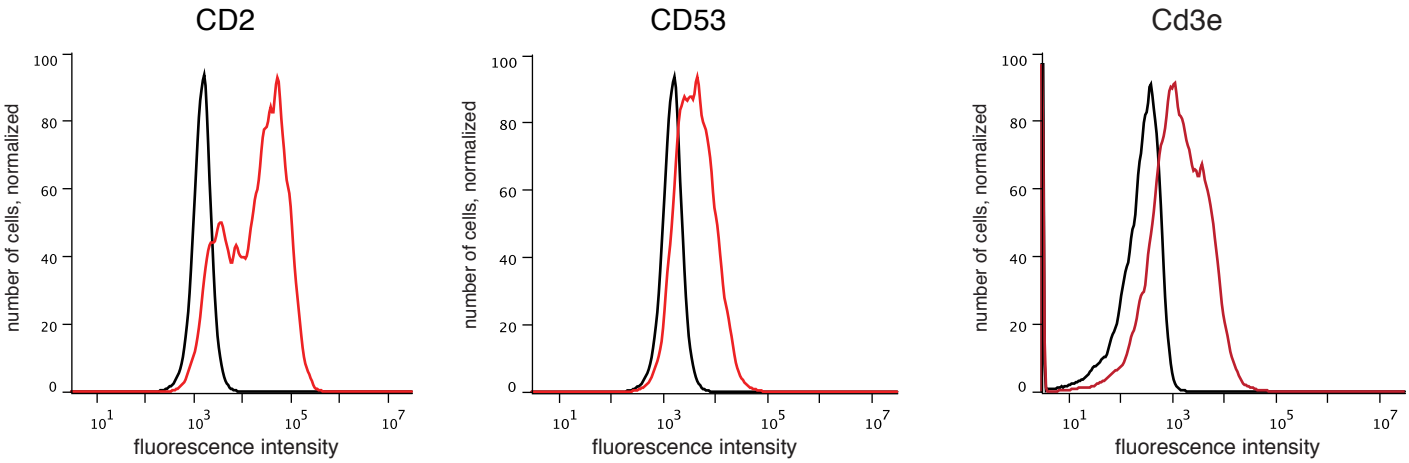
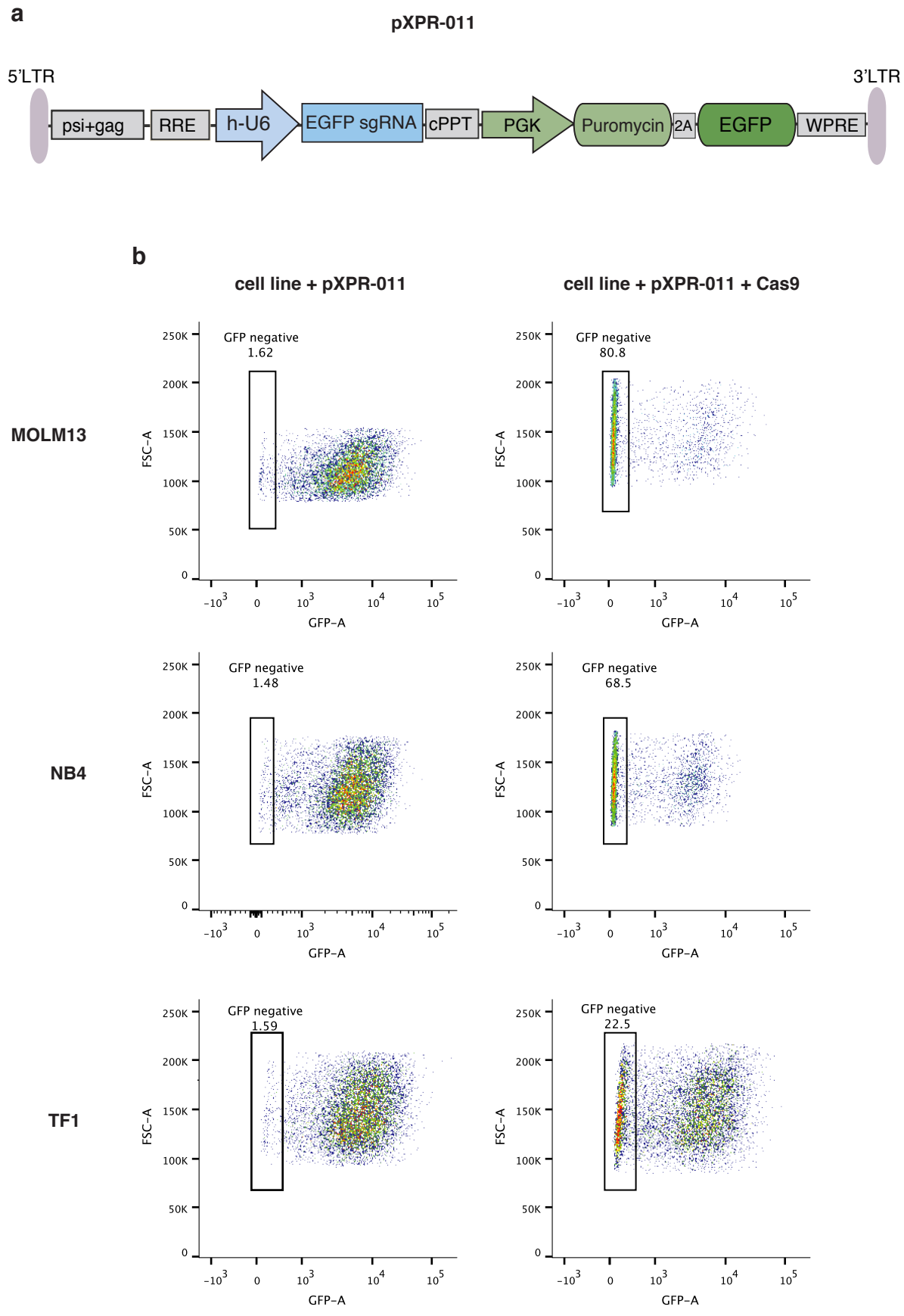
Supplementary Figure 1



