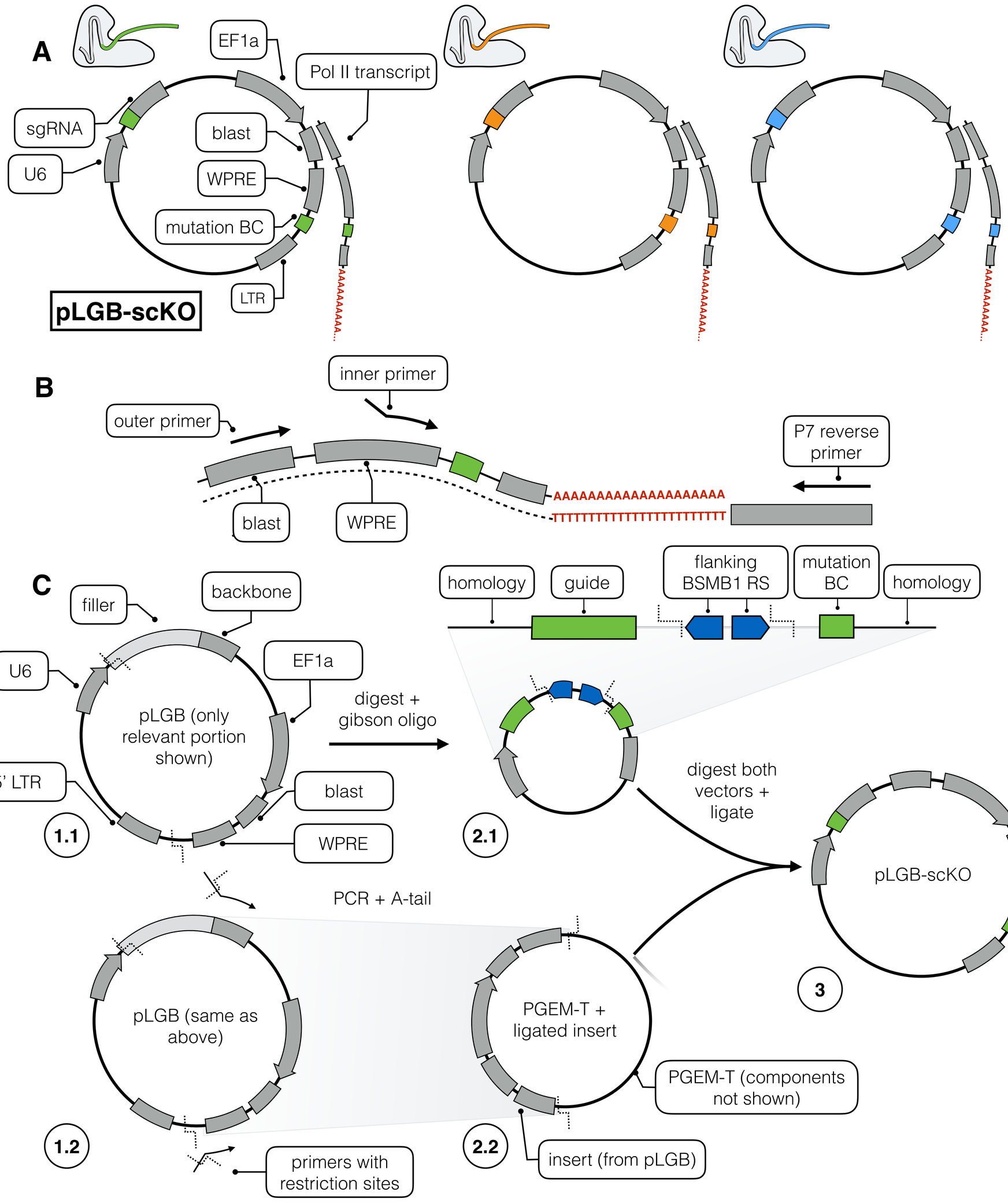
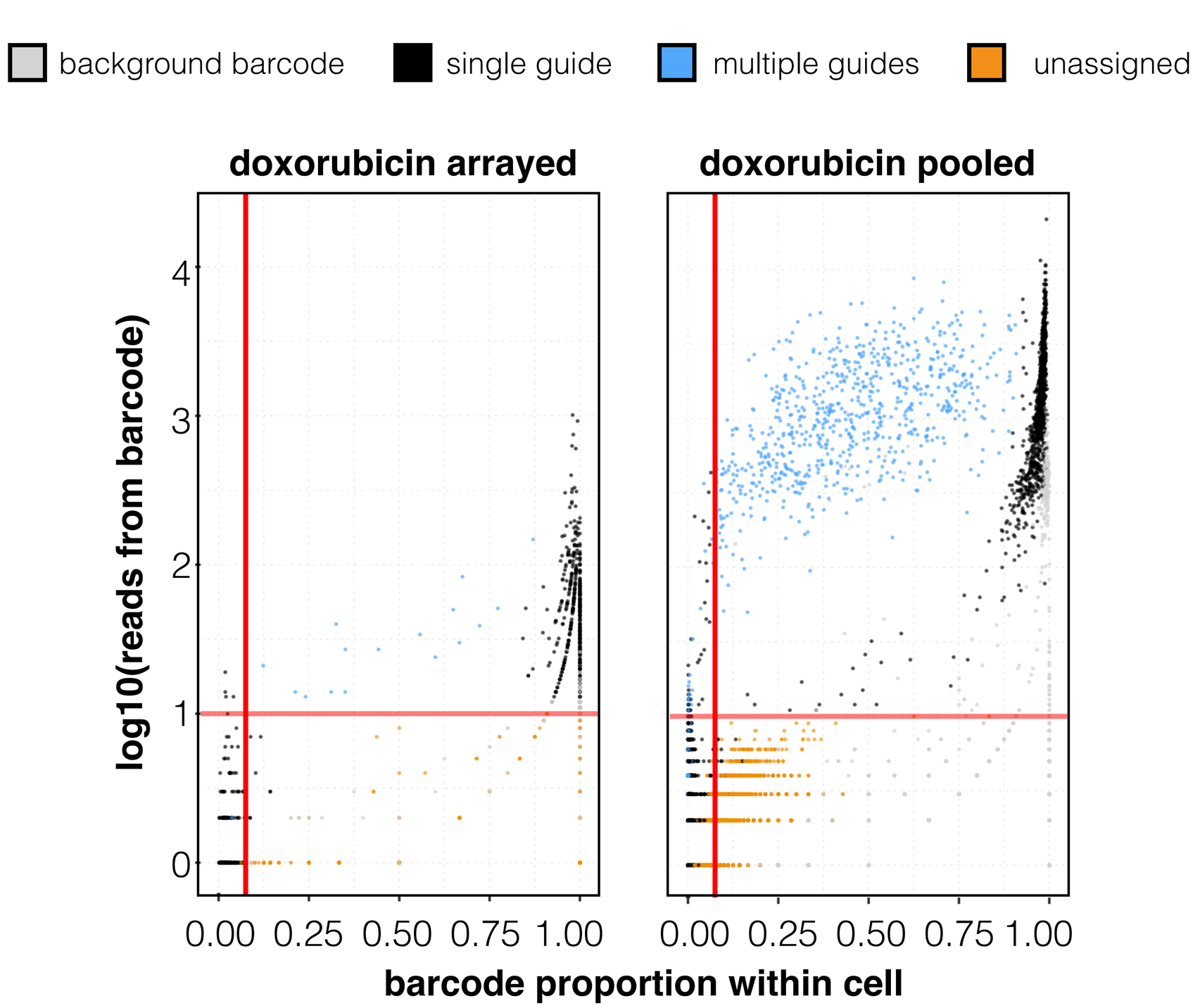


