

Main Figures

Figure 1. Heritable transcriptional states in expanded clonal T cells in vivo

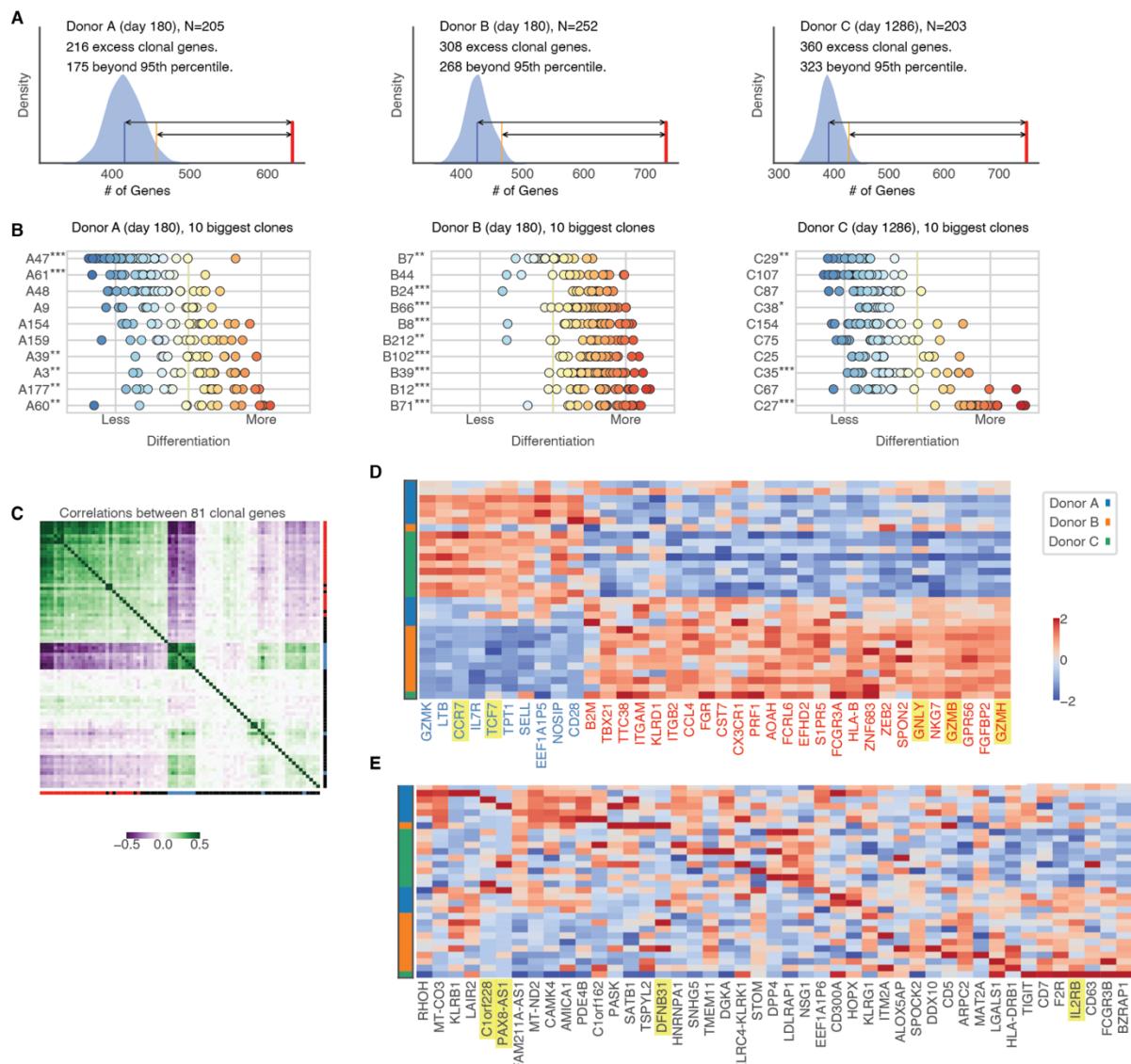


Figure 1 (legend):

