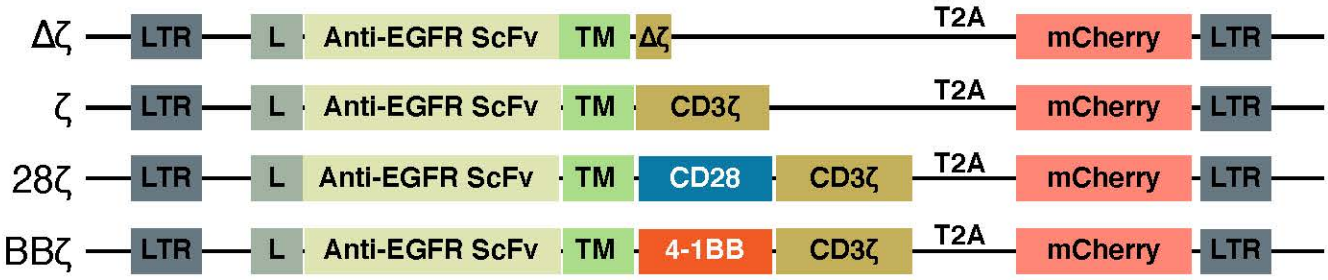
YMTHE, Volume *28*

Supplemental Information

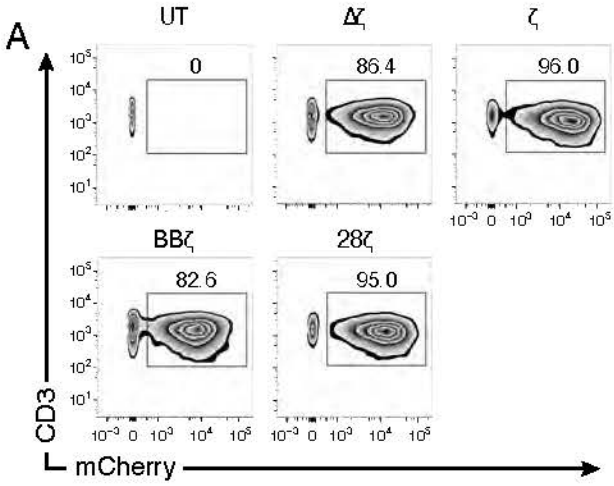
A Distinct Transcriptional Program in Human CAR T Cells Bearing the 4-1BB Signaling Domain Revealed by scRNA-Seq

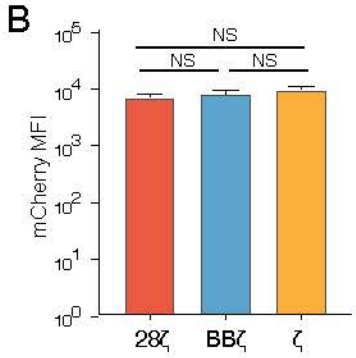
Angela C. Boroughs, Rebecca C. Larson, Nemanja D. Marjanovic, Kirk Gosik, Ana P. Castano, Caroline B.M. Porter, Selena J. Lorrey, Orr Ashenberg, Livnat Jerby, Matan Hofree, Gabriela Smith-Rosario, Robert Morris, Joshua Gould, Lauren S. Riley, Trisha R. Berger, Samantha J. Riesenfeld, Orit Rozenblatt-Rosen, Bryan D. Choi, Aviv Regev, and Marcela V. Maus