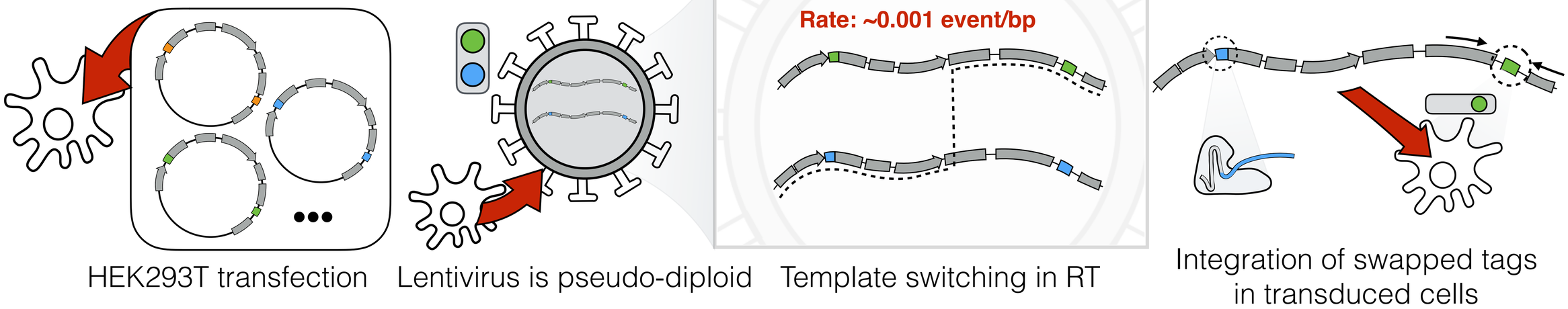
Supplementary Figure 1 Diagram of cloning protocol and barcoded transcript enrichment strategy relying on *cis* pairing of sgRNAs and barcodes (pLGB-scKO). Details on exact restriction sites and primer sequences used can be found in the methods. **A)** Schematic of our final vector relying on *cis* pairing of an sgRNA and a distal barcode. **B)** Nested PCR enrichment strategy to maximize recovery of barcoded transcripts from single-cell RNA-seq data. **C)** Pooled cloning protocol. In 1.1 we start with pLentiguideBlast and digest near the final locations of the sgRNA and paired barcode. In 2.1 an engineered library of oligos containing programmed pairs of sgRNAs and corresponding barcodes are inserted into the digested vector. In 1.2 a portion of pLentiguideBlast is amplified. In 2.2 this fragment is cloned into PGEM-T. Finally, in step 3 vectors resulting from 2.1 and 2.2 are digested with Bsmb1 and the insert from 2.2 is ligated into the backbone in 2.1 to produce the final library of sgRNAs and paired barcodes.



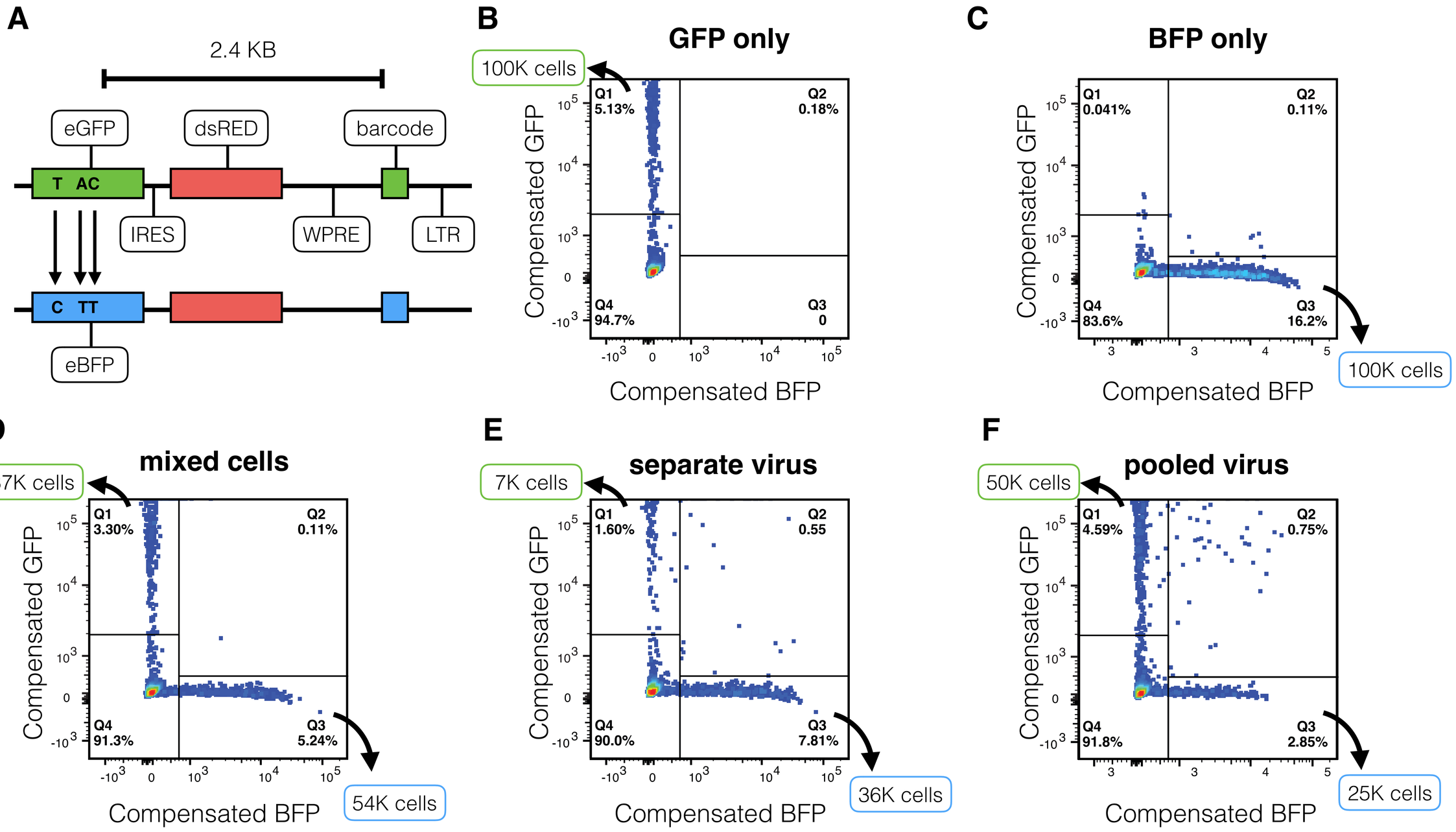
Supplementary Figure 2 Barcoded transcript enrichment quality control for arrayed and pooled pLGB-scKO experiments. Each dot represents a barcode sequence observed in a given cell. Plot of reads for a given barcode against the proportion of all barcode reads observed in a given cell for every barcode/cell pair colored according to whether the cell barcode is determined to be background from whole transcriptome scRNA-seq data, a cell with a single guide assignment, a cell with multiple guide assignments, or a cell that ultimately receives no assignment given the applied thresholds. Red lines indicate the lower-bound thresholds used to distinguish noise from true barcode observations (10 reads and 0.075 proportion within cell). All barcodes observed above the red lines are assigned to their respective cells. Left, doxorubicin treated sample from arrayed experiment. Right, Doxorubicin treated sample from pooled experiment. Note that the rate of cells with single guide assignments in the doxorubicin treated sample is approximately 81%, which is in line with the 80% expected value at an MOI of 0.45 in a selected population.

