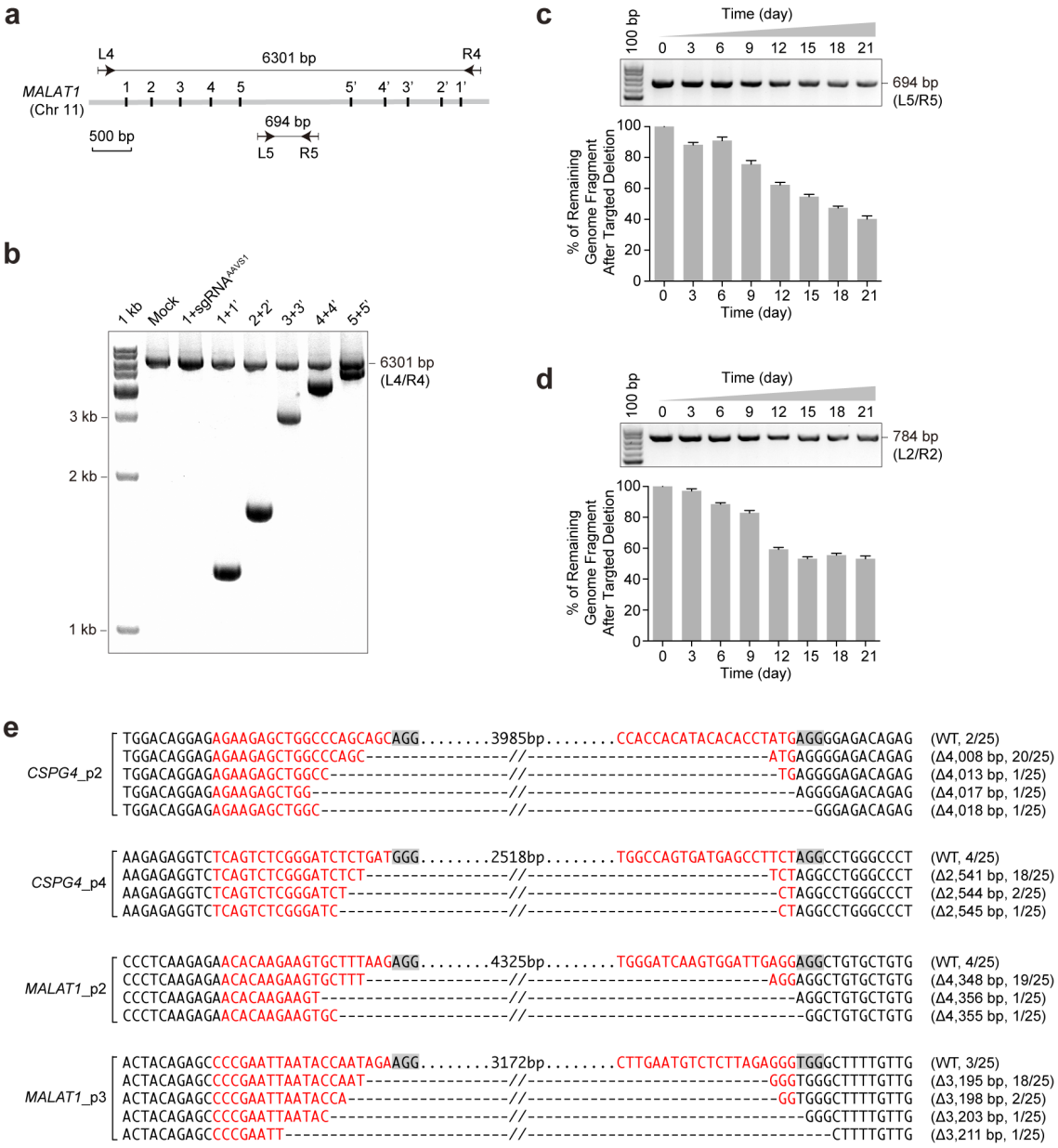
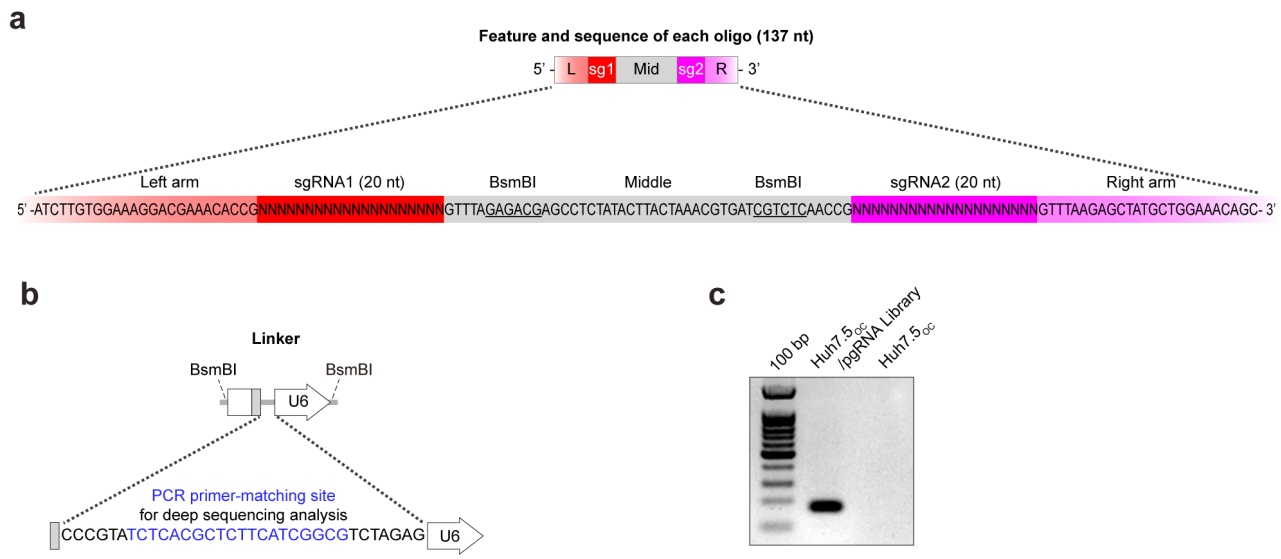


Supplementary Figure 1

Large-fragment deletions induced by lentivirally-delivered pgRNAs.

(**a**,**b**) Five pairs of gRNAs targeting the *MALAT1* gene with respective U6 promoters ($U6_2$) inducing a large-fragment deletion ranging around 1–5 kb were chosen and primers L4/R4 were used for the genomic PCR reactions. All infected Huh7.5_{OC} cells were enriched by

