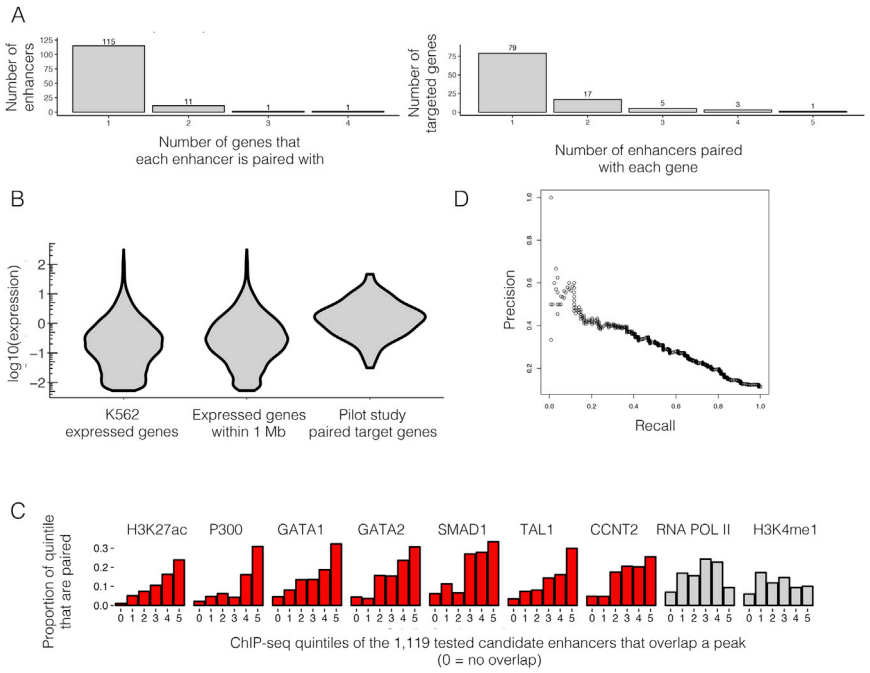

Figure S1. Details of 145 Enhancer-Gene Pairs Originally Identified in the Pilot Screen, Related to Figure 2

(A) Histogram per enhancer of the number of genes paired with that enhancer (3.5% empirical FDR in pilot screen, *left*). Histogram per gene of the number of enhancers paired with that gene (3.5% empirical FDR in pilot screen, *right*). (B) Expression of target genes paired with candidate enhancers in the pilot screen. expression = mean transcript UMIs/cell in the entire 47,650 cell pilot dataset for: K562 expressed genes; those that fell within 1 Mb of a targeted candidate enhancer in the pilot experiment; and for the 105 genes targeted in the pilot experiment's pairs. In the pilot screen, tested candidate enhancers were required to fall within TADs that contained genes highly expressed in K562s. As these were then only tested for pairing with genes within 1 Mb, the pilot screen's target genes are potentially biased toward being highly expressed. This enrichment for highly expressed genes is not seen in the at-scale experiment, where tested candidate enhancers were not required to be in the same TAD a highly expressed gene (Figure 6C). (C) Relative to the 1,119 candidate enhancers tested, the 128 paired candidate enhancers from the pilot experiment (3.5% empirical FDR) tend to fall in enhancer-associated ChIP-seq peaks that show stronger signals. All ChIP-seq peaks that overlap the 1,119 candidate enhancers were divided into quintiles of strength, defined as the average enrichment in ChIP-seq peak region (0 = no such peak overlaps the candidate enhancer, 1 = lowest, 5 = highest). Histograms of the proportion of each 1,119-quintile that were called as enhancer-gene pairs are shown. Red coloring = *P*-value < 0.005 for independent logistic regression for predicting a candidate enhancer as paired based on this peak type. (D) Precision-recall curve for a multivariate logistic regression classifier based on ENCODE enhancer-associated biochemical features that differentiates the 128 paired candidate enhancers from the remaining of the 1,119 candidate enhancers. The median AUPR from five-fold cross-validation was 0.31.

