Supplementary Table S1-S4, Supplementary Figures S1-S11

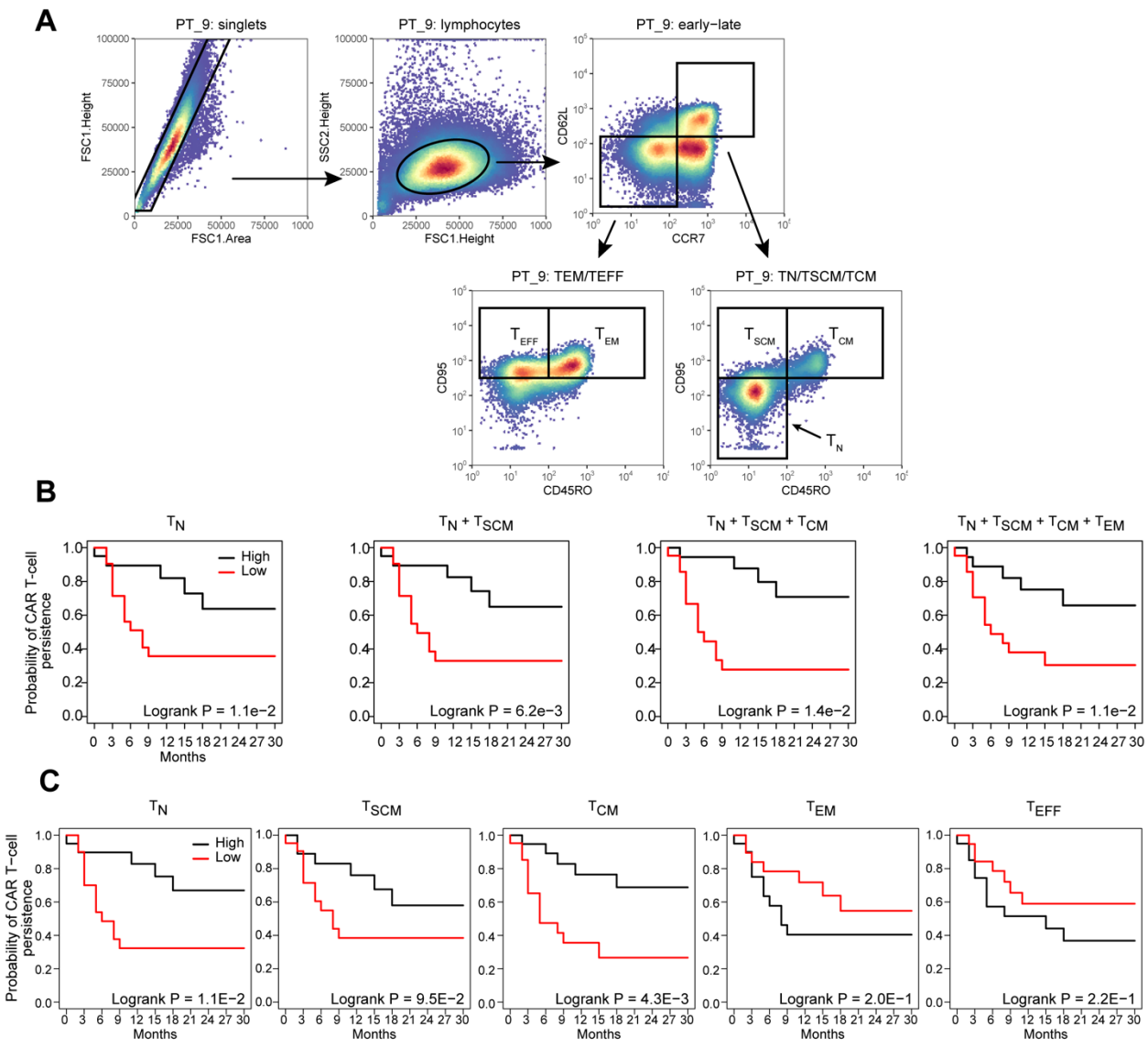
Characteristic	Number of Patients	Percent of Patients
Primary malignancy		
B-cell Acute Lymphoblastic Leukemia	70	98.6% [91.3–99.9]
Hodgkin Lymphoma	1	1.4% [0.1–8.7]
Age		
0-1	1	1.4% [0.1–8.7]
2-10	31	43.7% [32.1–55.9]
11-17	23	32.4% [22.0–44.7]
≥ 18	16	22.5% [13.8–34.3]
Sex		
Female	34	47.9% [36.0–60.0]
Male	37	52.1% [40.0–64.0]
Duration of B-cell Aplasia		
< 6 months	27	38.0% [27.0-50.4]
≥ 6 months	33	46.5% [34.7-58.6]
Censored before 6 months	11	15.5% [8.4-26.5]

Supplementary Table S1. Clinical characteristics of 71 children and young adults enrolled to receive anti-CD19 CAR T-cell therapy. T-cells from 71 patients were acquired by apheresis with the intent to manufacture and deliver anti-CD19 CAR T-cell therapy. Mean values and 95% confidence intervals are shown.

