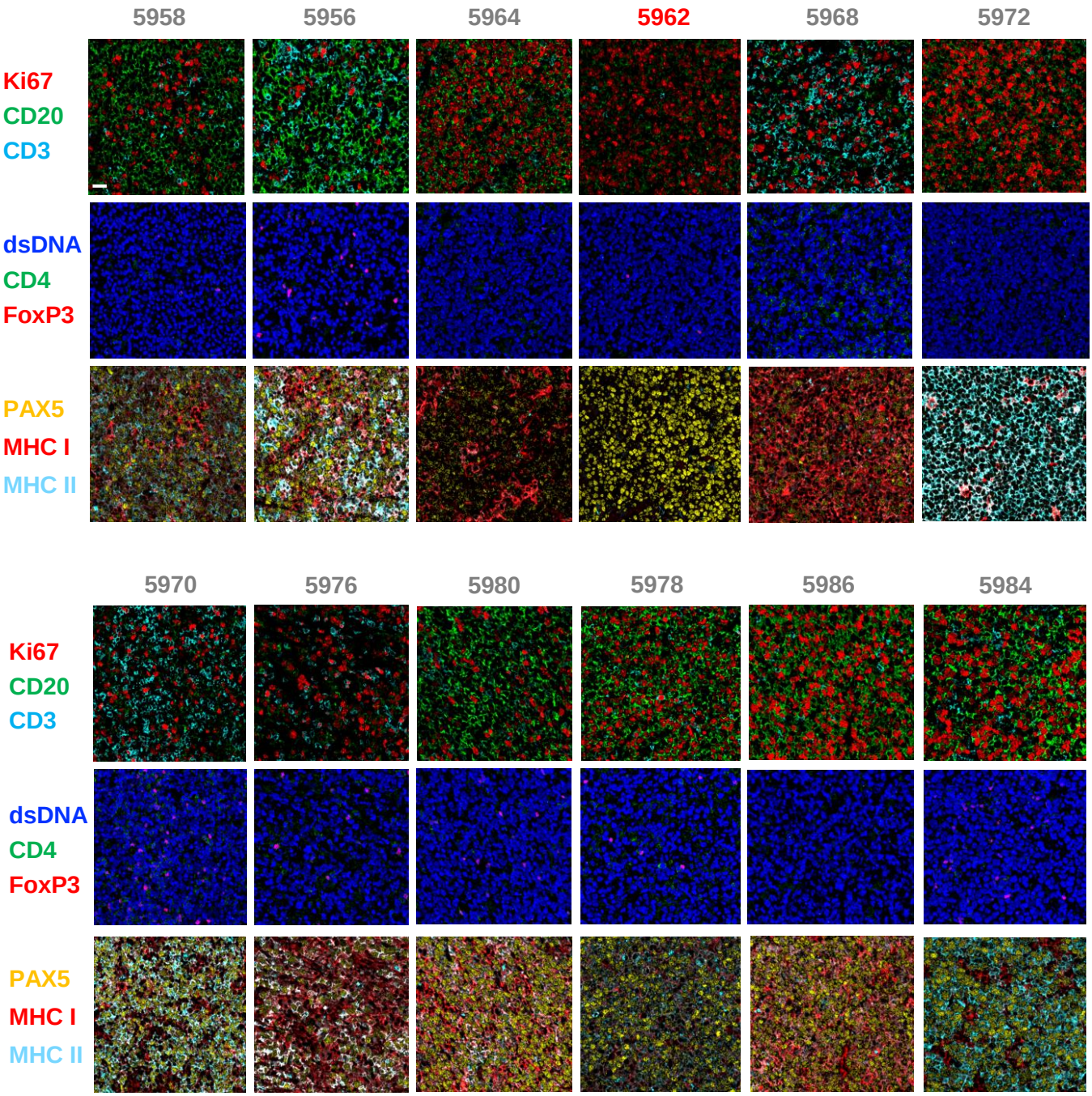
Supplemental Figure S15 – A7 genomic features in cell lines. A7-specific CNV features in cell line data. We use the number of genes with CNV gain as a measure of arm-level gain to select cell lines consistent with ABC biology and the A7-enriched arm-level gains in A) chr18 and B) chr3. Only RIVA and U2932 are ABC lines exhibiting both features. Source data are provided as a Source Data file.