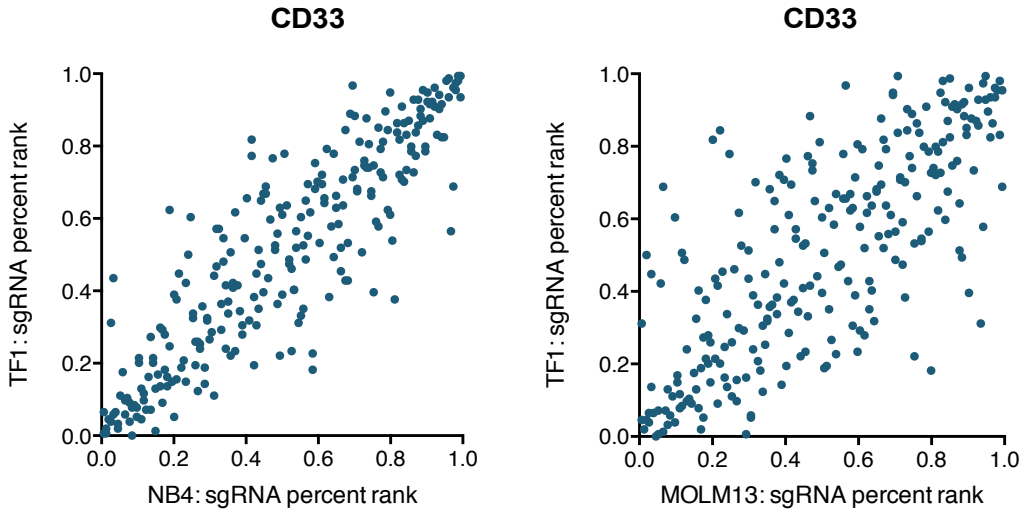
Supplementary Figure 1 Antibody staining for excluded surface markers. Stained EL4 cells (red) compared to unstained cells (black) for the three cell surface markers expressed at low levels in EL4 cells and therefore not included in subsequent analyses.

