

**Supplemental Figure S4.** Sequence quality and barcode representation in the array-synthesized lentiMPRA library. (A) Positioning of errors/differences to the designed oligo sequences as observed from consensus-called BWA alignments. Thick horizontal lines near the top indicate the different oligo parts (insert - blue, spacer - red, barcode - green). Deletions towards read ends are misrepresented in this plot as those that result in altered outer alignment coordinates rather than the identification of a deletion. Filtering for sequence mapability results in artifacts around the barcode sequence which is required to disambiguate oligos. (B) Number of times designed reporter barcodes are observed in the oligo validation sequencing experiment. The frequency axis (y) has been log-transformed to show an over dispersion effect in the library, where a minority of barcodes contribute many of the observations. This effect can be observed from a change in gradient between barcodes observed below 150 times and barcodes observed more than 150 times. The 11,889 barcodes (5.5%) observed with more than 150 observations account for 24.5% of all observations.

