nature biotechnology

Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library

Shiyou Zhu[1,2,10], Wei Li[3,4,10], Jingze Liu[1,2], Chen-Hao Chen[3,4], Qi Liao[5], Ping Xu[1], Han Xu[6], Tengfei Xiao[4,7], Zhongzheng Cao[1,8], Jingyu Peng[1], Pengfei Yuan[1], Myles Brown[4,7,9], Xiaole Shirley Liu[3,4,11] & Wensheng Wei[1,11]

CRISPR–Cas9 screens have been widely adopted to analyze coding-gene functions, but high-throughput screening of non-coding elements using this method is more challenging because indels caused by a single cut in non-coding regions are unlikely to produce a functional knockout. A high-throughput method to produce deletions of non-coding DNA is needed. We report a high-throughput genomic deletion strategy to screen for functional long non-coding RNAs (lncRNAs) that is based on a lentiviral paired-guide RNA (pgRNA) library. Applying our screening method, we identified 51 lncRNAs that can positively or negatively regulate human cancer cell growth. We validated 9 of 51 lncRNA hits using CRISPR–Cas9-mediated genomic deletion, functional rescue, CRISPR activation or inhibition and gene-expression profiling. Our high-throughput pgRNA genome deletion method will enable rapid identification of functional mammalian non-coding elements.

The CRISPR–Cas system seen in bacteria and archaea[1] has been developed into a genome editing tool with wide-ranging applications[2–4]. Functional screens of coding genes have been widely adopted, in which pooled libraries of single-guide RNAs (sgRNAs) that target the coding regions of genes associated with specific phenotypes can be selected using cell growth or specific markers as a readout[5–12]. Although similar strategies have been used to tile across regulatory elements in order to investigate *cis*-element function[13], such strategies may not work as well for non-coding elements, since indels caused by one gRNA are unlikely to produce loss-of-function phenotypes. Although two gRNAs have been used to generate a large genomic deletion to investigate the function of individual lncRNAs[14,15], a high-throughput screening using such approach has not been reported. We developed a CRISPR–Cas9 strategy using paired gRNAs (pgRNAs) to produce large-fragment deletions and enable the identification of functional long non-coding RNAs in cancer cells.

RESULTS

Lentivirally delivered paired-guide RNA system

We constructed a CRISPR pgRNA library such that the genomic sequences between two gRNA-targeting sites could be deleted. First, we tested two approaches to express the pgRNAs in one lentiviral backbone—two U6 promoters driving the two gRNAs separately ($U6_2$) and single U6 promoter driving two gRNAs linked consecutively ($U6_1$) (**Fig. 1a**). We compared these two approaches using six pairs of gRNAs that were predicted to delete 2–4.5 kb of the human *CSPG4* locus (*CSPG4* encodes an integral membrane chondroitin sulfate proteoglycan) (**Fig. 1b**,**c** and **Supplementary Table 1**). In the liver cancer cell line Huh7.5$_{OC,}$ which stably expresses the *Cas9* and *OCT1* genes[8,16], all six pgRNAs in a $U6_2$ vector produced genomic deletions with the correct sizes, whereas only two pgRNAs in a $U6_1$ vector produced the correct deletion, and at a much lower efficiency (**Fig. 1c**). Five pgRNAs in $U6_2$ targeting the lncRNA *MALAT1* also produced genomic deletions of the correct sizes with high efficiency (**Supplementary Fig. 1a**,**b** and **Supplementary Table 2**). This suggests that $U6_2$ has superior deletion efficiency, and we therefore adopted it for subsequent experiments. We next investigated whether the post-transduction culture time of lentivirally delivered pgRNAs affected the efficiency of genomic deletion, and observed continued genomic deletion over time that reached a plateau around 15 d post transduction (**Fig. 1d**). Similar results were observed when genomic deletions were induced using different pgRNAs targeting *CSPG4* (2+2′; **Fig. 1b**) or *MALAT1* (2+2′; **Supplementary Fig. 1a**,**c**,**d**). Therefore, culturing library cells for at least 2 weeks post transduction is desirable to allow sufficient time to produce genomic deletions in mammalian cells at a level that is optimal for screening. Genomic sequencing of