Supplementary Figure 2



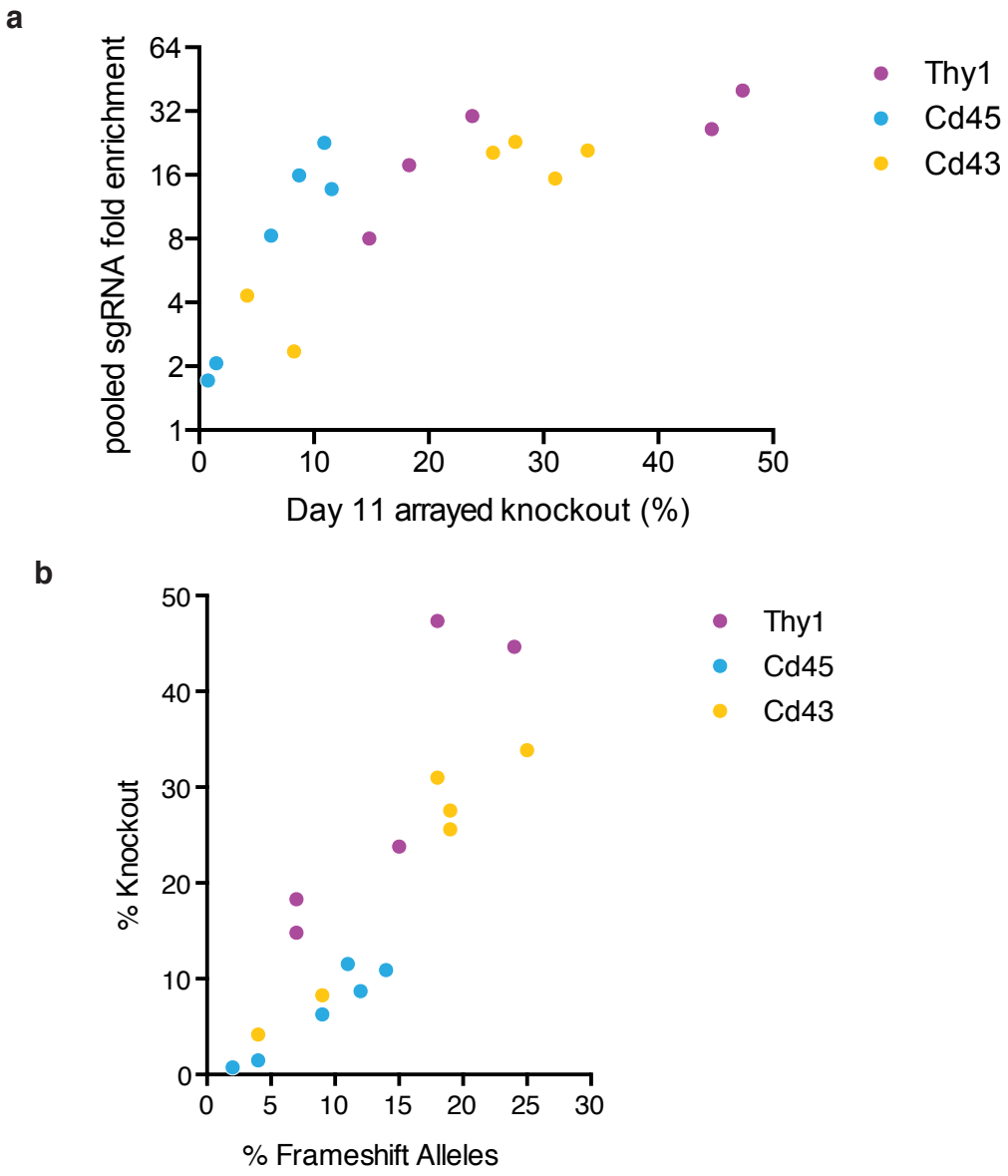
Supplementary Figure 2 pXPR-011, Cas9 activity test vector. (a) Vector schematic of pXPR-011. (b) FACS plots indicating the percentage of GFP negative cells in parental cells transduced with pXPR011 compared to Cas9 and pXPR-011 expressing cells.

Supplementary Figure 3



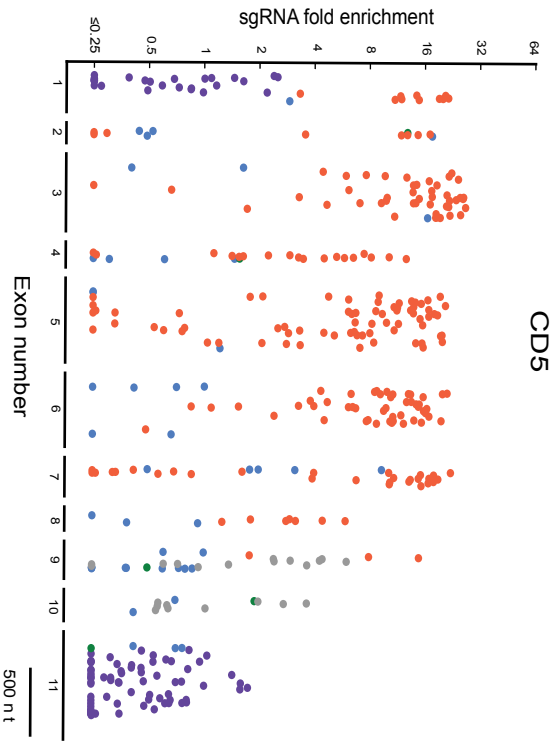
Supplementary Figure 3 sgRNA concordance across cell lines. Spearman rank correlation of 0.90 (left panel) and 0.76 (right panel).