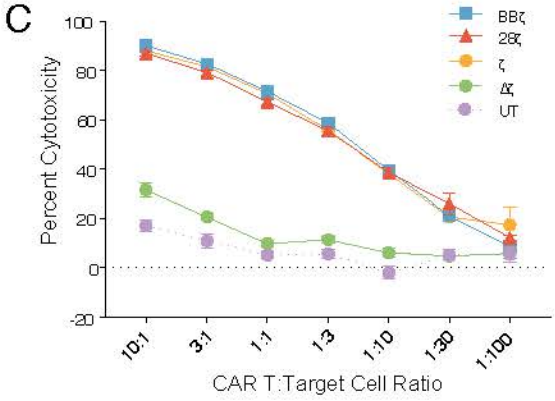
Supplemental Fig 1

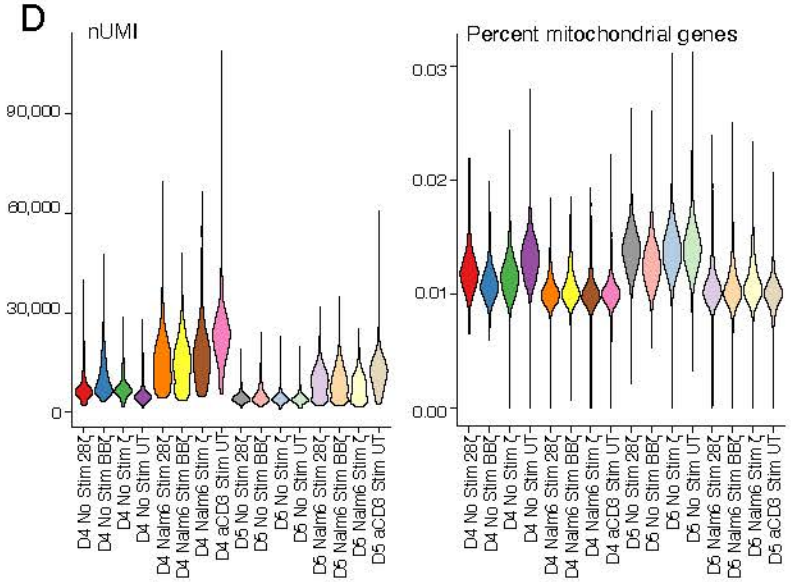


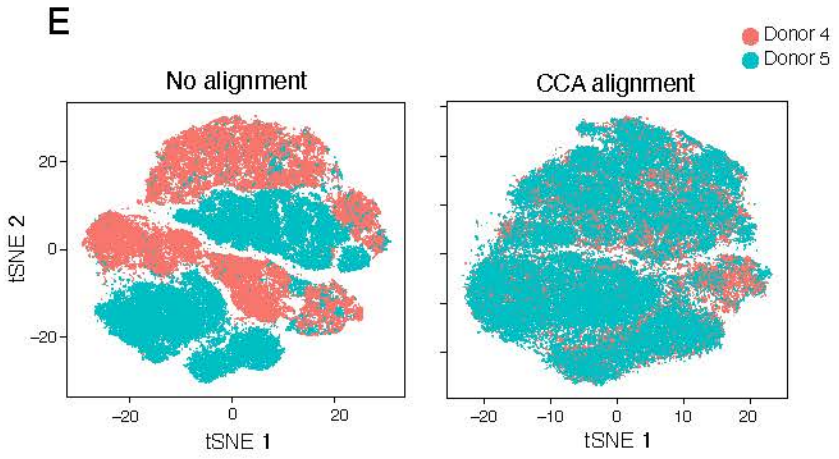
Supplemental Fig 2

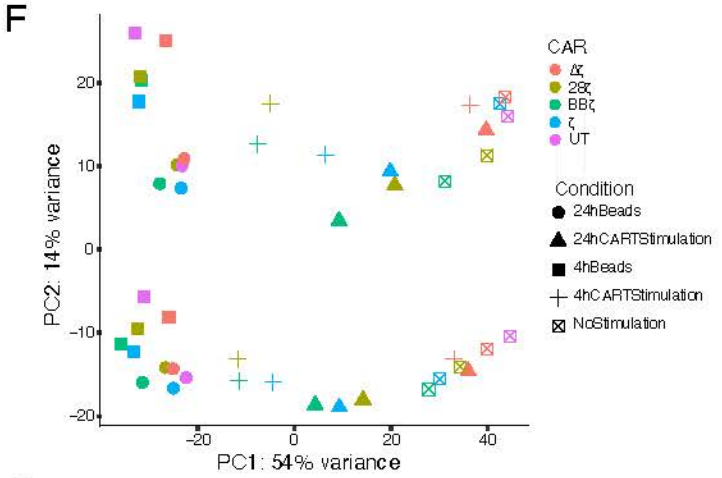


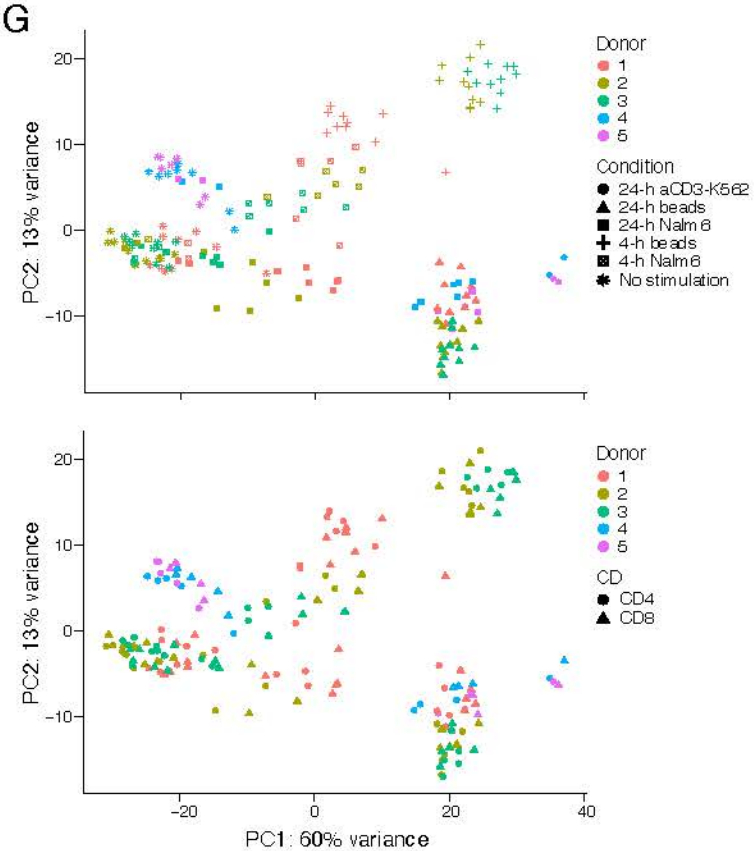




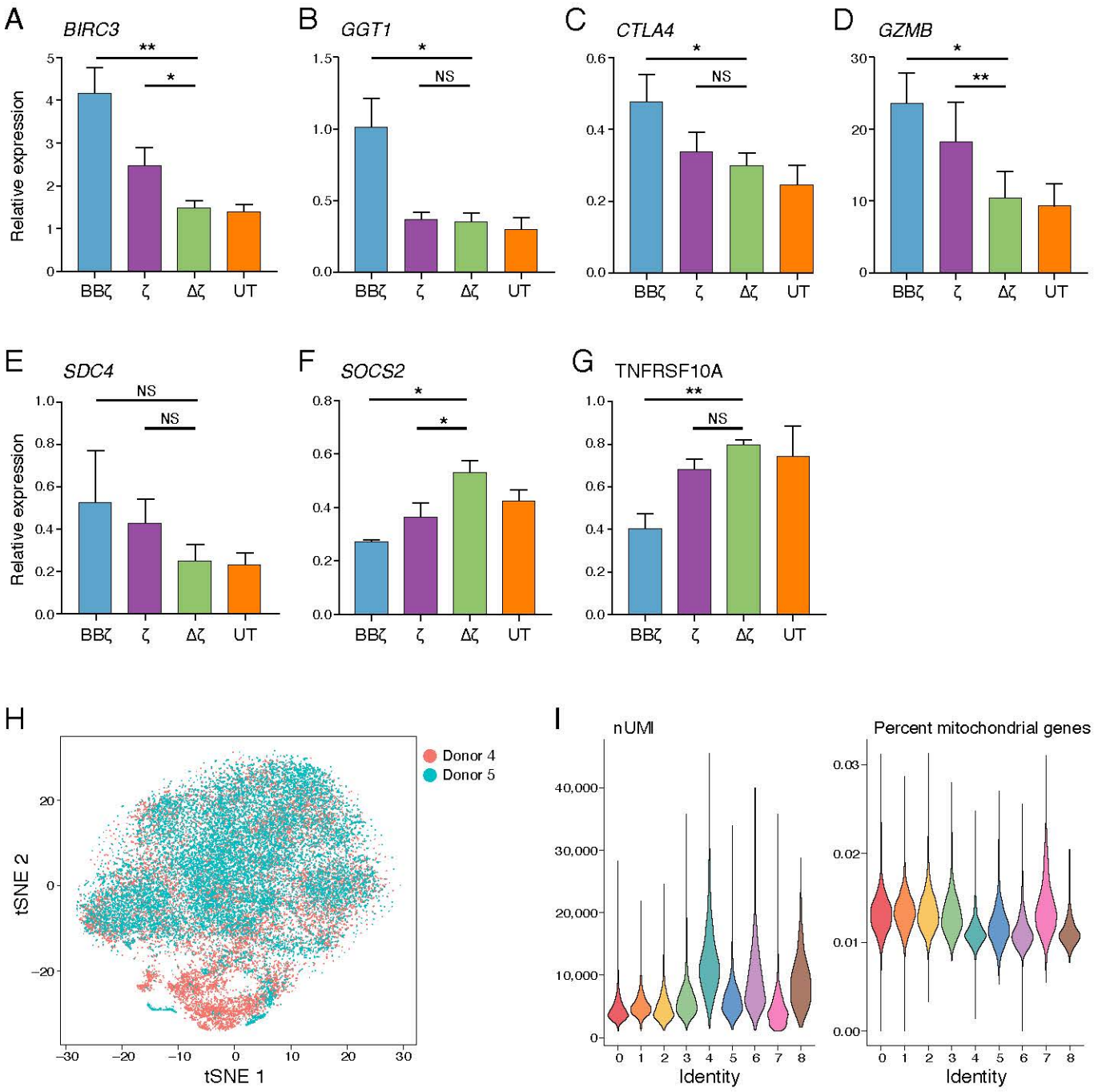




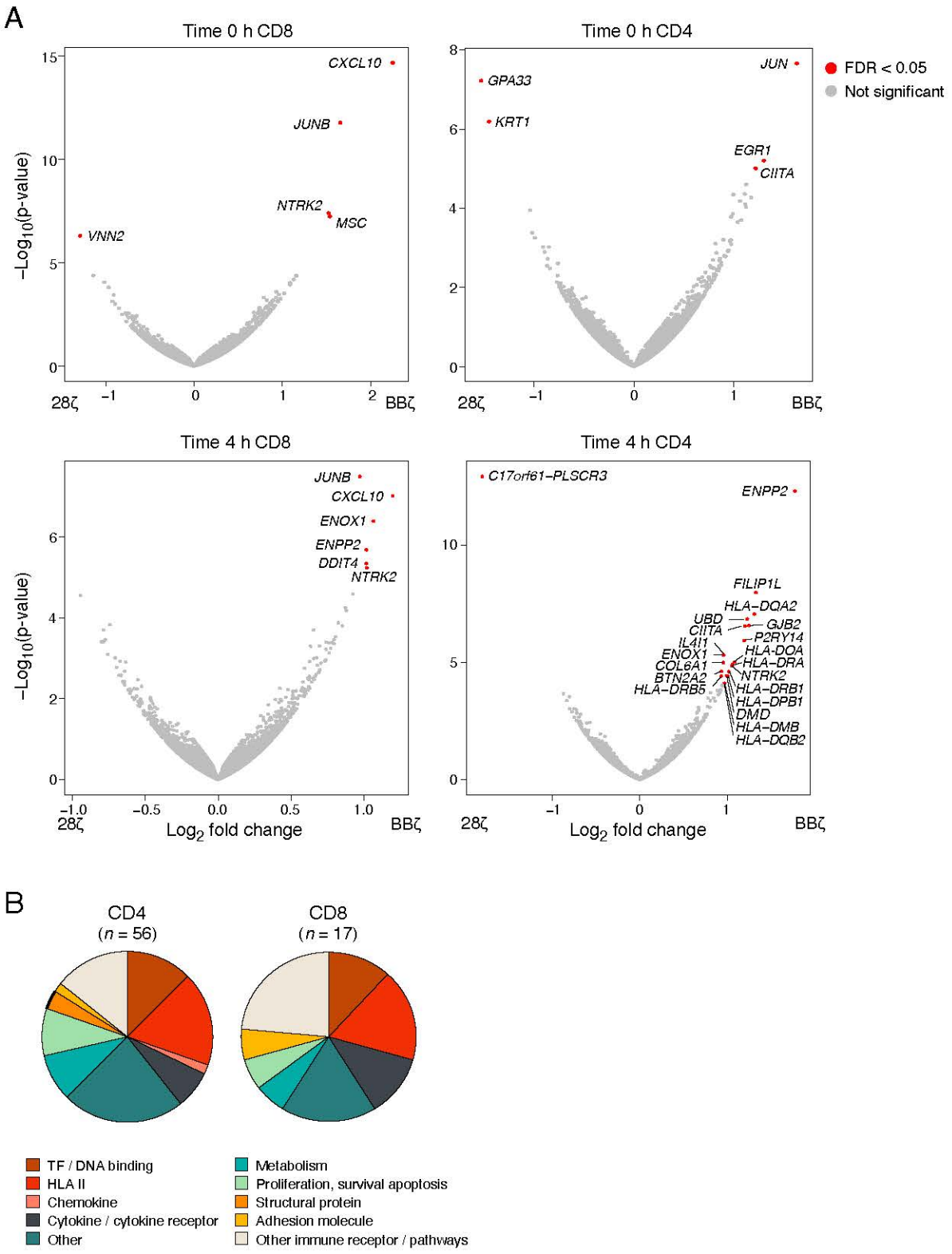




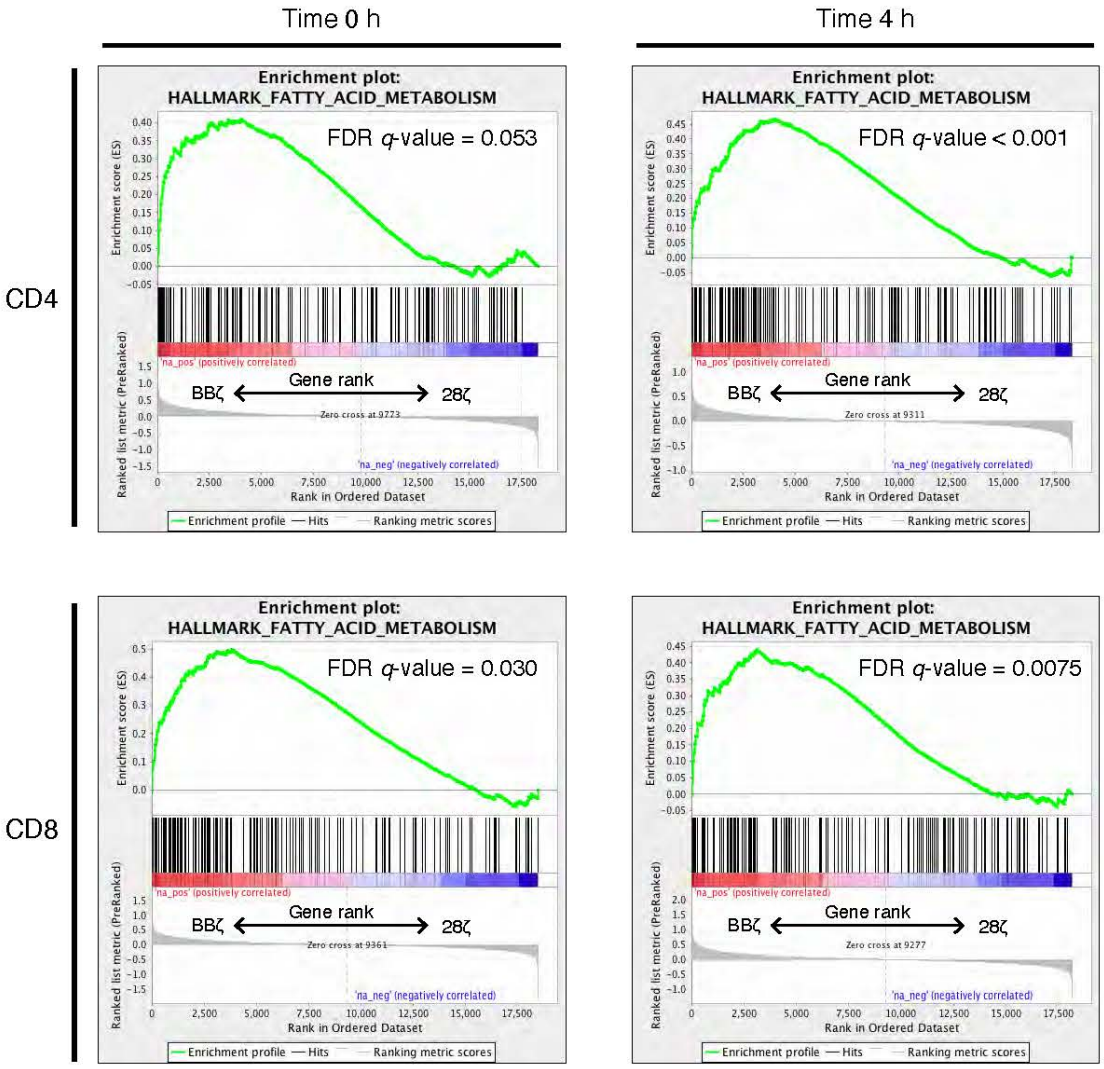
Supplemental Fig 3



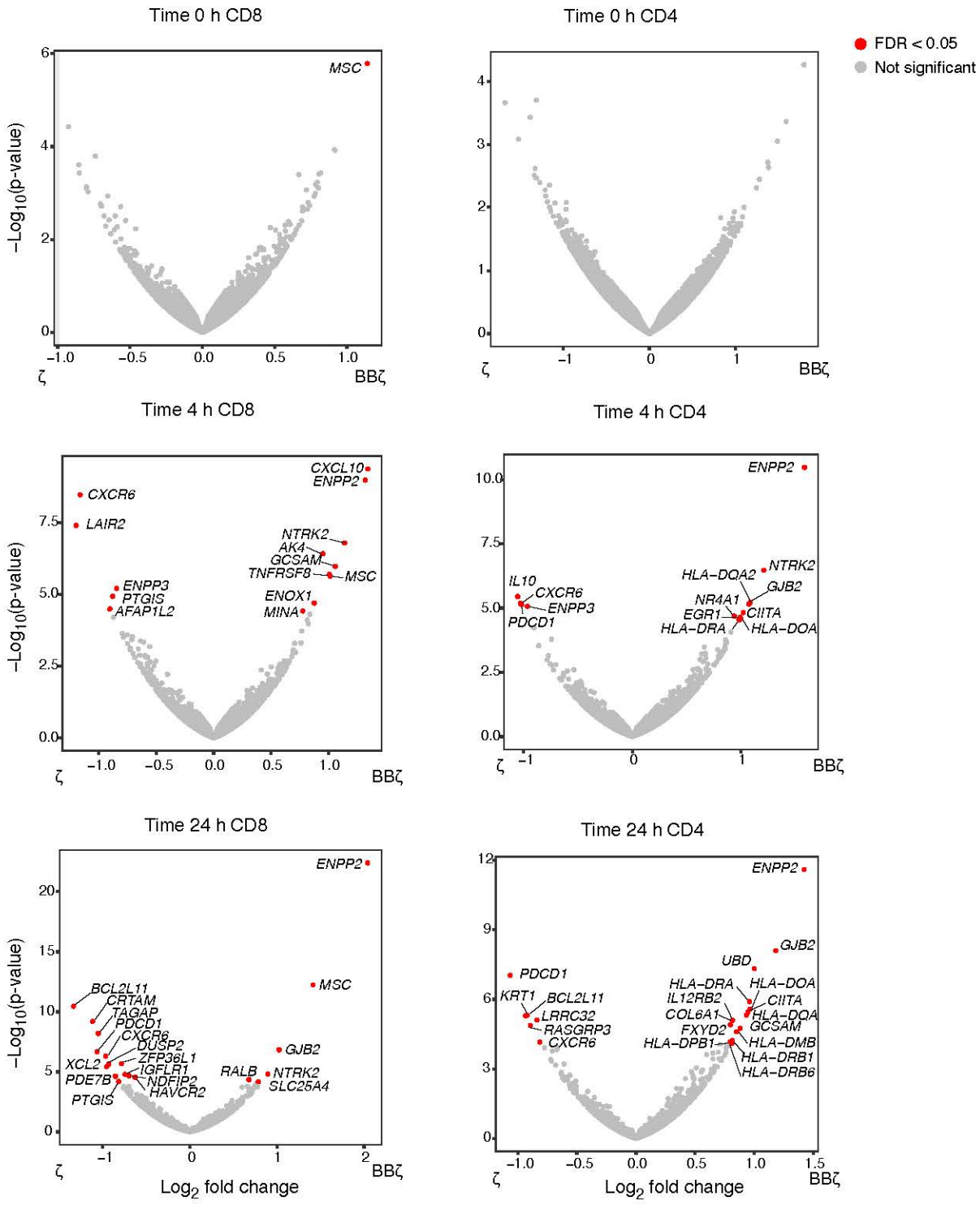