(A) Numbers of clonally variable genes found in top 10 largest clones for Donors A-C based on ANOVA F-statistic. Timepoint post-vaccination and cell numbers are shown near donor ID. Here, clonally variable genes are those with unadjusted $p < 0.05$, and their number is estimated by comparison to the 95th percentile among 1000 permutations of clone labels (blue KDE-smoothed histogram). **(B)** Distribution of cells from each clone and donor, according to differentiation state (top genes contributing to PC1, Figures S1B and S1C). Clones showing strong bias as compared to the full donor population are labeled (one-sample Kolmogorov-Smirnov test, two-sided p-value, * < 0.05 , ** < 0.01 , *** < 0.001). **(C)** Correlation plot indicating highly correlated (green) and anti-correlated (purple) modules among the most clonally variable genes (excluding 25 RPL/RPS genes). Genes are marked with red/blue if they are associated with differentiation state, and black otherwise. **(D,E)** Heatmaps indicating average clonal expression levels of genes with high levels of contribution to PC1 (**D**) and other clonal

genes not associated with differentiation states (**E**). Clones are ordered according to average PC1 score of all individual cells. Donors are indicated by color (Blue: A, Orange: B, Green: C) and rows depict average gene expression level of each clone (z-score). Highlighted genes are those mentioned in the main text.

Figure 2. Heritable transcriptional states in expanded clonal T cells in vitro

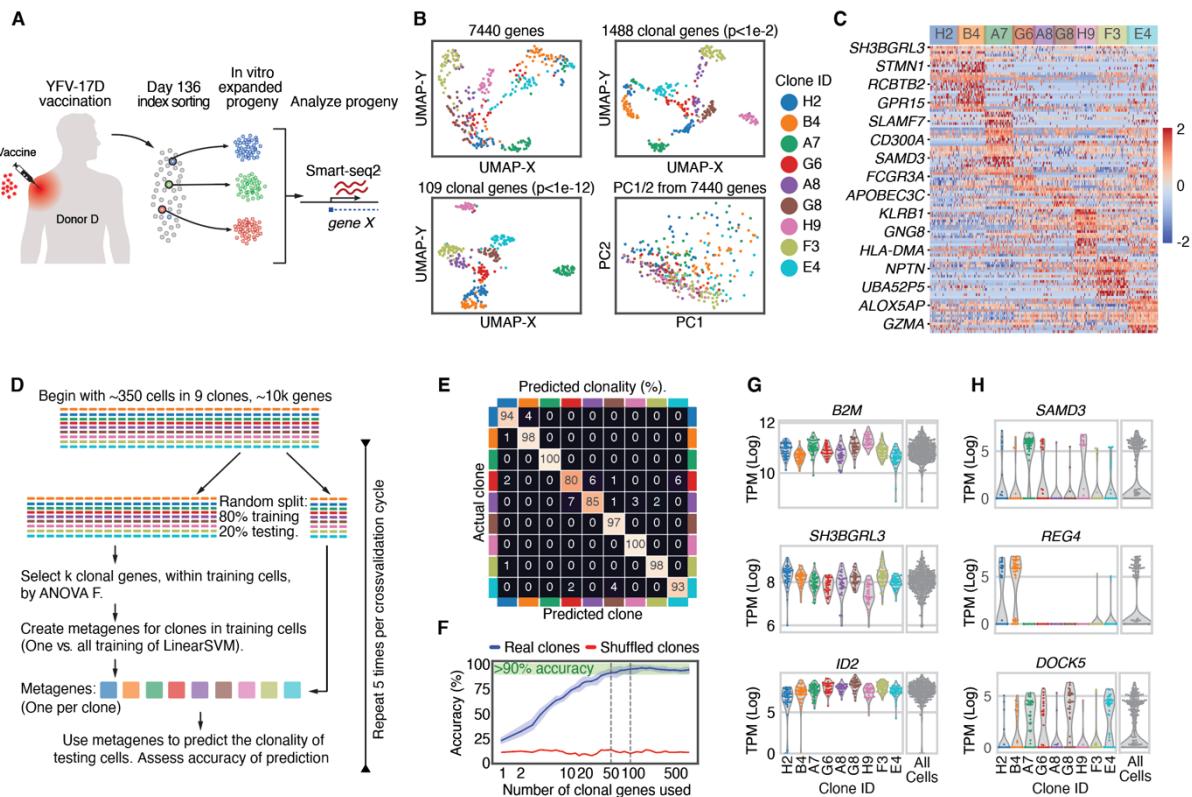


Figure 2 (Legend):