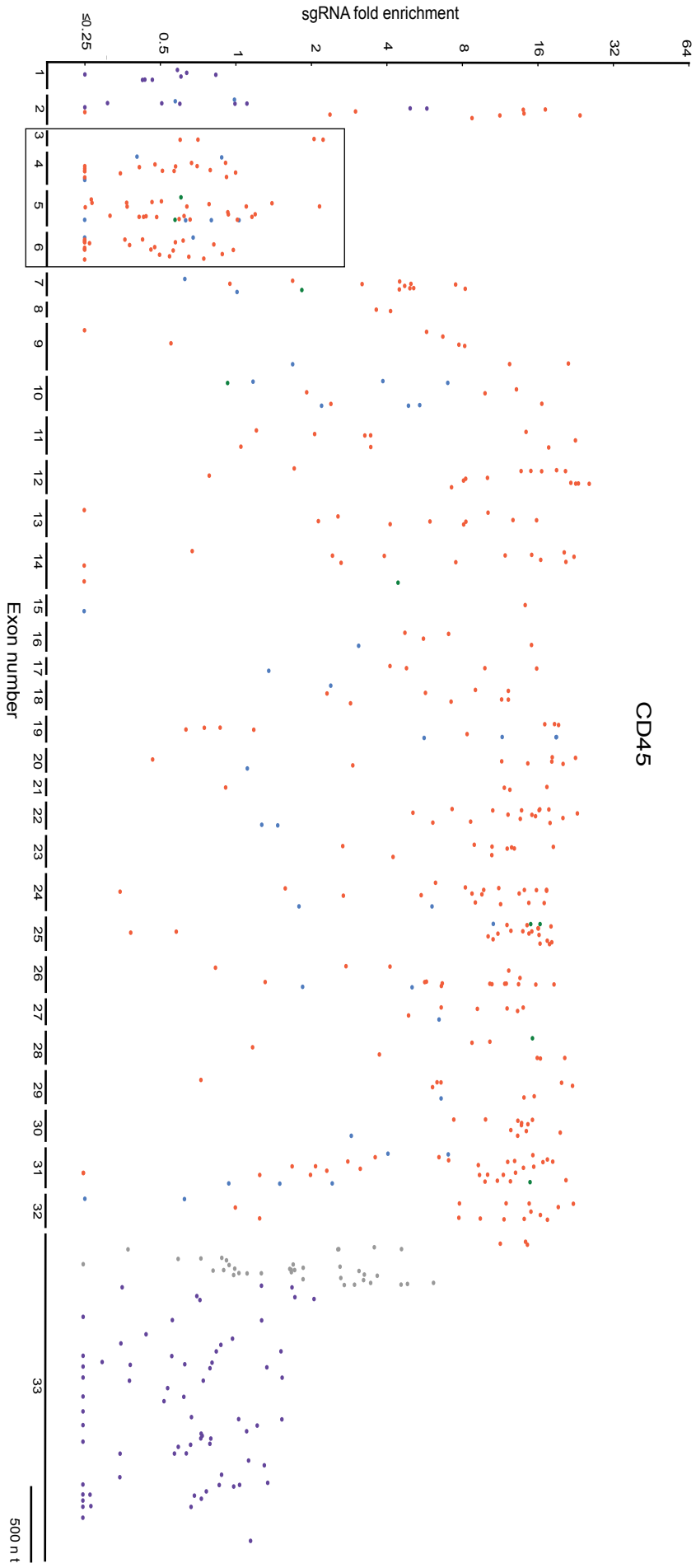
Supplementary Figure 4



Supplementary Figure 4 Activity of 17 sgRNAs in arrayed format. (a) Scatter plot of marker-negative fold enrichment values from the mouse pooled screen plotted against percent gene knockout as measured by flow cytometry in arrayed format. (b) Scatter plot of percent gene knockout as measured by flow cytometry plotted against the percentage of total sequenced alleles containing a frameshift, as measured by next generation sequencing of gDNA following introduction of the sgRNA.