A
500 nM doxorubicin arrayed
CDKN1A
TP53I3
expression
10.0
1.0
0.1
CBFB
NCOR1
PTEN
TP53
target
B
500 nM doxorubicin pooled
C
DEG count
300
200
100
0
arrayed
pooled
experiment
D
CCNB2
arrayed
cluster average CCNB2 expression
0.3
0.2
0.1
tsne2
tsne1
E
TP53I3
arrayed
F
99.4% of assigned cells are TP53
cluster average TP53I3 expression
3
2
1
G
CCNB2
pooled
cluster average CCNB2 expression
0.09
0.06
0.03
H
TP53I3
pooled
cluster average TP53I3 expression
1.5
1.2
0.9
0.6
I
41% of assigned cells are TP53
TP53
other target
unassigned

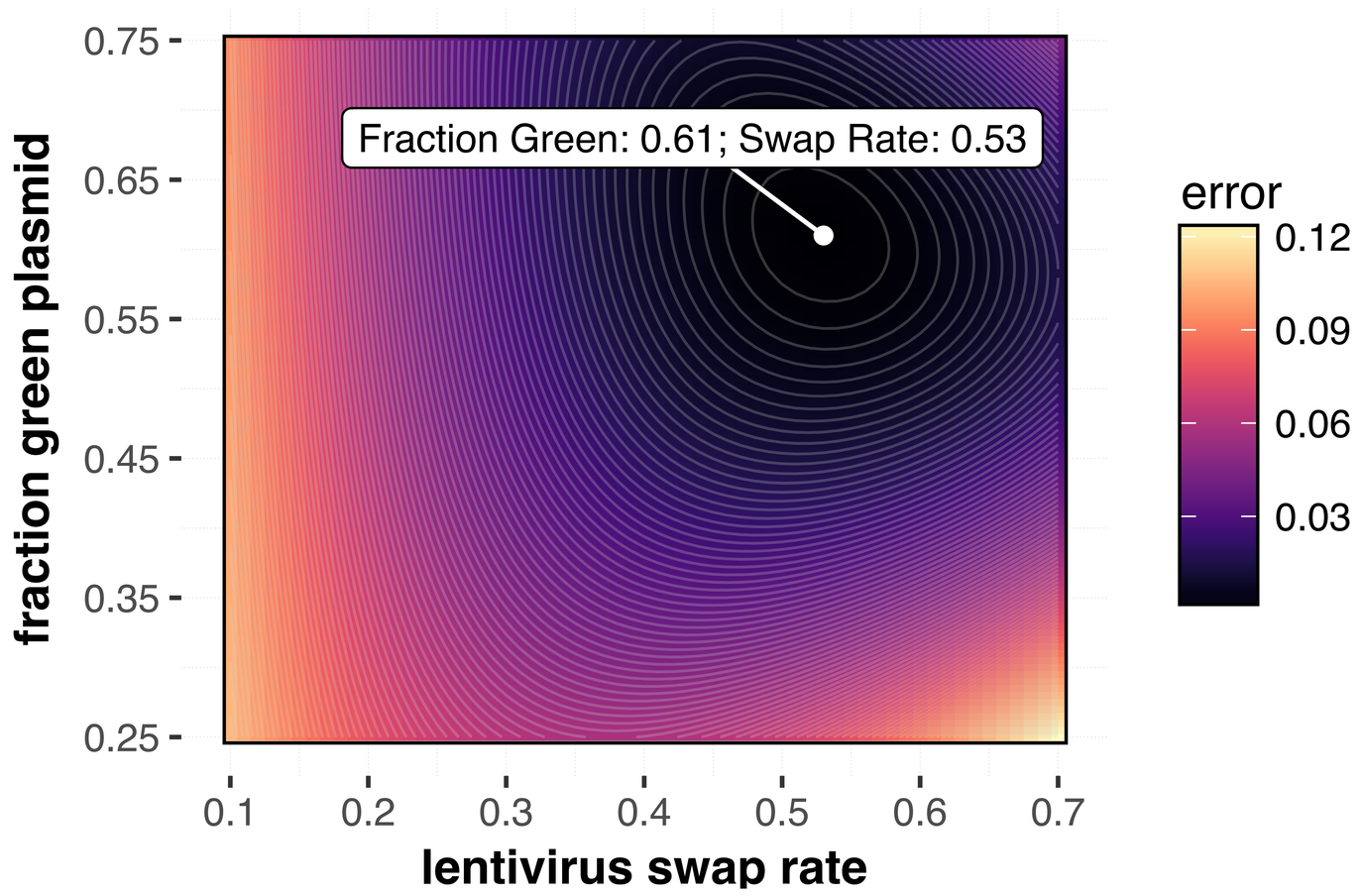
Supplementary Figure 3 Comparison of screens performed with arrayed and pooled lentivirus production using a vector that relies on *cis* pairing of sgRNAs and barcodes. Experiments were performed at different times but under the same conditions. The arrayed experiment was performed as a pilot experiment with four targets and observed an overall low rate of cells with detected barcodes. The pooled experiment was performed afterwards with 10 targets and a set of non-targeting controls and we observed a high proportion of cells with detected barcodes and good coverage of the library. To compare these experiments, only the four overlapping targets were considered for differential expression testing and the number of cells containing an sgRNA to each target and sequencing depth were matched between samples to control for power differences. **A)** Size-factor normalized *CDKN1A* and *TP53I3* expression across *TP53* and the three other targets in arrayed screen. **B)** *CDKN1A* and *TP53I3* expression across *TP53* and three other targets in pooled screen that overlap with the arrayed screen. Values for all violin plots capped to a minimum of 0.1 to facilitate plotting. **C)** Comparison of the number of differentially expressed genes detected at an FDR of 5% for arrayed across the target label in the arrayed and pooled experiments (n = 259 cells per dataset sampled 10 times such that cell counts per target match between datasets; see methods). For each boxplot from right to left the min, 1st quartile, median, mean, 3rd quartile, and maximum are 242.0, 246.8, 275.5, 277.7, 297.8, 325.0; and 5.00, 6.25, 9.00, 8.90, 10.75, 14.00 respectively. **D-I** show t-SNE embeddings of all cells collected in arrayed and pooled experiments (n = 2274 cells and n = 2191 respectively). **D and G)** Average size-factor normalized expression of *CCNB2* for the arrayed and pooled screens. **E and H)** Same as D and E but for *TP53I3*. There is a single distinct *CCNB2*-high, *TP53I3*-low cluster one expects for *TP53*-null cells in this experiment. **F and I)** The same t-SNE embeddings, colored by their knockout assignments; *TP53*, other targets, and cells without any assignment are shown as separate colors. The arrayed experiment shows that 99.4% of the cells with an assigned target within the *CCNB2*-high, *TP53I3*-low cluster are *TP53* knockouts compared to only 41% in the corresponding cluster from the pooled experiment.