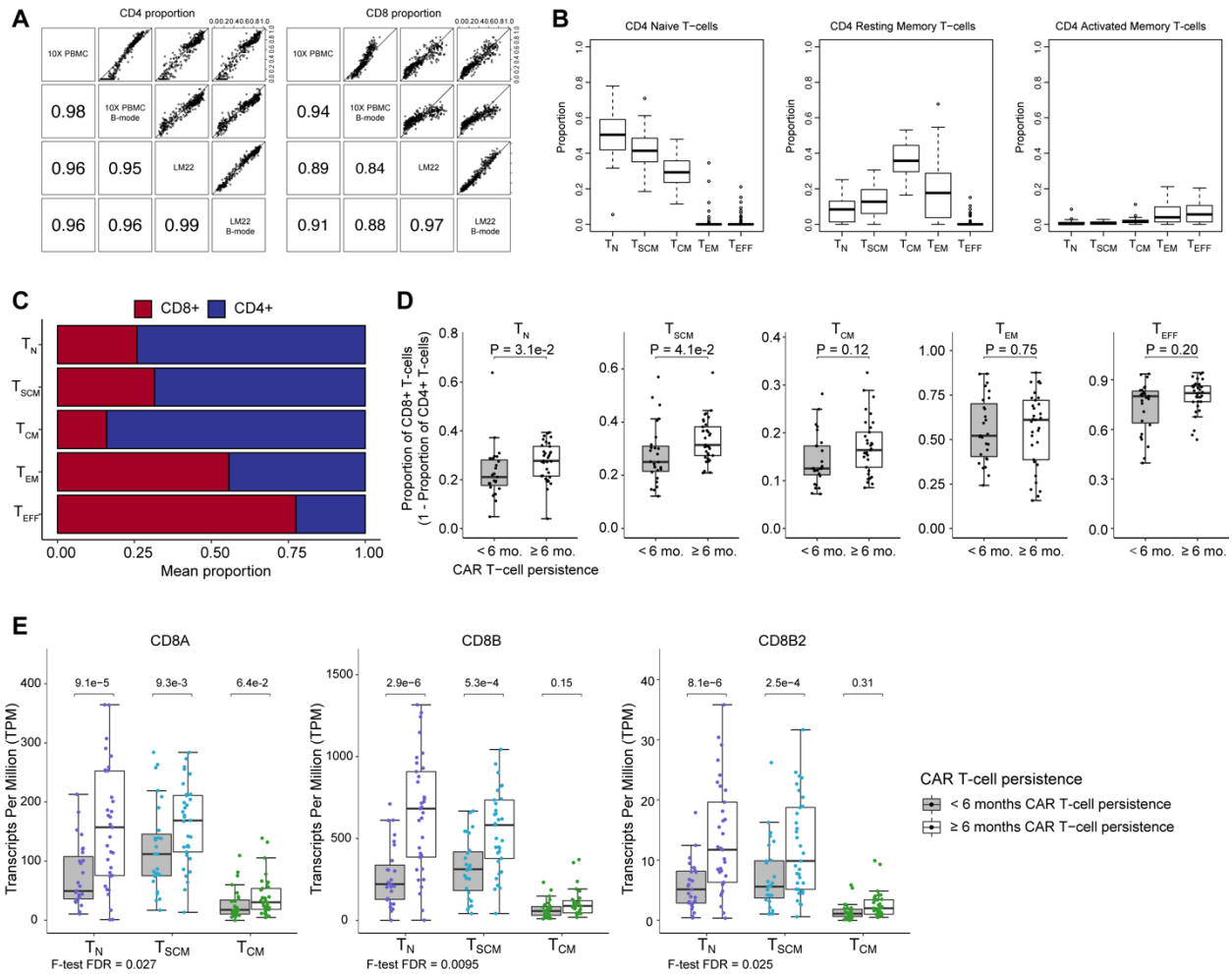
Supplementary Table S2. Differentially expressed genes between T-cell subtypes and clinical CAR T-cell persistence groups. Differential expression analysis was performed on RNA-Seq data with Limma. The first sheet contains genes ranked by p-value from the ANOVA model with the null hypothesis of equal expression across the five T-cell subsets. Subsequent sheets show the output of a mixed effects model including T-cell subtype and clinical CAR T-cell persistence group. Sheets 2-10 consist of comparisons between T-cell subsets, and sheets 11-19 consist of comparisons between patients with long (≥ 6 month) vs short (< 6 month) CAR T-cell persistence.

Supplementary Table S3. Gene sets for the TCF7 regulon and interferon response. For each gene set, three annotations are provided: the Ensembl gene id, the Entrez gene id, and HGNC symbol based on annotation from Ensembl version 94.

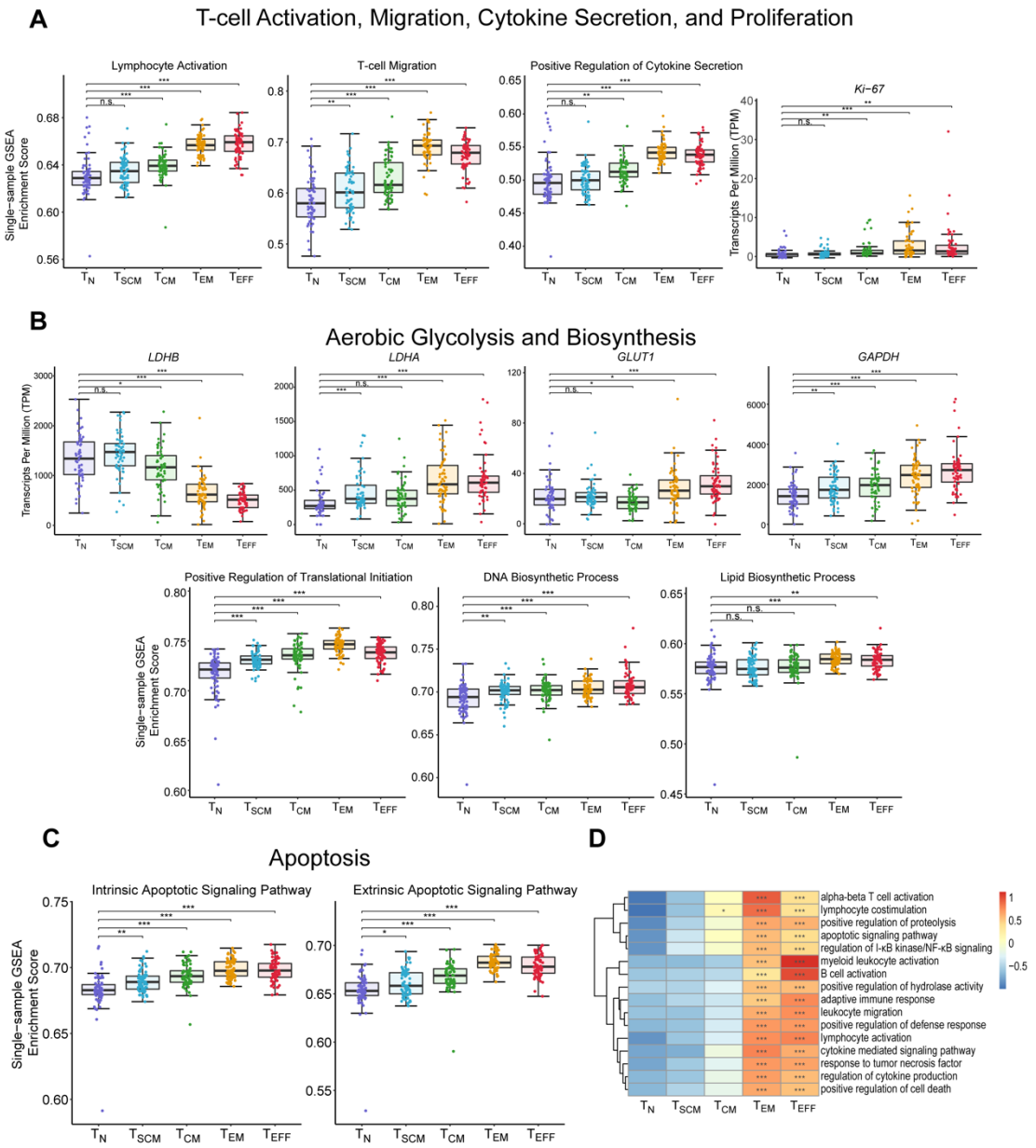
Supplementary Table S4. Primers for 3C-qPCR and RT-qPCR.



Supplemental Figure S1. T-cell subtypes associate with clinical response.
(**A**) Representative flow cytometry plot for sorting T-cell subpopulations. (**B**) Association of relative T-cell proportion with clinical CAR T-cell persistence, as assessed by duration of B-cell aplasia (BCA). Patients were median-dichotomized based on proportions of each T-cell subset, or summed proportions of T-cell subsets. (**C**) Association between proportion of T_N, T_{SCM}, T_{CM}, T_{EM} and T_{EFF} at time of leukapheresis with long-term CAR T-cell persistence. P-values were computed using the log-rank test.