Supplemental Fig 4



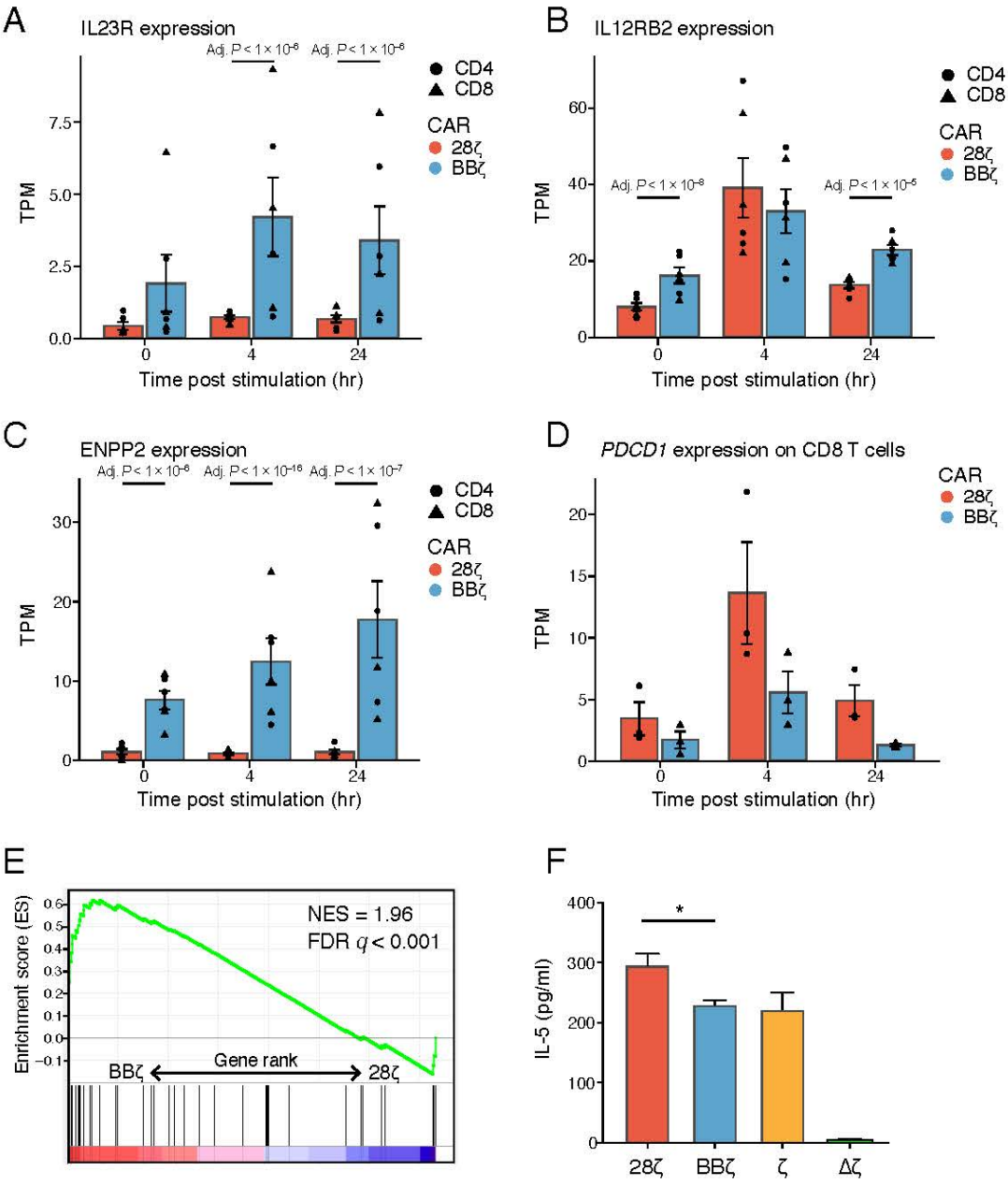
Supplemental Fig 5



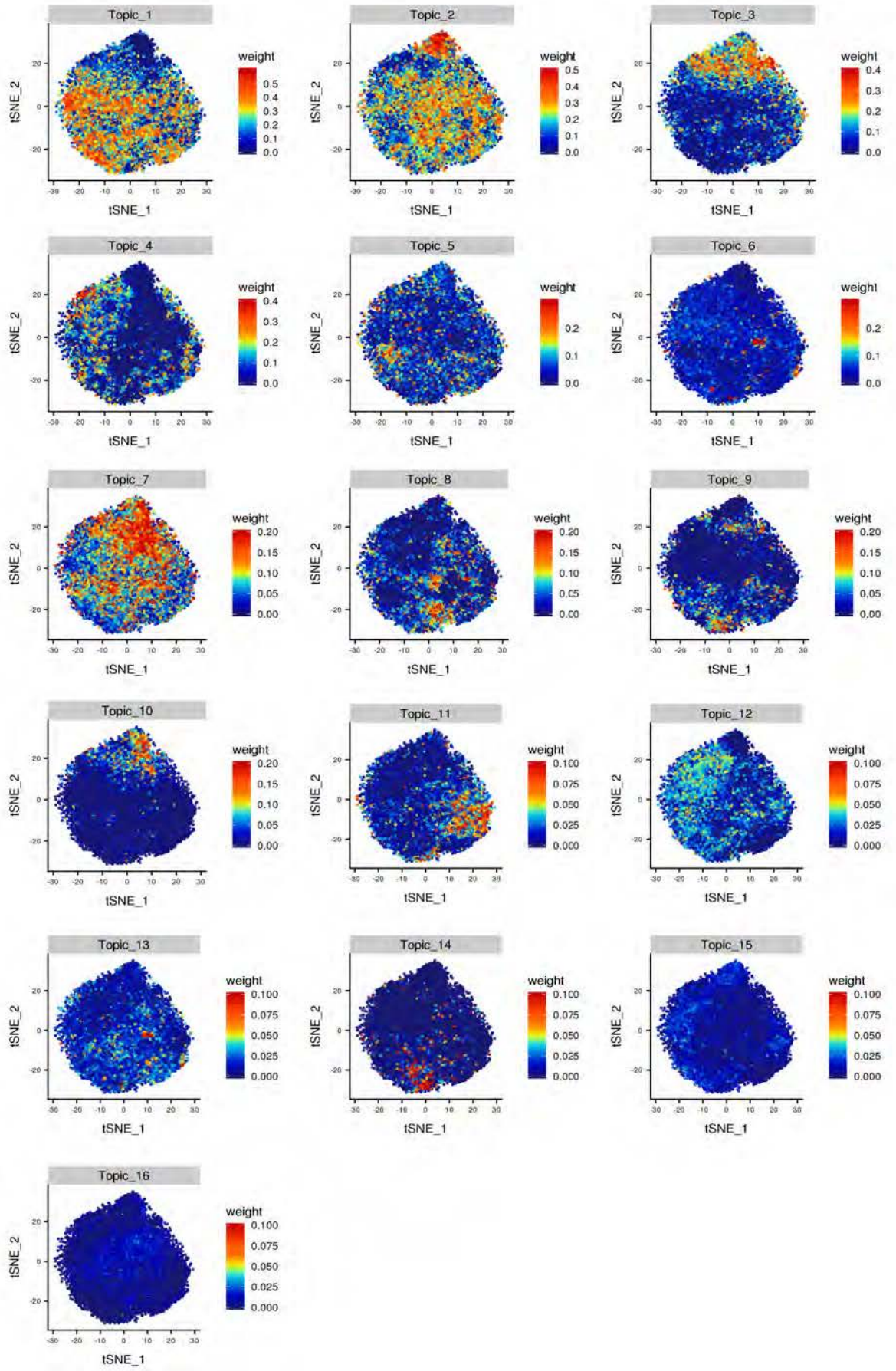
Supplemental Fig 6



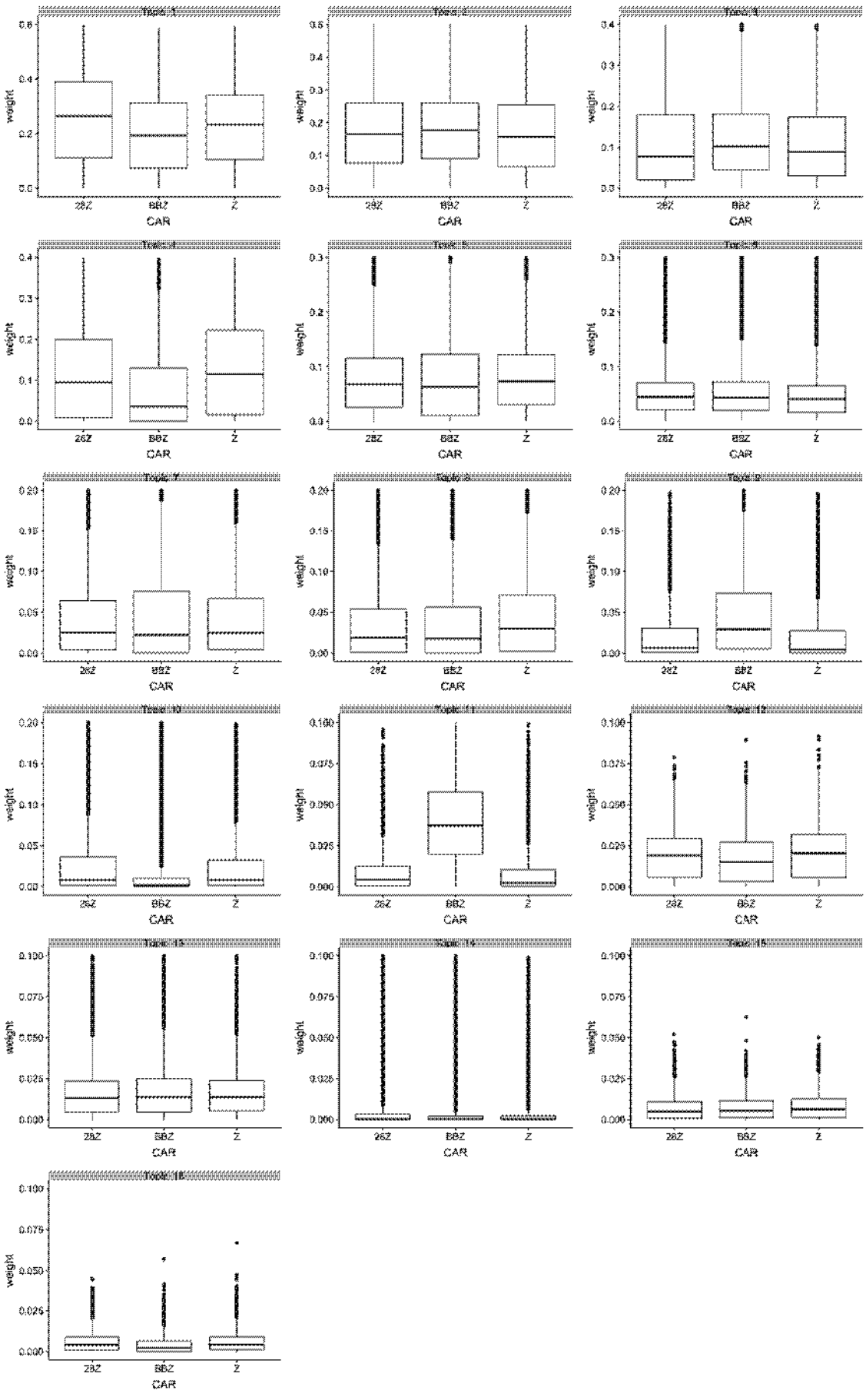
Supplemental Fig 7



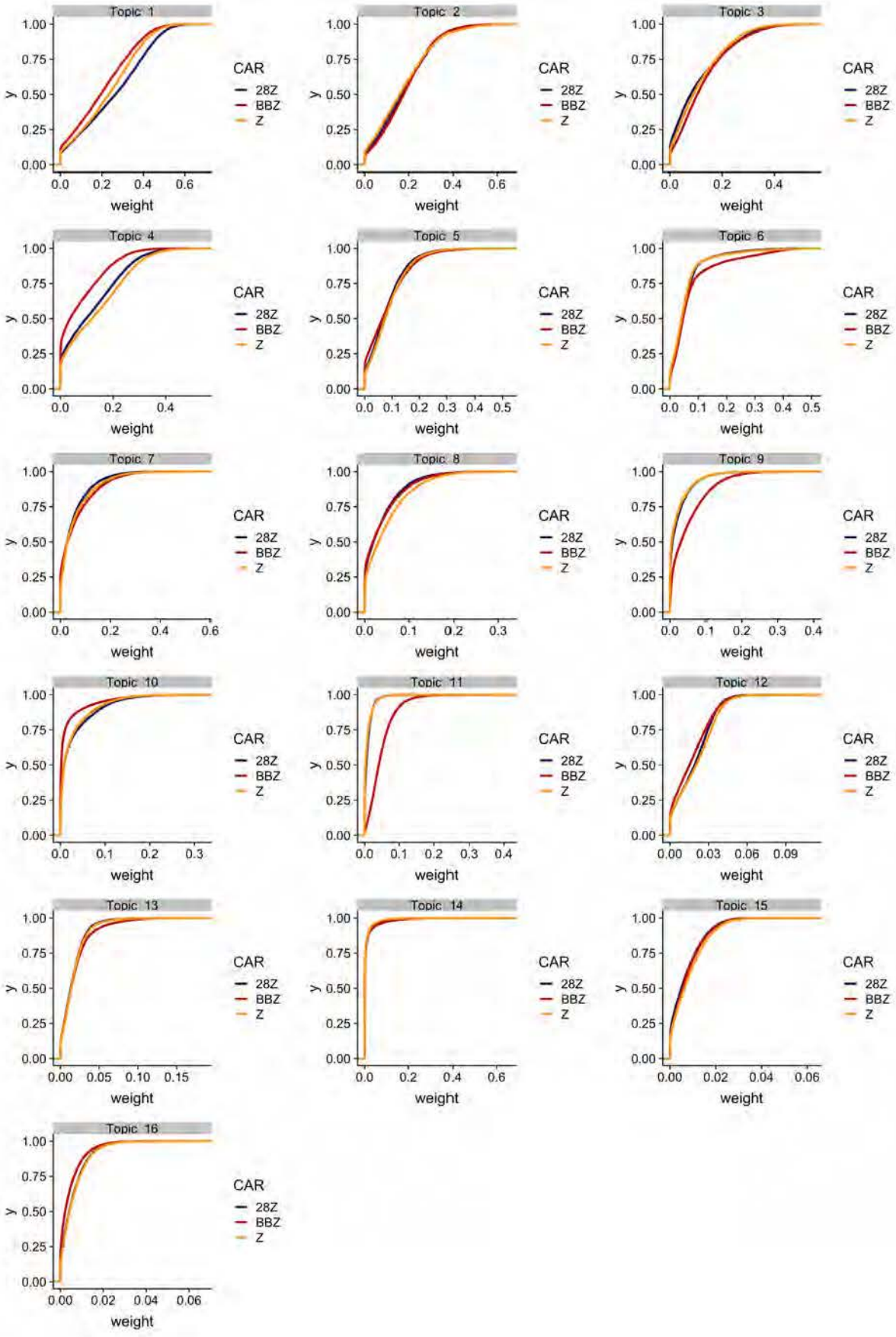
Supplemental Fig 8



Supplemental Fig 9



Supplemental Fig 10



Supplemental Information

Supplemental Figure Legends

Figure S1. Vector maps of EGFR CAR constructs.