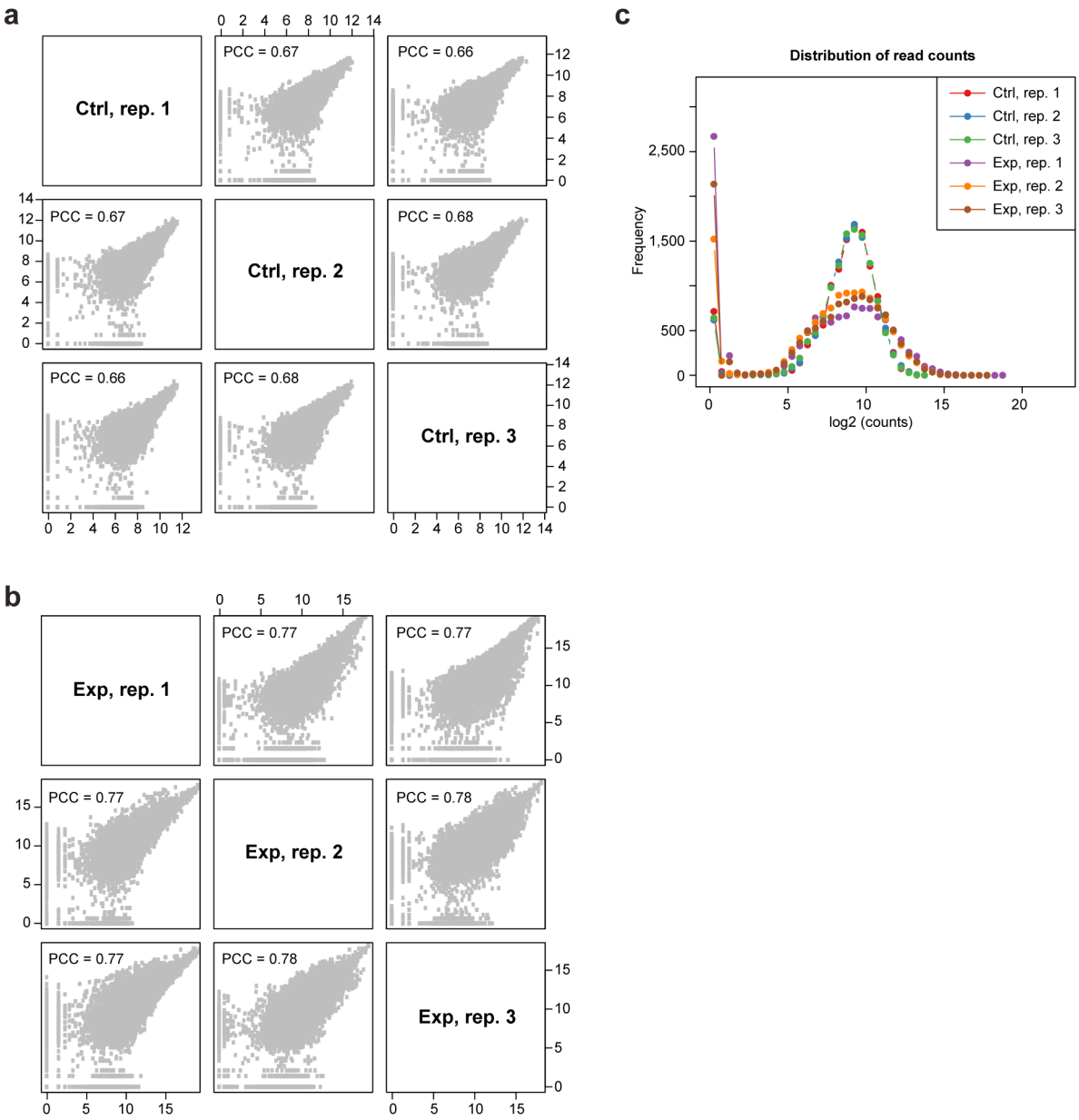
FACS and incubated for 6 days. The control is a pair of gRNAs with one targeting the *MALAT1* locus and the other targeting AAVS1 region. (**c**) Quantification for the efficiency of large-fragment deletions on *MALAT1* locus using genomic PCR over the course of time post transduction. The pgRNAs (2+2' in **Supplementary Fig. 1b**, generating 4.3 kb deletion by design) were delivered into Huh7.5$_{OC}$ cells through lentiviral infection, and genomic DNA was extracted from different time points as indicated. The primers L5/R5 corresponding to sequences flanking pgRNA targeting sites were used for quantification. (**d**) Quantification for the efficiency of large-fragment deletions on *CSPG4* locus using genomic PCR over the course of time post transduction. The pgRNAs (2+2' in **Fig. 1c**, generating 4.0 kb deletion by design) were delivered into Huh7.5$_{OC}$ cells through lentiviral infection, and genomic DNA was extracted from different time points as indicated. The primers L2/R2 corresponding to sequences flanking pgRNA targeting sites were used for quantification, and primers L3/R3 corresponding to sequences farther away from the targeting sites were used for normalization. All primer sequences are listed in **Supplementary Table 11**. Images were analyzed by ImageJ software and data are presented as the mean ± s.d. (n = 3). (**e**) DNA sequencing analysis of large-fragment deletion induced by two *CSPG4*-targeting pgRNAs (p2 and p4 stand for 2+2' and 4+4', respectively) in **Fig. 1d** and two *MALAT1*-targeting pgRNAs (p2 and p3 stand for 2+2' and 3+3', respectively) in **Supplementary Fig. 1c** from pooled cells 3 weeks post infection. Partial sequences of targeted gene in the genome containing the two gRNAs' targeting regions are labelled in red and the shaded nucleotides represent the PAM sequences. The dashes indicate deletions.



Supplementary Figure 2

DNA sequences of designed oligo and linker between the two gRNAs of each pair, and PCR amplification of the barcode-gRNA region from the pgRNA library for deep-sequencing analysis, as indicated by electrophoresis.

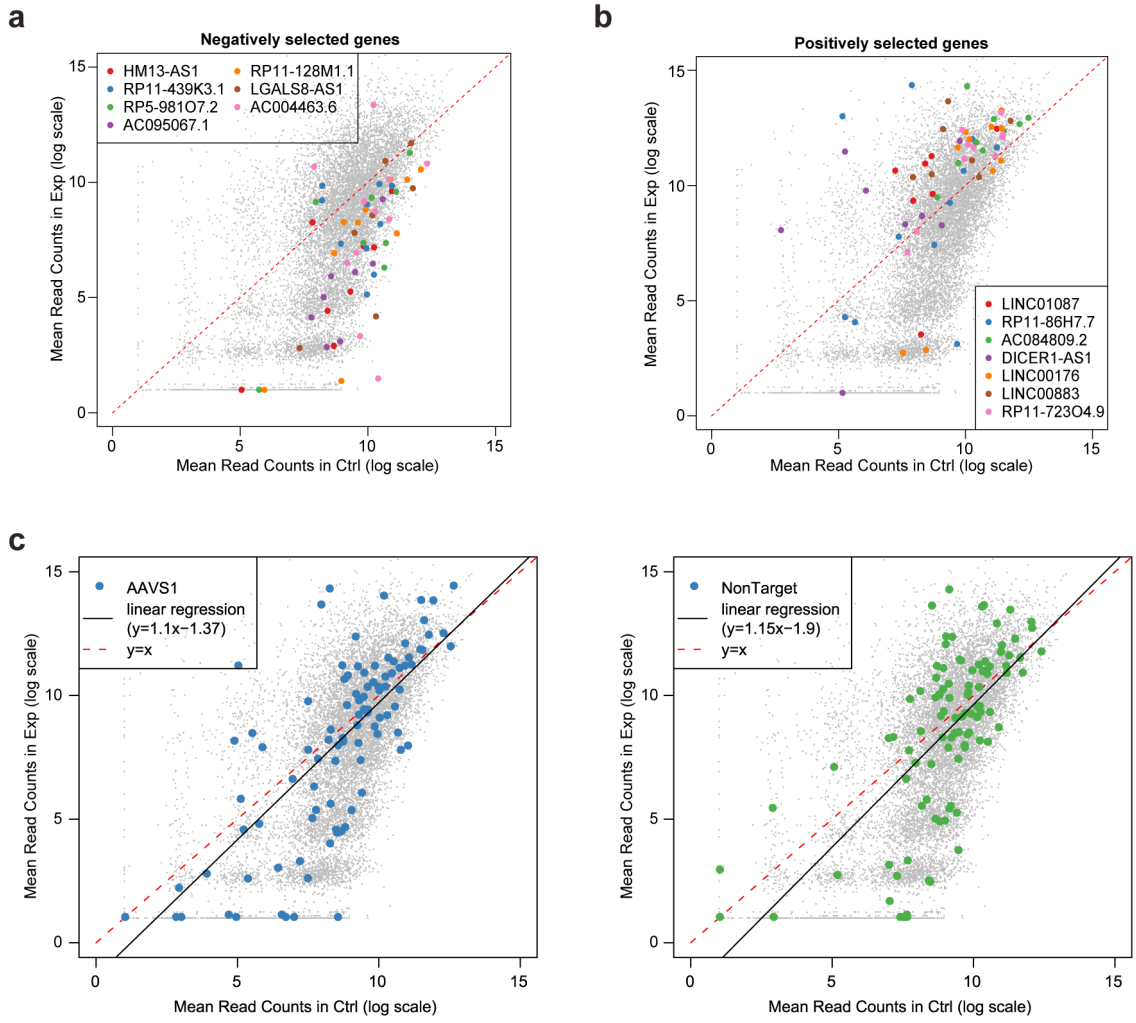
(**a**) Feature and sequences of each oligo. The left arm and right arm are used for primer targeting for amplification. (**b**) Feature of the linker for the construction of pgRNA plasmid and its unique sequence between the end of first gRNA scaffold and the beginning of the second U6 promoter. (**c**) The genomic DNAs isolated from Huh7.5$_{OC}$ library and wild-type cells were used as templates for PCR amplification.