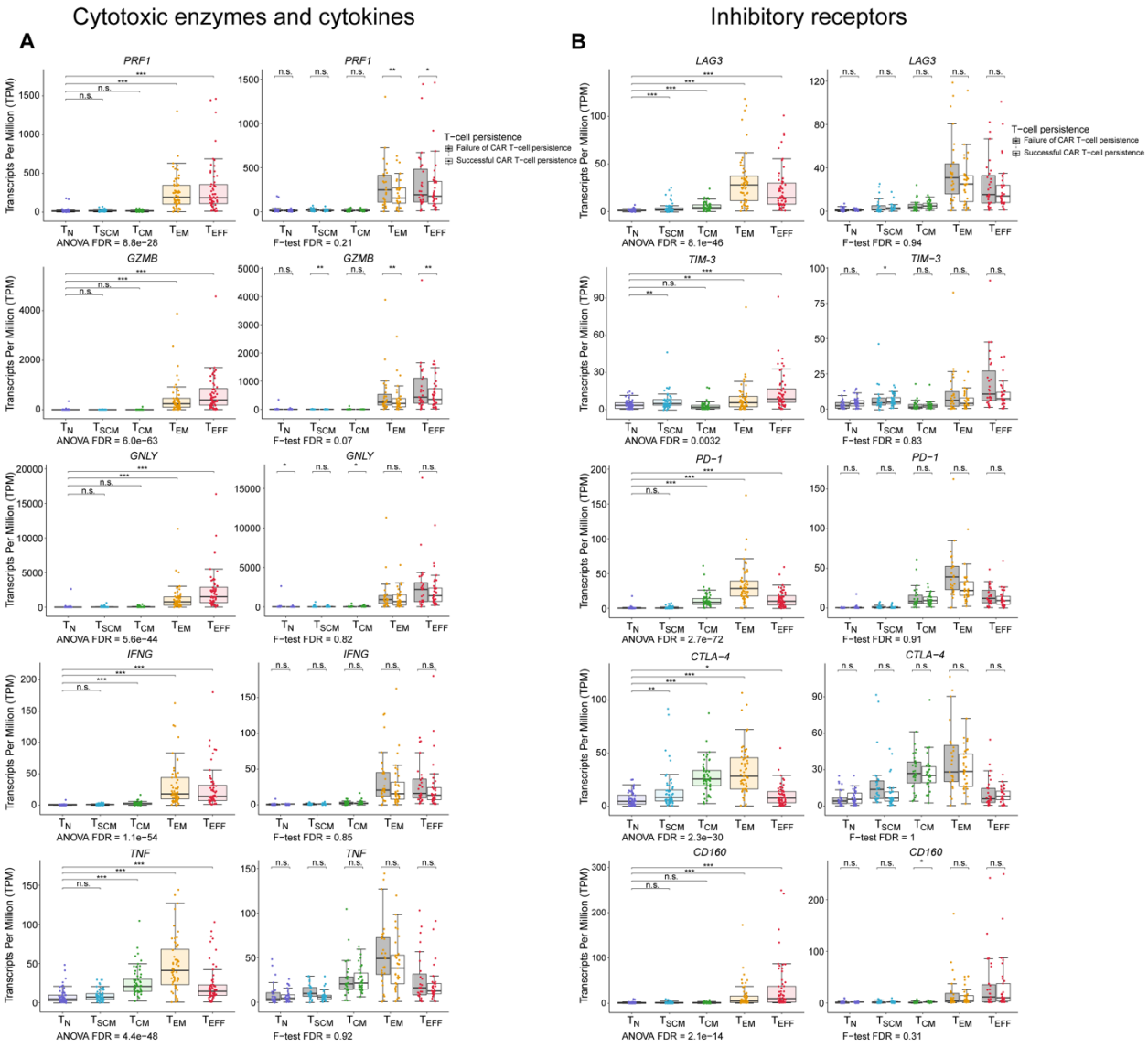


Supplementary Figure S2. Deconvolution of CD4$^+$/CD8$^+$ T cell proportions. (A) Assessment of robustness in CIBERSORTx estimates of CD4 and CD8 proportion based on a bulk (LM22) and single-cell (10X) reference, with and without the B-mode parameter. The lower diagonal contains Pearson correlation coefficients. (**B**) Association with annotated T-cell subsets in the LM22 reference set with T_N, T_{SCM}, T_{CM}, T_{EM}, and T_{EFF} populations. (**C**) Relative proportions of CD8$^+$ and CD4$^+$ T-cells estimated by CIBERSORTx compared across T-cell subtypes (ANOVA P < 2.2e-16). (**D**) Association between the proportion of CD8$^+$ T-cells and CAR T-cell persistence. Patients were dichotomized into those with long-term CAR T-cell persistence (≥ 6 months, n=33) and those with failed CAR T-cell persistence (< 6 months, n=27). Among T_N and T_{SCM}, greater proportions of CD8$^+$ T-cells associate with longer CAR T-cell persistence; conversely, greater proportions of CD4$^+$ T-cells among T_N and T_{SCM} associate with shorter CAR T-cell persistence. P-values were computed using the Wilcoxon rank-sum test, and multiple testing correction was performed using the Benjamini-Hochberg procedure. (**E**) Expression of CD8 genes in T_N, T_{SCM}, and T_{CM} populations compared between patients with long-term (≥ 6 month) vs failed (< 6 month) CAR T-cell persistence.

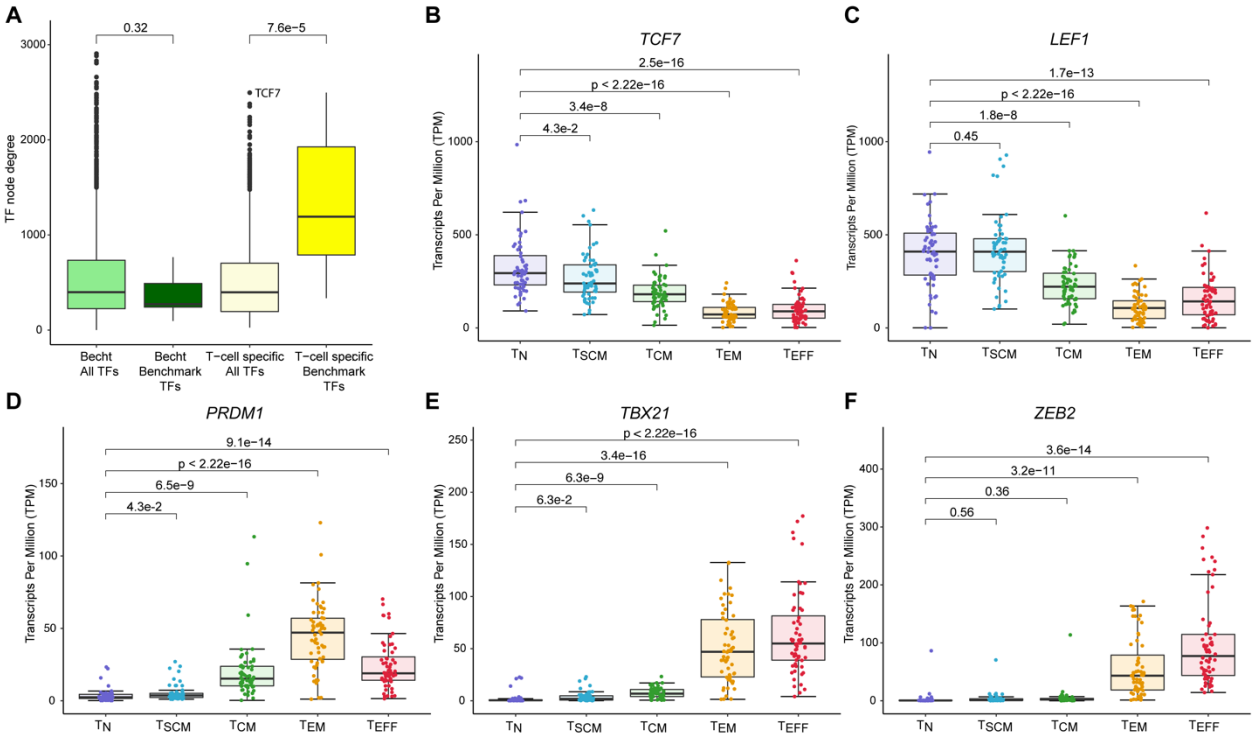


Supplemental Figure S3. T-cell subtypes differ in activation and proliferation pathways, metabolic reprogramming, and apoptotic pathways.

