OMTON, Volume *32*

Supplemental information

Bulk and single-cell transcriptomics identify gene signatures of stem cell-derived NK cell donors with superior cytolytic activity

Amanda A. van Vliet, Mirjam G.C.N. van den Hout, Daniëlle Steenmans, Adil D. Duru, Anna-Maria Georgoudaki, Tanja D. de Gruijl, Wilfred F.J. van IJcken, Jan Spanholtz, and Monica Raimo

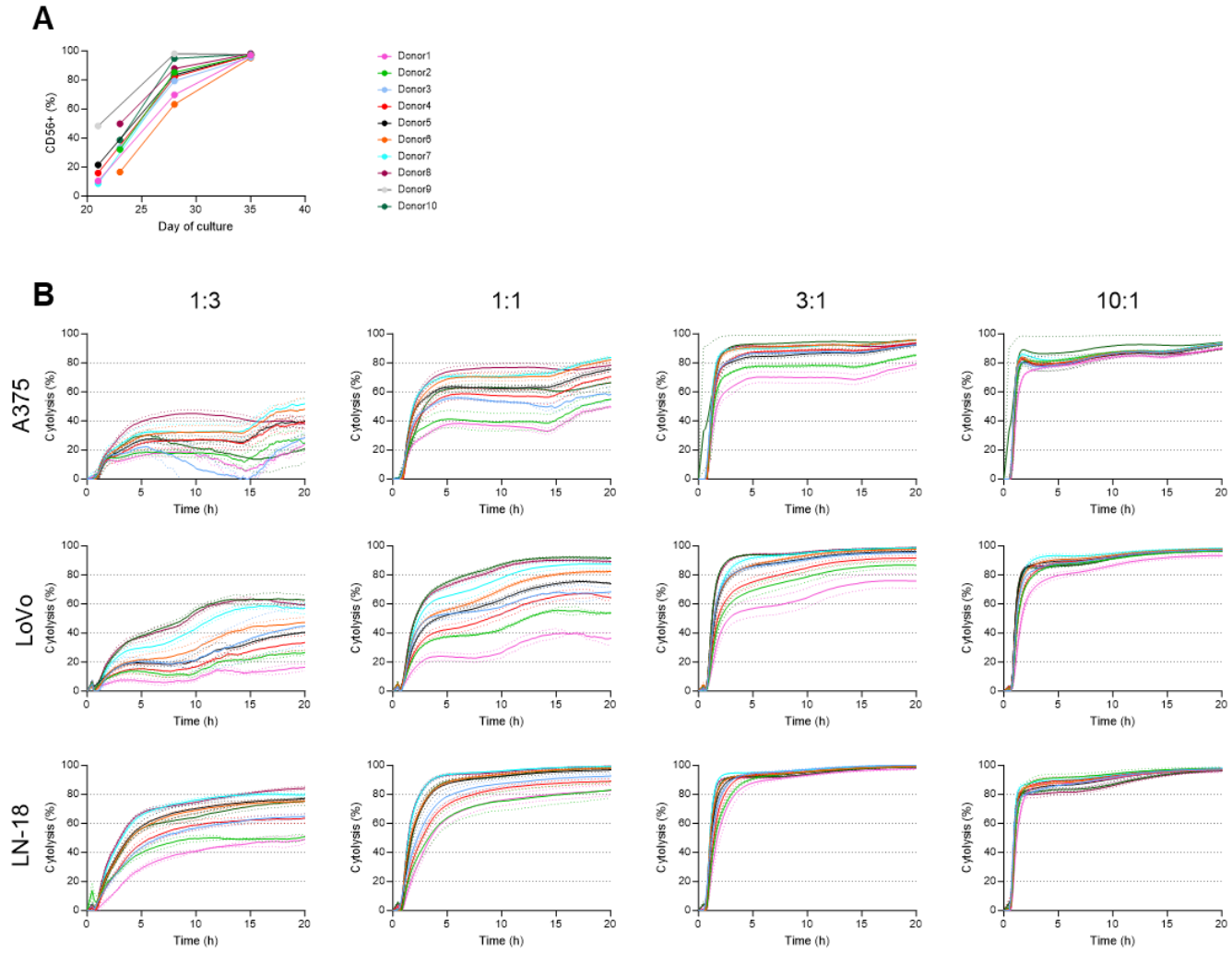

Figure S1. Differentiation and functional assessment of NK donors. **(A)** NK differentiation from progenitor cells between days 21-35 of ex vivo cell culture, expressed as percentage of $CD56^{+}$ cells (n=10 donors). **(B)** Impedance-based kinetic cytolysis assessment against A375, LoVo and LN-18 tumor cell lines at multiple E:T ratios (n=9 donors). Each line indicates 1 donor, shown as mean (solid line) ± SD (dotted lines) of technical triplicates (dotted line). SD: standard deviation.