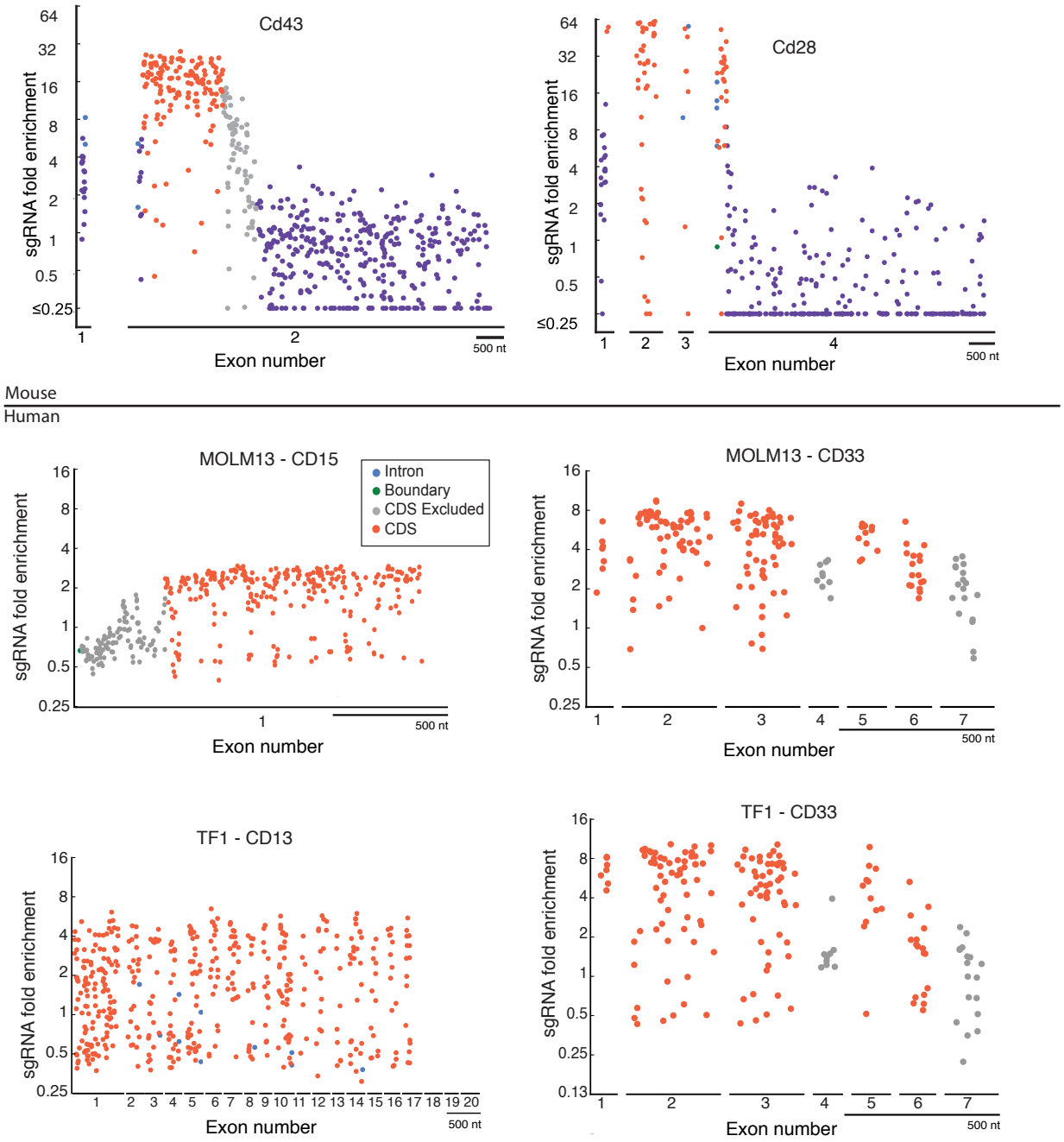
Supplementary Figure 5



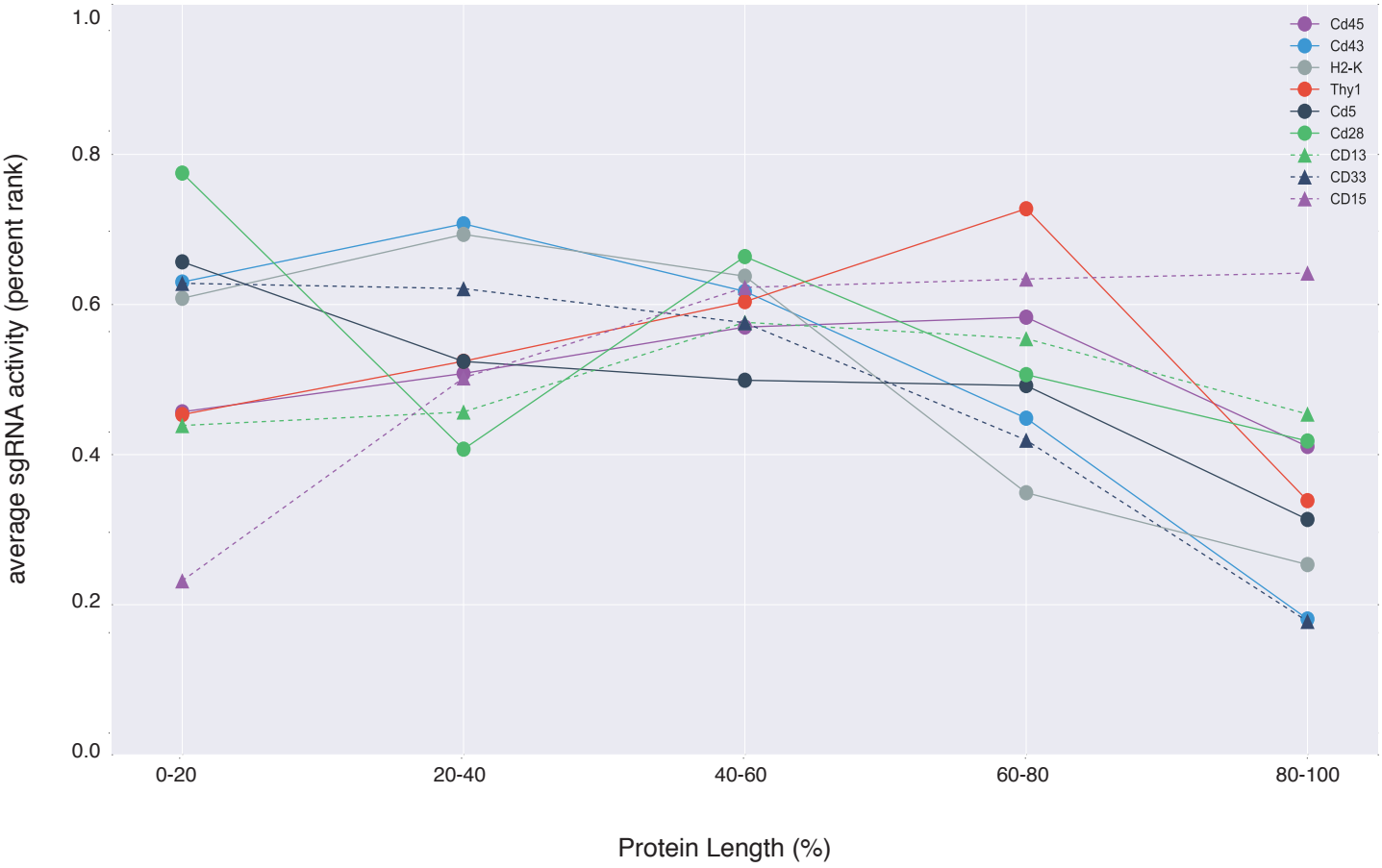


Supplementary Figure 5 Activity maps of sgRNA by cut site position. (a) Cd45 and Cd5 in EL4 cells. All sgRNAs with fold enrichment ≤ 0.25 are grouped at the bottom of the y-axis. The box around the sgRNAs in Cd45 exons 3-6 indicate exons with 0 sgRNAs enriched >10 fold; a splice isoform of this gene, Ref Seq NM_001268286.1, is known to exclude these exons. (b) CD15, CD33 in MOLM13 cells and CD13, CD33 in TF1 cells. For all plots, exons and 100 nts (Cd5) or 38 nts (Cd45) of flanking intronic sequence are shown as lines on the x-axis with gaps marking the remainder of the introns. sgRNAs targeting close to the C' terminus or other regions of low activity that were excluded from activity modeling are indicated in gray. Boundary sgRNAs (green) are those where the cut site, between nts 17 and 18, falls between annotated regions (e.g. CDS/intron).