Supplementary Figure 3

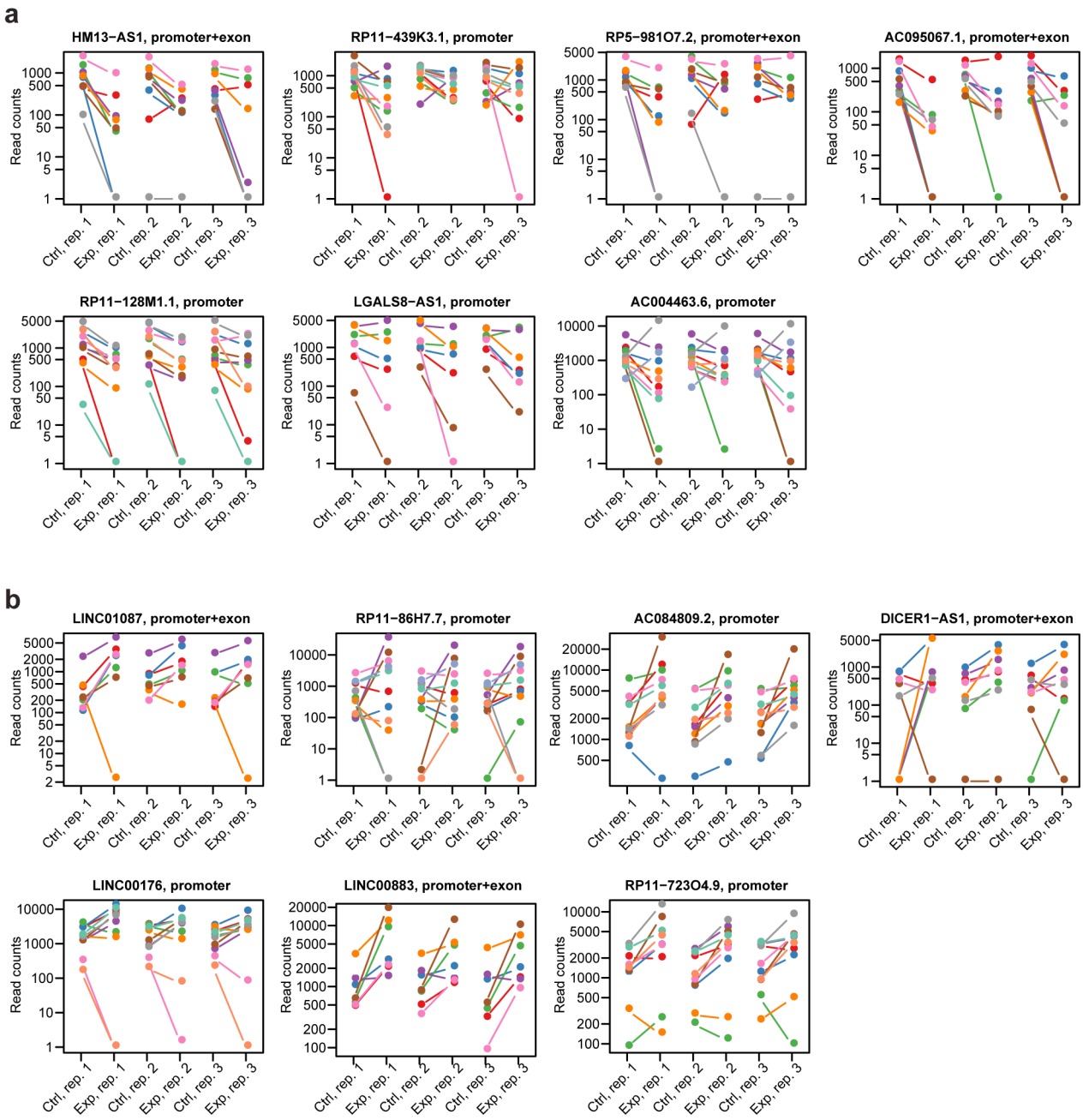
The correlations and read distribution of independent experimental replicates.

(**a**) The correlations between replicates in control sample. (**b**) The correlations between replicates in 30-day enrichment samples. The Pearson Correlation Coefficient (PCC) between replicates is also reported. (**c**) The distribution of normalized read counts in 3 replicates of control (Ctrl) and experimental (Exp) enriched (30-day culturing) Huh7.5$_{OC}$ cells.



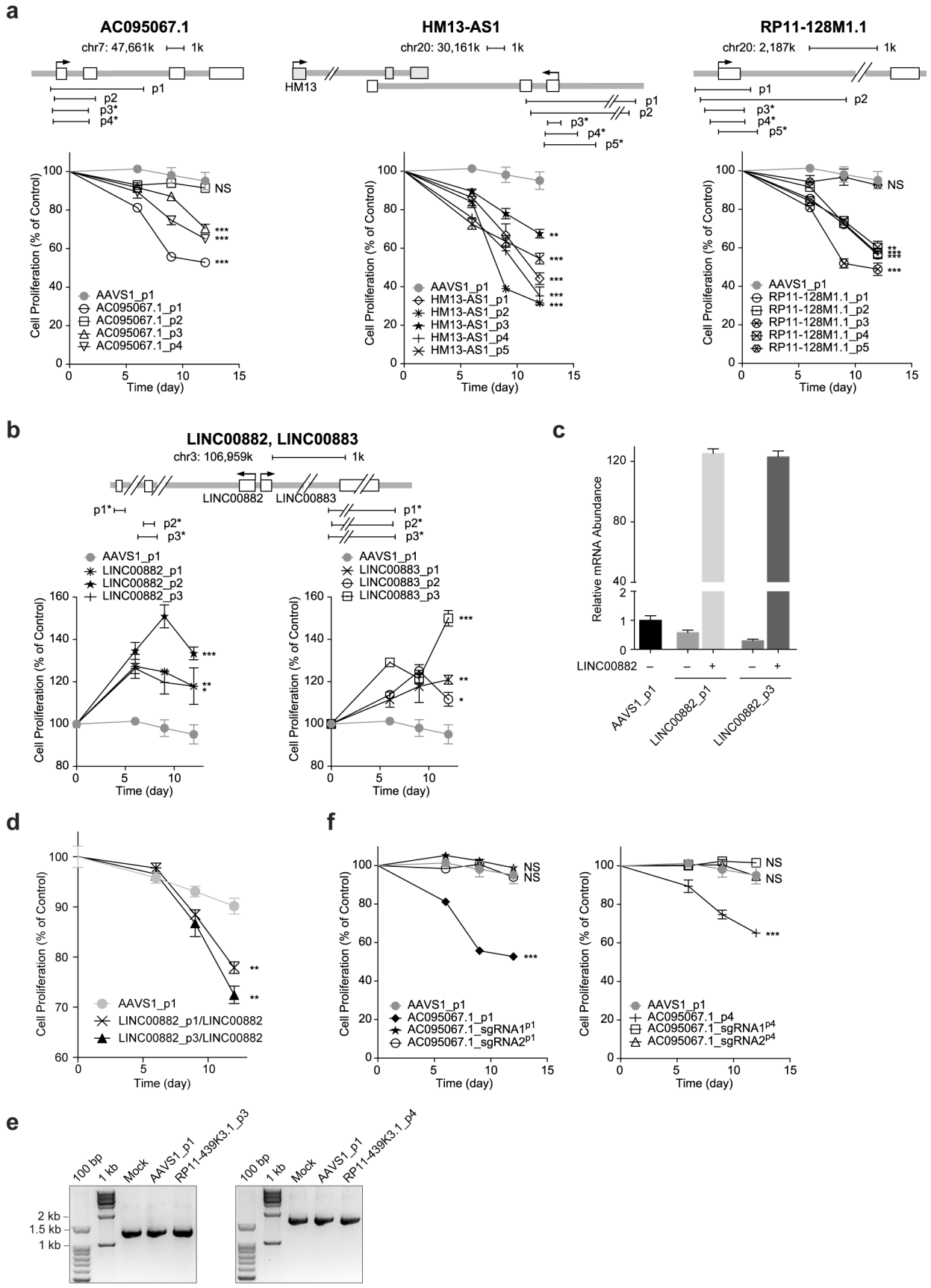
Supplementary Figure 4

The mean read counts of pgRNAs targeting top negatively selected lncRNAs (**a**) and positively selected lncRNAs (**b**). (**c**) The mean read counts of pgRNAs targeting the AAVS1 loci (left) and non-targeting controls (right).