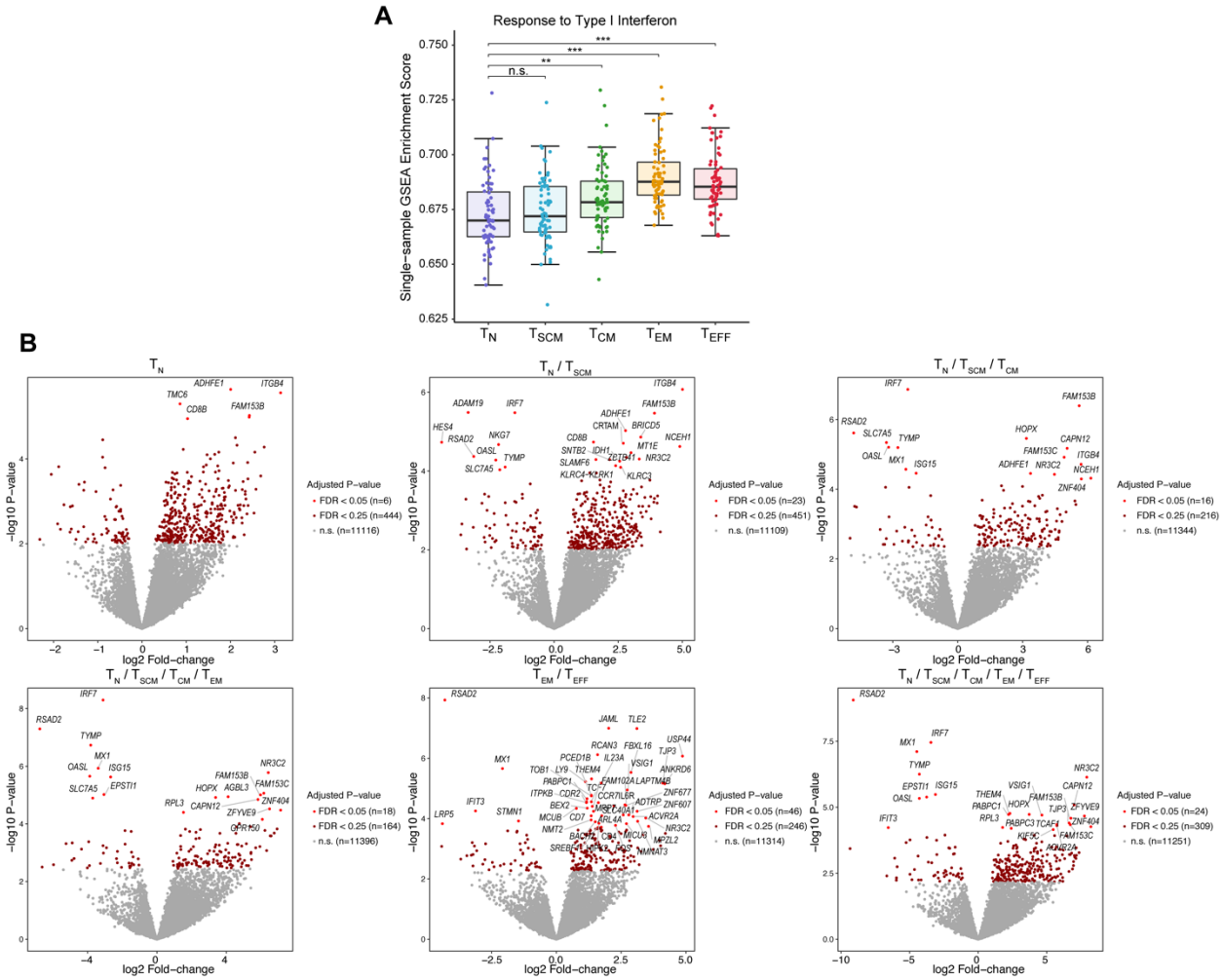
(**A**) Single-sample Gene Set Enrichment Analysis on representative pathways associated with T-cell subtype, and differential expression of *Ki-67*. (**B**) Single-sample Gene Set Enrichment Analysis and differential gene expression for key markers of aerobic glycolysis and biosynthetic pathways. Pairwise statistical significance was assessed through Welch's t-test. (**C**) Single-sample Gene Set Enrichment Analysis on intrinsic and extrinsic apoptotic signaling pathways associated with T-cell subtype. Pairwise statistical significance was assessed through Welch's t-test. (**D**) Heatmap of z-score normalized single-sample gene set enrichment scores of top significantly differentially activated pathways between T-cell subsets. Shown are the gene sets from Fig. 1D that were in the Gene Ontology Biological Pathways database. Asterisks show FDR-corrected p-values. ***: $p < 0.001$; **: $p < 0.01$; *: 0.05, n.s.: $p \geq 0.05$.