Supplementary Figure 5, continued

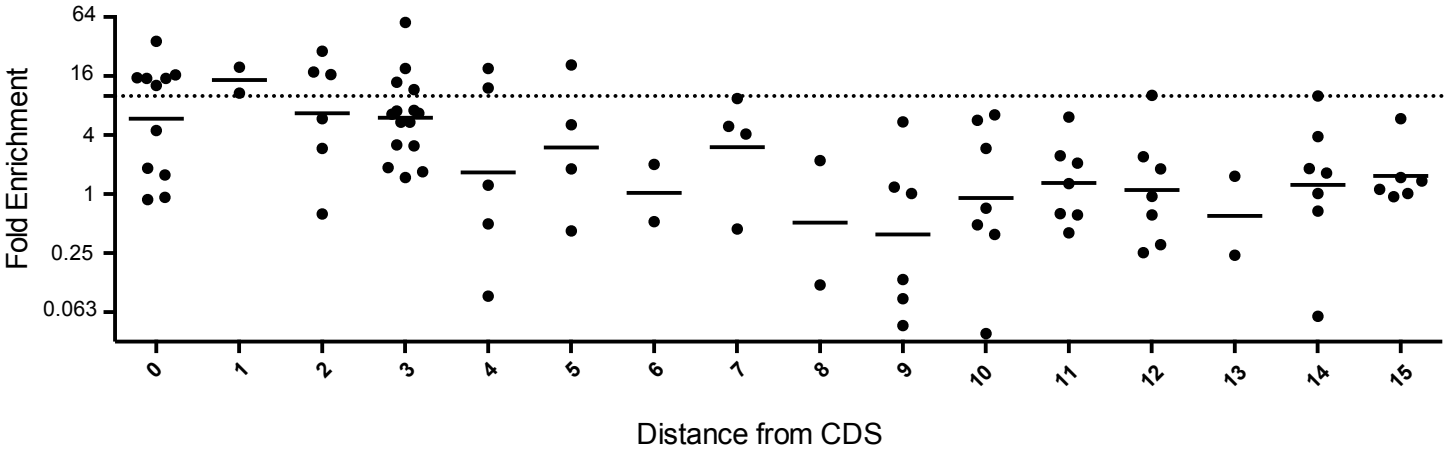


Supplementary Figure 6



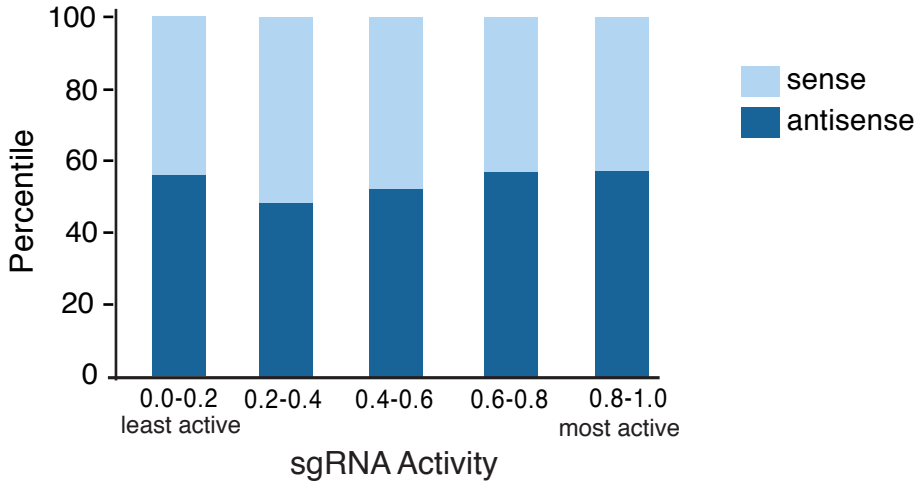
Supplementary Figure 6 Activity as a function of protein length. The average sgRNA percent rank is displayed for each protein quintile. For Cd45, the alternative transcript NM_001268286.1 was used to annotate the coding sequence and thus exons 3 – 6 were excluded (see Supplementary Figure 5).

Supplementary Figure 7



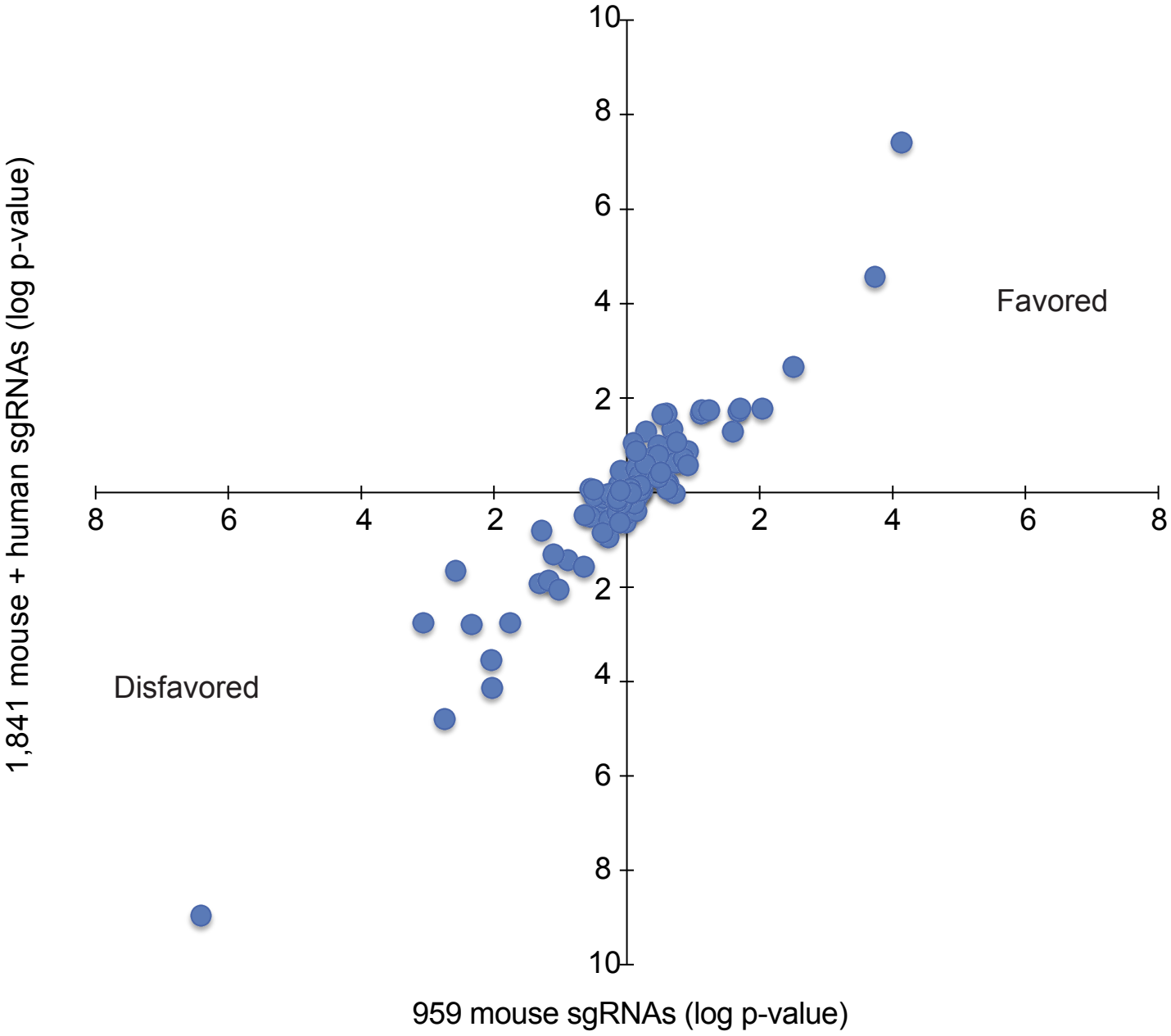
Supplementary Figure 7 sgRNA activity as a function of cut site distance from exon/intron boundary in the mouse pool. Each dot indicates an sgRNA with a cut position at the indicated distance from the CDS. The solid horizontal line at each position indicates the mean enrichment score for that distance from the CDS. The dotted horizontal line indicates 10-fold enrichment.