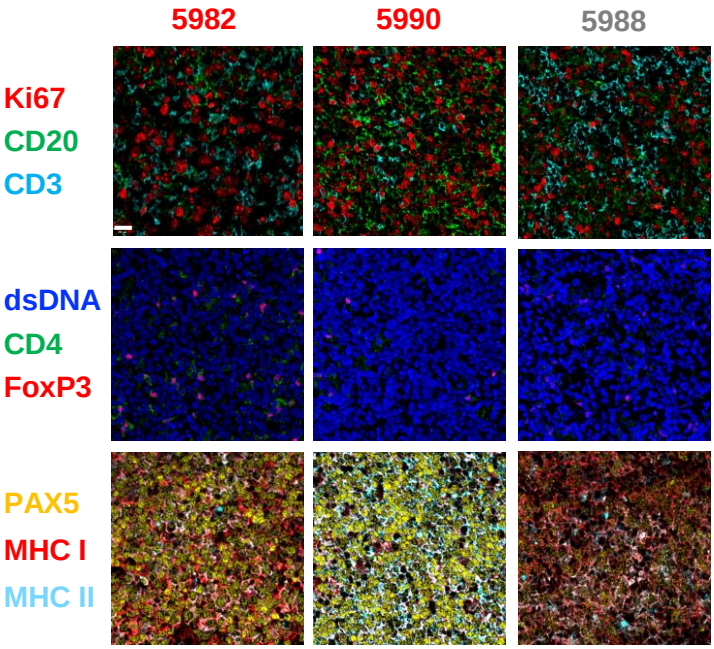
Supplemental Figure S16 – A, B) Western blot analyzing expression of TCF4 and c-myc in ABC-like DLBCL cell lines. Representative blots from N=1 experiment C) qPCR quantification of MYC and multiple direct transcriptional targets of myc. D, E) Annexin V/ToPro-3 apoptosis staining and quantification of control (shNT) and TCF4 knockdown (shTCF4) in ABC-like DLBCL cell lines. F) Western blot analyzing expression of c-myc and GAPDH in ABC-DLBCL cell lines. Representative blots from N=1 experiment I) Cell proliferation assay of control (shNT) and TCF4 knockdown (shMYC) in ABC-DLBCL cell lines. Each point represents the mean of technical triplicates with standard error of the mean too small to visualize. Source data are provided as a Source Data file.