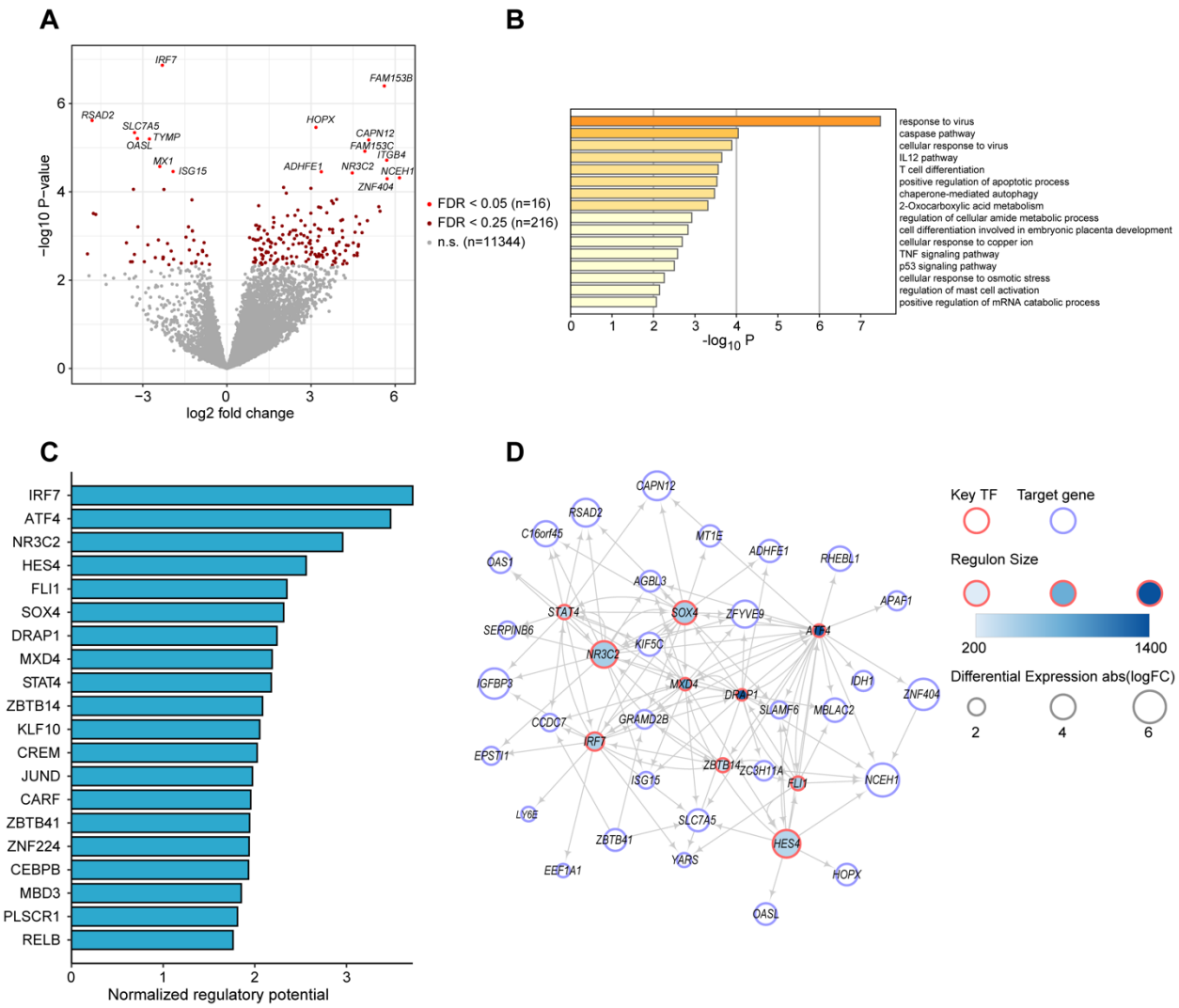
Supplemental Figure S4. Expression of inhibitory receptors, cytotoxic enzymes, and cytokines across T-cell subtypes and between clinical CAR T-cell persistence groups. (**A**) Expression of selected cytokines and genes involved in cytotoxicity. (**B**) Expression of the inhibitory receptors *LAG-3*, *TIM-3*, *PD-1*, *CTLA-4*, and *CD160* differed between T-cell subsets and were highly expressed among T_{EM} and T_{EFF}, but did not differ not between patients with long-term (≥ 6 month) vs failed (< 6 month) CAR T-cell persistence. Pairwise statistical significance between T-cell subtypes was assessed through Welch's t-test. Pairwise statistical significance between clinical CAR T-cell persistence groups was assessed through the Limma mixed-effects model. ***: $p < 0.001$; **: $p < 0.01$; *: 0.01, n.s.: $p \geq 0.05$.



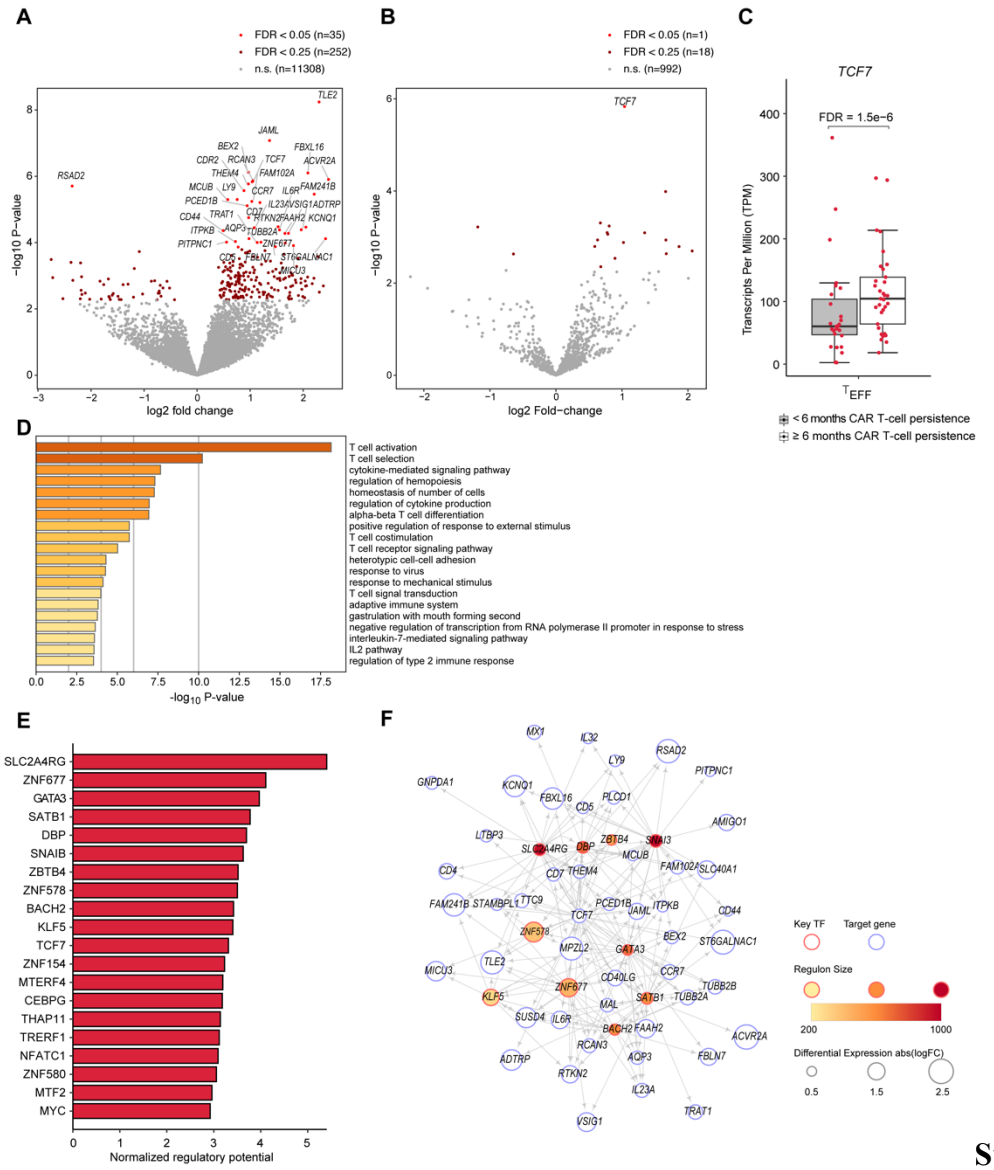
Supplemental Figure S5. T-cell specific network analysis captures key transcriptional regulators of T-cell subset. (**A**) Validation of the T-cell network construction method. Network construction was performed using a compendium of non-T-cell immune populations in Becht et al. (green boxes), as well as with our T-cell RNA-Seq data (yellow boxes). The distribution of TF node degrees was compared between the null population (light-colored boxes) and a set of benchmark TFs consisting of a set of known T-cell transcriptional regulators in Chang et al. (dark-colored boxes). Pairwise comparisons were performed with the Wilcoxon rank-sum test. (**B-F**) Expression of predicted key regulators of T-cell state: *TCF7*, *LEF1*, *PRDM1*, *TBX21*, and *ZEB2*. Pairwise statistical significance was assessed with Welch's t-test.