Supplementary Figure 8



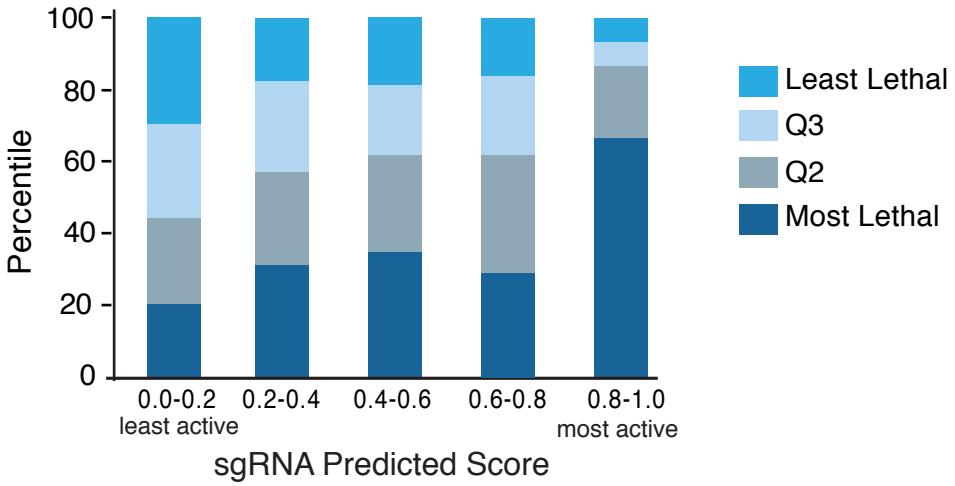
Supplementary Figure 8 Activity as a function of DNA target strand. sgRNA activity bins by quintile. After correcting for multiple comparisons, no bin achieves statistical significance (p-values, Chi-squared test: Q1 = 0.54; Q2 = 0.03; Q3 = 0.47; Q4 = 0.26; Q5 = 0.21).

Supplementary Figure 9



Supplementary Figure 9 Stability of observed base preferences. Comparison of the p-values for all single nucleotide features using only the data from the 6 mouse genes (959 sgRNAs) to the p-values obtained for all 9 human and mouse genes (1,841 sgRNAs, Fig. 3a), showing that nearly doubling the data size and adding a different species does not appreciably alter the observed preferences (r = 0.95, Pearson's correlation coefficient).

Supplementary Figure 10



Supplementary Figure 10 Predicted activity scores of sgRNAs targeting essential genes in A375 cells. Sequence feature weights (Supplementary Table 9) were used to compute predicted activity scores for a set of sgRNAs targeting essential gene classes in a previously-published genome-wide CRISPR screen9. Predicted-activity scores are binned by quintile where a score of 0.8 - 1.0 are the sgRNAs predicted to be most active. sgRNA effect on cell viability is binned by quartile where Q1 indicates most-lethal sgRNAs.