Vector maps of EGFR CAR constructs (**TM**, hinge and transmembrane domain; **L**, leader sequence).

Figure S2. Generation of CAR T cell profiles for Figure 1 with quality and batch correction.
A) Representative transduction efficiency of CAR constructs from **Figure 1** determined by mCherry expression and CD3 surface expression on day 13. **B**) mCherry MFI of CAR T cells measured on day 13, N = 6 normal donors. Paired-ratio student t-test, NS - p>0.1. **C**) Luciferase-based killing assay with different numbers of CD19 CAR T cells co-cultured with 10,000 CD19+ target cells (Nalm6 CBG-GFP) at indicated CAR T cell-to-target ratio for 16 hours. N = 3 normal donors with technical triplicates. **D)** Number of individual unique molecular identifiers (UMIs) and the percent mitochondrial genes sequenced per sample from Figure 1C loaded on the 10x. **E**) *t*SNE of scRNA-seq T cell profiles from Figure 1C colored by donor pre- and post-CCA alignment (batch correction). **F**) Principal component analysis performed on a representative donor (donor 2) bulk RNAseq samples. **G**) Principal component analysis of bulk (donor 1-3; Figure 1B) and summed single cell data (donor 4 and 5; Figure 1C) after LIMMA correction for donor/sequencing method batch effects. Samples (data points) colored by donor and condition (top panel) or CD4 *vs.* CD8 (bottom panel).

Figure S3. Tonic Signaling signature in EGFR and CD19 CAR T cells. EGFR CAR T cells made with the constructs from **Figure S1** were sorted on $CD8^+mCherry^+$ cells after 7 days of bead expansion and 7 days of rest. After sorting, RNA was isolated followed by reverse transcription to cDNA. Digital droplet PCR of genes upregulated (**A** to **E**) and downregulated (**F** and **G**) in our signature for tonic signaling from CD3ζ (**Figure 2A**) was performed. N=4 normal donors, mean and SEM plotted. Significance was determined with a paired-ratio student t-test comparing BBζ and ζ to Δζ and correcting for two comparisons with Holm-Bonferroni method adjustment. Genes are expressed relative to internal reference gene TBP. * adj-p<0.05 **adj-p<0.01. **H**) *t*SNE of scRNA-seq T cell profiles from no stimulation conditions colored by donor after CCA batch correction. **I**) Number of individual UMIs and the percent mitochondrial genes sequenced per cluster from the unstimulated samples described in **Figure 2B**.

Figure S4. DE genes from bulk RNA-seq between BBζ and 28ζ CAR T cells. **A**) Volcano plot of log fold-change genes expression on the x axis and –log10(p value) on the y axis between CD19 BBζ and 28ζ CAR T cells at 0 and 4 hours post-Nalm6 activation in $CD4^+$ and $CD8^+$ cells. Genes with FDR<0.05 are plotted in red. Positive x-axis is up in BBζ *vs*. 28ζ (see **Figure 3A** for 24 hour time point; **Table S5** for gene lists). **B**) Classification of types of significantly differentially expressed genes at 24 hours after Nalm6 stimulation detected by bulk RNA-seq with an FDR<0.1 between BBζ and 28ζ CARs using GO annotation.

Figure S5. BBζ CARs have increase fatty acid metabolism before activation. GSEA of hallmark fatty acid metabolism genes in rank fold-change list of DE genes between bulk RNA-seq profiles of CD19 BBζ and 28ζ CAR T cells at 0 hours and 4 hours post CAR activation with irradiated Nalm6 cells.

Figure S6. DE genes from bulk RNA-seq between BBζ and ζ CAR T cells Volcano plot of log fold-change genes expression on the x axis and –log10(p value) on the y axis between CD19 BBζ and ζ CAR T cells at 0, 4 and 24hours post-Nalm6 activation in $CD4^+$ and $CD8^+$ cells. Genes with FDR<0.05 are plotted in red. Positive x-axis is up in BBζ *vs*. ζ.

Figure S7. Bulk RNAseq profiles indicate cytokine and cytokine receptor differences in BBζ and 28ζ CAR T cells. Normalized gene expression in $CD4^+$ and $CD8^+$ in BBζ and 28ζ CD19 CAR T cells activated with irradiated Nalm6 of **A**) *IL23R* - IL-23 receptor, **B**) *IL12RB2* - IL-12 receptor and **C**) *ENPP2* – autotaxin. N= 3 normal donors. Mean and SEM plotted with points showing individual donor values. Adj-p values were calculated by DEseq2 using Holm-Bonferroni correction (see also **Figure 3F**). **D**) Bulk gene expression of *PDCD1* (encoding the PD1 protein*)* in $CD8^+$ CD19 CAR T cells with Nalm6 stimulation over

time. Dots represent the individual donor samples. Mean and SEM plotted, * adj-p<0.05. **E**) GSEA of early polarizing T_H1 signature genes scored across the rank fold-change list of DE genes between BBζ(+) and 28ζ CAR T cells at the 0 hour time point (see also **Figure 4A**). **F**) IL-5 measured in the supernatants of bulk $CD4^+/CD8^+$ EGFR CAR T cells stimulated for 24 hours with irradiated U87 at a 2:1 effector to target ratio (see **Figure C-D** for IL-4). N=3 normal donors, mean and SEM plotted. * p<0.05.

Figure S8. Topics analysis of Nalm6 stimulated CAR T cells. Topics discovered by LDA (setting *K* parameter to 16) of single cell data from CD19-CAR BBζ, ζ, and 28ζ T cells 24 hours after Nalm6 stimulation. The *t*SNEs are colored by the weight of the given topic in each cell (see also **Figure 5B**, **Table S7**).

Figure S9. Topics enrichment across Nalm6 stimulated CAR T cells with different costimulation domains. Box and whiskers plot of topic weights of a given topic in scRNAseq profiles of cells separated by CAR T cell group after 24 hour stimulation with Nalm6 cells at a 1:1 effector:target ratio.

Figure S10. CDF of topics across Nalm6 stimulated CAR T cells. ScRNAseq profiles were used to plot the cumulative distribution plot of each topic weights per cell across the different CAR T cell groups after 24 hour stimulation with Nalm6 cells at a 1:1 effector:target ratio.

Supplemental Tables

See Excel files online for full Supplemental Tables.

Supplemental Methods

Raw data "scrubbing" for removal of donor specific SNPs

Removal of donor specific SNPs was required by the IRB to de-identify the sequencing information from anonymous human donors, since we obtained discarded tissues under an IRB-approved protocol. Both single cell and bulk RNA-Seq reads were scrubbed to maintain gene expression levels whilst simultaneously removing any donor specific SNPs before uploading to the uploading to the SRA database, accession number PRJNA554339. To this end, fastq files were aligned with STAR aligner[1] to output a BAM file of aligned reads and a splice junction file. Splice junctions were then added to the reference genome for each sample and then realigned with STAR resulting in a BAM file with splice aware alignment.

The file was used for variant calling using the Genome Analysis Tool Kit (GATK)[2] such that each variant could be identified. Reads including variants were replaced with a corresponding reference read containing no identifiable SNPs. Post processing of the alignment was done by the SplitNCigarReads tool in GATK to split and trim intronic reads. Freebayes[3] was run using the GRCh38 as the reference genome to identify the sites of variants. These sites were then identified in the STAR aligned BAM file and a vcf file created from variants being called from the BAM file. A custom script took the vcf file and identified each point of variation in the BAM file. If the position had an alternative allele it was flipped to the reference nucleotide. If there is no variant in the position or the reference allele is present the read was written as is. This output a new BAM file which was converted to a newly scrubbed fastq file.

Bulk RNA-Seq PCA and differential gene expression

Differentially expressed genes were identified using the DEseq2 R package[4], after correcting for the effect of different patient donors. P-values were corrected for multiple hypotheses testing using the Benjamini & Hochberg (1995) method. Genes with an FDR (q-value) less than 0.05 were considered significant. PCA plots were constructed by DESeq using linear model batch corrected data from LIMMA R package[5]. The contribution of each covariate on the principle components was calculated using the SWAMP R package[6]. Row normalized (Complete linkage, average linkage, ward method), heat maps were constructed using the

gene expression (TPM) data generated by RSEM. Genes were classified using their Gene ontology annotation[7, 8] and the gene cards database (www.genecards.org).

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed at each time point against a gene list of early polarizing T_H1 genes[9] with the Deseq2 generated DE gene lists ranked by $\log_2$(fold change). Analysis was performed using the desktop GSEA (v3.0). To identify MSigDB[10] gene sets enriched in significantly differentially expressed genes (FDR<0.05) between any two CARs we ran the GSEA software[11].

Bulk expression from single cell samples

Bulk expression levels were generated form single cell transcriptomes by first binning cells into a CD8 group based on the cell's expression of *CD8A* and not *CD4* or vice versa for the CD4 group. The gene expression was then summed across 1000 randomly chosen cells per condition. The bulked expression from single cells for donor 4 and 5 were combined with the true bulk expression profiles from donors 1-3 and PCA was run with LIMMA correction for donor variation[5].

Cytokine detection of stimulated T cells

For cytokine release assays, T cells were stimulated in a 96 well plates with 100 000 T effector cells/well combined with irradiated target cells at a CAR T cell-to-target ratio of 1:1 for Nalm6 targets and 2:1 for U87 targets. Supernatants were harvested after 24 hours and frozen at –80C. Supernatants were analyzed for cytokine levels using FLEXMAP 3D® platform from Lumina Instrumentation (Thermo Fisher Scientific) according to manufacturer's instructions with a panel of the following cytokines: IL-1β, IL-2, IL-4, IL-5, IL-6, IL12p70, IL-13, IL-18, IFN-γ, GM-CSF, TNF-α, IL-10 and IL-21. Plates were read using xPONENT Software 4.1. All samples were in measured in technical triplicates and with N=3 normal donors. Triplicates measured were averaged before graphing with Prism (Graphpad software).

Cytotoxicity Assay

CD19 CAR T cells were normalized to equal %CAR+ by adding UT cells. CAR T cells were titrated in 96 well plates co-cultured with 10,000 Nalm6 CBG-GFP tumor cells/well. After 16 hours in culture, tumor cells were lysed and luciferin (Promega) was added. Luciferase activity was measured with a Synergy Neo2 luminescence microplate reader (Biotek). Percent cytotoxicity was calculated by the following equation: %cytotoxicity = (total RLU / target cells only RLU) x100.

Digital Droplet PCR

EGFR CAR T cells were transduced, expanded and rested for 7 days. $5e^5$ cells were collected by FACs and resuspended in 350µl RLT buffer with 1% 2-mercaptoethanol. RNA was extracted and purified using RNAeasy kit (Qiagen) and cDNA was generated from 270ng of RNA/ 20µl reaction using iScript Reverse Transcription supermix (Bio-Rad). Digital Droplet PCR was performed using ddPCR supermix with no dUTPs (Bio-Rad) with a QX200 Droplet Digital PCR (ddPCR™) System (Bio-Rad) platform for quantification. Droplet generation, PCR and detection of positive droplets were performed according to manufacturer's instructions (Instruction Manual, QX200™ Droplet Generator – Bio-Rad).

The cycling protocol was according the manufacturer's instructions with a 57°C melting temperature. Human TBP was used at as the reference gene in each reaction, (HEX fluorophore : TBP PrimePCR™ ddPCR™ Expression Probe Assay:**Unique Assay ID:** dHsaCPE5058363 (Bio-Rad)). The following FAM fluorophore primer probes were used (IDT PrimeTime Std® qPCR Assay).

Gene	Primer/Probe
CTLA4	PrimeTime Primer 1: CGG ACC TCA GTG GCT TTG PrimeTime Primer 2: TTC ATC CCT GTC TTC TGC AA PrimeTime Probe: /56-FAM/CG CCA GCT T/Zen/T GTG TGT GAG TAT GC/3IABkFQ
GZMB	PrimeTime Primer 1: CAG AGA CTT CTG ATC CCA GAT PrimeTime Primer 2: TCC TGA GAA GAT GCA ACC AAT PrimeTime Probe: /56-FAM/CC CGC CCC T/Zen/A CAT GGC TTA TCT /3IABkFQ/
SOCS2	PrimeTime Primer 1: GAT ATT GTT AGT AGG TAG TCT GAA TGC PrimeTime Primer 2: GGA GCT CGG TCA GAC AG PrimeTime Probe: /56-FAM/AA AGA GGC A/Zen/C CAG AAG GAA CTT TCT TGA /3IABkFQ/
SDC4	PrimeTime Primer 1: GGT ACA TGA GCA GTA GGA TCA G PrimeTime Primer 2: GCA GCA ACA TCT TTG AGA GAA C PrimeTime Probe: /56-FAM/CC ACG ATG C/Zen/C ACC CAC AAT CAG A/3IABkFQ/
GGT1	PrimeTime Primer 1: TTC AGG TCC TCA GCT GTC A PrimeTime Primer 2: TGG CTG ACA CCT ACG AGA C PrimeTime Probe: /56-FAM/CC GCC TGG A/Zen/T GTC CTT CAC AAT CT/3IABkFQ/
TNFRSF10A	PrimeTime Primer 1: GTC CAT TGC CTG ATT CTT TGT G PrimeTime Primer 2: GTC AGT GCA AAC CAG GAA CT PrimeTime Probe: /56-FAM/AT TCT GCT G/Zen/A GAT GTG CCG GAA GT/3IABkFQ/
BIRC3	PrimeTime Primer 1: GTA GAT GAG GGT AAC TGG CTT G PrimeTime Primer 2: GGT GTT GGG AAT CTG GAG ATG PrimeTime Probe: /56-FAM/CC TTG GAA A/Zen/C CAC TTG GCA TGT TGA /3IABkFQ/

Supplemental References:

1. Dobin, A, Davis, CA, Schlesinger, F, Drenkow, J, Zaleski, C, Jha, S, Batut, P, Chaisson, M, and Gingeras, TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*: 15-21.
2. McKenna, A, Hanna, M, Banks, E, Sivachenko, A, Cibulskis, K, Kernytsky, A, Garimella, K, Altshuler, D, Gabriel, S, Daly, M*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*: 1297-1303.
3. Garrison, AaM, G (2012). Haplotype-based variant detection from short-read sequencing.
4. Love, MI, Huber, W, and Anders, S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*: 550.
5. Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, and Smyth, GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*: e47.
6. Lauss, M (2018). swamp: Visualization, Analysis and Adjustment of High-Dimensional Data in Respect to Sample Annotations. . 1.4.1 ed.
7. Ashburner, M, Ball, CA, Blake, JA, Botstein, D, Butler, H, Cherry, JM, Davis, AP, Dolinski, K, Dwight, SS, Eppig, JT*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*: 25-29.
8. The Gene Ontology, C (2017). Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res *45*: D331-D338.
9. Aijo, T, Edelman, SM, Lonnberg, T, Larjo, A, Kallionpaa, H, Tuomela, S, Engstrom, E, Lahesmaa, R, and Lahdesmaki, H (2012). An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation. BMC Genomics *13*: 572.
10. Liberzon, A, Subramanian, A, Pinchback, R, Thorvaldsdottir, H, Tamayo, P, and Mesirov, JP (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*: 1739-1740.
11. Subramanian, A, Tamayo, P, Mootha, VK, Mukherjee, S, Ebert, BL, Gillette, MA, Paulovich, A, Pomeroy, SL, Golub, TR, Lander, ES*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*: 15545-15550.