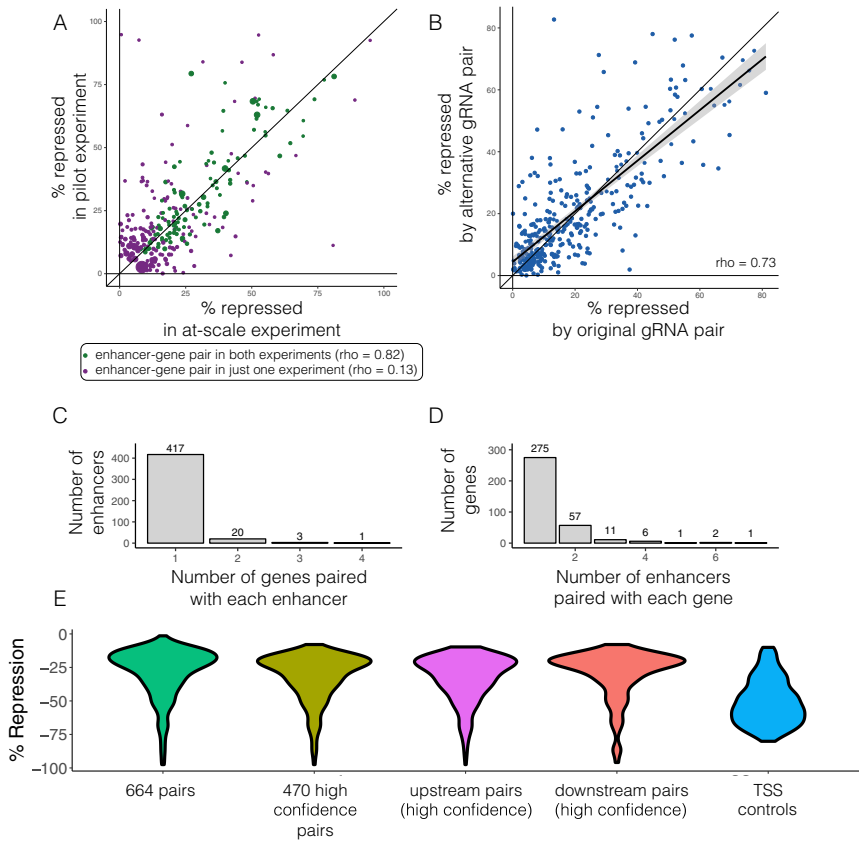

Figure S2. Replication of Effect across Experiments and Alternative gRNA Pairs, Related to Figure 3

(A) The percent target gene repression of an enhancer-gene pair in the pilot versus the scaled experiments (green: called as a pair in both experiments; purple: called as pair in only one experiment).

(B) The effect sizes on the most highly repressed gene for each pair of gRNA pairs targeting the same candidate enhancer (see STAR Methods).

(C) Histogram per enhancer of the number of genes paired with that enhancer (high confidence pairs of the at-scale screen).

(D) Histogram per gene of the number of paired with that gene (high confidence pairs of the at-scale screen).

(E) Effect sizes from enhancer-gene pairs identified in the at-scale screen. % repression of target transcript for the 664 enhancer-gene pairs that pass a < 0.1 empirical FDR, the 470 high confidence enhancer-gene pairs, the high confidence pairs in which the enhancer is upstream of the target gene, high confidence pairs in which the enhancer is downstream, and the 97% of 381 TSS controls that are detected as repressing their target genes in the at-scale screen.