(A) Schematic illustrating experimental strategy for isolating and expanding individual T cells in vitro. Single HLA-A2/YFV NS4b-specific memory CD8+ T cells were index-sorted and expanded with irradiated autologous feeder cells, IL-2 and NS4b peptide for 21 days prior to analyses. **(B)** Visualization of 352 cells from 9 different clones based on UMAP analysis using the ten first principal components (PCs) by PCA on all genes (7,440 genes, excluding TCR components and low expressed genes) (top left square), on clonally variable genes defined by both ANOVA F-statistic and Kruskal-Wallis as significant ($n=1,488$ genes; $p < 1e-2$, top right) or highly significant ($n=109$ genes; $p < 1e-12$, bottom left). The clonal distribution based on the two first PCs is shown in lower right plot. **(C)** Heatmap displaying 109 clonally variable genes defining distinct clonal transcriptome profiles. **(D)** Schematic illustrating strategy to test identification of single cells based on SVM classifier. **(E)** Confusion matrix displaying accuracy for each clonal population. **(F)** Prediction accuracy for all clones (real clones) compared to prediction accuracy of test performed on randomly assigned ‘clones’ (shuffled clones) relative to numbers of genes included for prediction. **(G)** Examples of highly expressed genes which show clonal variability (‘tunable’ genes) and **(H)** genes which are either ON or OFF in the majority of cells from each clone. All 352 cells are shown together in gray on right hand of each plot.

Figure 3. Clonal transcriptional signatures are robust and persists through subclonal diversification

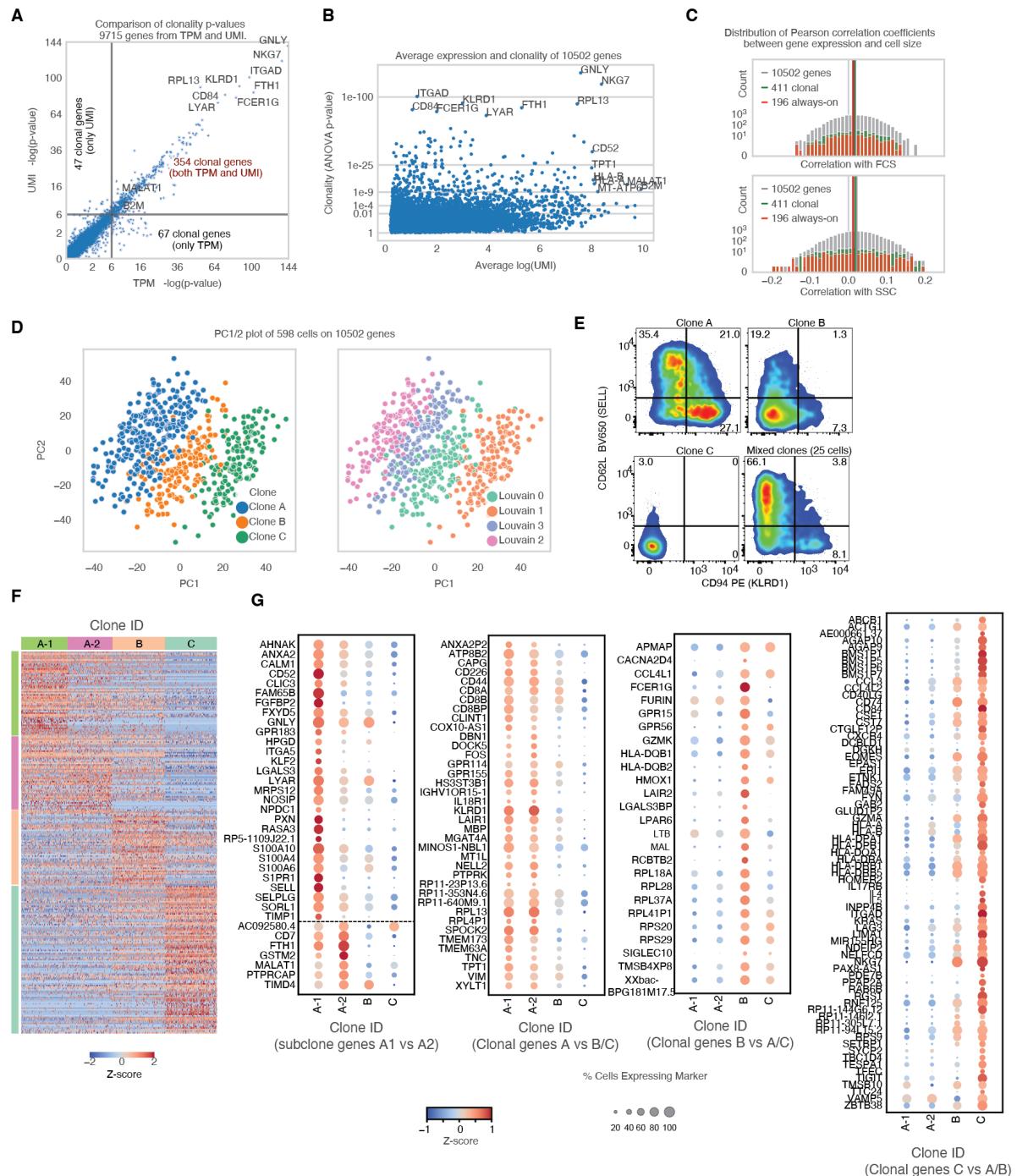


Figure 3 (Legend):