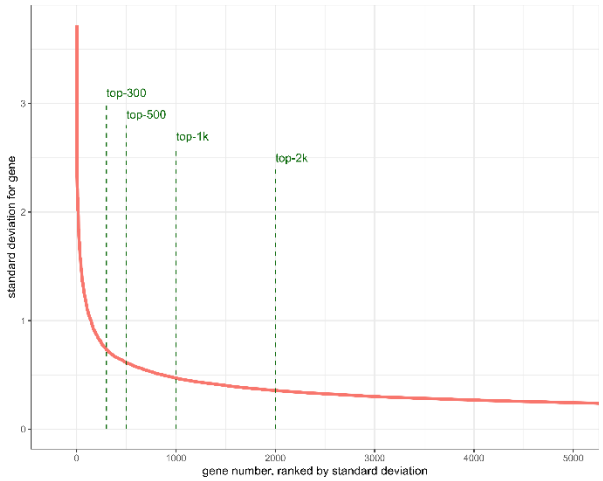

Figure S2. **Quantitative analysis of variable features for hierarchical clustering of bulk RNA-Seq data.** Elbow plot of the standard deviation for each gene detected with bulk RNA-Seq. The top 1000 genes (top-1k) captured the majority of the variation, therefore they were used for hierarchical clustering.