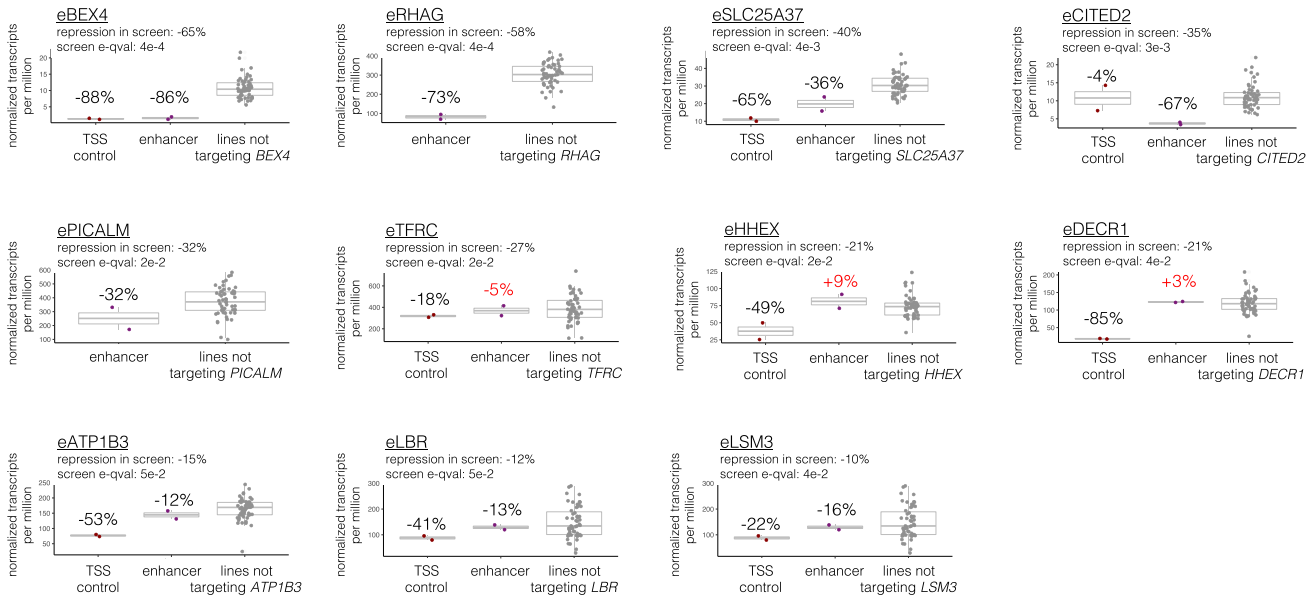

Figure S3. Eleven Further Singleton CRISPRi Experiments (in Addition to the Four Featured in Figures 4A–4D), Related to Figure 4

For each singleton replication experiments of enhancer-gene pairs, bulk RNA-seq was performed on CRISPRi-positive K562 cells transduced with CRISPRi-optimized CROP-seq gRNAs targeting the labeled paired-enhancer (purple, denoted with an "e" prefix) or the TSSs (dark red) of their respective target genes. Target gene transcript expression in the singleton-target cell lines (dark red/purple) as compared to 'non-targeting' lines (gray; singleton experiments in which the other 10 candidate enhancers or TSSs were targeted plus a line transduced with non-targeting gRNAs). Repression in screen = differential expression from at-scale screen. Screen e-qval = Benjamini-Hochberg corrected empirical *P*-value from at-scale screen. Normalized transcripts per million (tpm) from sleuth. %s above boxplots = sample's % repression in bulkRNA-seq calculated from (transcript's mean tpm between the sample's two technical replicates) / (transcript's mean tpm from all the 'non-targeting' lines). % repression labeled light red if in disagreement with the enhancer-gene pair in the at-scale screen (Table S3A).

