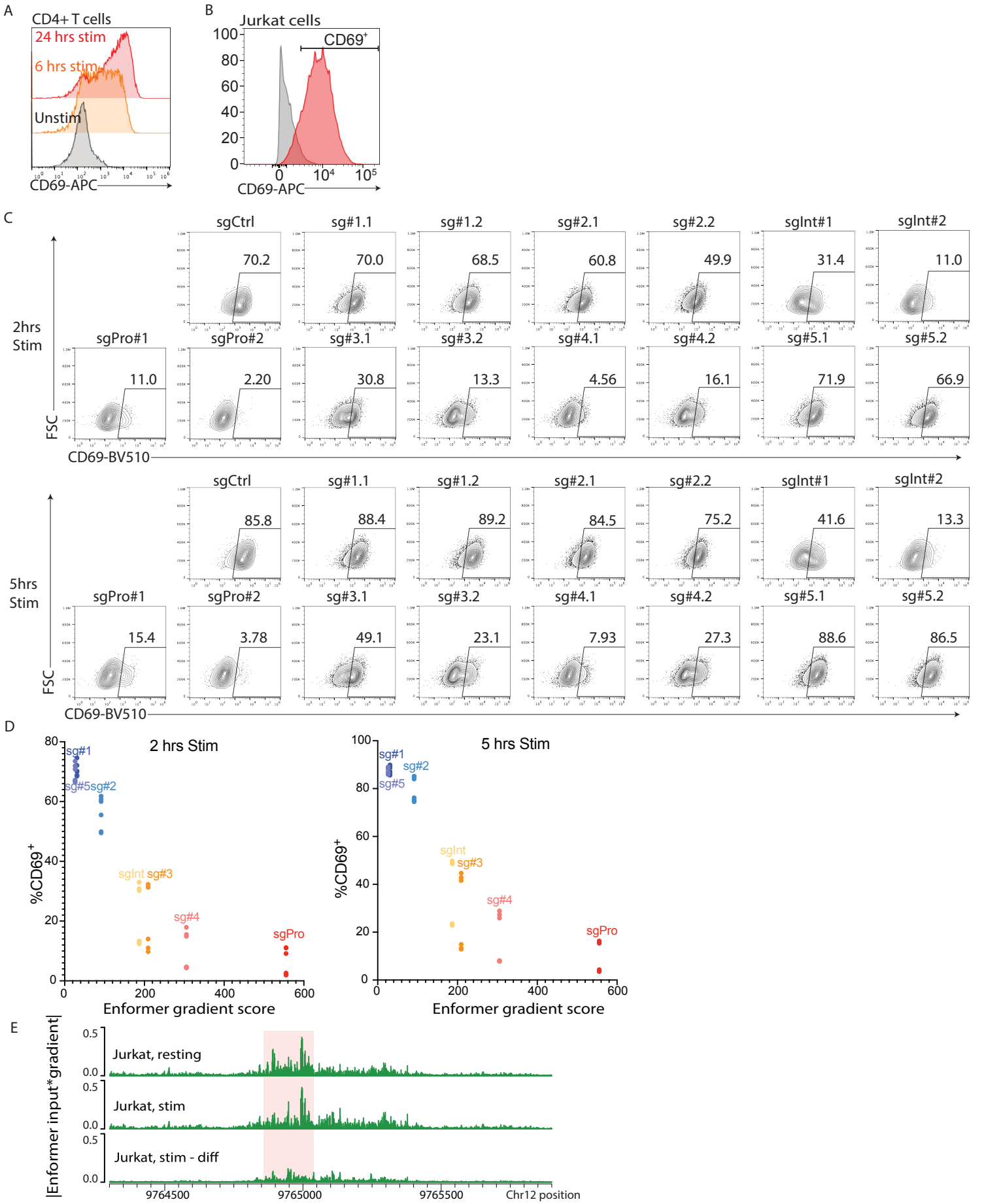
Cell Genomics, Volume *3*

Supplemental information

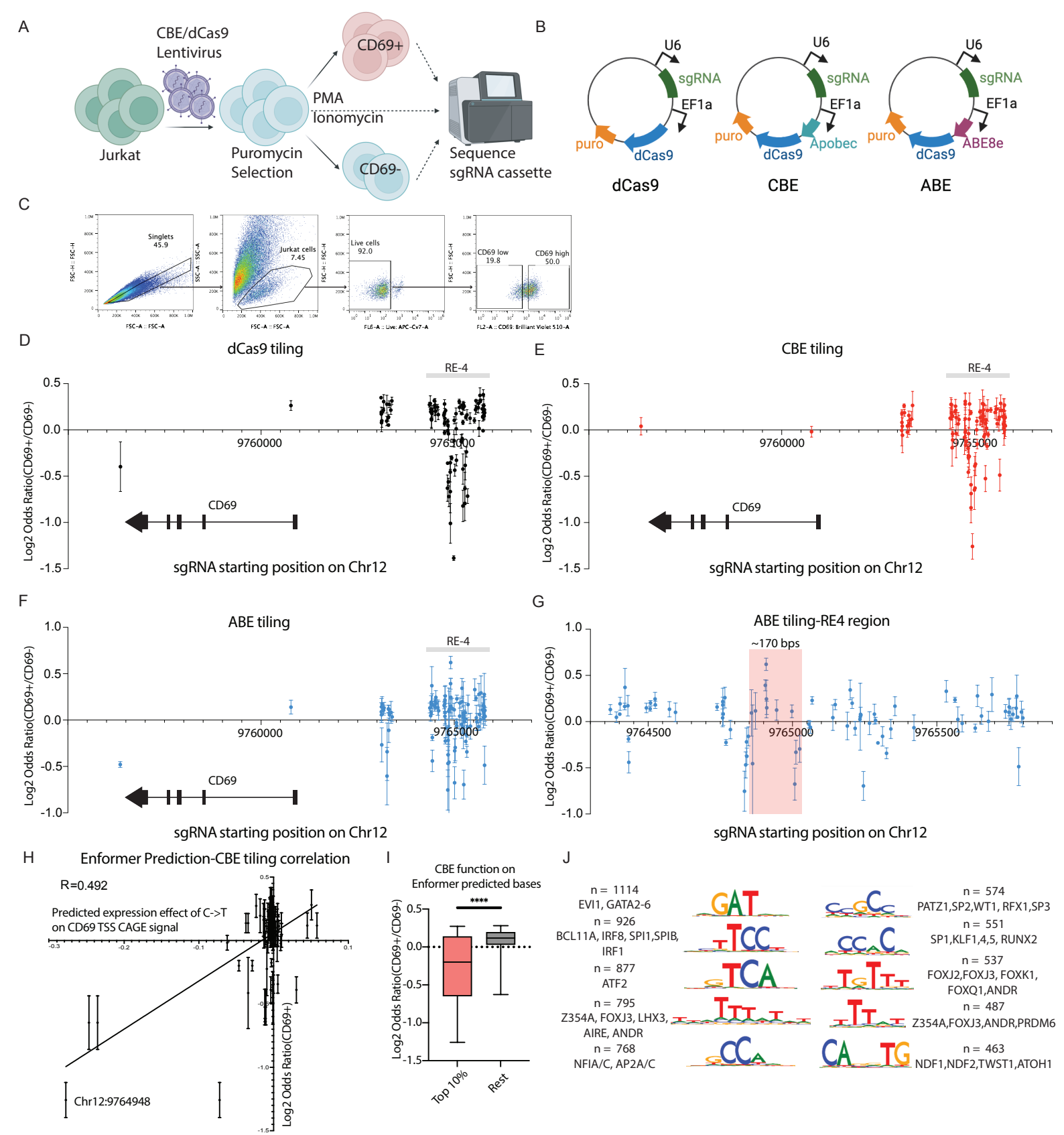
Integrative dissection of gene regulatory elements at base resolution

Zeyu Chen, Nauman Javed, Molly Moore, Jingyi Wu, Gary Sun, Michael Vinyard, Alejandro Collins, Luca Pinello, Fadi J. Najm, and Bradley E. Bernstein



Supplementary Figure 1. Dissecting CD69 regulatory elements upon stimulation, related to Figure 1