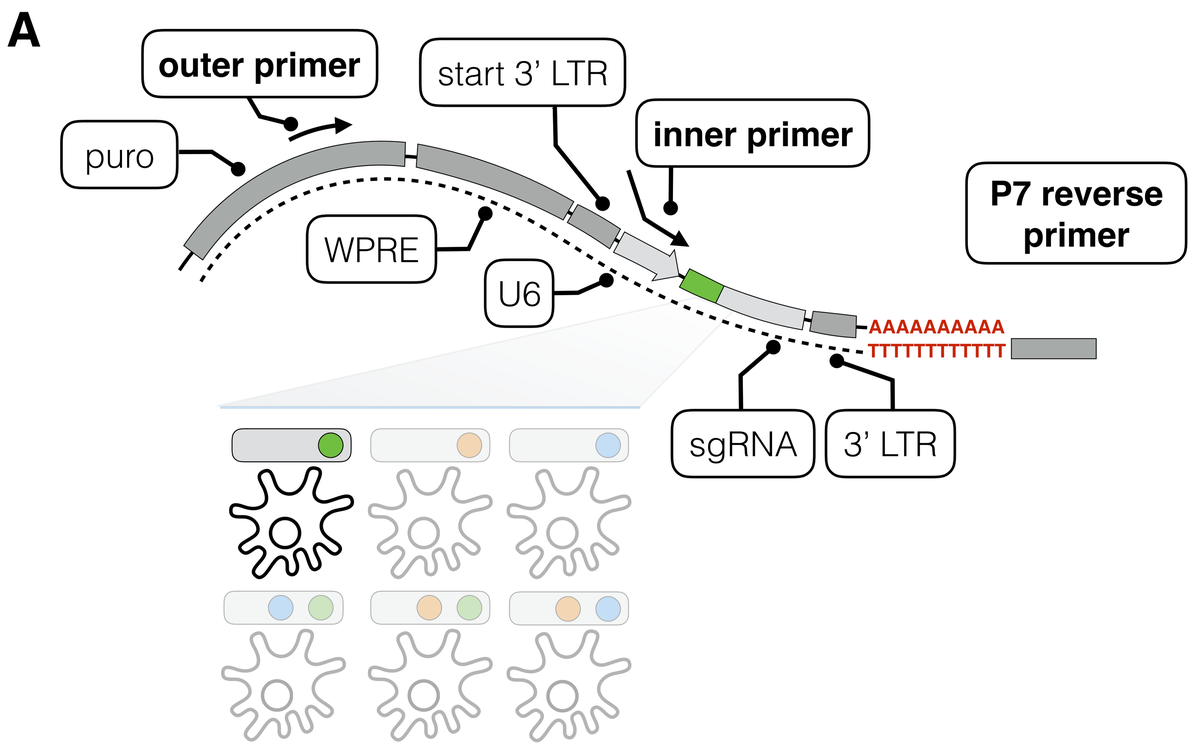
Supplementary Figure 4 Schematic illustrating how template switching in lentivirus leads to the swapping of associations between the sgRNAs and their corresponding barcode.

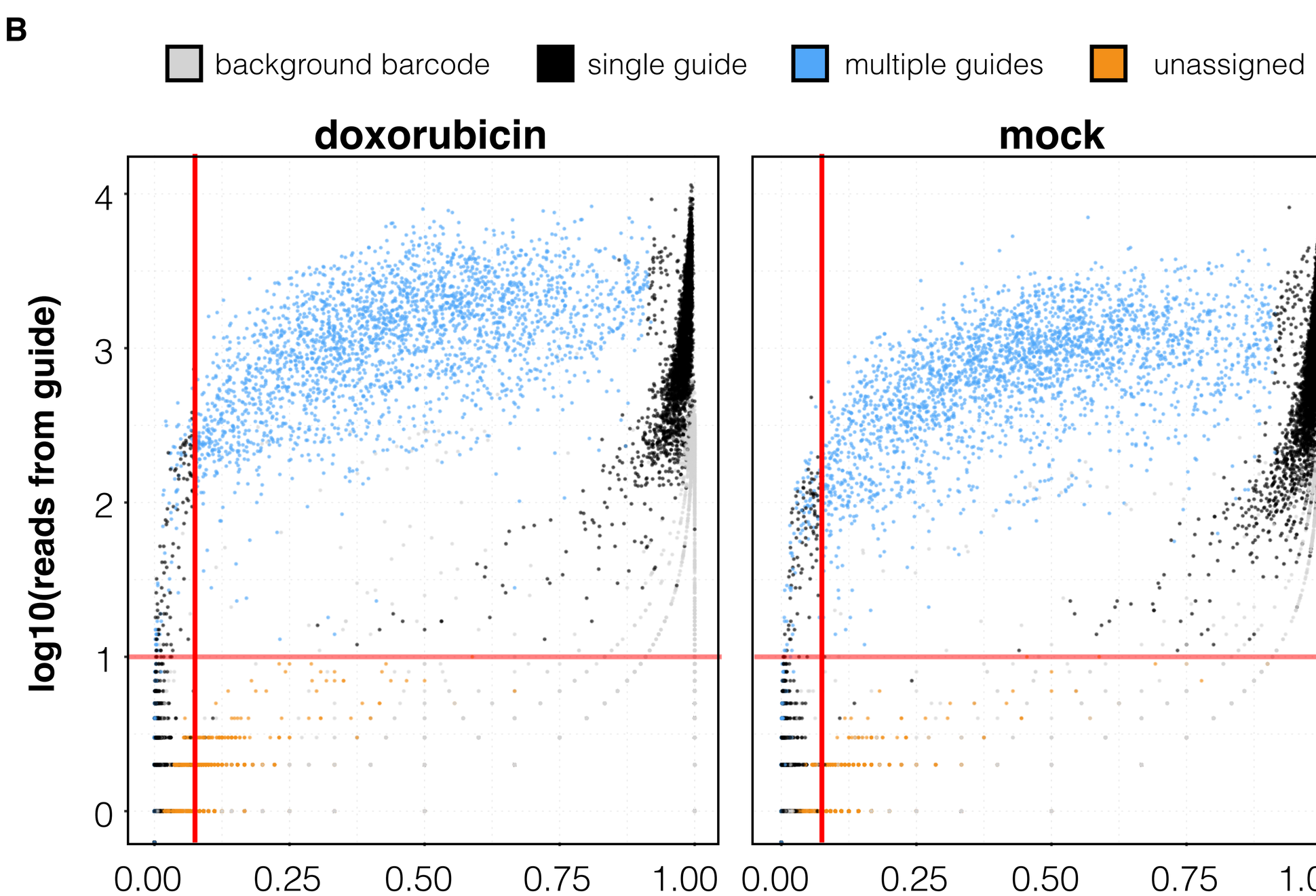


Supplementary Figure 5 Design and sorting of GFP and BFP positive fractions in lentivirus barcode swapping experiment. **A)** Schematic of vectors (pHAGE-GFP and pHAGE-BFP) designed to quantify template switching rate at 2.4 kb using a FACS readout. FACS plots are shown for sorted cells in samples corresponding to **B)** GFP only transduced cells **C)** BFP only transduced cells **D)** GFP and BFP only transduced cells mixed just prior to FACS as a control **E)** cells transduced with BFP and GFP virus that was generated separately but pooled prior to transduction **F)** cells transduced with BFP and GFP virus that was generated from pooled plasmids. The fraction of green plasmids assumed in the determination of lentivirus swap rate from FACS experiments is taken as the fraction of GFP+ cells relative to the total GFP+ and BFP+ cells from this sort (4.59 / (4.59 + 2.85) or 61.7%). This accounts for the fact that plasmids were likely not completely equimolar. The approximate number of total cells cells sorted in each fraction is indicated along the appropriate axes on each plot. Similar results were obtained for **B-F** in n = 2 independent viral transductions.