Supplementary Figure 5

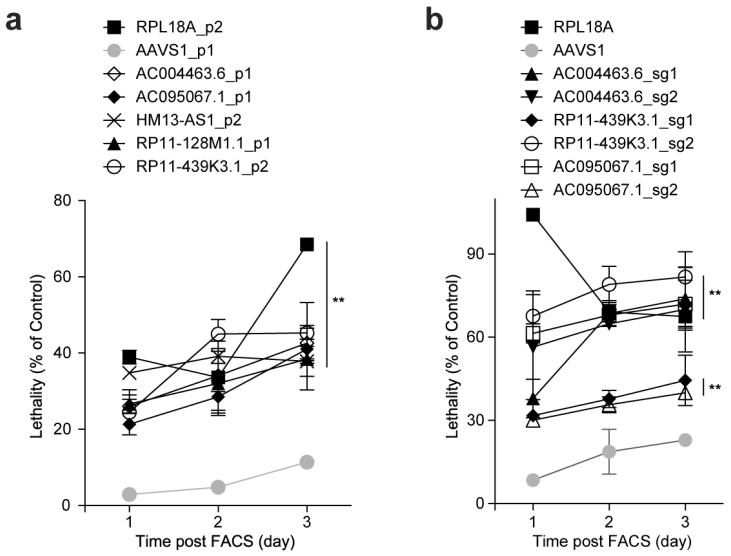
The pgRNA read counts of negatively (**a**) and positively (**b**) selected lncRNAs chosen for validation.


a
AC095067.1
chr7: 47,661k 1k
p1
p2
p3*
p4*
Cell Proliferation (% of Control)
Time (day)
NS
AAVS1_p1
AC095067.1_p1
AC095067.1_p2
AC095067.1_p3
AC095067.1_p4
HM13-AS1
chr20: 30,161k 1k
HM13
p5*
HM13-AS1_p1
HM13-AS1_p2
HM13-AS1_p3
HM13-AS1_p4
HM13-AS1_p5
RP11-128M1.1
chr20: 2,187k 1k
RP11-128M1.1_p1
RP11-128M1.1_p2
RP11-128M1.1_p3
RP11-128M1.1_p4
RP11-128M1.1_p5
b
LINC00882, LINC00883
chr3: 106,959k 1k
LINC00882
LINC00883
p1*
p2*
p3*
LINC00882_p1
LINC00882_p2
LINC00882_p3
LINC00883_p1
LINC00883_p2
LINC00883_p3
c
Relative mRNA Abundance
LINC00882
AAVS1_p1
LINC00882_p1
LINC00882_p3
d
LINC00882_p1/LINC00882
LINC00882_p3/LINC00882
f
AC095067.1_p1
AC095067.1_sgRNA1p1
AC095067.1_sgRNA2p1
AC095067.1_p4
AC095067.1_sgRNA1p4
AC095067.1_sgRNA2p4
e
100 bp
1 kb
Mock
AAVS1_p1
RP11-439K3.1_p3
RP11-439K3.1_p4
2 kb
1.5 kb
1 kb

Supplementary Figure 6

Genetic validation of candidate lncRNAs in Huh7.5 cells and genomic deletions induced by RP11-439K3.1 targeting pgRNAs.

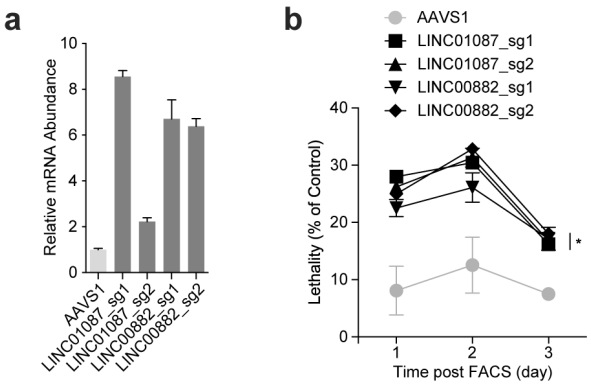
(**a**) Effects of large-fragment deletions of negatively selected lncRNAs on cell proliferation in Huh7.5$_{OC}$ cells. (**b**) Effects of large-fragment deletions of positively selected lncRNAs on cell proliferation in Huh7.5$_{OC}$ cells. (**c**) The LINC00882 mRNA levels (normalized to *GAPDH*) of pooled Huh7.5$_{OC}$ cells containing indicated LINC00882-targeting pgRNAs, with (+) or without (−) the ectopic expression of LINC00882. All primers used for quantitative PCR are listed in **Supplementary Table 12**. (**d**) Effects of the ectopic expression of LINC00882 on cell proliferation in pooled Huh7.5$_{OC}$ cells containing indicated LINC00882-targeting pgRNAs. Data are presented as the mean ± s.d. (n = 3). *P* values are calculated using Student's t test and are corrected for multiple comparison using Benjamini Hochberg procedure, ***P* < 0.01. (**e**) RP11-439K3.1_p3 and RP11-439K3.1_p4 were delivered into Huh7.5$_{OC}$ cells through lentivirus and the deletions were assayed by genomic PCR. (**f**) Effects of sgRNA and pgRNAs targeting negatively selected lncRNAs as indicated on cell proliferation in Huh7.5$_{OC}$ cells. Cell proliferation was determined the same as described in **Fig. 4**. The newly designed pairs different from those used in the original library are marked as asterisk (*) for this and the rest figures. The arrows indicate the transcriptional starting sites. Open and shaded boxes refer to exons of non-coding and coding genes, respectively. Data are presented as the mean ± s.d. (n = 3). *P* values are calculated using Student's t test and are corrected for multiple comparison using Benjamini Hochberg procedure, **P* < 0.05; ***P* < 0.01; ****P* < 0.001; NS, not significant.



Supplementary Figure 7

Relative cell viability for Huh7.5 cells transduced with indicated pgRNA constructs and transcriptional repression.