(A) Comparison of ANOVA statistics measured for clonal gene expression variability between 3 large clones using either transcripts per million (TPM) or UMI-based quantification strategies for Smart-seq3 data (Table S4). (B) Relationship between gene expression levels (based on UMI counts) and clonal variability of gene expression. (C) Pearson correlation coefficients for gene expression levels (UMI) vs cell size (FSC, top

graph) and cell granularity (SSC, bottom graph) (10,502 detected genes, 411 highly variable clonal genes, or 196 ‘always on’ genes expressed by all cells). **(D)** Principal component analysis (PCA) performed on all genes (10,502 genes) from all high quality single cell libraries (598 cells) from 3 clones. Cells are colored according to clonality. Louvain clustering (right plot) indicating that clusters correspond to unique clones with clone A having two distinct clusters (Louvain 2 and 3, later denoted as A-1 and A-2). **(E)** Protein expression for established differentiation/activation markers on clones A, B, C and a 25-cell mixed clonal bulk population generated in parallel. Clone A shows a clear split in the population according to these two markers which typically are associated with more highly differentiated cells (CD94, gene ID: *KLRD1*) and less differentiated cells (CD62L, gene ID: *SELL*). **(F)** Heatmap highlighting clonally distinct gene expression profiles, 411 genes and 598 cells. **(G)** Among 411 genes, those which identify each clonal population including genes which separate sub-clonal populations A-1 and A-2 (left) and genes enriched in Clone A vs B or C, and genes enriched in Clone B vs A or C, and genes enriched in Clone C vs A or B.

Figure 4. Shared transcriptional identities of progeny from sister clones separated in vivo