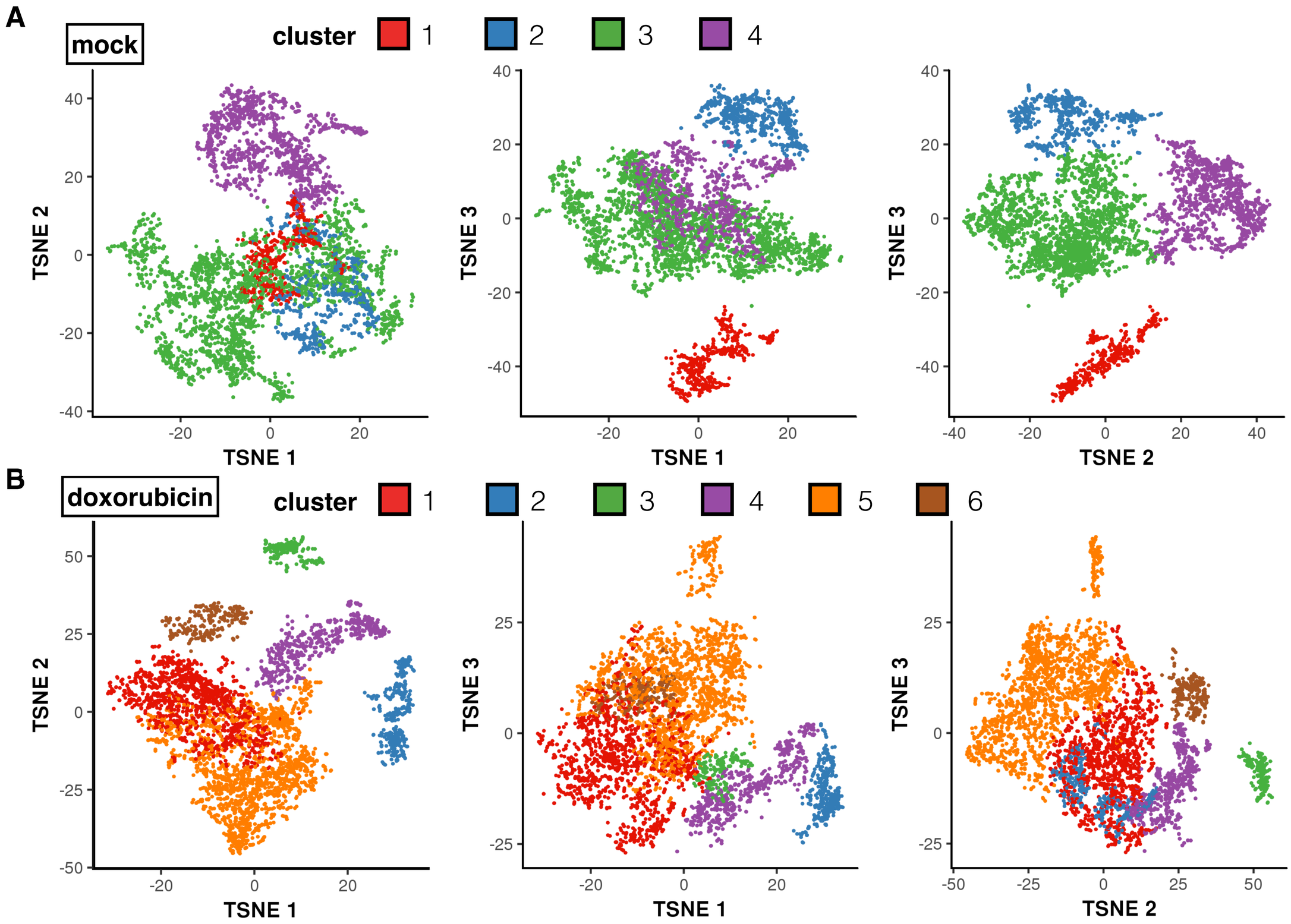


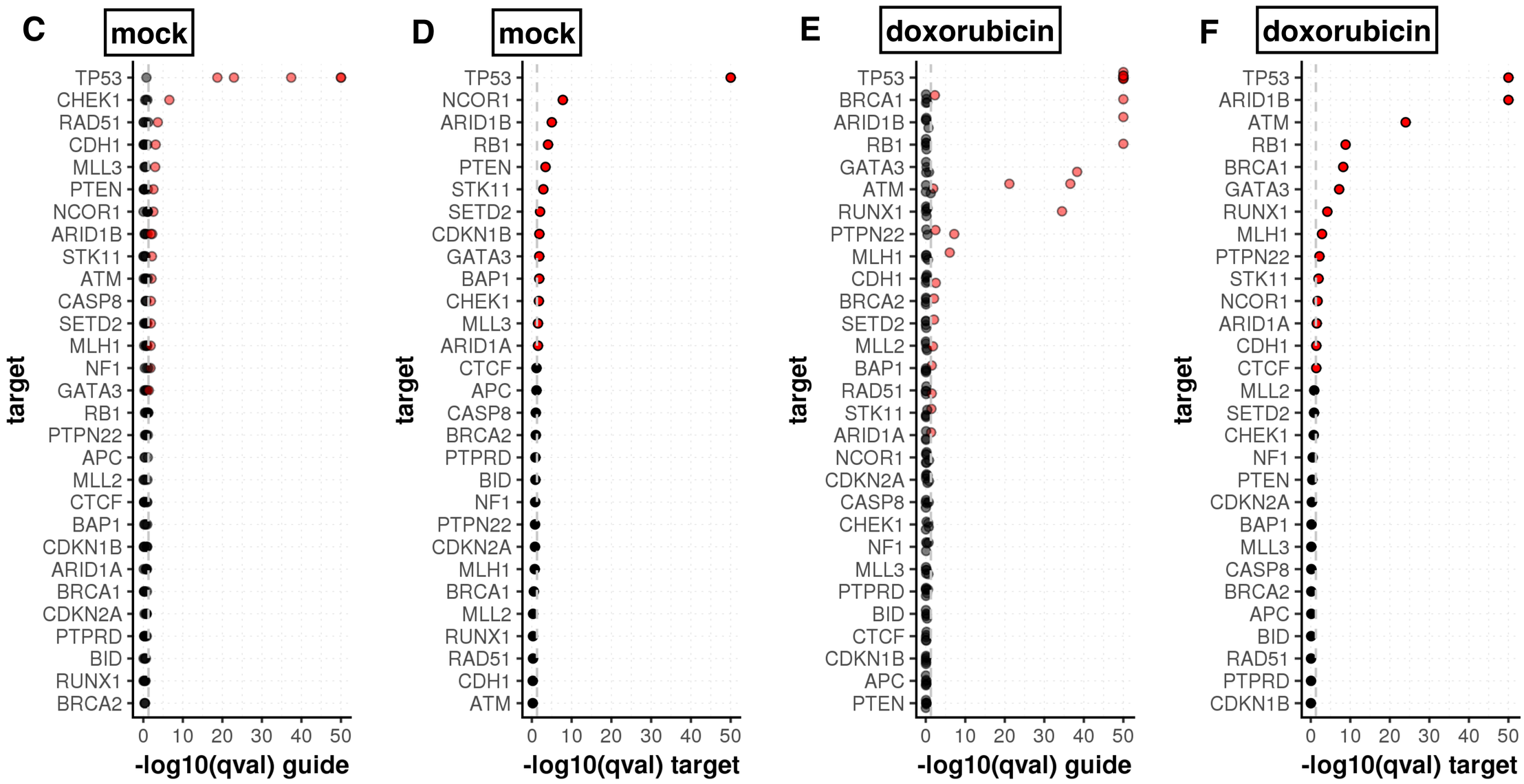
Supplementary Figure 6 Simulation of concordance between observed and expected data obtained from FACS experiment in **Supplementary Figure 5** to quantify template switching rate at 2.4 kb separation between paired sequences. **Figure 1e** assumed a fraction of 0.617 of GFP plasmid in the original green plasmid / blue plasmid mix as determined from FACS in **Supplementary Figure 5**. In this figure, both the fraction of GFP plasmid and lentivirus swap rate are varied to obtain the set of parameters that best fit the collected fraction GFP and BFP barcodes in the green and blue sorted samples (n = 4 measurements). The sum of the squared error between expected and observed values from FACS given each combination of parameters is shown.