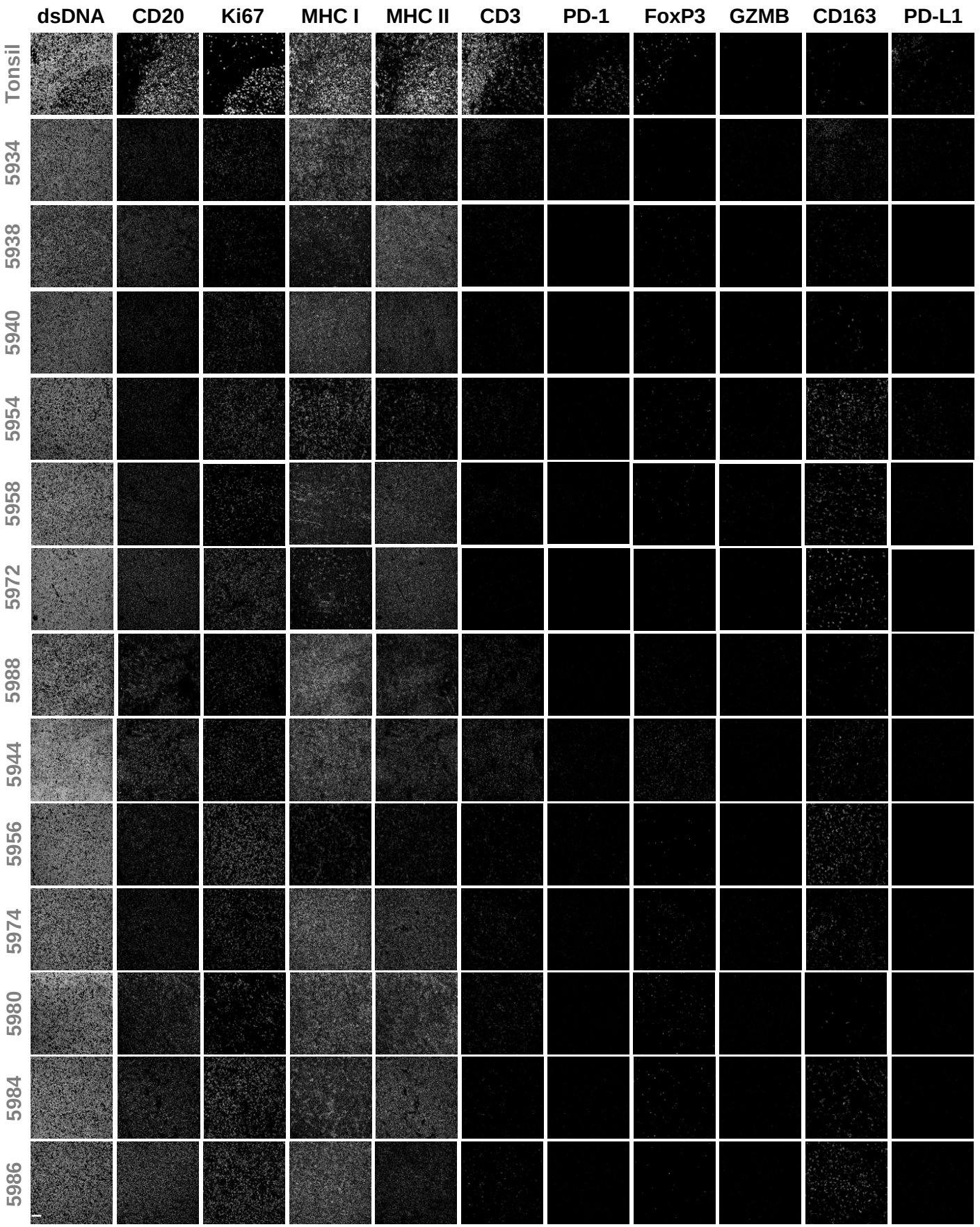
Supplemental Figure S17 - Multiplexed MIBI images for indicated markers in representative A7 (red) and non-A7 DLBCL cases (gray). The white bar in the image is 20um.