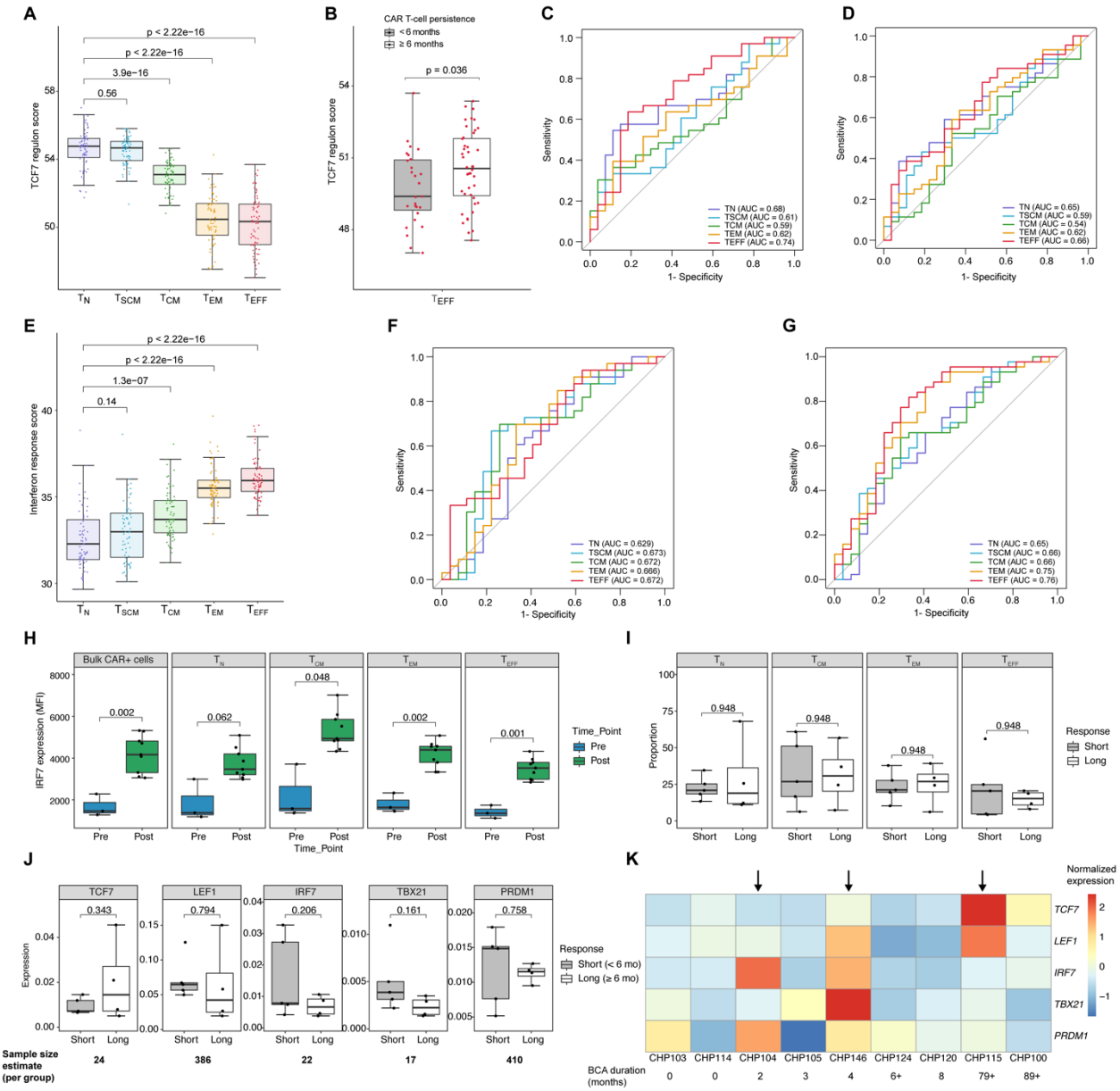
Supplementary Figure S6. Interferon response genes are up-regulated in patients with poor CAR T-cell persistence across T-cell subtypes. (A) Single-sample Gene Set Enrichment Analysis comparing the enrichment score of the Gene Ontology Type I Interferon response pathway across T-cell subtypes. P-values were computed using Welch's t-test. (**B**) Volcano plots showing differentially expressed genes in patients with long-term vs short-term CAR T-cell persistence using gene expression data from specific T-cell subsets.



Supplementary Figure S7. Interferon response genes are up-regulated among patients with poor CAR T-cell persistence in naive and early memory T-cells, and IRF7 is the top predicted transcriptional regulator in these subsets. (A) Volcano plot showing differentially expressed genes in patients with long-term vs short-term CAR T-cell persistence using gene expression data from T_N, T_{SCM}, and T_{CM} subsets. (**B**) Enriched pathways among differentially expressed genes between patients with long-term vs short-term CAR T-cell persistence (FDR < 0.25). (**C**) Top 20 predicted key transcription factors (TFs) associated with the long-term vs short-term persistence ranked by normalized regulatory potential (FDR < 0.05). (**D**) Transcriptional regulatory network of top 10 predicted key TFs from (E) and top 50 target DEGs.