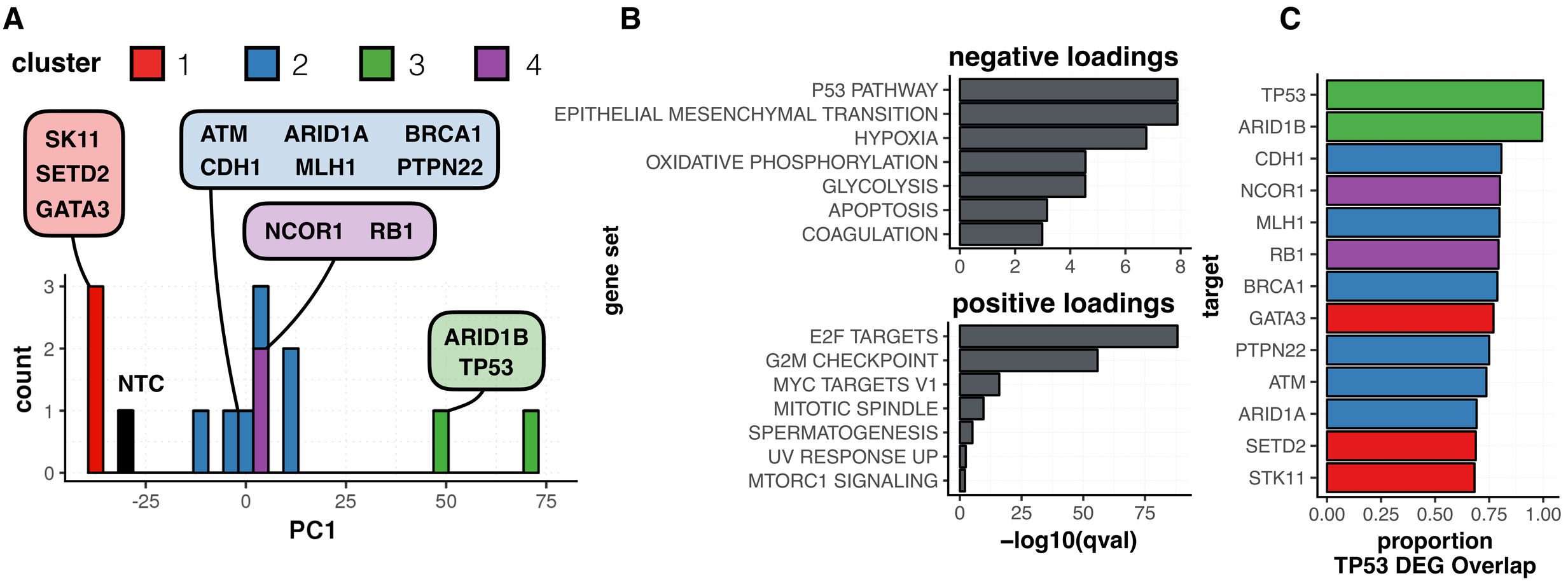


Supplementary Figure 7 A) Schematic of PCR enrichment of barcoded transcripts from CROP-seq samples. **B)** Guide transcript enrichment quality control plot for tumor suppressor knock-out screen performed with CROP-seq. Each dot represents a barcode sequence observed in a given cell. Plot of reads for a given barcode against the proportion of all barcode reads observed in a given cell for every barcode/cell pair colored according to whether the cell barcode is determined to be background from whole transcriptome scRNA-seq data, a cell with a single guide assignment, a cell with multiple guide assignments, or a cell that ultimately receives no assignment given the thresholds applied. Red lines indicate the lower-bound thresholds used to distinguish noise from true guide observations (10 reads and 0.075 proportion within cell). All guide observed above the red lines are assigned to their respective cells. Left, Doxorubicin treated sample from CROP-seq experiment. Right, Mock sample from CROP-seq experiment. Note that the rate of cells with single guide assignments across both experiments above is approximately 81%, which is in line with the 80% expected value at an MOI of 0.45 in a selected population.