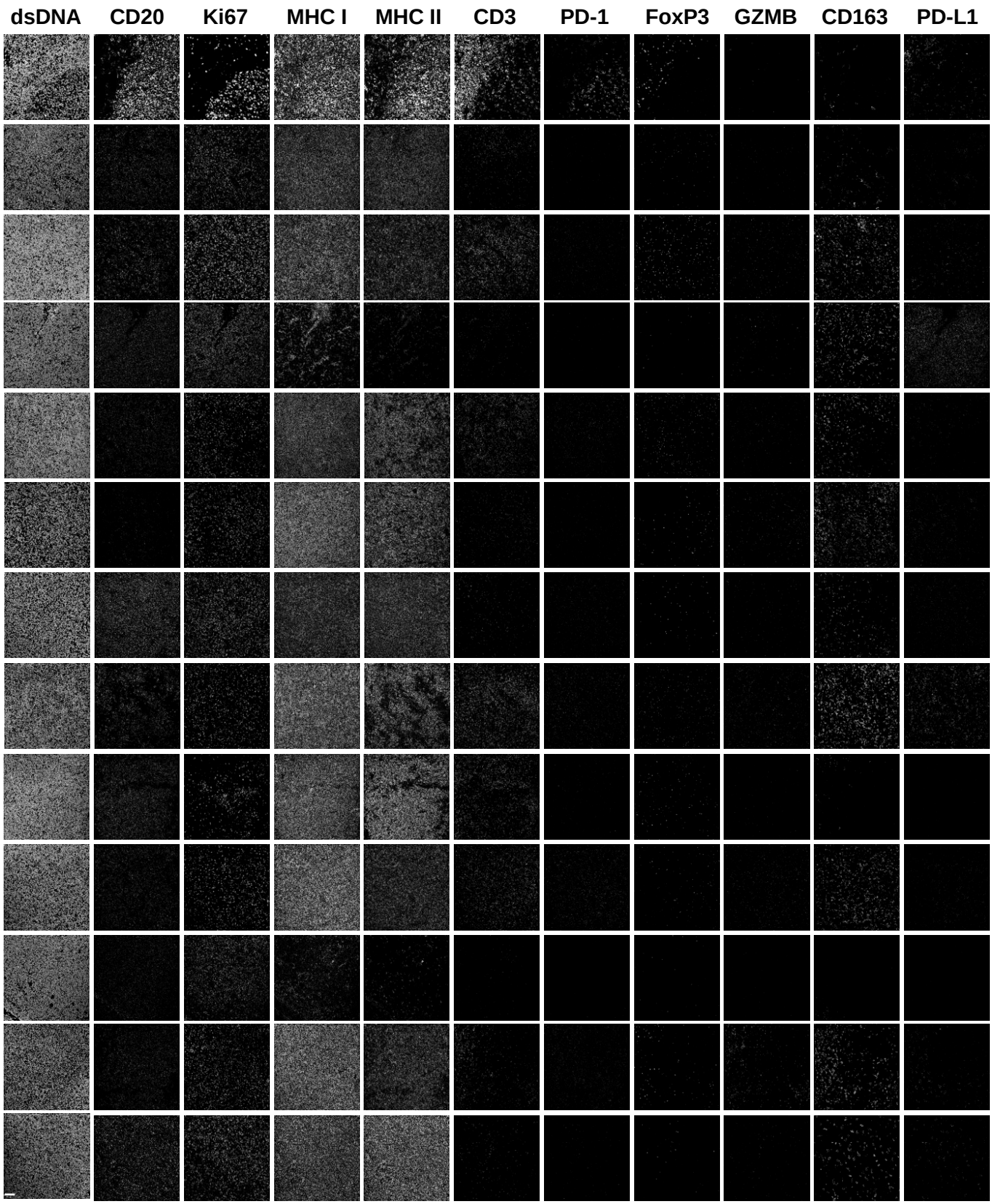
Supplemental Figure S18 - Multiplexed MIBI images for indicated markers in representative A7 (red) and non-A7 DLBCL cases (gray). The white bar in the top-left image is 20um.