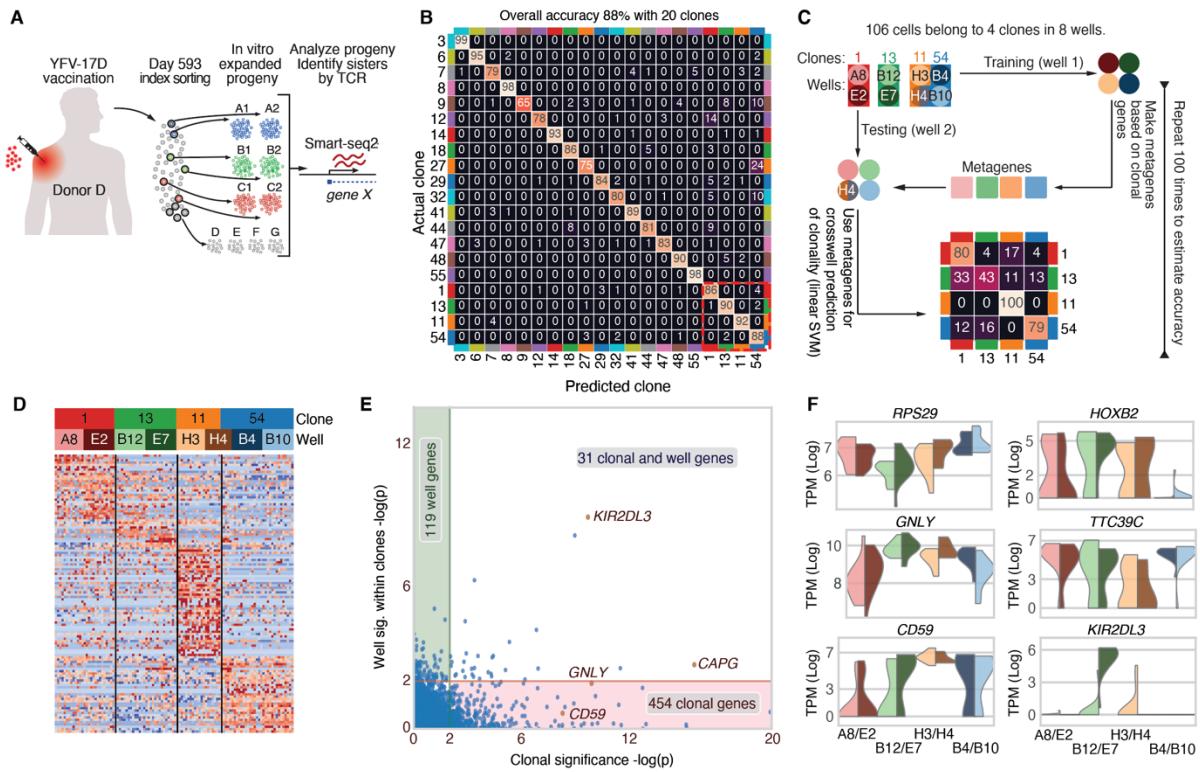


Figure 4 (Legend):

(A) Schematic illustrating experimental strategy for isolating and expanding sister clones in vitro. (B) Confusion matrix showing SVM prediction accuracies for all clonal cells (defined by TCR sequence, combining sisters in separate wells (clones 1, 13, 11 and 54)). (C) Schematic illustrating strategy to measure prediction accuracy by SVM algorithm trained on clonal T cells from one sister (in a single well) to predict cells from the second sister (in a separate well). One well (H4) contained a mixture of two clones (cl11 and cl54) and was used as a test well. Prediction accuracies are reported in confusion matrix shown underneath schematic illustration. (D) Heatmap showing clonal genes which were shared by clonally related cells derived from each sister in separate wells. Well H4 is bisected to indicate cells from clones 11 and 54 respectively. (E) Nested ANOVA test to estimate the number of genes which show significant variability according to which well they are from versus which clonal origin they share. Genes on the Y-axis (well genes: 119) show significant transcriptional differences arising during activation in a given well independent of clonal relationships between wells. X-axis genes (clonal genes: 454) are clonally variable and shared by cells from distinct sisters. (F) Split violin plots showing clonally variable gene expression patterns by sister clones. *KIR2DL3* is a rare example of a clonally variable gene which also varies between sisters.