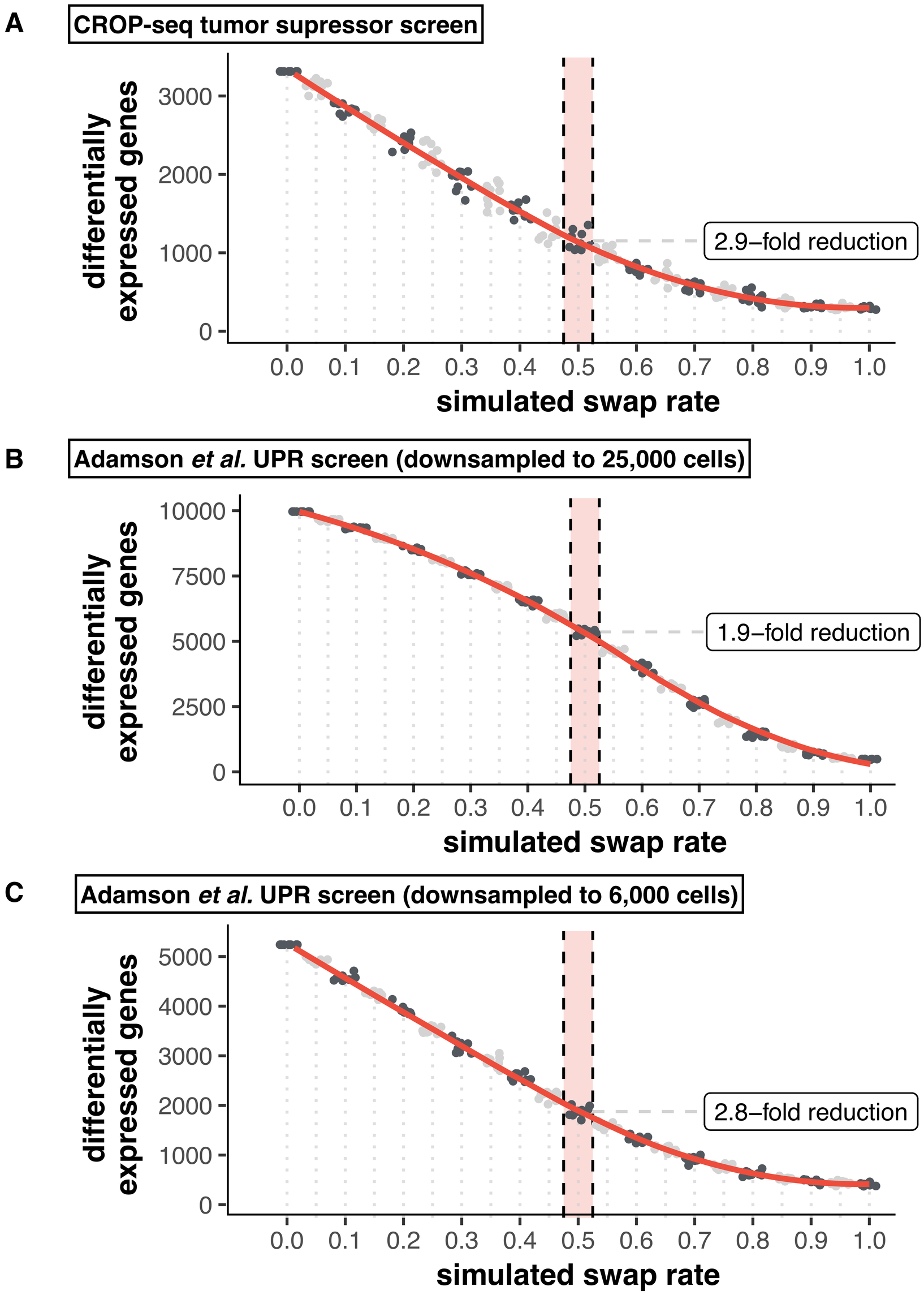




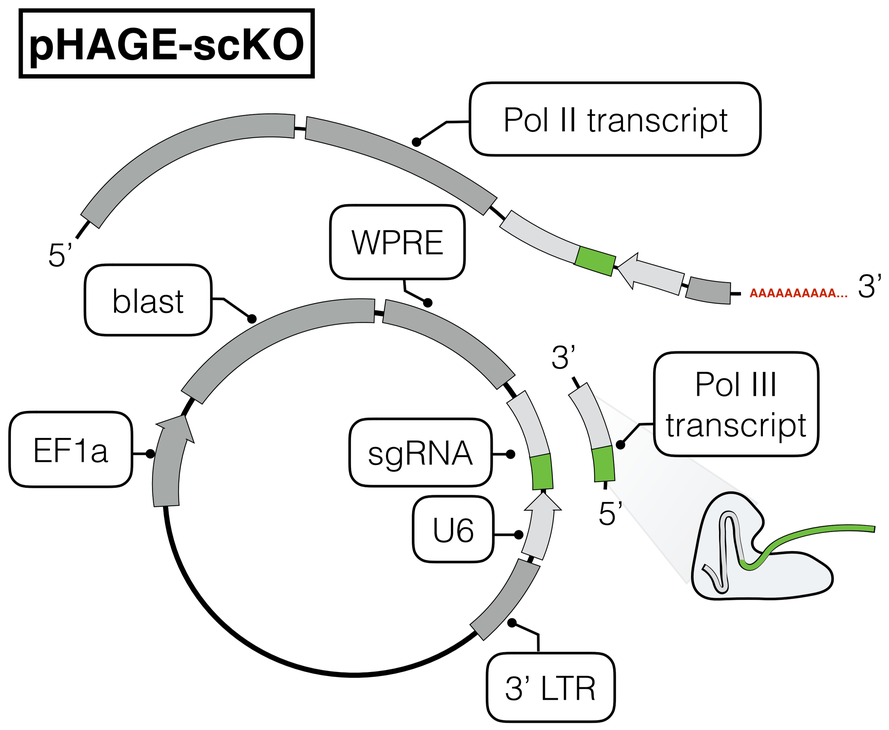
Supplementary Figure 8 Loss of several targets alter the distribution of mock and doxorubicin exposed cells within tSNE clusters. **A-B)** 3D tSNE embedding and clustering of all cells with single target assignment in mock and doxorubicin treated samples (n = 4740 and n = 4467 cells respectively). **C-F)** Chi-squared test qvalues resulting from testing for differences in the distribution of targets in our screen at both the individual sgRNA **(C and E)** and overall target levels **(D and F).** Comparisons are relative to the distribution of non-targeting controls across tSNE clusters for mock and doxorubicin treated samples, respectively (qvalues were capped to 1e-50 for visualization). Significant differences below a qvalue of 0.05 are colored in red (boundary marked by the grey dashed line).



Supplementary Figure 9 Enriched target-cluster pairs highlight tumor suppressors that share various degrees of a *TP53* deficient signature **A)** Fisher's exact with weights applied to guides according to an expectation maximization procedure were performed for the doxorubicin treated sample to find clusters from Supplementary Figure 8 panel B where particular targets were found to be enriched. Cells with target-cluster pairs that showed enrichment were used to generate an aggregate expression profile for every target within genes that are differentially expressed between *TP53* and non targeting controls (NTC). A PCA was performed on these average expression profiles and a distribution of targets across PC1 is shown colored by the cluster in which they were found to be enriched. **B)** Gene set enrichments for top positively and negatively loaded (less than -0.02 or greater than 0.02; n = 491 and n = 347 genes respectively) genes along PC1 (qval < 0.01). **C)** Differential expression tests were performed for cells within each enriched target-cluster pair, comparing each target to all NTC cells. The proportion of overlap between these differentially expressed genes and the genes differentially expressed between *TP53* and NTC is shown.

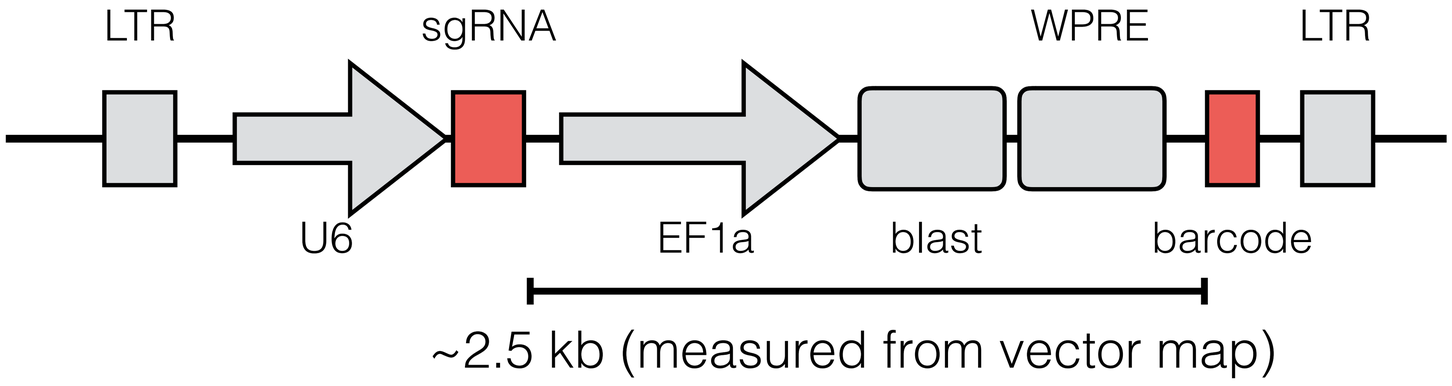


Supplementary Figure 10 Swap rate simulations for our own CROP-seq tumor suppressor screen and the unfolded protein response screen from Adamson *et al.* Each dataset was subjected to simulation of progressively higher fractions of target assignment swapping to mimic the impact of template switching. Number of differentially expressed genes across the target label at FDR of 5% is plotted at each swap rate for 10 samplings per rate. 0.5 corresponds to the 50% swap rate determined via FACS. **A)** CROP-seq tumor suppressor screen from our study (doxorubicin treated condition; n = 4467 cells used in tests). **B)** Unfolded protein response screen from Adamson et al. downsampled from ~50,000 to n = 25,000 cells to make simulations computationally feasible. **C)** Unfolded protein response screen from Adamson *et al.* downsampled to n = 6000 cells to illustrate how reduced power impacts the observed impact from simulated swapping.