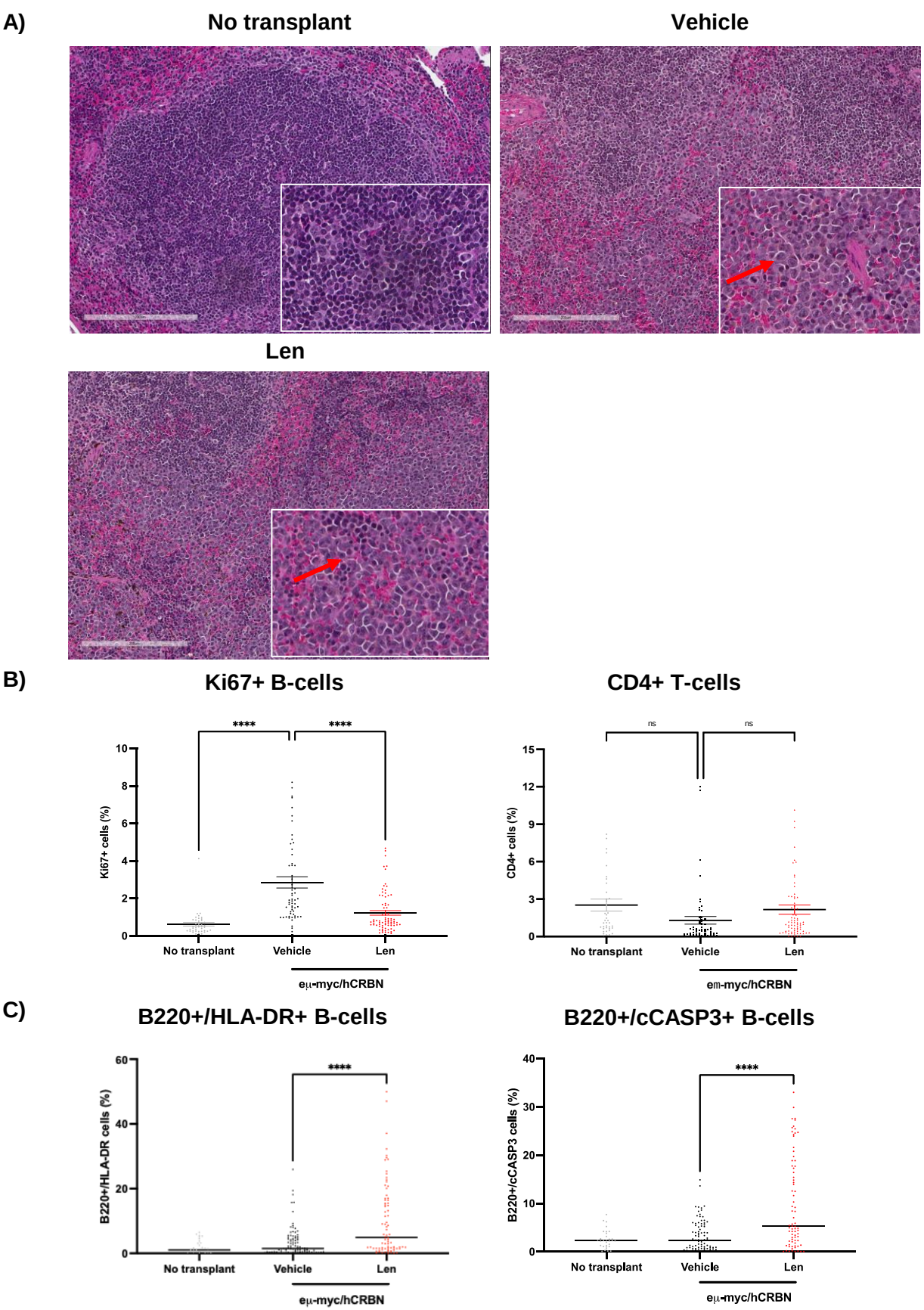
Supplemental Figure S19 - Multiplexed MIBI images for indicated markers in representative A7 (red) and non-A7 DLBCL cases (gray). The white bar in the top-left image is 20um.