Figure 5. Clonally heritable gene expression in the mouse central nervous system

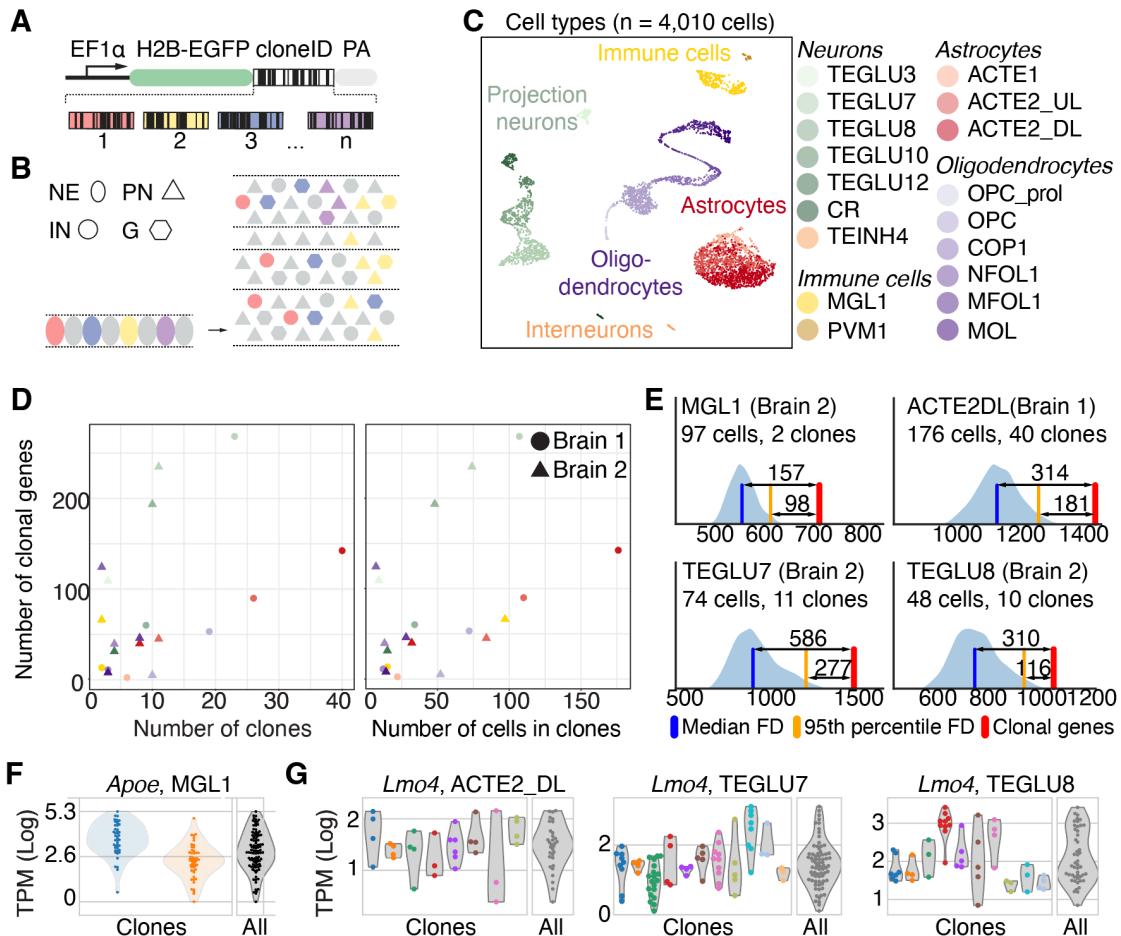


Figure 5 (Legend):

(A) A lentivirus library encoding nuclear-localized EGFP and about 1 million expressed barcodes (cloneID) for unique labeling of progenitor cells and high-throughput clonal tracing. This approach enables simultaneous clonal tracking and gene expression profiling. **(B)** Mouse cortical development from embryonic age 9.5 (E9.5) to post-natal day 14 (P14). Neuroepithelial cells (NE) generate a large diversity of cell types including excitatory projection neurons (PN), inhibitory interneurons (IN) and non-neuronal glia cells (G). Each color represents a distinct barcode. **(C)** Visualization of identified cell classes using UMAP. In total 4,010 single cell transcriptomes were collected from the somatosensory cortex of two P14 mouse brains that were classified into 18 cell types. Capital black letters indicate a unique identifier for each cell type take from www.mousebrain.org. Colors indicate five broader cell type classes: astrocytes (reds), immune (yellows), interneurons (oranges), projection neurons (greens), and oligodendrocytes (purples). **(D)** Scatter plots showing that the number of clonal genes positively correlates with the number of clones (left) and the number of cells in clones (right). **(E)** KDE-smoothed histograms displaying the results of clonal shuffling experiments to identify clonal genes per cell type and brain. Blue KDE-smoothed histogram displays the number of false discoveries (genes with ANOVA F, $p < 0.05$) among 1000 shuffles of clone labels. Blue line is the median number of false discoveries, yellow line the 95th percentile in the count of false discoveries. Red line the number of clonal genes that were found with real

clone labels. **(F,G)** Examples of clonally-variable genes *Apoe* (**F**) and *Lmo4* (**G**) in different cell types.

Figure 6. Clonal maintenance of parental chromatin accessibility patterns and clone-specific chromatin accessibility