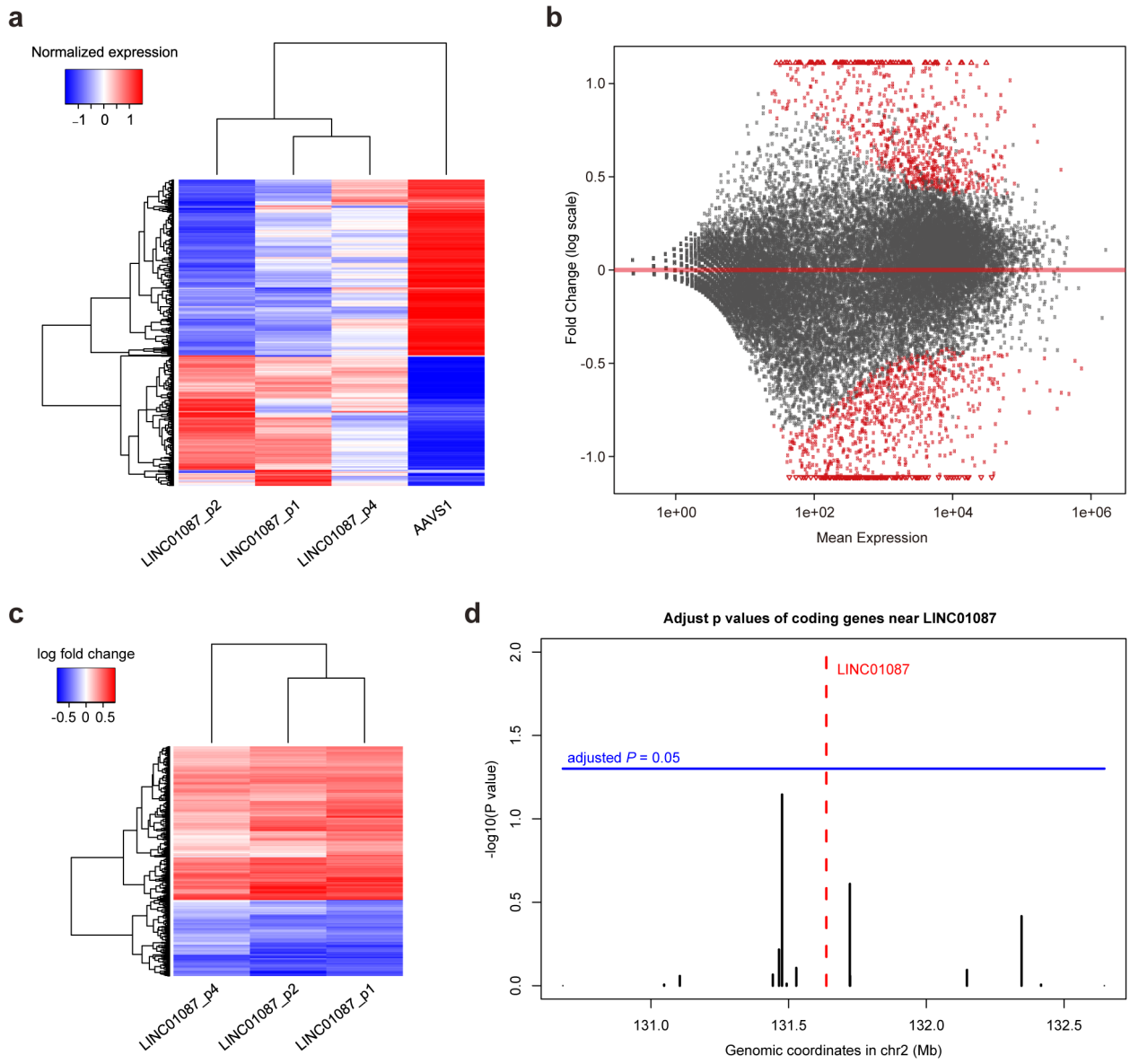
Effects of large-fragment deletions (**a**) and transcriptional repression (**b**) of negatively selected lncRNAs on cell viability in Huh7.5$_{OC}$ cells. All the pgRNAs and sgRNAs were delivered into Huh7.5$_{OC}$ cells through lentiviral infection. The cells were conducted with FACS enrichment 72 h after infection, and the LDH lethality assay were performed from 1-3 days post FACS. Data are presented as the mean ± s.d. (n = 3). *P* values are calculated using Student's t test and are corrected for multiple comparison using Benjamini Hochberg procedure, **P* < 0.05; ***P* < 0.01.



Supplementary Figure 8

Effects of transcriptional activation of positively selected lncRNAs on cell viability.

(**a**) LINC01087 and LINC00882 mRNA levels (normalized to *GAPDH*) were quantified. All primers used for quantitative PCR are listed in **Supplementary Table 12**. (**b**) All the sgRNAs were delivered into Huh7.5 cells through transient transfection. The cells were conducted with FACS enrichment 72 h after transfection, and the LDH lethality assay were performed from 1-3 days post FACS. Data are presented as the mean ± s.d. (n = 3). *P* values are calculated using Student's t test and are corrected for multiple comparison using Benjamini Hochberg procedure, **P* < 0.05.

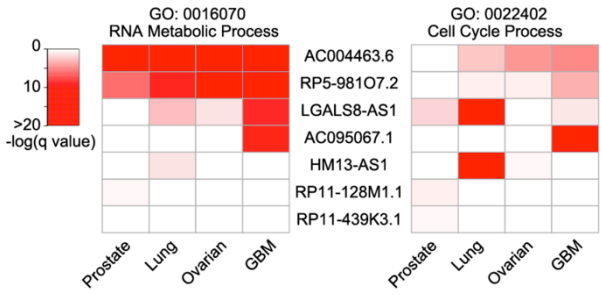


Supplementary Figure 9

RNA-seq profiling of cells knocking out LINC01087, one positively selected lncRNA in the screen and the effect of knocking out LINC01087 on the expressions of nearby coding genes.

(**a**) Hierarchical clustering of the expressions of 500 genes showing the highest variance across samples. (**b**) The mean expression and log fold change (MA plot) of all genes (grey dots), and differentially expressed genes (red dots) detected using DESeq2. (**c**) Comparison of differentially expressed genes between three knockout samples. Genes that are differentially expressed (defined as log fold change > 0.5 or < −0.5) in at least one

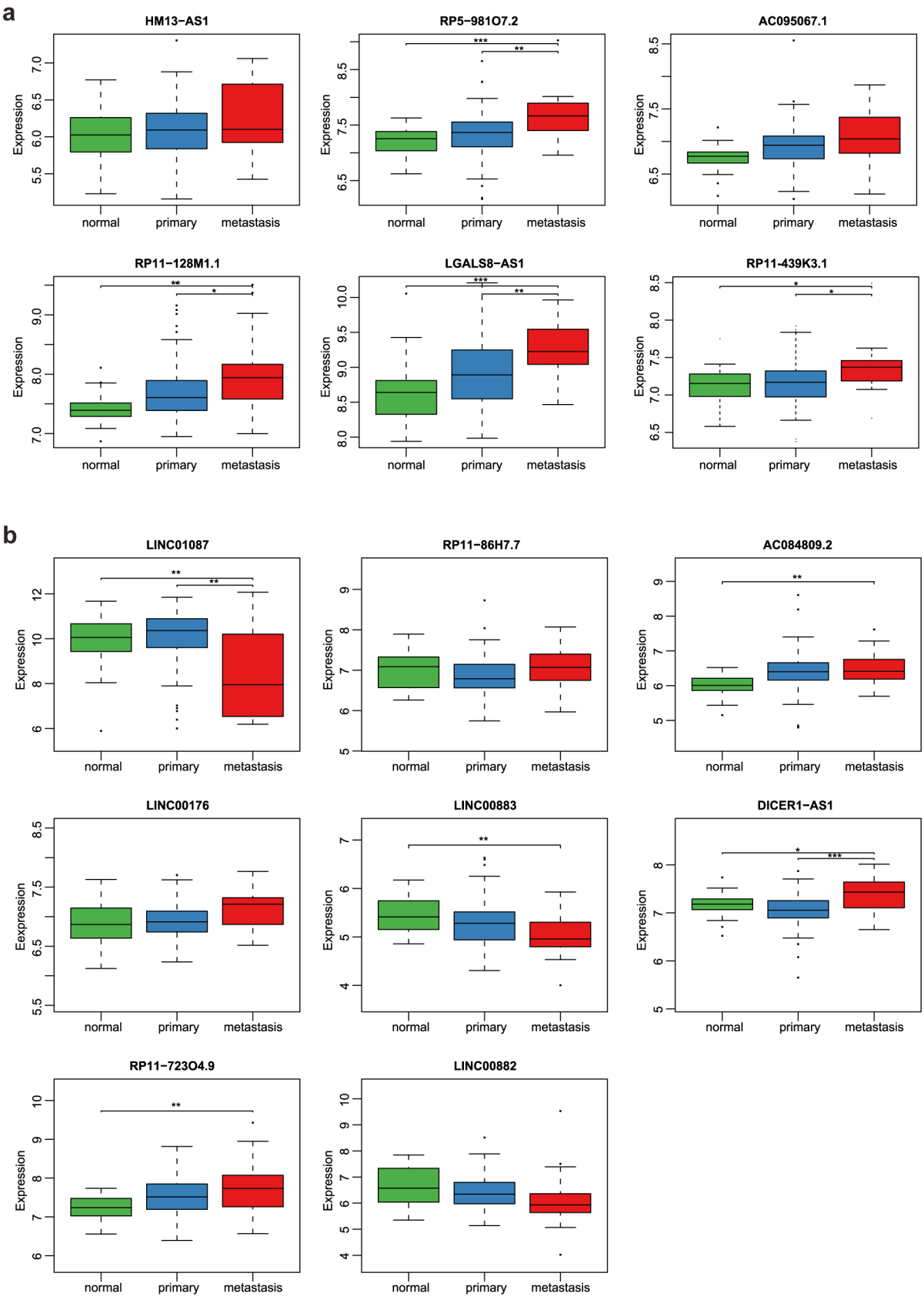
sample are selected, and their log fold changes across all samples are shown. The majority of these genes (99%) are consistently up-regulated (or down-regulated) in all three samples. (**d**) For each coding gene proximal to LINC01087 (distance <1 Mbp), the genomic location and the corresponding differential expression statistics (adjusted *P* value) is reported.