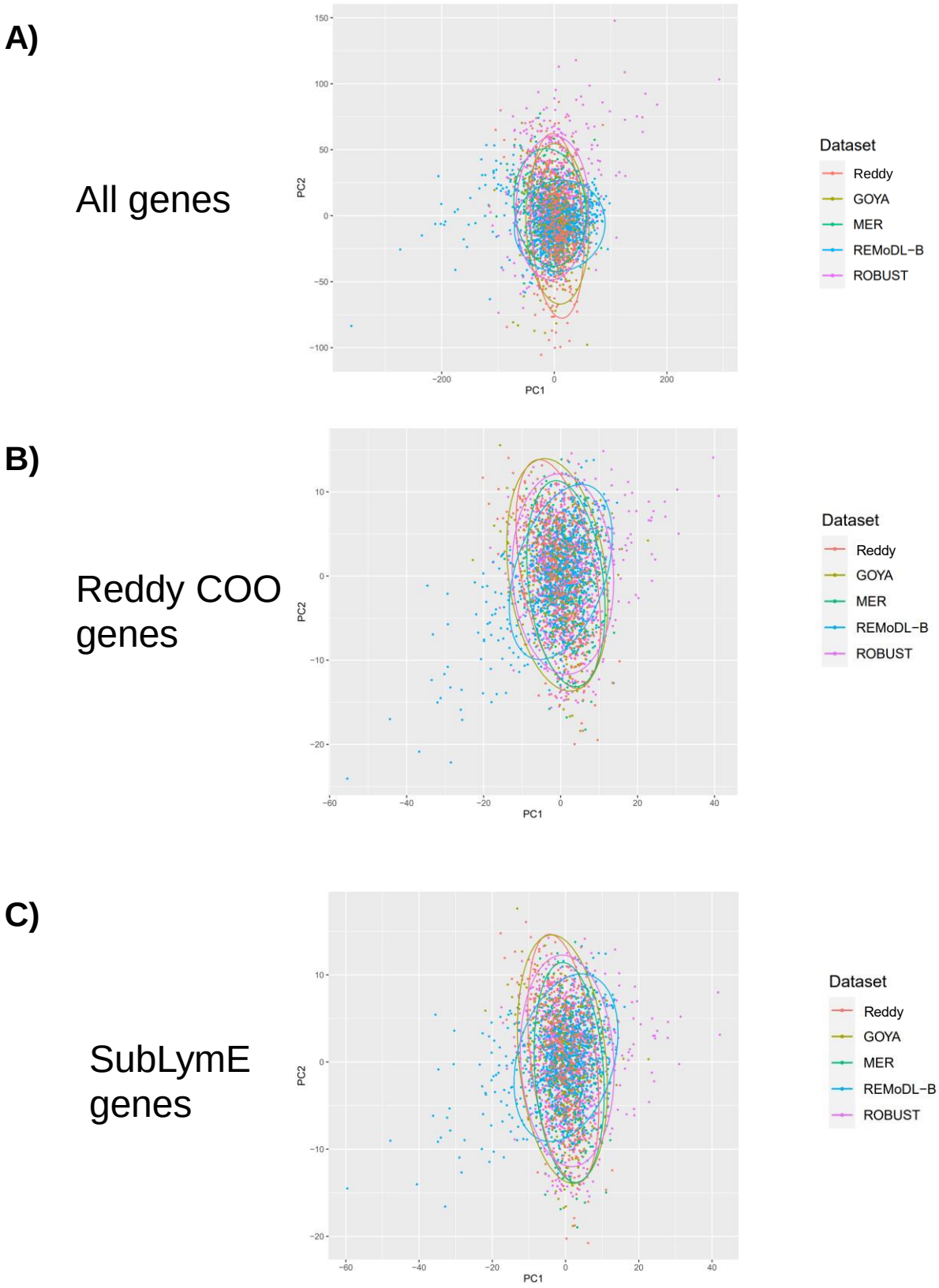
Supplemental Figure S20- Single channel MIBI images for indicated markers in representative non-A7 DLBCL cases (gray). The white bar in the bottom-left image is 95um.



Supplemental Figure S21- Single channel MIBI images for indicated markers in representative A7 (red) and non-A7 DLBCL cases (gray). The white bar in the bottom-left image is 95um.