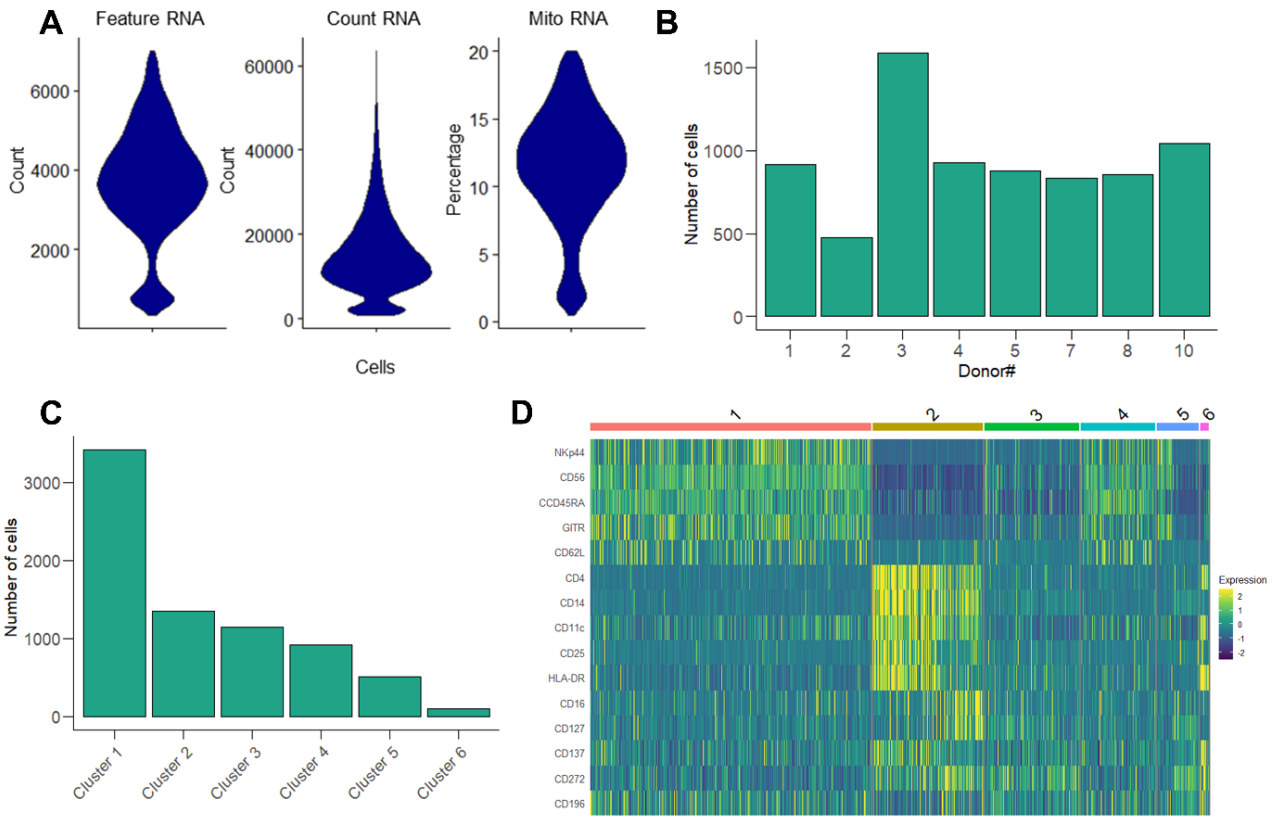

Figure S3. QC metrics of single-cell RNA-Seq libraries and expression of surface antigens, assessed with AbSeq profiling. (A) Quality control of scRNA-Seq libraries. Violin plots show the number of feature RNAs per cell (number of genes expressed), the total count of RNA molecules per cell (library size) and the mitochondrial gene proportion (%) per cell. Low-quality cells were defined as expressing fewer than 7000 genes and having a mitochondrial DNA content above 20%, then removed. **(B)** Distribution of the number of cells analyzed per donor after QC analysis. **(C)** Distribution of the number of cells per cluster after QC analysis. **(D)** Heatmap showing the AbSeq-based surface antigen enrichment in each cluster, identified by differential expression analysis. QC: quality control; sc: single-cell.