Supplementary Figure 10

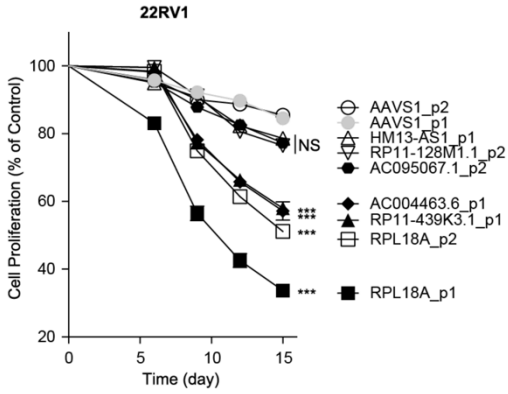
In different cancer types, genes that are co-expressed with top negatively selected.

lncRNAs are enriched in RNA metabolic process and cell cycle process.


a
HM13−AS1
RP5−981O7.2
AC095067.1
RP11−128M1.1
LGALS8−AS1
RP11-439K3.1
b
LINC01087
RP11−86H7.7
AC084809.2
LINC00176
LINC00883
DICER1−AS1
RP11−723O4.9
LINC00882
Expression
normal
primary
metastasis

Supplementary Figure 11

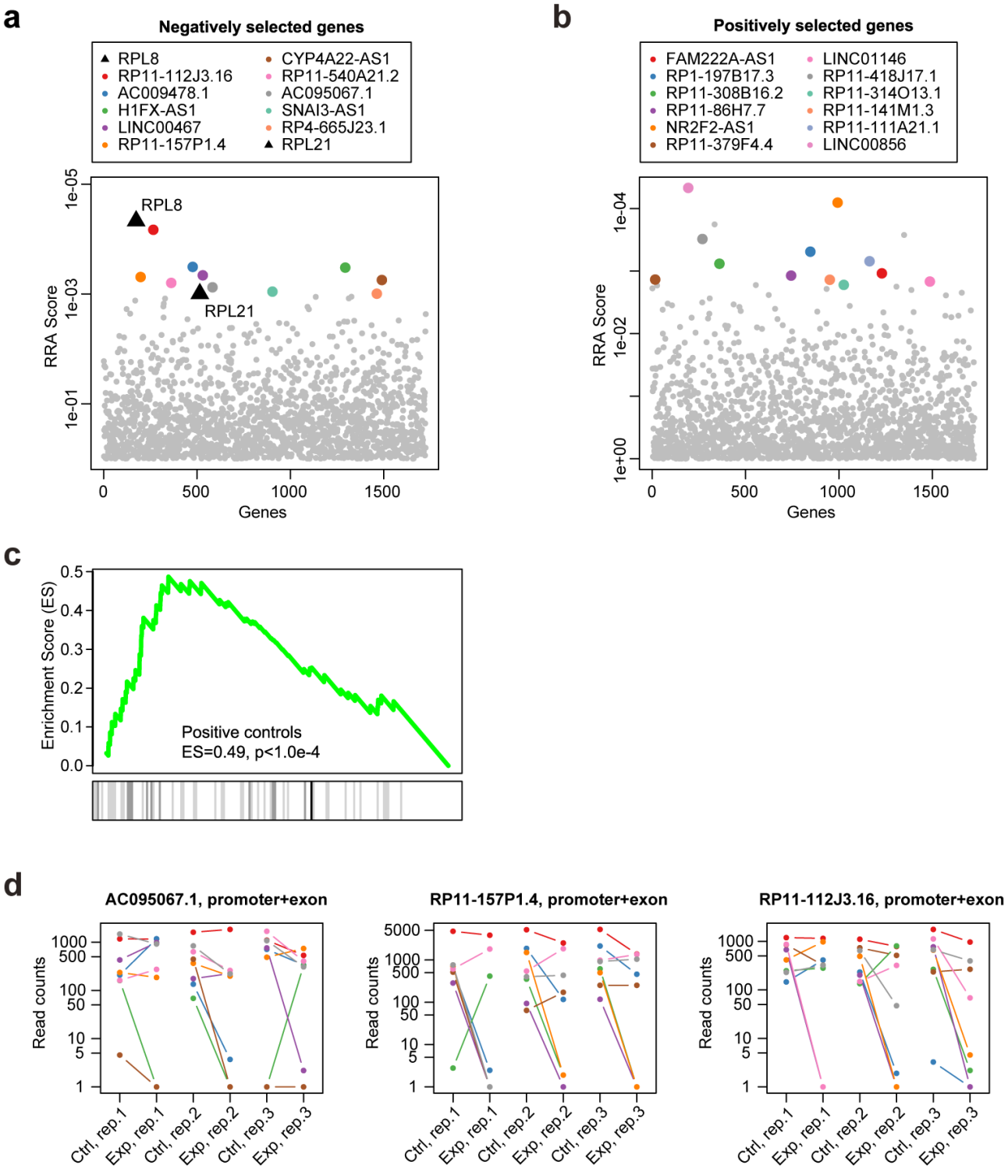
The expression distribution of top negatively (**a**) and top positively (**b**) selected lncRNAs in different stages of prostate cancer. **P* < 0.05; ***P* < 0.01; ****P* < 0.001, Student's t test. Center lines represent median values; box limits represent the interquartile range; whiskers extend each 1.5 times the interquartile range; dots represent outliers.



Supplementary Figure 12

Effects of genomic deletion of the negatively selected lncRNAs on cell proliferation in 22RV1, an advanced stage prostate cancer cell line.

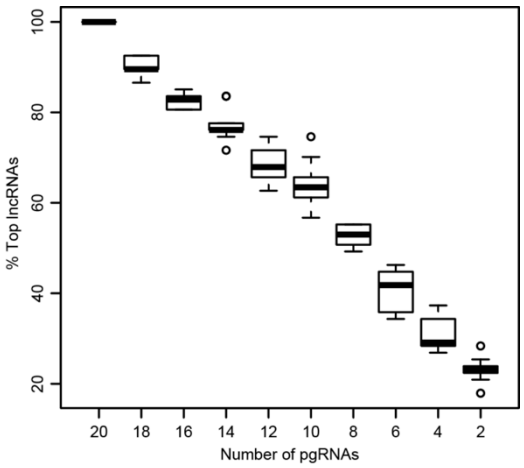
Data are presented as the mean ± s.d. (n = 2). *P* values are calculated using Student's t test and are corrected for multiple comparison using Benjamini Hochberg procedure, **P* < 0.05; ***P* < 0.01; ****P* < 0.001; NS, not significant.



Supplementary Figure 13

The lncRNA screening in HeLa cell line.