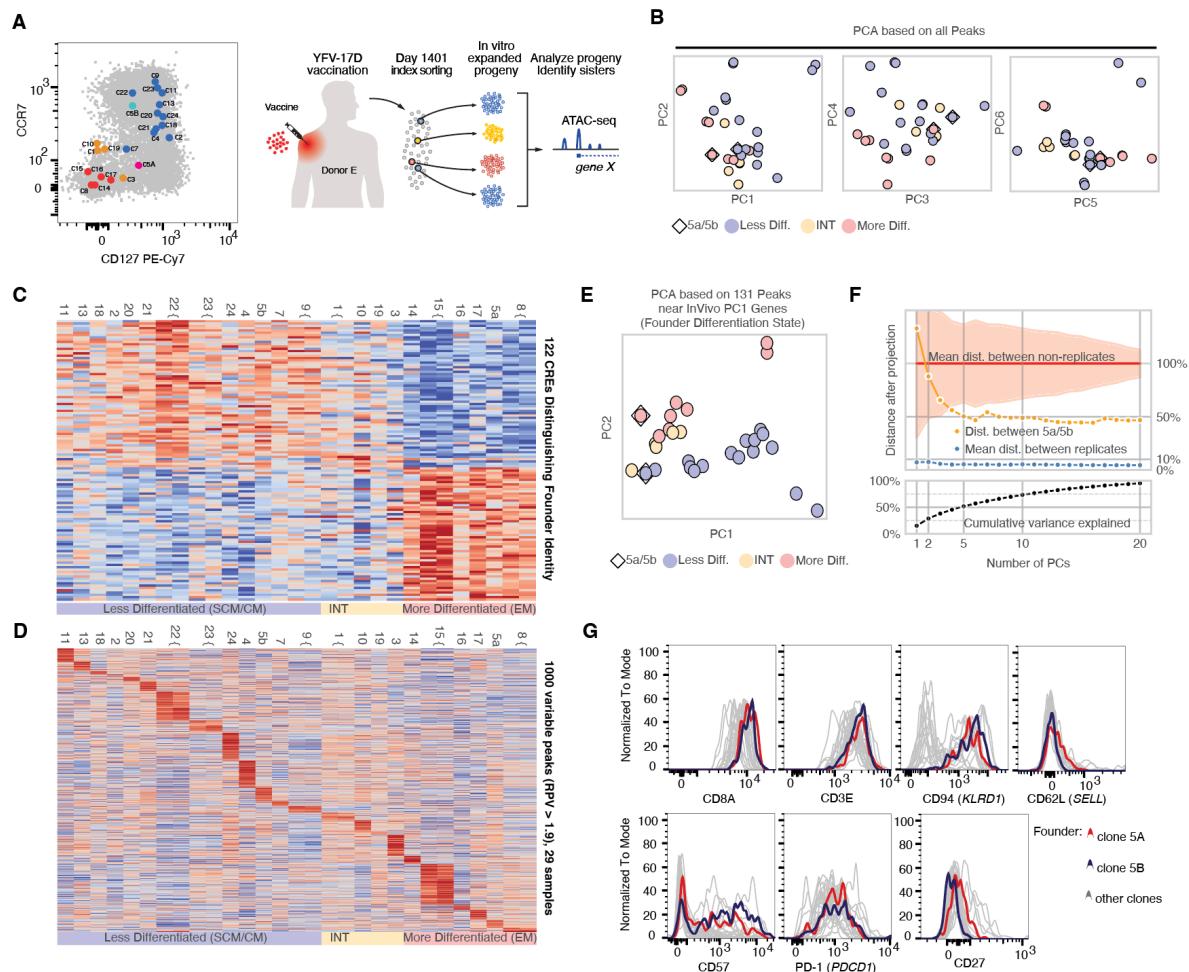


Figure 6 (Legend):

(A) Memory T cells sorted from Donor E at day 1,401 labelled on a scatter plot showing total distribution of all CD8+ T cells. Markers of differentiation state are included to distinguish clones occupying different differentiation states in vivo. Less differentiated clones (blue, top right) express high levels of CCR7 and IL7R (both contribute significantly to PC1 in Fig. S1B, C). More differentiated clones are negative for these markers and colored either yellow (intermediate) or red (most differentiated). Each population is labelled according to standard nomenclature in the CD8+ T cell field: Less differentiated – SCM (stem cell memory), Intermediate – INT, and more differentiated – EM (effector memory). Sister clones (5a and 5b) were found to occupy distinct memory differentiation states in vivo (5a = EM, pink and 5b = SCM, light blue). **(B)** Clonally expanded progeny displayed in the first six principal components based on all 26,040 peaks. Founder identity contributes to differences in the first few PCs where it separates sister progeny (shown in boxes in each plot). Sisters appear closer in PC5,6. **(C)** Progeny of SCM vs EM memory T cells have clear signatures based on founder cell differentiation states with INT cells exhibiting intermediate identities. **(D)** Clones exhibit specific enrichment for sets of CRE independent of founder identity. **(E)** PCA performed on all clones labelled based on founder identity using peaks nearby genes contributing to memory T cell differentiation differences in vivo (extended figure 1b, PC1

genes). Note 5a and 5b separated to their appropriate founder group. **(F)** Distance between unrelated clones (shaded area, all clones range), replicate clones (blue), and sisters (yellow) according to different PCs ranging from PC1-PC20. Black line indicates contribution of each PC to total variance of peak heights in dataset. **(G)** Progeny of sister clones 5a and 5b exhibit remarkably similar protein expression profiles for all activation and differentiation markers profiled, despite reactivation in separate environments.