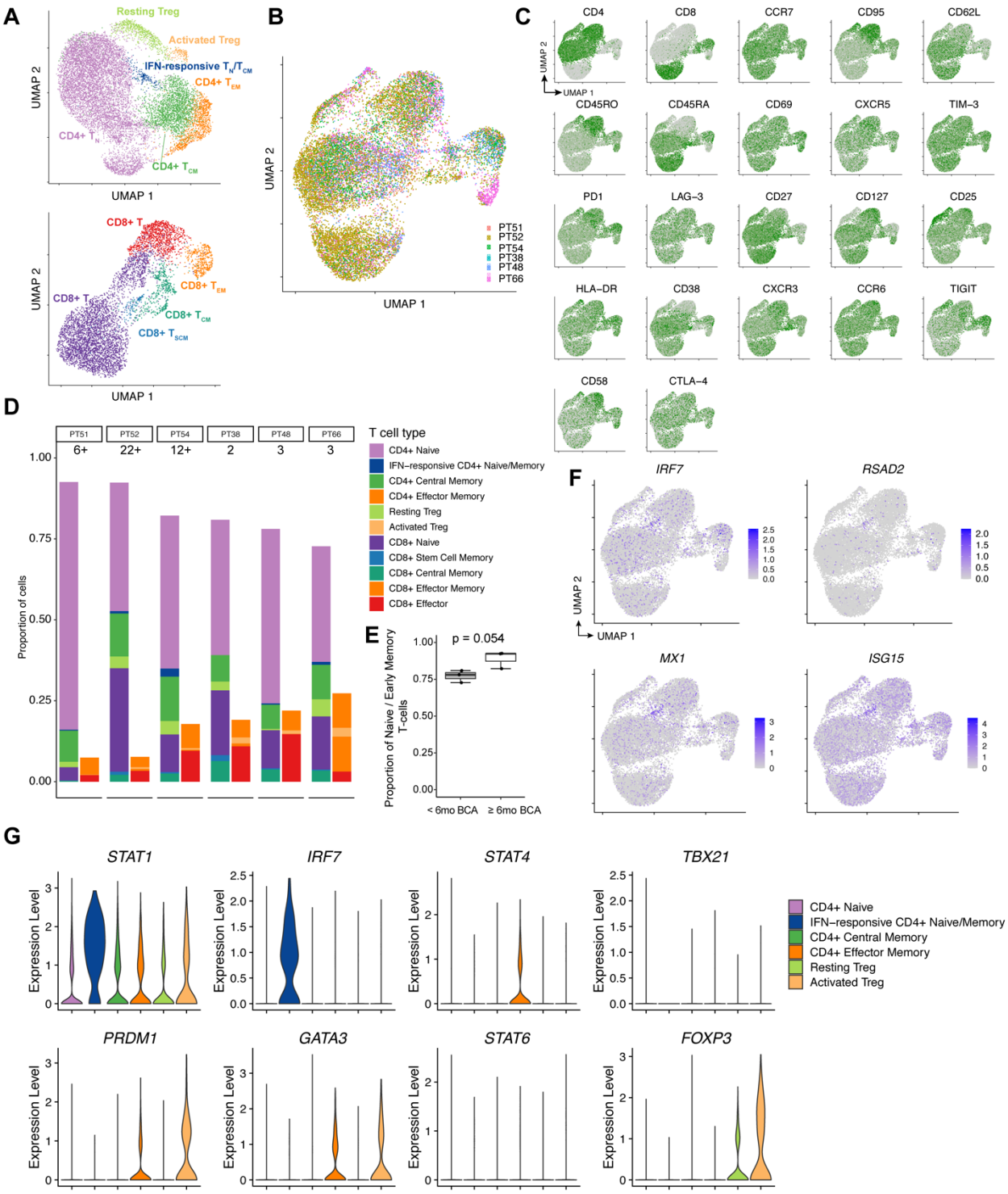


Supplementary Figure S8. Network analysis of T_{EFF} alone highlights TCF7 as a regulator of long CAR T-cell persistence. (A) Volcano plot showing differentially expressed genes in patients with long-term vs short-term CAR T-cell persistence using gene expression data from the T_{EFF} subset. **(B)** Volcano plot showing differentially expressed transcription factors in patients with long-term vs short-term CAR T-cell persistence using gene expression data from the T_{EFF} subset. *TCF7* was the most significantly up-regulated transcription factor among patients with long-term CAR T-cell persistence (FDR=0.0015). (**C**) Differential expression of *TCF7* between patients with long-term (≥ 6 month) vs short-term (< 6 month) CAR T-cell persistence. Statistical significance was assessed with the FDR-adjusted P-value from the Limma mixed-effects interaction model. (**D**) Enriched pathways among differentially expressed genes between patients with long-term vs short-term CAR T-cell persistence (FDR < 0.25). (**E**) Top 20 predicted key transcription factors (TFs) associated with the long-term vs short-term persistence ranked by normalized regulatory potential (FDR < 0.05). **(F)** Transcriptional regulatory network of top 10 predicted key TFs from (E) and top 50 target DEGs.



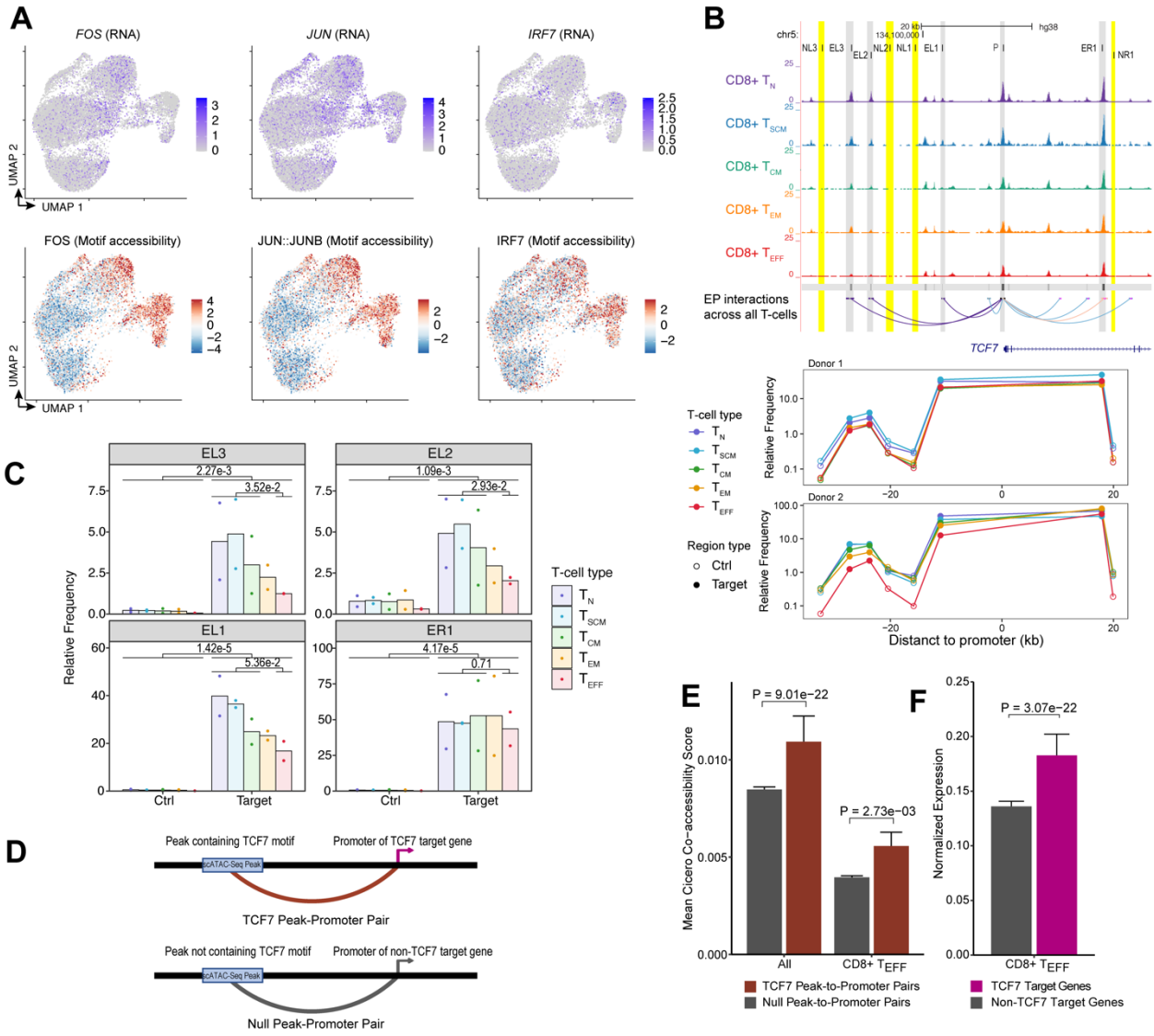
Supplementary Figure S9. Internal validation of prognostic gene signatures. (A) Association of the *TCF7* network score with T-cell subset, and with (**B**) clinical CAR T-cell persistence in T_{EFF}. Pairwise comparisons were assessed for significance with Welch's t-test. (**C**) Receiver-operating characteristic (ROC) curves to validate the TCF7 regulon gene signature using leave-one-patient out and (**D**) leave-one-T-cell-type-out cross-validation. Each curve represents the performance of this gene signature when evaluated on left-out samples. (**E**) Association of the interferon response gene signature with T-cell subsets. Pairwise comparisons were assessed for significance with Welch's t-test. (**F**) Receiver-operating characteristic (ROC) curves to validate of the interferon response gene signature discovery method using leave-one-patient-out cross validation and (**G**) leave-one-T-cell-type-out cross validation. Each curve represents the performance of this gene signature when evaluated on left-out samples. (**H**) IRF7 expression assessed by median fluorescent intensity (MFI) from FACS comparing 3 pre-manufacture T-cell