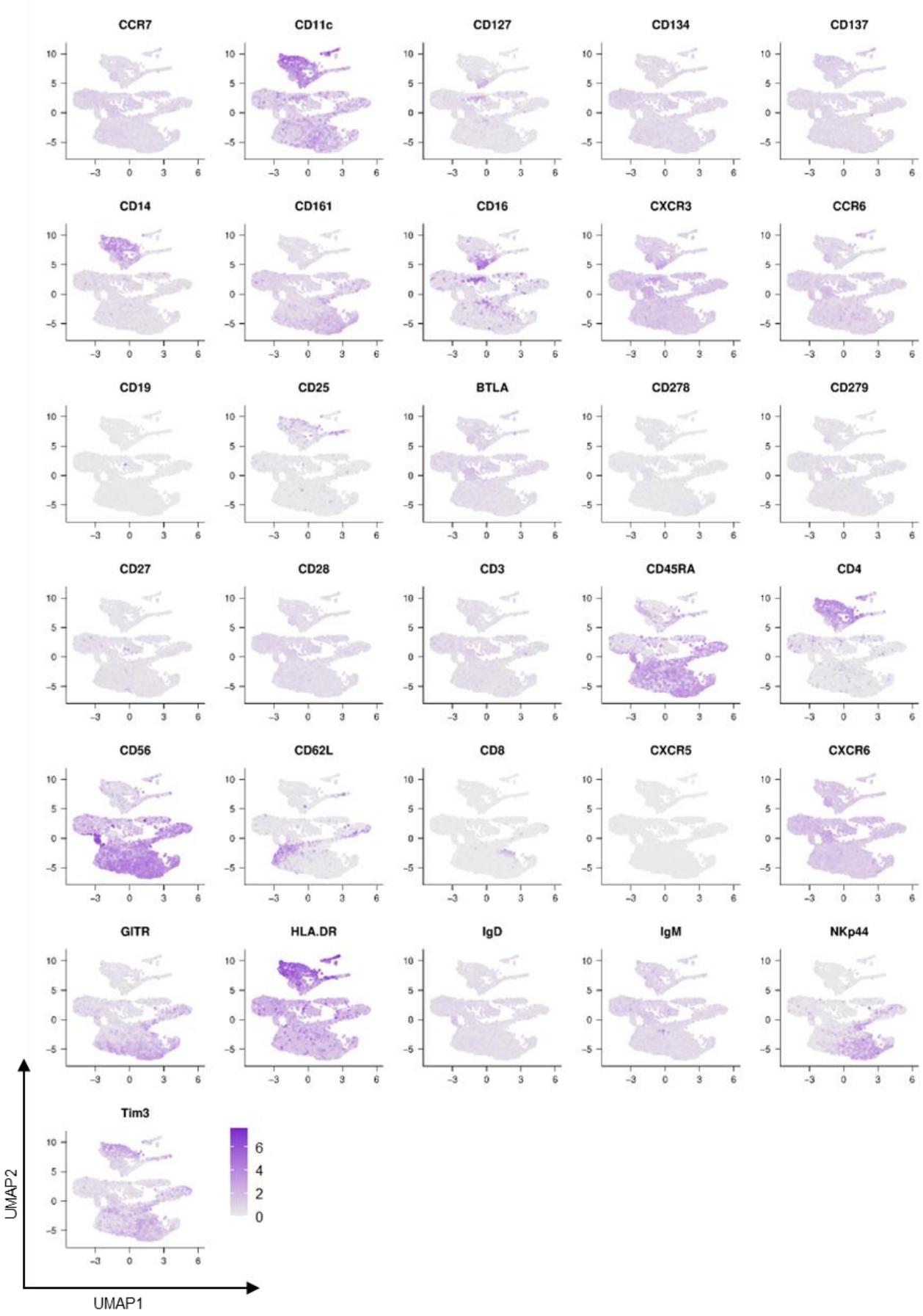

Figure S4. Expression and distribution of surface proteins, assessed via AbSeq profiling and single-cell RNA-Seq. UMAP of the detection of all 31 AbSeq antibodies used in the study. UMAP: Uniform manifold approximation and projection.