[1]Biodynamic Optical Imaging Center (BIOPIC), Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [2]Peking University–Tsinghua University–National Institute of Biological Sciences Joint Graduate Program (PTN), Peking University, China. [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [4]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [5]Department of Prevention Medicine, School of Medicine, Ningbo University, Ningbo, Zhejiang, China. [6]Broad Institute of MIT and Harvard, Cambridge Center, Cambridge, Massachusetts, USA. [7]Division of Molecular and Cellular Oncology, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [8]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [9]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [10]These authors contributed equally to this work. [11]These authors jointly directed this work. Correspondence and should be addressed to W.W. (wswei@pku.edu.cn) or X.S.L. (xsliu.dfci@gmail.com).

Received 15 December 2015; accepted 30 September 2016; published online 31 October 2016; corrected online 9 November 2016 (details online); doi:10.1038/nbt.3715

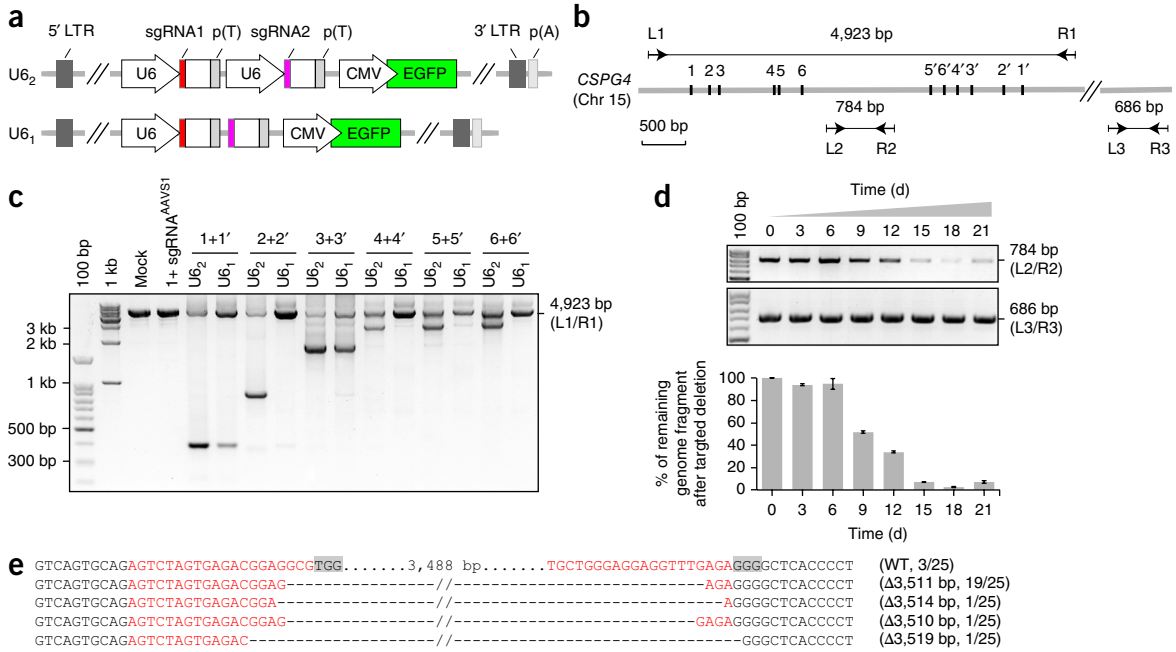

Figure 1 Lentivirally delivered paired-guide RNAs create large-fragment deletion with high efficiency in human cells stably expressing Cas9. (**a**) Structures of the lentiviral plasmids expressing paired-guide RNAs (pgRNAs). The U6 promoter(s) and gRNA coding sequences were cloned into an LL3.7 lentiviral backbone. Amplified DNA fragments encoding customized pgRNAs were ligated into the lentiviral backbone with U6 promoters ($U6_2$) or only one mutual U6 promoter ($U6_1$) using the Golden Gate method. (**b**,**c**) pgRNA vectors were delivered into human cells that express Cas9 through lentivirus. Large-fragment deletions induced by pgRNAs targeting the *CSPG4* gene were identified by PCR. Six pairs of gRNAs that produced large-fragment deletions of 2–4.5 kb were chosen (**b**), and primers L1/R1 were used for the genomic PCR reactions (**b**,**c**). All infected Huh7.5_{OC} cells were enriched by FACS and incubated for 6 d. '$U6_2$' and '$U6_1$' represents two tandem structures in **a**, and the control is a pair of gRNAs with one targeting the *CSPG4* locus and the other targeting *AAVS1* region. (**d**) Quantification of the efficiency of large-fragment deletions using genomic PCR over the course of time post transduction. The pgRNAs (3+3′ in **c**, generating 3.5-kb deletion by design) were delivered into Huh7.5_{OC} cells through lentiviral infection, and genomic DNA was extracted at different time points as indicated (upper). The primers L2/R2 corresponding to sequences flanking pgRNA targeting sites (**b**) were used for quantification, and primers L3/R3 corresponding to sequences farther away from the targeting sites (**b**) were used for normalization. Primer sequences are listed in **Supplementary Table 11**. Images were analyzed using ImageJ software, and data are presented as the mean ± s.d. ($n = 3$) (below). (**e**) DNA sequencing analysis of large-fragment deletions in the human *CSPG4* locus targeted by pgRNAs (3+3′) from pooled cells 3 weeks post infection (**d**). Partial sequences of targeted genes containing the two gRNAs' targeting regions are labeled in red, and the shaded nucleotides represent the PAM sequences. Dashes indicate deletions.