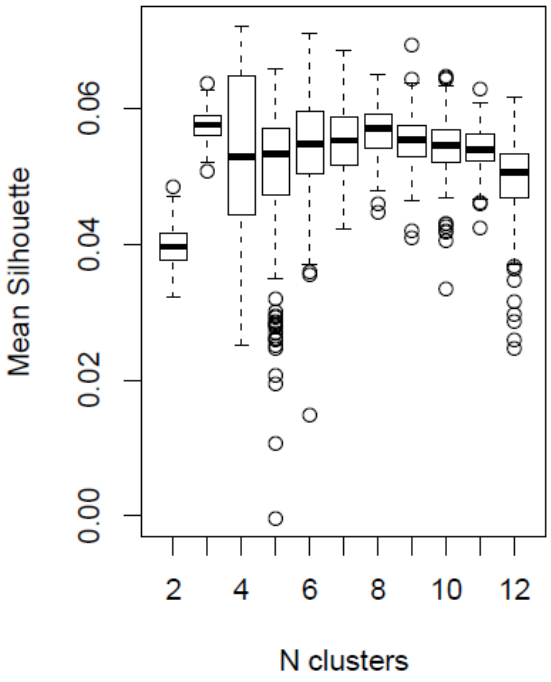


Supplemental Figure S22 – Effect of lenalidomide treatment in mouse imaging data. A) H&E staining of spleen tissues collected from hCRBN mice that were not transplanted or were transplanted with eu-myc/hCRBN splenocytes. Red arrows indicate the presence of large B-cells. B,C) Intratumoral cell counts in non-transplanted mice and transplanted mice treated with vehicle or lenalidomide by multiplex immunofluorescence for indicated cell population. N=5.

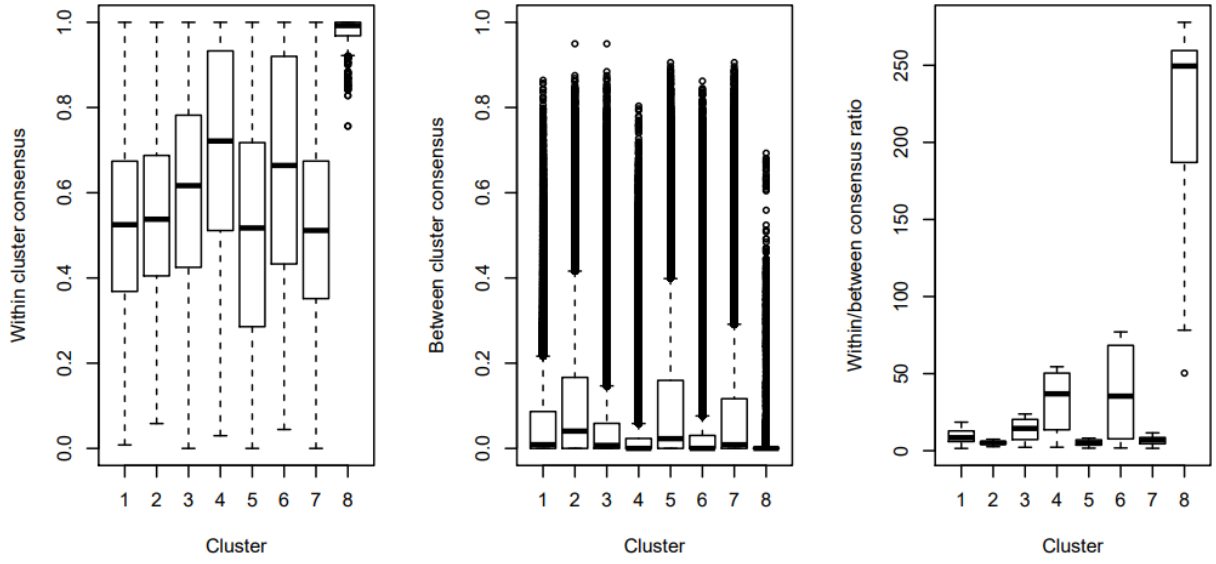


Supplemental Figure S23 – PCA of normalized datasets. A principal component projection of the normalized datasets shows no major batch effects when considering A) the full gene expression space, or more restricted gene expression spaces related to DLBCL biology, including B) the genes that define Reddy COO, and C) the genes that define the SubLymE clusters. Although there are some dataset-specific outlier samples along the PC1 dimension, the centered and overlapping multinomial density ellipses indicate that the cohorts are largely comparable with one another. Source data are provided as a Source Data file.