A) CD69 expression in primary CD4+ T cells following anti-CD3/CD28 stimulation. Unstimulated, 6 hours stimulation and 24 hours stimulation are shown in grey, orange and red color. Data represents 2 independent experiments.
B) CD69 expression of Jurkat cells with and without stimulation. Cells were stimulated with PMA/Ionomycin for 5 hours. Data represent 3 independent experiments.
C) Flow cytometry plots of CD69 expression in Jurkat cells targeted with the indicated CRISPR-i sgRNA following PMA/Ionomycin stimulation. Samples were gated on live populations after puromycin selection. Data represent 2 independent experiments with triplicates per experiment.
D) Scatter plot showing Enformer gradient score (Methods) for different CREs versus %CD69+ after CRISPRi targeting. Each dot represents an individual sgRNA, colored by the targeted CRE. Data is taken from the same experiment as Fig. 1C and S1C. Enformer gradient score corresponds to the summed absolute value of the gradient with respect to CD69 transcriptional output over the targeted RE (Methods).
E) Magnitude of input*gradient across RE-4 of the fine-tuned Enformer model with respect to the indicated output heads (Methods). Red highlight corresponds to the 170bp interval highlighted in Figure 1D.



Supplementary Figure 2 dCas9, CBE and ABE tiling for CD69 loci and Enformer predicted motifs during Jurkat cell stimulation, related to Figure 1

A) Experimental workflow for dCas9 and CBE tiling of the CD69 loci.

B) dCas9+sgRNA, CBE+sgRNA and ABE+sgRNA vectors used for the experiment (details in Methods).

C) Example gating strategy of cytidine base editor tiling experiment. Top 50% CD69 expressed cells are sorted as CD69+; bottom 20% CD69 expressed cells are sorted as CD69-.

D) Enrichment/depletion plot of dCas9 sgRNAs in CD69+ Jurkat cells, relative to CD69- cells (y-axis; log2 odds ratio of normalized sgRNA reads). sgRNAs along the x-axis are plotted according to their 5' starting position on the positive strand. Each data point represents mean±s.e.m.

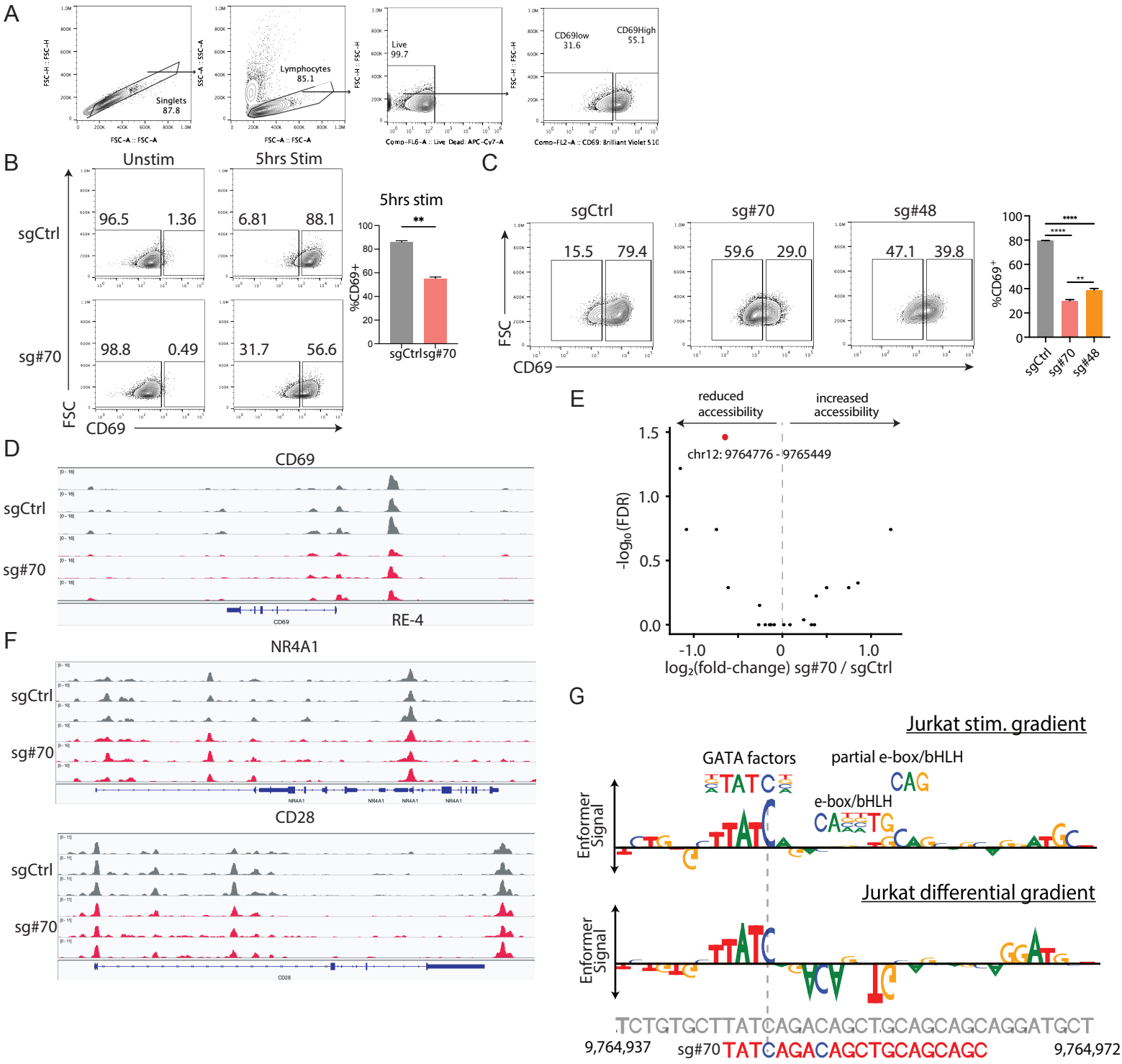
E) Enrichment/depletion plot of Cytidine Base Editor (CBE) sgRNAs in CD69+ Jurkat cells, relative to CD69- cells (as in panel D).