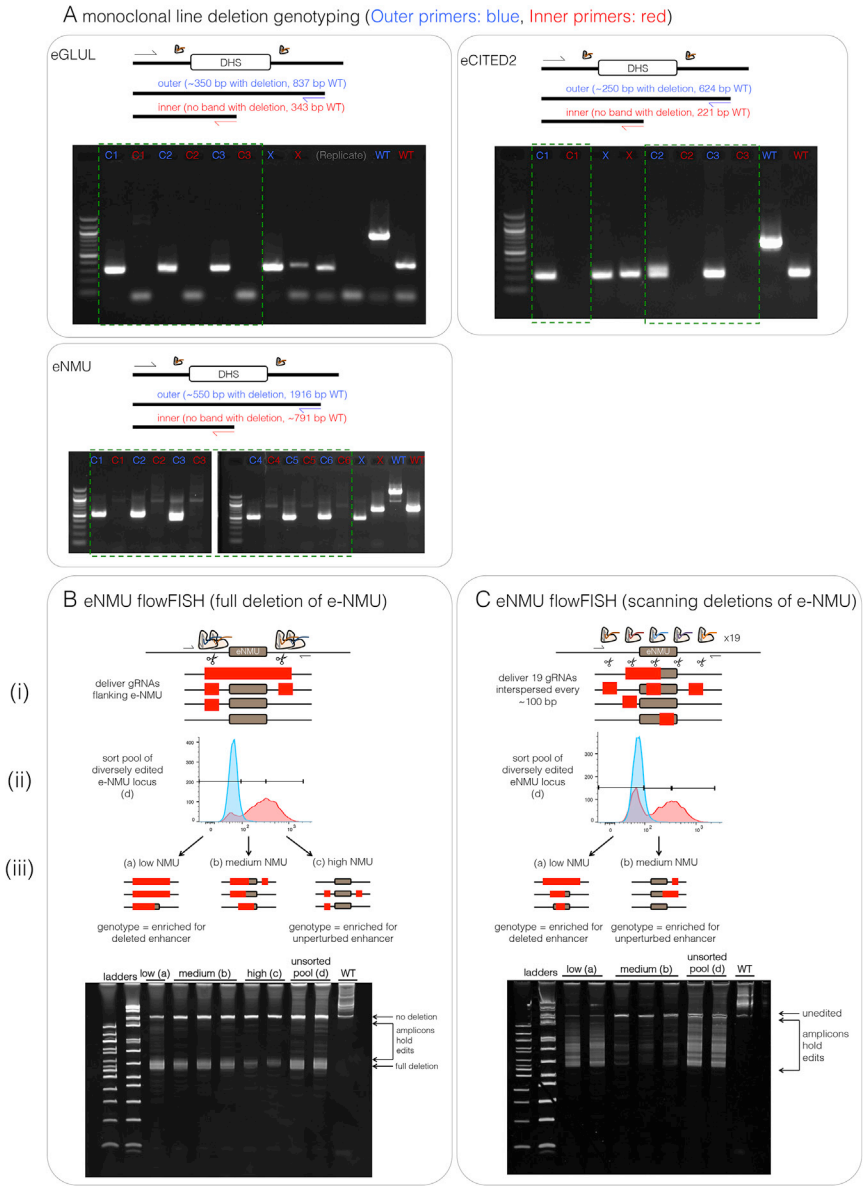

Figure S4. Details of Sequence Deletion Validation, Related to Figures 4E–4H

(A) Genotyping PCR design and gels for the homozygous sequence deletion monoclonal lines, as featured in Figures 4E–4G. Outer primers were designed to amplify the entire candidate enhancer locus; shorter band in these 'outer' lanes (blue label, as compared to 'WT' lane) represents presence of a full deletion. Inner primers were designed to amplify only if a wild-type allele remained ("red" labeled lanes); presence of a band indicates a remaining wild-type locus. Primers design is schematized at the top. Clones with a deletion band (in the "outer" PCR lane) *and* no wild-type band (in the "inner" PCR lane) were submitted to bulkRNA-seq. Green dashed outline represents the clones used in Figure 4. Nomenclature of "C1" and "C2" etc correspond to "clone 1" and "clone 2" *et cetera* as labeled in Figure 4. 'WT' lanes = same parental K562 cell line that was transfected with gRNA targeting *HPRT1*. Ladder = NEB 100 bp (N3231L). "X" = cell line did not harbor homozygous deletions.

(B and C) e-NMU sequence-disrupted cells were phenotyped by *NMU* RNA flowFISH, as featured in Figure 4H. First, K562 cells were transfected with nuclease-active Cas9 and gRNAs either flanking (B) or scanning (C) the e-NMU locus to create a heterogeneous population of cells (*i*) in which a portion (based on editing efficiency) harbor full or partial deletion of e-NMU. Then, intracellular *NMU* expression was labeled via flowFISH (*ii*) and cells were sorted into bins of low (a), medium (b), or high (c) *NMU* expression (as to sort genotypes based on the effect upon disruption of e-NMU function, *iii*). Last, gDNA was extracted from the cells in each bin, and the e-NMU locus was amplified (primers diagrammed at the top of the figure). Unsorted pool (d) = unsorted-but-edited cells to demonstrate original distribution of genotypes in the original heterogenous pool. Each lane is a replicate PCR of gDNA (10 ng per reaction) from that same sorted sample. Ladder L = 100 bp (NEB), Ladder R = 1 Kb ext (Invitrogen). WT = untreated parental K562s. Remaining full-length alleles in the 'low' expression bins could correspond to inaccuracy of flowFISH, alleles with very small edits, or (as K562s are pseudotriploid) heterozygous cells that still retained a largely uninterrupted copy of e-NMU on one or two alleles.