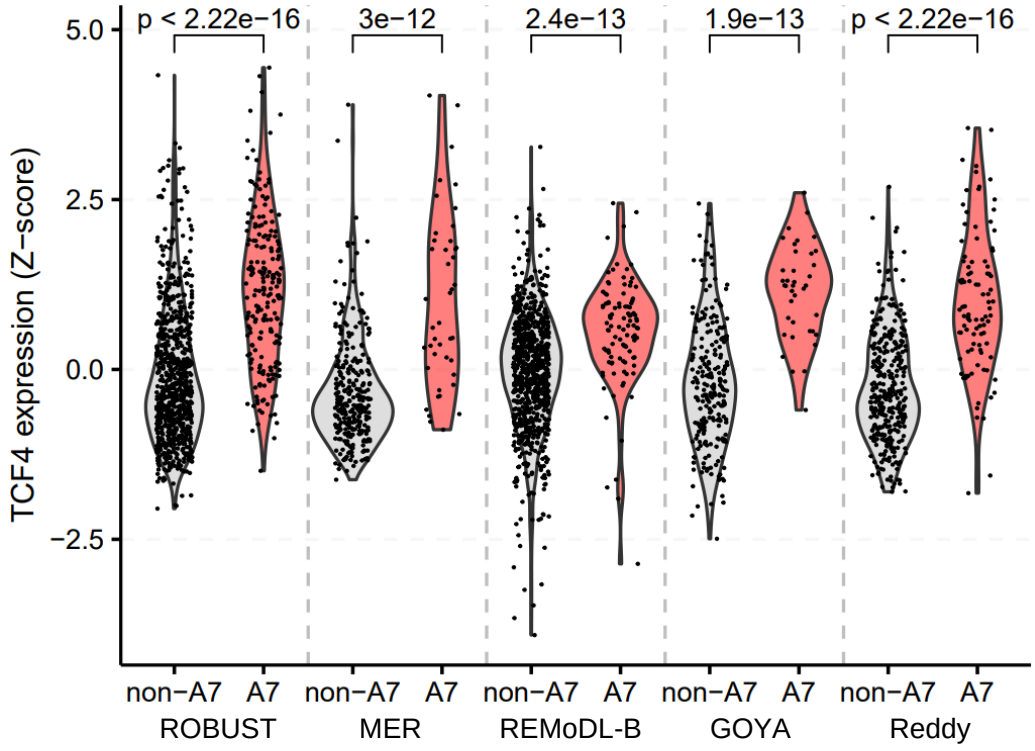
A)



B)



Supplemental Figure S24 – Cluster metrics. A) iClusterPlus was applied to the Discovery dataset 200 times with random sample and feature dropout. The distribution of mean silhouette metrics peaked at 8 clusters, indicating a good fit. B) Boxplots show consensus measured as the frequency with which cases in each subtype clustered with samples from the same cluster (within cluster consensus) or with samples from other clusters (between cluster consensus). Cluster A8 was found to be abnormally distinct, likely due to its comparatively poor technical quality, and was removed from further analysis. Source data are provided as a Source Data file.



Supplemental Figure S25 – TCF4 expression by A7. Shown is the gene expression value of TCF4 in each cohort, split by A7 versus non-A7 cases. In all cohorts, TCF4 was significantly upregulated in A7 cases. Source data are provided as a Source Data file.

Primer	Sequence (5' to 3')
MYC RT F	TTTCGGGTAGTGGAAAACCA
MYC RT R	CACCGAGTCGTAGTCGAGGT
CDC45 RT F	GTGGTCTTTGCCACCATGTCT
CDC45 RT R	TGTCCCTGAGCCATCCTTCT
MAD2L1 RT F	CGGGAGCGCCGAAATC
MAD2L1 RT R	TGCTGTTGATGCCGAATGA
MCM4 RT F	ATGAGAAACCTGAATCCAGAAG
MCM4 RT R	ATCAGCTGGGATGTCCTGAT
MCM7 RT F	GCGTCACTGGTATTTTCTTGC
MCM7 RT R	ATGGGCTTCCAGGTAGGTTT
MCM2 RT F	AGAGAGGCCTGGCTCTGG
MCM2 RT R	CACCACGTACCTTGTGCTTG
BUB3 RT F	GGAGTCCAGCCTGAAATACC
BUB3 RT R	CTCGGCCTTCAATAGAGCTT
18S RT F	CCGCAGCTAGGAATAATGGA
18S RT R	CCCTCTTAATCATGGCCTCA

Supplemental Table 1 – Primer sequences used