samples and 11 post-manufacture CAR T-cell infusion products from a separate group of patients from those studied in our bulk RNA-Seq analysis. T-cell subsets were defined based on CD62L and CD45RO gating. T_N: $CD62L^+CD45RO^-$; T_{CM}: $CD62L^+CD45RO^+$; T_{EM}: $CD62L^-CD45RO^+$; T_{EFF}: $CD62L^-CD45RO^-$. (**I**) Assessment of CAR T-cell subset proportion from FACS analysis comparing five patients with short clinical CAR T-cell persistence and four with long-term CAR T-cell persistence. Statistical significance was assessed using Welch's t-test and Benjamini-Hochberg multiple-testing correction. The equal FDR-adjusted p-values in some plots were a result of ties in the Benjamini-Hochberg procedure. (**J**) Assessment of transcription factor expression by RT-qPCR relative to *ACTB* control between the five patients with short clinical CAR T-cell persistence and four with long-term CAR T-cell persistence. Statistical significance was assessed using Welch's t-test. Sample size estimates per group are shown for a significance level of 0.05 and 80% power. FDR-adjusted p-values were 0.572, 0.794, 0.515, 0.515, and 0.794 respectively. (**K**) Heatmap indicating transcription factor expression by RT-qPCR with arrows indicating CHP104 (a patient with failed CAR T persistence at 2 months and high *IRF7* and *PRDM1* expression); CHP146 (a patient with failed CAR T persistence at 4 months and high *TBX21* expression); and CHP115 (a patient with > 6.5 years of CAR T persistence and high *TCF7*/*LEF1* expression in the CAR T infusion product). For plotting purposes, expression values for each gene were z-score normalized across patients.



Supplementary Figure S10. Analysis of CITE-Seq data. (A) UMAP of $CD4^+$ (top panel) and $CD8^+$ (bottom panel) T-cells, colored by T-cell cluster. **(B)** UMAP of 17,750 cells from CITE-Seq data, colored by patient. **(C)** UMAP of CITE-Seq data, colored by cell surface protein expression of 22 markers based on CITE-Seq antibody-derived tags. **(D)** Proportions of T-cell subsets within each patient, determined by relative numbers of cells per cluster. The number below each patient identifier is the duration of CAR T-cell persistence for each respective

patient, with censored values indicated with a plus symbol, and T-cell types are stratified by naive and early memory (purple, blue, and green shades) compared to effector memory and effector T-cells (orange and red shades). **(E)** Boxplot showing the proportion of naive and early memory T-cells ($CD4^+$ and $CD8^+$ T_N, T_{SCM}, T_{CM}, and resting T_{reg}) among the three patients with ≥ 6 months BCA compared to the three patients with < 6 months BCA. Statistical significance was assessed with Welch's t-test. **(F)** Normalized RNA expression of representative type I interferon response genes. **(G)** Expression of transcription factors associated with $CD4^+$ subsets in our single-cell data. *STAT1* and *IRF7* were up-regulated among the IFN-responsive Naive/Memory subset, with low *STAT1* expression among other subsets; *STAT4*, *PRDM1*, and *GATA3* were expressed among the $CD4^+$ T_{EM} subset; and *PRDM1*, *GATA3*, and *FOXP3* were expressed among the T_{reg} cells.



Supplementary Figure S11. Analysis of scATAC-Seq data to uncover gene regulatory interactions. (A) Normalized RNA expression and chromVAR motif accessibility of signaling transcription factors *FOS*, *JUN*, and *IRF7*. **(B)** Top, schematic for primers used in the chromosome conformation capture (3C) experiments of predicted enhancer-promoter interactions involving the *TCF7* promoter. Predicted enhancers (EL3, EL2, EL1, ER1) are highlighted in yellow. Adjacent negative control regions (NL3, NL2, NL1, NR1) and the *TCF7* promoter (P) are highlighted in grey. Bottom, 3C relative interaction frequencies of the tested interactions using T-cell samples from two healthy donors. **(C)** Bar plots for relative interaction frequencies from 3C experiment comparing each adjacent negative control region (Ctrl) and predicted enhancer (Target) as well as comparing E-P interactions between two groups of T cell subtypes (T_N+T_{SCM}+T_{CM} vs T_{EM}+T_{EFF}). Each point represents the mean of three technical replicates performed on one of two healthy donor samples sorted by T-cell subtype. Pairwise statistical significance was assessed using Welch's t-test. **(D)** Schematic indicating how *TCF7* and null peak-to-promoter pairs were determined. *TCF7* peak-promoter pairs were defined as pairs of chromatin peaks from Cicero for which one peak contained the *TCF7* motif and did not overlap with a promoter or exon, and the other peak overlapped with a promoter from a gene

within the *TCF7* regulon defined by our bulk RNA-Seq analysis. Null peak-promoter pairs were defined as pairs of chromatin peaks from Cicero for which one peak did not contain the *TCF7* motif and did not overlap with a promoter or exon, and the other peak overlapped with a promoter not within the *TCF7* regulon. **(E)** Enrichment of Cicero co-accessibility scores among *TCF7* peak-promoter pairs compared to null, both in the pan-T-cell analysis and when restricted to the T_{EFF} subset. In the pan-T-cell analysis (under the "All" label on the x axis), Cicero was run on all of the T-cells in the scATAC-Seq data, capturing regulatory interactions with and between T-cell subsets. In the T_{EFF} analysis, Cicero was run strictly on the $CD8^+$ T_{EFF} subset, capturing only interactions occurring within the T_{EFF} cells. Error bars indicate 95% confidence intervals of the mean estimate, and pairwise statistical significance was assessed using the Wilcoxon rank-sum test. **(F)** Normalized RNA expression of *TCF7* target genes compared to non-*TCF7* target genes based on the peak-promoter pairs defined for T_{EFF} as described in (B). Genes were included in this analysis if the RNA count was nonzero in greater than 1% of cells. Error bars indicate 95% confidence intervals of the mean estimate, and pairwise statistical significance was assessed using the Wilcoxon rank-sum test.