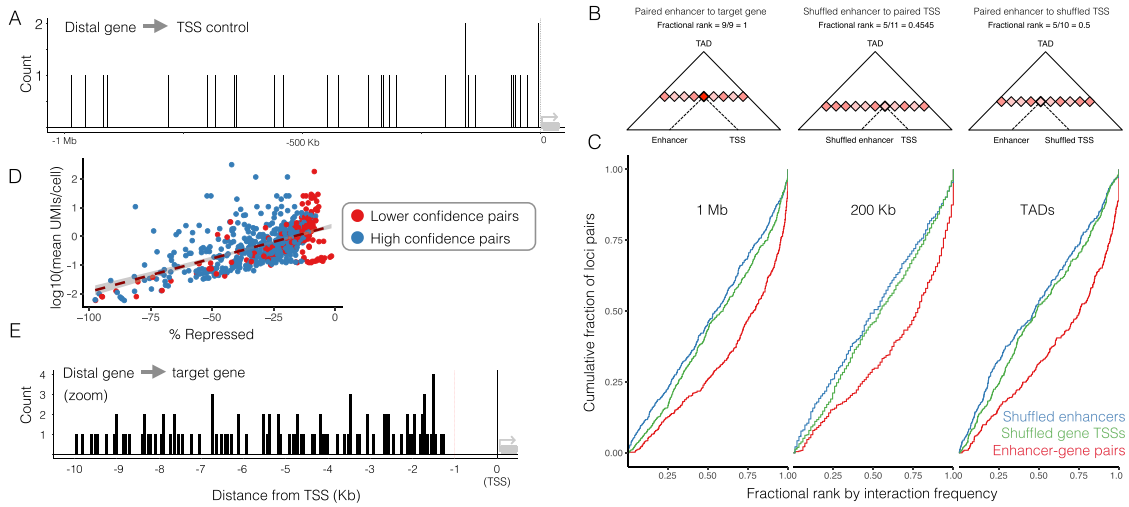

Figure S5. Details on Characteristics of K562 Enhancer-Gene Pairs, Related to Figure 6 and Discussion

(A) Distribution of distances between "positive control" TSSs and any secondarily repressed genes. Of our 359 'positive control' TSSs whose targeting successfully repressed the expected gene in both experiments, 35 reduced expression of 1+ additional genes (45 apparent promoter-promoter relationships in total). 15 of these 45 involved overlapping promoters (TSSs within 1 Kb) and are not shown here as the observed effect of CRISPRi is likely direct. The distances that the remaining 30 secondarily repressed genes fall upstream of the targeted TSS are shown. In contrast with enhancer-gene pairs (Figure 6A), these 30 are largely *not* enriched for proximity to affected genes. Dashed line = target gene TSS.

(B and C) Hi-C interaction frequency analysis, (B) Example schematic of fractional ranking by interaction frequency analysis. The interaction frequency of each loci pair (color of pixel) is ranked within the interaction frequencies of all distance-matched genomic-pairs in the same TAD (the stripe of pixels shown in schematic). For the two null distributions in Figure 6E, each pair's target gene's TSS is given a shuffled enhancer (and then ranked again within this new distance distribution), or the pair's candidate enhancer is given a shuffled TSS (and then ranked again within this new distance distribution). Shuffled TSSs and enhancers are drawn from the same distance distribution as the actual enhancer-gene pairs. (C) The same fractional rank by interaction frequency analysis within the same TAD as shown in Figure 6E, but also comparing ranking to all pairs within 1 Mb or 200 Kb of the chosen enhancer-TSS pair. Red = enhancer-gene pairs, blue = hit-gene to shuffled enhancer pair null distribution, green = hit enhancer to shuffled gene TSS pair null distribution.

(D) Correlation of effect size of enhancer-gene pair versus expression level of target gene. Effect size (% transcript repressed) was correlated with expression level of targeted gene (Spearman's rho for 664 inclusive pairs: 0.56; Spearman's rho for 470 high confidence pairs: 0.53). This is likely consequent to power, as small effects (less than −25%) are not detected on lowly expressed genes (less than 0.12 UMIs/cell). log10 of the mean UMIs/cell is denoted per target gene transcript.