Figure 7. Heritable differences in chromatin accessibility underlie clonally variable gene expression

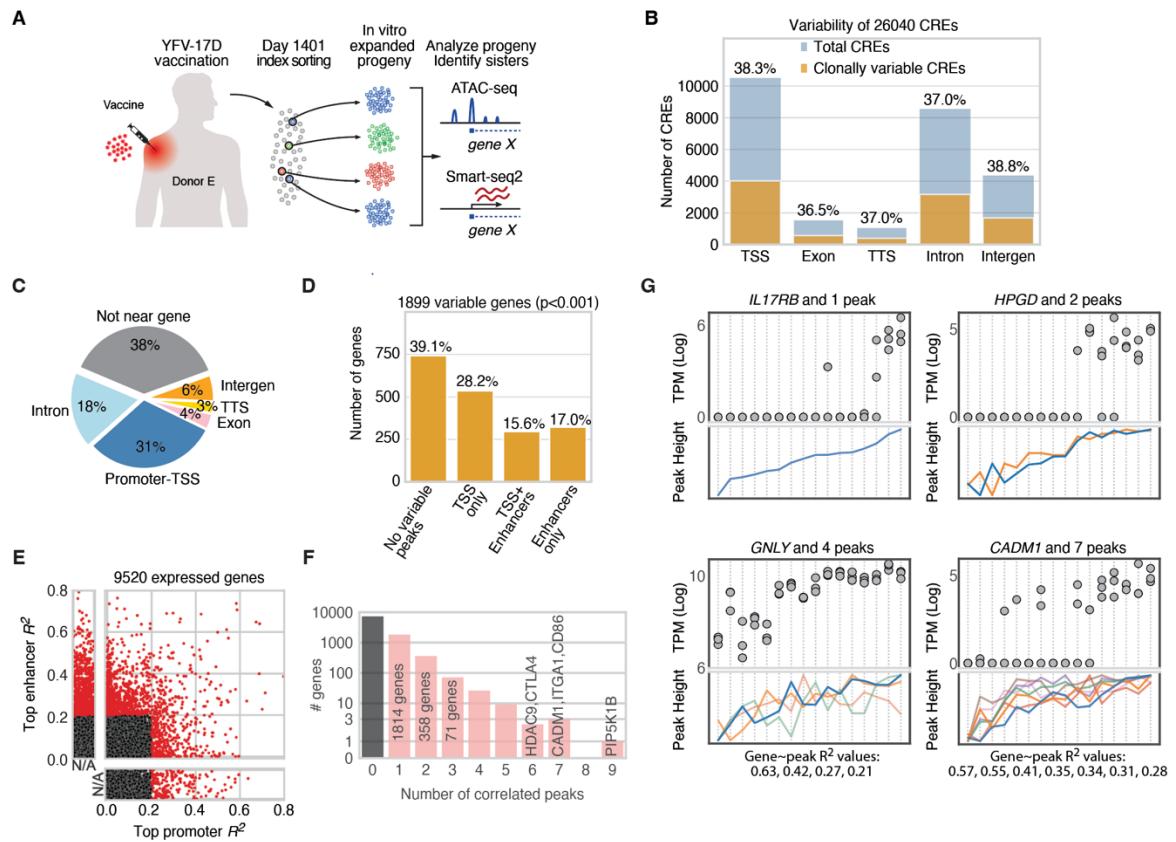


Figure 7 (Legend):

(A) Experimental approach to collect matched RNA and ATAC-seq datasets from clonally expanded T cells. **(B)** Fraction of peaks showing evidence of clonal variability (RPV>1) in genomic positions annotated as promoter (TSS) and enhancer regions. The 26,040 CREs analysed reach peak height ≥ 30 for at least one clone. **(C)** Distribution of 9,846 clonally variable CREs (RPV>1). CREs not located within 50kbp of an expressed gene in matched RNA-seq dataset (in any clone) are considered ‘not near gene’. **(D)** Relationship between variable CRE locations and genes which show clonally variable expression patterns in matched RNA-seq analysis. **(E)** Scatterplot showing correlation between peak heights and gene expression for all expressed genes (9,520 genes) in the RNA-seq dataset. 16 clonal populations are included in this analysis (including 5a and 5b). Red dots indicate CRE-gene relationships with $R^2 > 0.2$ (Pearson Correlation). **(F)** Numbers of highly correlated peaks plotted for each gene (2,284/9,520 expressed genes have at least 1 highly correlated peak). For a given gene, ‘highly correlated’ peaks are those whose R^2 with gene expression exceeds the 95th percentile among all peaks on the same chromosome. **(G)** Relationships between RNA-seq measurements (top frames, dots represent triplicate RNA-seq measurements/clone) and peak heights (bottom frames, line graphs show individual CREs, colored separately) for select genes.