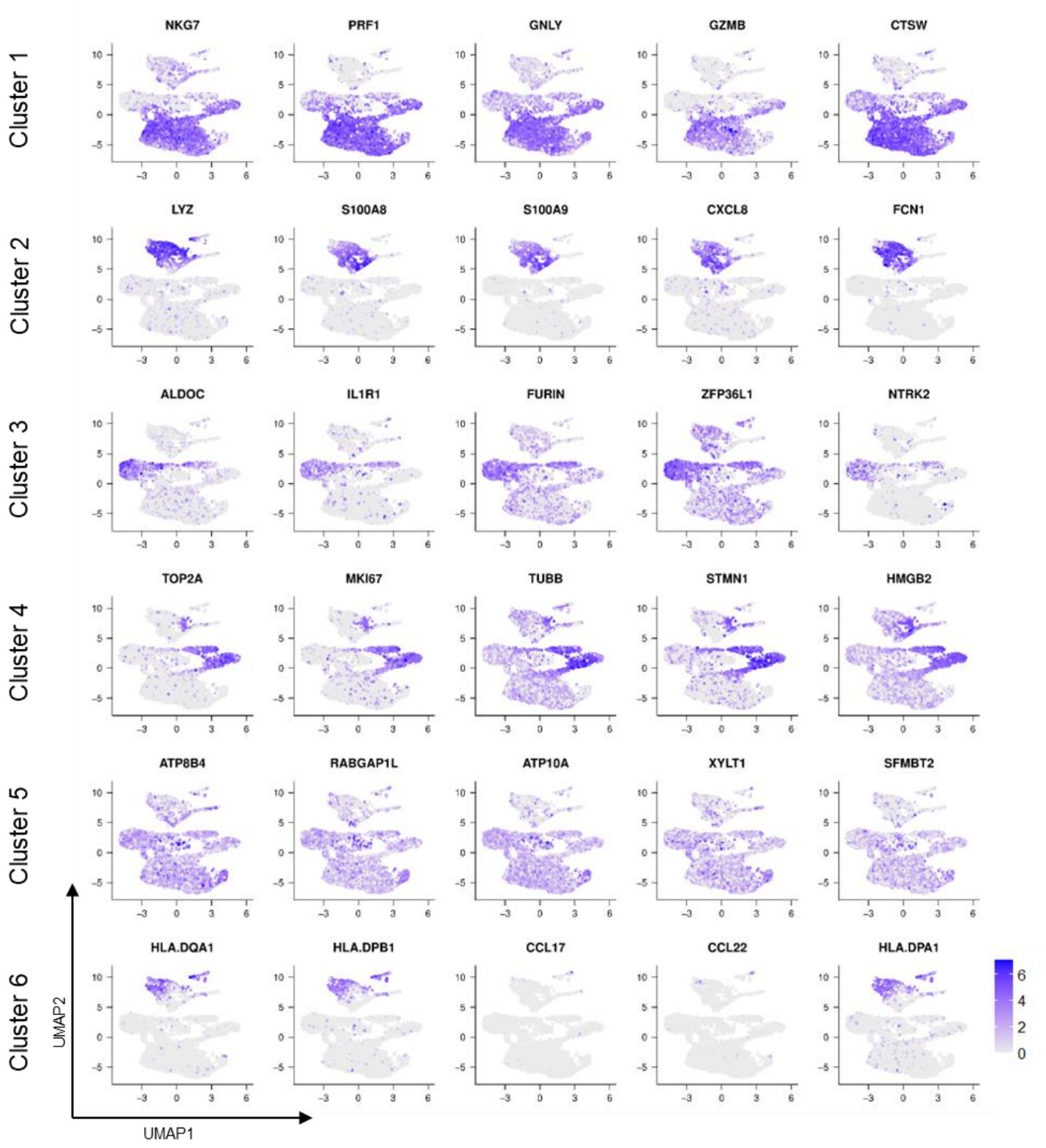

Figure S5. Expression and distribution of the top 5 markers of each cluster, identified with single-cell RNA-Seq. UMAP of the 5 highest-ranking genes distinguishing each cluster, identified by differential expression analysis. UMAP: Uniform manifold approximation and projection.

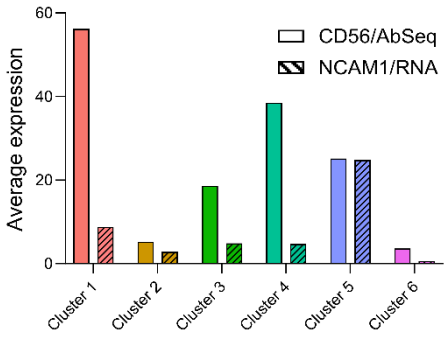

Figure S6. Cluster expression of CD56/NCAM1, analyzed with single cell RNA-Seq. Average CD56 surface protein expression per cluster from AbSeq profiling and average NCAM1 RNA expression per cluster from gene profiling.

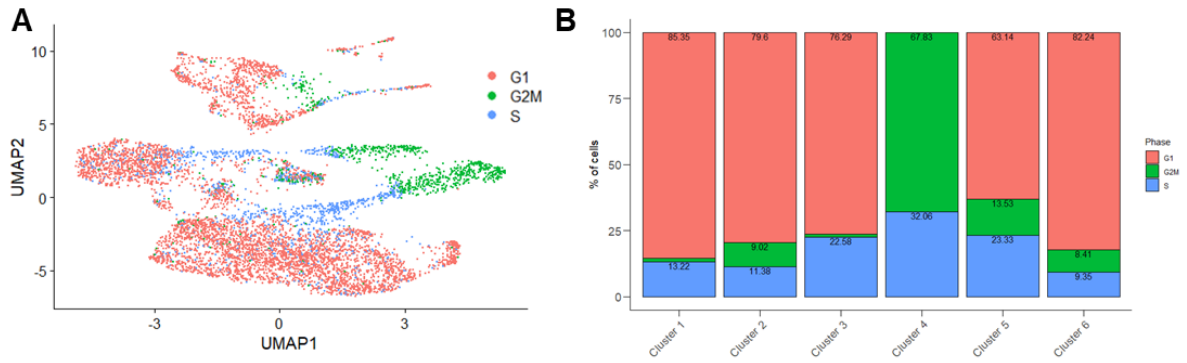

Figure S7. Cell cycle analysis via single cell RNA-Seq. **(A)** UMAP visualization of cell cycle (G1, G2/M, S) distribution, after scoring each cell based on the expression of phase-specific genes. **(B)** Relative distribution of cell cycle phases per cluster, expressed as percentage of total.