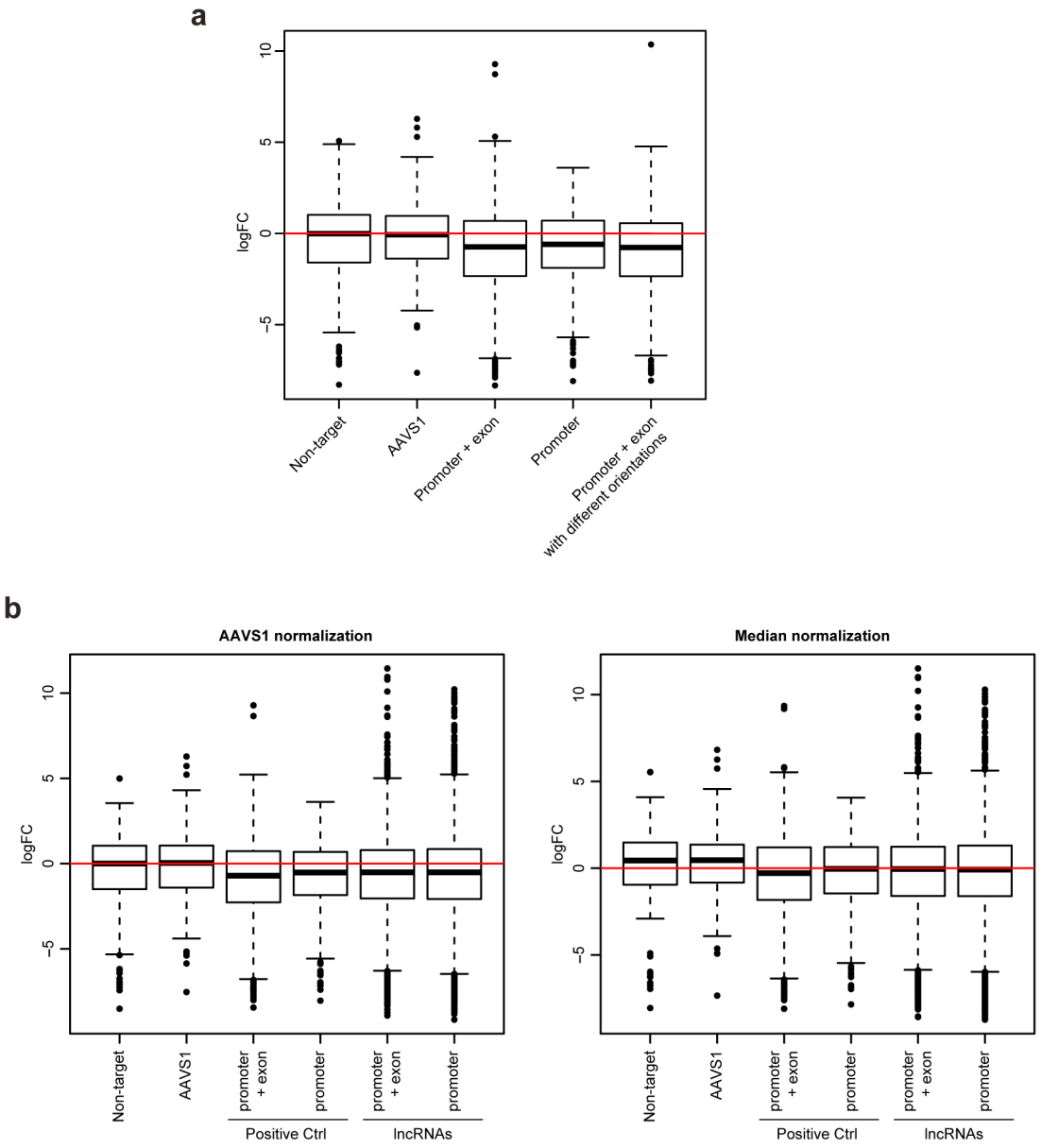
(**a**) The Robust Rank Aggregation (RRA) scores of top ranked negatively selected lncRNAs (marked as dots) and positive control genes (marked as triangles) in HeLa screen. (**b**) The Robust Rank Aggregation (RRA) scores of top ranked positively selected in HeLa screen. (**c**) Essential genes served as positive controls are enriched in negative selections of HeLa screen, using Gene Set Enrichment Analysis (GSEA). The degree of enrichment is measured as Enrichment Score (ES), a measurement of over-representation of a gene set commonly used in the Gene-Set Enrichment Analysis (GSEA). (**d**) The sgRNA read counts of some negatively selected lncRNAs in HeLa screen.



Supplementary Figure 14

The sensitivity of lncRNA screening library in terms of the number of pgRNAs targeting each lncRNA.

For sensitivity evaluation, we use lncRNAs that are ranked top 10% using 20 pgRNAs/lncRNA as “gold standard”, and randomly down-sample the number of pgRNAs for each lncRNA. The percentage of “gold standard” lncRNAs that are still ranked on top 10% is reported. For each number of pgRNAs, 10 independent sampling processes are performed and the distribution of the percentage is reported. Center lines represent median values; box limits represent the interquartile range; whiskers extend each 1.5 times the interquartile range; dots represent outliers.



Supplementary Figure 15

The effect of sgRNA orientation on phenotype and effects of different normalization methods.

(**a**) The distributions of log fold changes of pgRNAs targeting the promoters and exons of positive control genes are shown. These pgRNAs have at least one sgRNA with different orientation compared with the orientation of the target gene. The log fold changes of pgRNAs with the same orientation as target genes are also reported as a reference. (**b**) The distributions of log fold changes of pgRNAs with different groups are shown. Two

different normalization methods are used: AAVS1 normalization (used in the current study) and median normalization, which is a default normalization method for MAGeCK algorithm. Using AAVS1 normalization, the median log fold change of AAVS1 targeting pgRNAs and non-targeting pgRNAs is 0, while median normalization will have non-zero median log fold change. Center lines represent median values; box limits represent the interquartile range; whiskers extend each 1.5 times the interquartile range; dots represent outliers.

Supplementary Tables

Supplementary Table 1 | Design and sequences of paired gRNAs (pgRNAs) targeting *CSPG4* for deletion of DNA fragments

Supplementary Table 2 | Design and sequences of paired gRNAs (pgRNAs) targeting *MALAT1* for deletion of DNA fragments

Supplementary Table 3 | Summary of the pgRNA library for functional screening

Supplementary Table 4 | Qualitative summary of the pgRNA library

Supplementary Table 5 | MAGeCK results of negatively and positively selected lncRNAs in Huh7.5 cell line

Supplementary Table 6 | pgRNA design for functional validation of selected lncRNAs

Supplementary Table 7 | sgRNA design of CRISPR-inhibitor and CRISPR-activator for functional validation of selected lncRNAs

Supplementary Table 8 | The Gene Ontology (GO) enrichment results of genes that are correlated with top lncRNA hits in liver cancer cell lines

Supplementary Table 9 | MAGeCK results of negatively and positively selected lncRNAs in HeLa cell line

Supplementary Table 10 | Oligos for the pgRNA library construction, the read counts of control and treatment samples, and the locations of targeting lncRNAs

Supplementary Table 11 | Primers used for PCR amplification of the genomic DNAs and library construction

Supplementary Table 12 | Primers for Quantitative PCR

Supplementary Table 1. Design and sequences of paired gRNAs (pgRNAs) targeting *CSPG4* for deletion of DNA fragments.

Items	pgRNAs^{CSPG4}	Predicted Size of Deletion (kb)	Predicted Size of Genome PCR using Primers L1/R1 (kb)
1	sgRNA1/sgRNA1’	4.531	0.392
2	sgRNA2/sgRNA2’	4.008	0.915
3	sgRNA3/sgRNA3’	3.476	1.447
4	sgRNA4/sgRNA4’	2.541	2.382
5	sgRNA5/sgRNA5’	2.035	2.888
6	sgRNA6/sgRNA6’	1.967	2.956

sgRNA^{CSPG4}	Sequence
sgRNA1	5’-AGGAGACTGGAGGTAAGACA
sgRNA1’	5’-TCACTCCTGTGCACAGCAGC
sgRNA2	5’-AGAAGAGCTGGCCCAGCAGC
sgRNA2'	5’-CCACCACATACACACCTATG
sgRNA3	5’-AGTCTAGTGAGACGGAGGCG
sgRNA3’	5’-TGCTGGGAGGAGGTTTGAGA
sgRNA4	5’-TCAGTCTCGGGATCTCTGAT
sgRNA4’	5’-TGGCCAGTGATGAGCCTTCT
sgRNA5	5’-GTGCTGGGACTTGCTGTGGT
sgRNA5’	5’-CAGAAAGGCAACTAAACAGA
sgRNA6	5’-ACACCTCTTGCCAGTCTGCT
sgRNA6’	5’-GTTGTAAGCTCCATGGGATT
sgRNA^{AAVS1}	5’-CGGAACCTGAAGGAGGCGGC

Supplementary Table 2. Design and sequences of paired gRNAs (pgRNAs) targeting *MALAT1* for deletion of DNA fragments.