F-G) Enrichment/depletion plot of Adenine Base Editor (ABE) sgRNAs in CD69+ Jurkat cells, relative to CD69- cells (as in panel D) across all sgRNAs(F) or sgRNAs targeting RE-4(G). Red region shows the machine learning predicted interval.

H) Correlation of the Enformer predicted impact of each CBE targetable base on the transcriptional output of CD69 as measured by CAGE-seq vs. CBE tiling data (Methods). R corresponds to Pearson's correlation computed with n=97 points ($p < 2.2e-16$). The targeting window of each sgRNA is set at 2-8 bases opposite the NGG PAM.

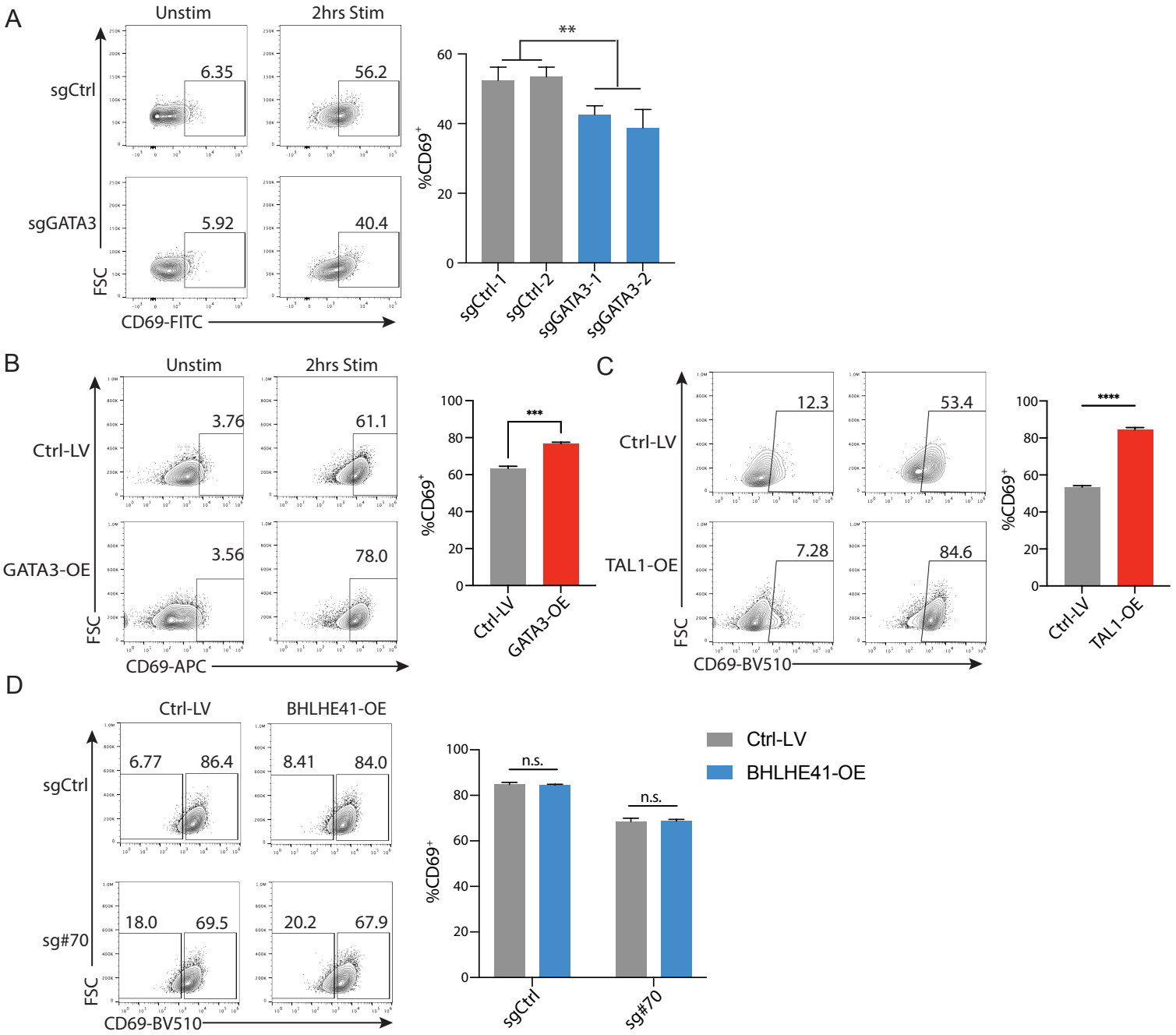
I) Experimentally measured CBE effect on CD69 expression between the top 10% Enformer predicted negative C->T artificial SNPs vs. the remaining 90 %(rest). A total of 97 Cytosines were considered in the analysis. **** corresponds to $p<0.001$ with unpaired t-test.

J) TFModisco nominated motifs and closest Tomtom matches from Enformer base importance scores genome-wide. For each differentially expressed gene between stimulated vs. resting Jurkat cells, gradients with respect to the CAGE output for unstimulated Jurkat cells at the TSS were calculated. Gradients were then intersected with ATAC-seq peaks in resting Jurkats and provided as input to TFModisco, and the resulting seqlet clusters were matched to known motifs in HOCOMOCO. Listed motifs are significant matches from HOCOMOCO using TomTom at $q < 0.05$, while n here refers to the total number of matched seqlets.



Supplementary Figure 3: CBE-sg#70 and CBE sg#48 targeting specifically affects CD69 expression via RE-4, related to Figure 2 and 3

A) Example gating strategies of CBE-sgRNA targeting experiment.
B) Flow cytometry plots show CD69 expression for CBE-sgCtrl+ and CBE-sg#70+ Jurkat cells under resting or stimulated conditions. Bar plot shows the proportion of CD69+ cells after stimulation. Data represent 4 independent experiments each with 2-3 technical replicates. **p<0.01 per unpaired t test. Error bars represent s.e.m. Jurkat cells were stimulated for 5 hours in this panel.
C) Flow cytometry plots show CD69 expression for CBE-sgCtrl+, CBE-sg#70+ and CBE-sg#48+ Jurkat cells under stimulated conditions. Bar plot shows the proportion of CD69+ cells after stimulation. Data represent 2 independent experiments each with 3 technical replicates. **p<0.01, ***p<0.001, ****p<0.0001 per unpaired t test. Error bars represent s.e.m. Jurkat cells were stimulated for 2 hours in this panel.
D) Chromatin accessibility shown over the CD69 locus for CBE-sg#70 Jurkat cells infected with either BHLHE40 overexpression lentivirus (red) or control (gray). Cells were stimulated with PMA/Ionomycin. Biological replicates are shown.
E) Volcano plot depicts chromatin accessibility changes between sgCtrl and sg#70 groups within a 1mb window around RE4. Each dot represents a called ATAC-seq narrowPeak (Methods) within 1Mb of RE-4. X-axis shows log2(fold-change) for sg #70 peaks relative to sgCtrl while Y-axis shows -log10(FDR). FDR is computed using the BH procedure for all p-values for peaks within a 1Mb radius of RE-4 (Methods). RE-4 is highlighted in red and shows a significant decrease in accessibility at FDR cutoff of 0.05.
F) Chromatin accessibility shown over the NR4A1 and CD28 locus for CBE-sg#70 Jurkat cells infected with either BHLHE40 overexpression lentivirus (red) or control (gray). Cells were stimulated with PMA/Ionomycin. Biological replicates are shown.
G) Enformer signal (letter height) corresponds to the gradient of the fine-tuned model with respect to predicted stimulated Jurkat accessibility in RE-4, as well as differential accessibility between the resting and stimulated states. The sgRNA coincides with a GATA motif and bHLH/e-box motifs, and incurs an edit that disrupts the form (vertical dashed line).

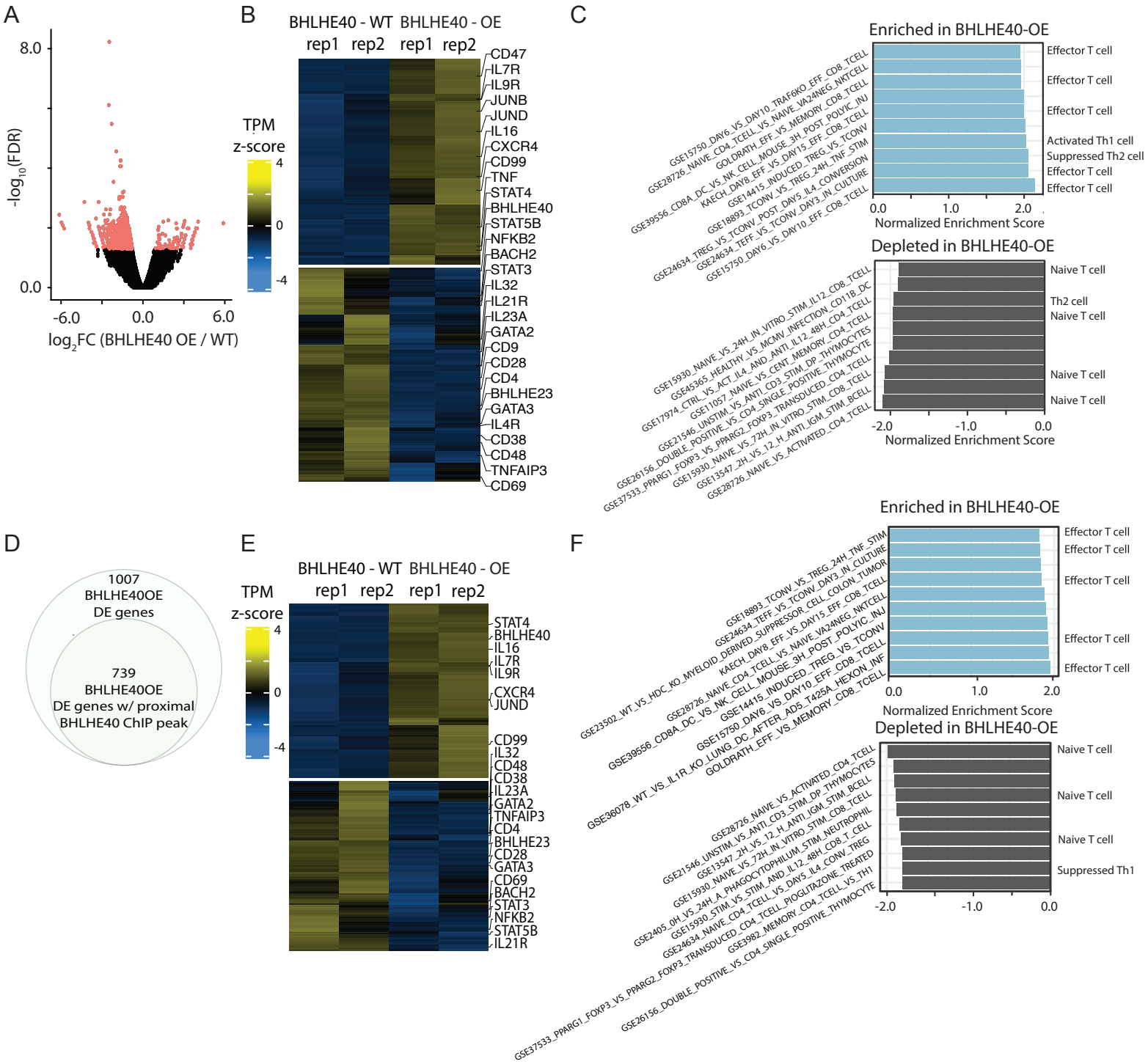


Supplementary Figure 4: Transcriptional control of CD69 expression, related to Figure 4

A) Flow cytometry plots show CD69 expression for Cas9-sgCtrl+ or Cas9-sgGATA3+ stimulated Jurkat cells. Bar plot shows the proportion of CD69+ cells in each condition. CD69+ Gate setup using unstimulated cells with the same construct. Data represent 2 independent experiments each with 2-3 technical replicates. P-value based on unpaired t test, **P<0.01.

B-C) Flow cytometry plots show CD69 expression for Ctrl-LV versus GATA3-OE(B) or TAL1-OE(C) stimulated Jurkat cells. Lentiviruses are transfected into the cells for 4 days(GATA3-OE) or 7 days(TAL1-OE) before the experiment. Bar plot shows the proportion of CD69+ cells in each condition. CD69+ Gate setup using unstimulated cells with the same construct. Data represent 2 independent experiments each with 2-3 technical replicates. P-value based on unpaired t test, ***P<0.001,****P<0.0001.

D) Flow cytometry plots of CD69 signal for stimulated Jurkat cells transduced with CBE-sg#70 and a BHLHE41 overexpression construct(BHLHE41-OE), or with corresponding controls (sgCtrl and Ctrl-LV, respectively). Bar plot depicts the proportion of CD69+ cells in each condition. P-value based on unpaired t-test. Data are from 2 independent experiments with 2-3 technical replicates, mean±s.e.m.



Supplementary Figure 5 BHLHE40 regulates Jurkat cell immune response via reshaping chromatin landscape, related to Figure 4

A) Volcano plot depicting global accessibility changes during BHLHE40OE relative to WT. Red points represent differentially accessible ATAC-seq peaks between the two conditions at FDR< 0.05.

B) Heatmap depicting differentially expressed genes during BHLHE40OE relative to WT at FDR<0.05, with selected immune genes labeled on the right.

C) Gene set enrichment analysis based on RNA-seq in BHLHE40-WT(Ctrl-LV) and BHLHE40-OE (over-expression) Jurkat cells using the ImmuneSigDb. Upper panel shows top 10 gene-sets enriched in BHLHE40-OE relative to WT, while the bottom panel shows top 10 depleted gene sets. Gene set names are provided on the left side of the figure while broad phenotypic labels are provided on the right.

D) Venn diagram where outer circle depicts the total number of differentially-expressed(DE) genes between the BHLHE40 OE conditions and WT, while inside circle depicts the number of DE genes with a BHLHE40 ChIP-seq peak(based on WT Jurkat cells BHLHE40 ChIP-seq) within 25 kb of the major TSS.

E) Heatmap depicting differentially expressed genes between BHLHE40OE and WT Jurkat cells with a BHLHE40 ChIP-seq peak within 25 kb of the major TSS. Select immune genes are labeled on the right.

F) Gene set enrichment analysis based on selected 739 genes in S5D in BHLHE40-WT(Ctrl-LV) and BHLHE40-OE (over-expression) Jurkat cells using the ImmuneSigDb. Upper panel shows top 10 gene-sets enriched in BHLHE40-OE relative to WT, while the bottom panel shows top 10 depleted gene sets. Gene set names are provided on the left side of the figure while broad phenotypic labels are provided on the right.