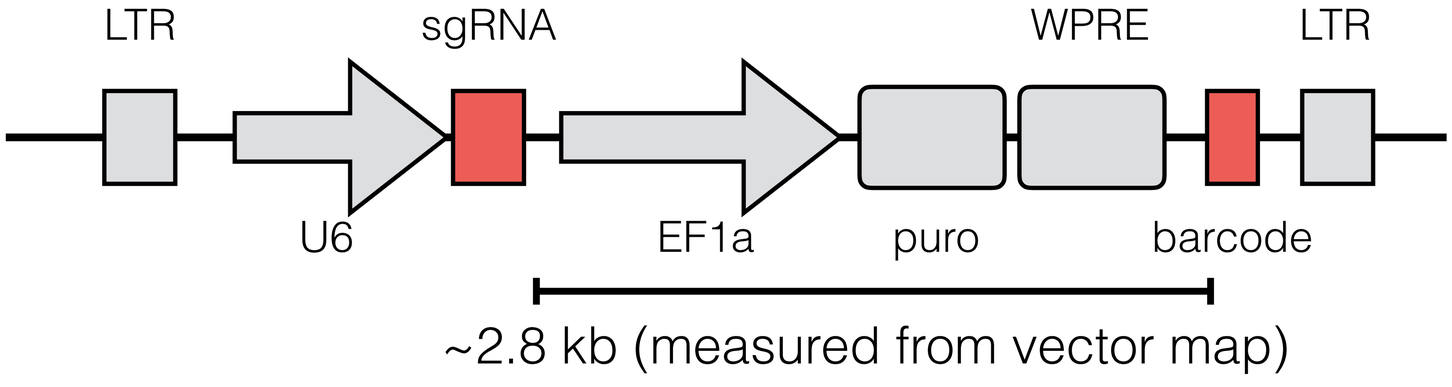


Supplementary Figure 11 Schematic of pHAGE-scKO design where guide is placed in the 3' UTR outside of the LTR such that a second copy is not generated during integration.

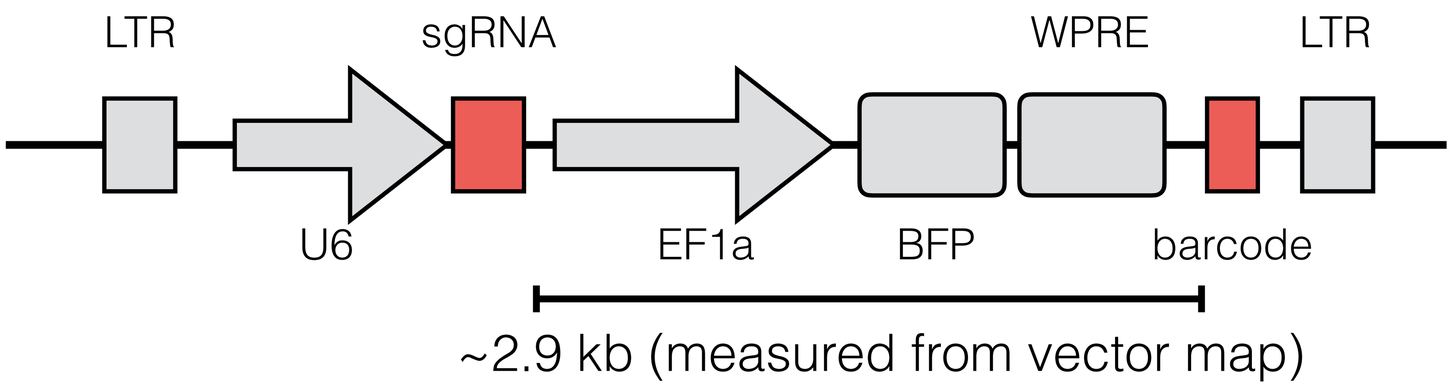
A **pLGB-scKO (our initial design)**



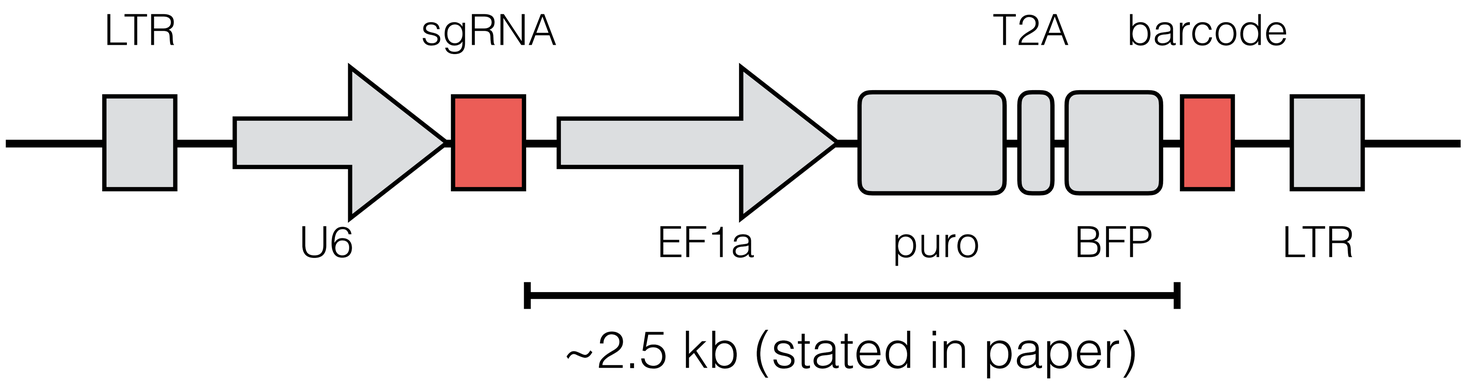
MOSAIC-seq



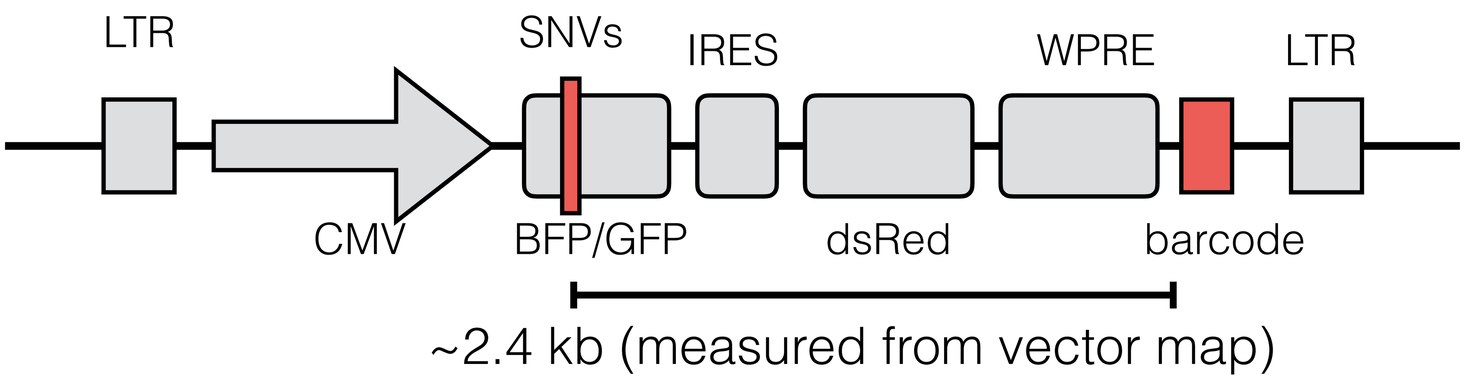
CRISP-seq



Peturb-seq



B **pHAGE-GFP/BFP (our vector to measure swap rate)**



Supplementary Figure 12 A comparison of all relevant vector designs that rely on the linkage of sgRNAs and distal barcode. **A)** vector maps for our own initial design (pLGB-scKO) and all three published designs that rely on linkage between an sgRNA and a distal barcode. Distances between the guide sequence and the barcode are indicated in the schematic of each vector. **B)** Vector map for pHAGE-GFP/BFP, the vector we used to quantify the rate of swapping that results during lentiviral packaging at a distance of 2.4 kb or greater.