Items	pgRNAs^{MALAT1}	Predicted Size of Deletion (kb)	Predicted Size of Genome PCR using Primers L1/R1 (kb)
1	sgRNA1/sgRNA1’	5.022	1.279
2	sgRNA2/sgRNA2’	4.348	1.953
3	sgRNA3/sgRNA3’	3.195	3.106
4	sgRNA4/sgRNA4’	2.012	4.289
5	sgRNA5/sgRNA5’	0.995	5.306

sgRNA^{MALAT1}	Sequence
sgRNA1	5’- CCGCAGATCAGAGTGGGCCAC
sgRNA1’	5’- GGATAGTACACTTCACTCAG
sgRNA2	5’- ACACAAGAAGTGCTTTAAG
sgRNA2'	5’- GGGATCAAGTGGATTGAGG
sgRNA3	5’- CCCGAATTAATACCAATAGA
sgRNA3’	5’- CTTGAATGTCTCTTAGAGGG
sgRNA4	5’- CCCATCAATTTAATTTCTGG
sgRNA4’	5’- CCAGTTTGAATTGGGAAGCT
sgRNA5	5’- GAGCCAGTGCGATTTGGTGA
sgRNA5’	5’- GGTCTTAACAGGGAAGAGAG

Supplementary Table 3. Summary of pgRNA library for functional screening.

Positive controls		**Negative controls**		**lncRNA targets**	
Genes	20	Non-targeting	100	lncRNA	671
Pairs targeting promoters	~20/gene	Pairs targeting AAVS1 loci	100	Pairs targeting promoters	~10/gene
Pairs targeting promoters + exons	~80/gene	Pairs targeting the introns of positive control genes	259	Pairs targeting promoters + exons	~10/gene
Total	1,708	Total	459	Total	10,305
					Total gRNA pairs: 12,472

Supplementary Table 4. Qualitative summary of pgRNA library.

Category	Plasmid Library	Cell Library
Successful rate	92.5%	84.2%
Mutation rate	7.5%	8.3%
Recombination rate	0%	7.5%

Note: gRNAs' sequences were verified for 80 and 120 randomly-picked clones from the plasmid and the cell libraries, respectively.

Supplementary Text 1

Algorithm for designing pgRNA barcodes

In this section, we describe the algorithm for designing barcodes for a given set of pgRNAs. Note that an intuitive approach is to randomly select one of the two gRNAs as a barcode. However, this intuitive approach may reduce the number of available pgRNAs in the library. The following figure provides an example. Here, three pgRNAs share 3 gRNAs, and a random assignment may lead to a solution with only 2 allowable pgRNAs.

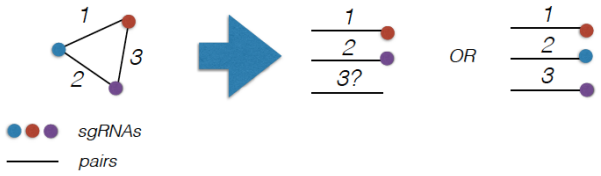


Figure 1: Randomly assigning gRNAs as barcodes may reduce the number of possible pgRNAs in the library.

We sought to develop an algorithm to design as many pgRNAs with barcodes. Here, we use a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ to describe the relationships between gRNAs and pgRNAs. $\boldsymbol{V}$ (vertex) represents all possible gRNAs, and $(\boldsymbol{v_1}, \boldsymbol{v_2}) \in \boldsymbol{E}$ represents one pgRNA, with $\boldsymbol{v_1}$ as a barcode. We first have the following lemma.

Theorem 1. The maximum number of pgRNAs in G is $|V|$, the number of sgRNAs.

Proof: This is true since each gRNA can be served as barcode at most once.

We notice that, if a vertex v is used by only one pgRNA e in G, then v can be served as a barcode of e without affecting the choice of barcodes for other pgRNAs. Therefore, we develop an iterative greedy algorithm to choose the barcodes:

Algorithm 1. The barcode selection algorithm

Input: a graph $G = (V, E)$ with V represents all possible gRNAs, and E represents all possible pgRNAs;

Output: $E' = \{(v_1, v_2)\} \in E$, where $\{v_1\}$ are barcodes of E'.

Procedure:

1. Set $E' = \{\}$;
2. Search for all vertices in V; if v_1 is connected by only one edge $e = (v_1, v_2)$, then set $E' = E' \cup \{(v_1, v_2)\}$, remove v_1 from V and remove (v_1, v_2) from E;

3. If all vertices in V are used by at least two edges, then randomly choose one edge $e = (v_1, v_2)$ and randomly select one vertex (like v_1) as the barcode of e, set $E' = E' \cup \{(v_1, v_2)\}$, remove v_1 from V and remove (v_1, v_2) from E;
4. Go to step 2 until $V = \{\}$.

We now show that this algorithm can indeed select the maximum number of pgRNAs ($|V|$).

Theorem 2. Algorithm 1 can select $|V|$ pgRNAs with barcode for a given graph $G = (V, E)$.

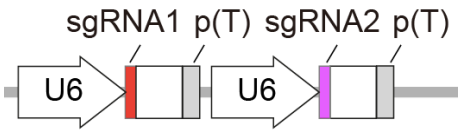
Proof: We prove this theorem using induction on $n = |V|$. Theorem 2 is obviously true for a simple case of $n = 2$, where two gRNAs can be served as barcodes for two pgRNAs. Suppose the algorithm is true for $|V| = n - 1$. Given graph $G = (V, E)$ with $|V| = n$, we have the following scenarios:

1. G includes a vertex v_1 that are connected by only one edge $e = (v_1, v_2)$. In this case, Algorithm 1 will include $(v_1, v_2) \in E'$, and generate a subgraph $G_{n-1} = (V_{n-1}, E_{n-1})$. Note that G_{n-1} has $n - 1$ vertices; therefore, Algorithm 1 can generate a barcode solution $\mathrm{E}'_{\mathrm{n-1}}$ with $\mathrm{n} - 1$ pairs. $\mathrm{E}' = \mathrm{E}'_{\mathrm{n-1}} \cup \mathrm{e}$ is thus a barcode solution with size n, as v_1 is not included in $\mathrm{G}_{\mathrm{n-1}}$.
2. All vertices are connected by at least two edges. Similar to the first scenario, Algorithm 1 will randomly choose an edge $\mathrm{e} = (\mathrm{v}_1, \mathrm{v}_2)$, and generate an optimal solution $\mathrm{E}'_{\mathrm{n-1}}$ for the subgraph $\mathrm{G}_{\mathrm{n-1}}$ without the vertex v_1 Therefore, $\mathrm{E}' = \mathrm{E}'_{\mathrm{n-1}} \cup \mathrm{e}$ is also a barcode solution with size n.

In practice, the number of sgRNAs is more than the desired number of pgRNAs in the design (***e.g.***, 10 pairs for lncRNA promoters and 10 pairs for lncRNA promoters/exons). We therefore run Algorithm 1 until a desired number of pairs is met for each lncRNA or gene.

Supplementary Text 2

The full sequence of the paired sgRNA cassette



TCTAGAGGTACCAAGGTCTGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATT
TGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACA
AAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGT
TTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGA
TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNN
NNNNGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTCCCGTATCTCACGCTCTT
CATCGGCGTCTAGAGGTACCAAGGTCTGGCAGGAAGAGGGCCTATTTCCCATGATTC
CTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACT
GTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTA
GTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA
GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGNNNNNNNN
NNNNNNNNNNNNGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT