Supplementary Methods and Figures

Datasets

Data	Source	Identifier
Genome assembly	Hg38	Nuclear analysis set
Genome annotation	GENCODE[16]	V20 V29 basic (for Figure 2C only)
lncRNA screen scores	Y. Liu et al.[1]	Table S8
lncRNA targeting sgRNAs	Y. Liu et al.	Table S6
Relative copy number	ENCODE[17]	GSM999287 (K562), GSM999302 (HeLa), GSM999285 (GM12878)
Protein-coding screen scores	Wang et al.[3]	Table S3, K562 CS
HeLa CRISPRi RNAseq	S. J. Liu et al.[5], Sequence read archive	SRR3987839 and SRR3987840 (NT sgRNA) SRR3987750 (*CCAT1* sg1), SRR3987751 (*CCAT1* sg2)
HPV-18 genome	NCBI Nucleotide	NC_001357.1

Code availability

All code used to perform the analysis and produce figures is included as a Jupyter Notebook in the Supplementary File.

Software/library	Version	Reference
Python	2.7	www.python.org
NumPy	1.16.1	[18]
SciPy	1.1.0	[19]
Pandas	0.24.1	[20]
Matplotlib	2.2.3	[21]
Plastid	0.4.6	[22]
BedTools	2.17.0	[23]
TopHat	2.1.0	[24]
Bowtie	1.0.0	[25]
Integrated Genome Viewer (IGV)	2.5.0	[26]
IGV tools	2.3.25	[26]
UCSC Genome Browser	v379	genome.ucsc.edu

Copy number amplification analysis

Gene level lncRNA screen scores were obtained from Table S8. Gene symbols were referenced against GENCODE v20 (as used by Y. Liu et al.) to obtain genome coordinates. BedLogR files containing relative copy number data (logR) as determined by Illumina BeadChip and circular binary segmentation were obtained for each cell line and loaded using Plastid. To produce Figures 1A, 1B, and 2A, genes were plotted by annotated gene start. For Figures 1C-D, the mean copy number data across the window

from gene start to end was calculated by finding all overlapping BedLogR features with Plastid and averaging logR weighted by feature length. Relative copy number data was plotted between -1.0 and +1.0 for clarity.

Protein-coding essential gene scores were obtained from Wang et al. 2015 and genome coordinates were obtained from GENCODE v20. Median K562 CRISPR scores were calculated in 1 Mb windows at 50kb steps and plotted in Figures 1A-B from the midpoint of the window. For Figure 1C, the corresponding median CRISPR score for each lncRNA and ribosomal subunit gene was obtained from the window with midpoint closest to the annotated gene start.

Overlapping protein-coding gene analysis

To determine the number of lncRNAs overlapping any protein-coding gene window, GTF files were generated of all GENCODE v20 gene entries with either "gene_type protein_coding" or with gene symbol in the set of genes targeted in the lncRNA screen. BedTools was used to intersect the two GTF files to yield all lncRNA screen genes with no overlap (-v flag). The fraction of lncRNA genes in the library not within the output gene set was then calculated to obtain the percentage of non-intergenic genes.

To determine the set of lncRNA-targeting sgRNAs overlapping a protein-coding gene exon, sgRNA target sites were first obtained by aligning all library sgRNA sequences in Table S6, appended with "NGG", to hg38 using bowtie. The -v 1 flag was used to require perfect matches besides the N of NGG and the -m 1 flag was used to filter out multi-mapping sgRNAs. The alignment output was converted to GTF format. A GTF file of all GENCODE v20 exons with "gene_type protein_coding" was generated. BedTools was used to intersect the two GTF files to yield all lncRNA-targeting sgRNAs with no overlap of protein-coding genes. Genes for which at least one sgRNA did not appear in this set were considered overlapping coding gene exons for calculation of the fraction of non-intergenic genes and Figure 1D. For Figure 1D and calculations in the text, fractions of non-intergenic hit genes were calculated after excluding genes in copy-number amplified regions.

University of California, Santa Cruz (UCSC) genome browser views in Supplementary Figures 1-2 were generated with GENCODE v20, the GTF of library sgRNAs generated above, and a GTF of validation sgRNAs (Y. Liu et al. Table S9).

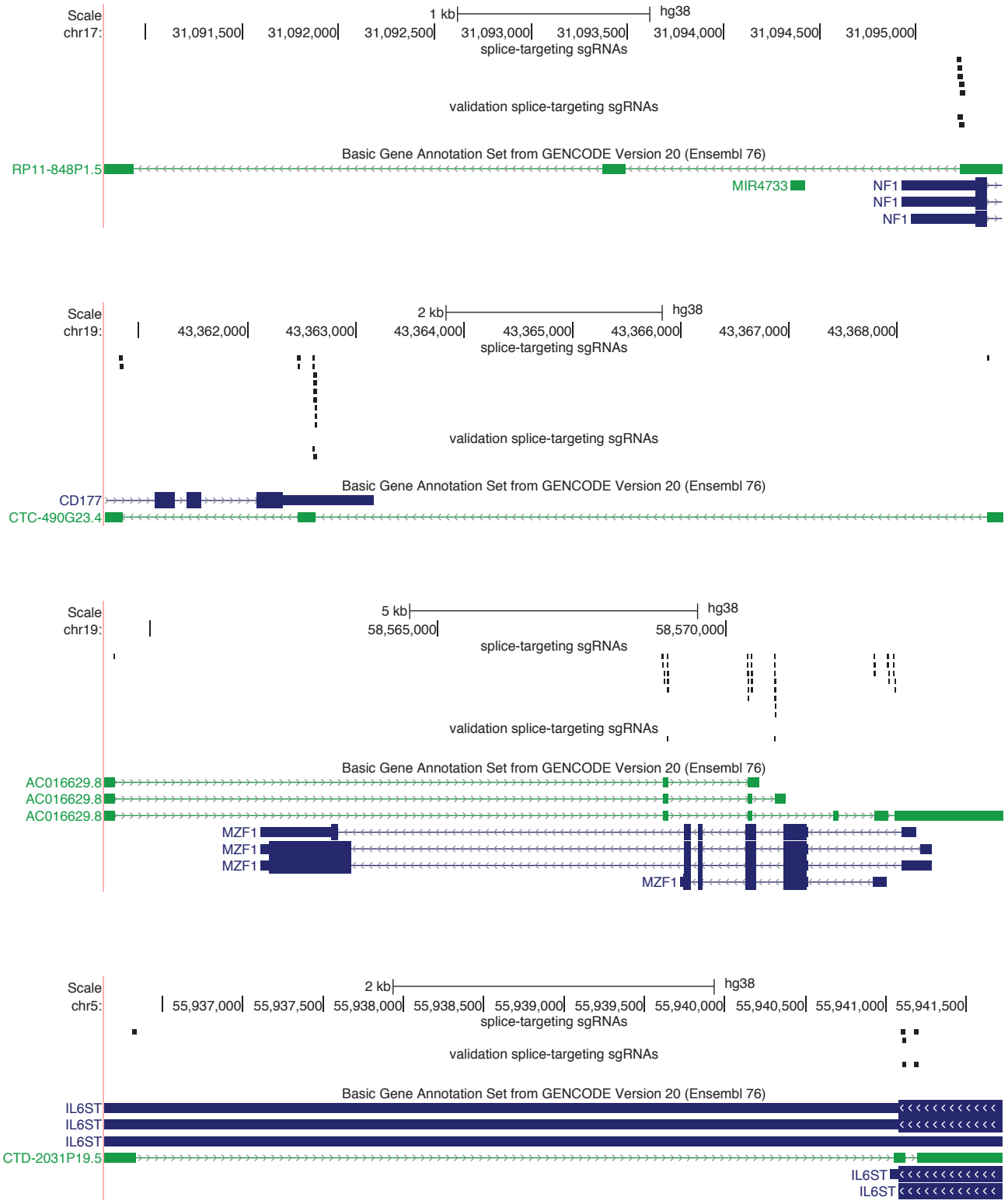
HPV18 fusion analysis

The RNAseq FASTQ files from the SRA entries (accession numbers in the table above) were accessed via SRA Tools Fasterq-dump. A combined genome index of hg38 and the HPV18 genome as an additional chromosome was generated with bowtie-build. The FASTQ files were aligned to this genome index using Bowtie v1 through TopHat-Fusion, run as TopHat with flags: --fusion-search, --max-intron-length 1000000, --fusion-min-dist 1000000, and --fusion-anchor-length 13. Other flags used were: --library-type fr-firststrand --fusion-search --keep-fasta-order --bowtie1 --no-coverage-search. To generate genome browser tracks in Figures 2B-C, coverage tracks were computed with IGVtools count and loaded into IGV along with the merged genome FASTA, GENCODE v20 and v29.basic GTF files, and TopHat accepted_hits.bam files. For Figure 2D, the fusions.out files from each TopHat-fusion result were loaded, the count of fusion-spanning reads were normalized to the total mapped reads of the corresponding RNAseq run, and averaged between non-targeting sgRNA replicates and sgRNAs targeting *CCAT1*.

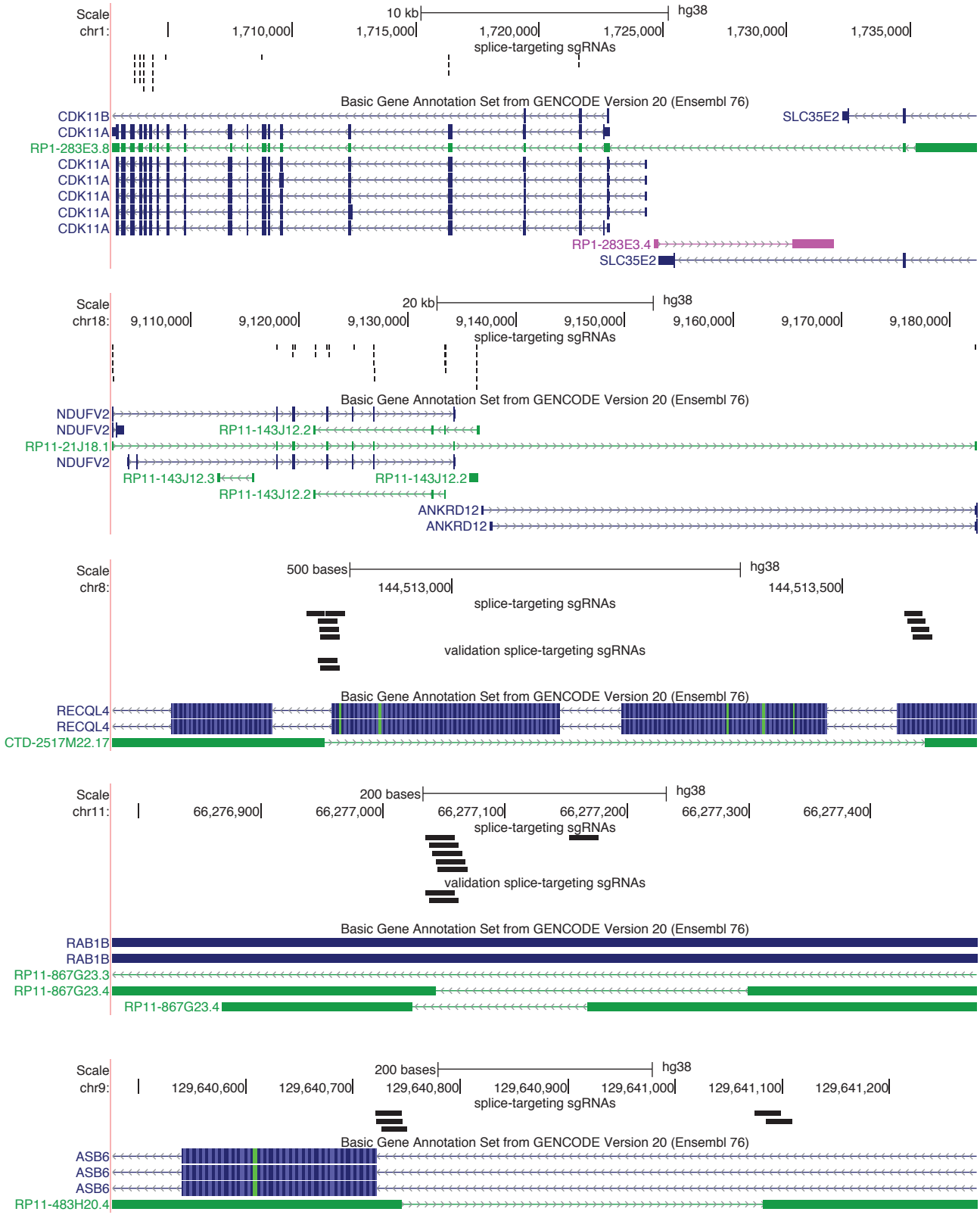
Supplementary References

16. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766–D773 (2019).
17. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
18. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **13**, 22–30 (2011).
19. Virtanen, P. *et al.* SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python. *arXiv:1907.10121 [physics]* (2019).
20. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. Walt, S. van der & Millman, J.) 51–56 (2010).
21. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
22. Dunn, J. G. & Weissman, J. S. Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* **17**, (2016).
23. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
24. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
25. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
26. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24–26 (2011).

Supplementary Figure 1. UCSC genome browser tracks of selected lncRNA loci. All genes displayed were considered non-intergenic (library sgRNAs overlap exons of protein-coding genes) and were validated with the indicated sgRNAs in Liu *et al.*[1] Supplementary Figures 8-9.



Supplementary Figure 2. UCSC genome browser tracks of lncRNA loci that were hits in all three cell lines screened. All genes displayed were considered non-intergenic (library sgRNAs overlap exons of protein-coding genes). Two hits were validated with the indicated sgRNAs in Liu *et al.*[1] Supplementary Figure 8.



Supplementary File. Jupyter notebook containing all analysis scripts. Notebook is provided in HTML and Jupyter notebook (.ipynb) formats.