(E) A "zoom-in" of Figure 6A to the 10 Kb upstream of the target gene's TSS (rather than 1 Mb). 101 of 354 upstream, high confidence enhancer-gene pairs fall within 10 Kb of the TSS. Same restrictions to enhancer-gene pairs plotted here as in Figure 6A. Gray line = TSS, red line = 1 Kb upstream of TSS (all protospacers within 1 Kb of a TSS were excluded from any candidate enhancer gRNA library)

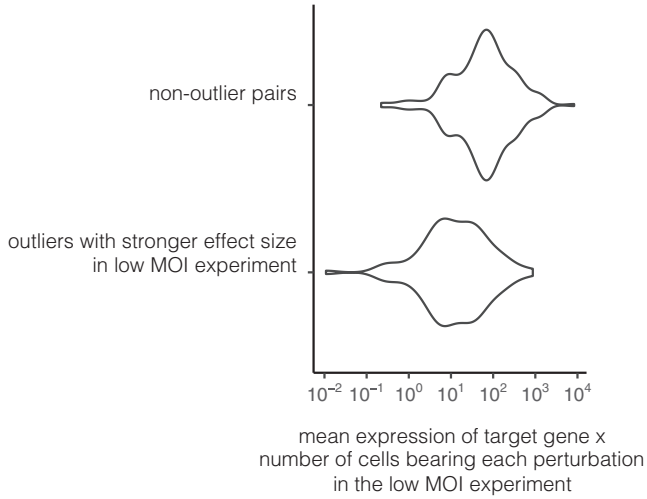

Figure S6. Outliers with Greater Effect Size in Low MOI Replicate Are Likely Due to Low Expression and Low Cell Count in Low MOI Replicate, Related to Figure 7

The mean expression of the target gene in the low MOI 41,284 cell dataset as a function of the number of cells bearing each perturbation in that experiment.

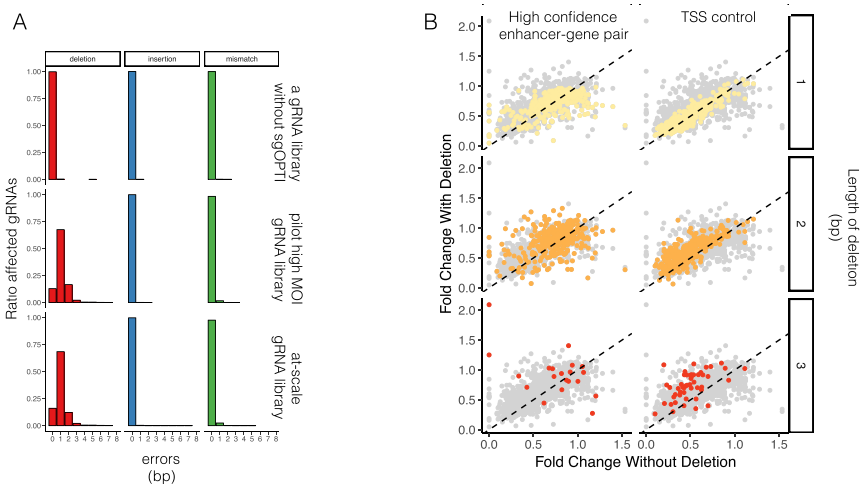

Figure S7. Supplementary Details, Related to STAR Methods

(A and B) Quantification of errors in synthesis of the sgOPTI gRNA backbone across scRNA-seq datasets. (A) Deletion (red), insertion (blue), or mismatch (green) rate in the 8-15 bp downstream of the spacer in the gRNA backbone as captured by gRNA-transcript enrichment from scRNA-sequencing data. Data is shown for scRNA-seq datasets of a gRNA library that does not have sgOPTI added to the backbone (but was cloned, amplified and sequenced in a similar manner), the pilot high MOI gRNA library, and the at-scale gRNA library. (B) The impact of indels on effect sizes for paired-candidate enhancers (high confidence set) and TSS positive controls. The effect size of gRNAs with versus without perfect backbones, stratified by length of deletion. Gray points = a unique dot is plotted for the subgroups of each paired enhancer/TSS gRNA, divided by if they harbor 0, 1, 2, or 3 bp deletions. Colored points = set of gRNAs bearing the specified deletion length. Only points for which there are > = 50 cells in a given deletion-length group are plotted to ensure reasonable estimates of fold change.