five pgRNAs targeting regions in total (3 pgRNAs targeting *CSPG4* and 2 pgRNAs targeting *MALAT1*) revealed that almost 80% of the deletions at each site were the result of the precise joining of two Cas9 cleavage sites 3 nucleotides (nt) upstream of the protospacer adjacent motifs (PAMs) (**Fig. 1e** and **Supplementary Fig. 1e**), consistent with previous findings[17]. Taken together, these data indicate that lentivirally delivered pgRNAs are capable of creating large genomic deletions with high efficiency in mammalian cells.

pgRNA library construction and genome-wide lncRNA deletion screen

A pgRNA library targeting around 700 human lncRNA genes (**Fig. 2a**; **Supplementary Table 3** and Online Methods) with known or putative roles in cancers or other diseases[18] was designed. For each lncRNA target, we first identified all possible 20-nt gRNAs adjacent to the canonical PAM, then filtered gRNAs that were predicted to have low cutting specificity[19] or efficiency[20] (Online Methods and **Supplementary Code**). We selected gRNA pairs with one unique gRNA as a barcode for each pair (Online Methods) and developed a rapid and accurate method to clone the pgRNAs into a lentiviral expression vector (**Fig. 2b** and **Supplementary Fig. 2a**,**b**). Since the two gRNAs in each pair are driven by the same type of U6 promoter and contain identical 3′ scaffold sequences, recombination might occur, which could result in erroneous pgRNA pairing. We tested the recombination rate in both pgRNA library plasmid constructs and chromosomal integrations in cells after transduction, and found that recombination occurred after viral transduction in approximately 7.5% of cells, which is comparable to error rates during oligosynthesis (**Supplementary Table 4**). This suggests that recombination should have a negligible effect on pgRNA library screening.

We constructed our pgRNA library in $U6_2$ at low multiplicity of infection (MOI) into Huh7.5_{OC} cells that had previously been used for functional screening for coding genes[16]. We cultured the cells for 30 d post transduction to try and maximize the identification of lncRNAs that either positively or negatively affect cell growth or viability. PCR-amplified barcode-gRNA regions from the extracted genomic DNA of cells before and after CRISPR screening were subjected to deep-sequencing analysis (**Fig. 2c** and **Supplementary Fig. 2c**). Overall, the read distribution of three independent experimental replicates within each condition showed a high level of correlation (**Fig. 3a** and **Supplementary Fig. 3**). After 30 d of culture, pgRNAs targeting either positive control genes (mostly ribosomal genes) or lncRNAs were depleted compared with negative control pgRNAs (non-targeting pgRNAs or pgRNAs targeting the non-functional adeno-associated virus integration site 1 (AAVS1) loci) (**Fig. 3b**), indicative of their effect on cell survival or proliferation.

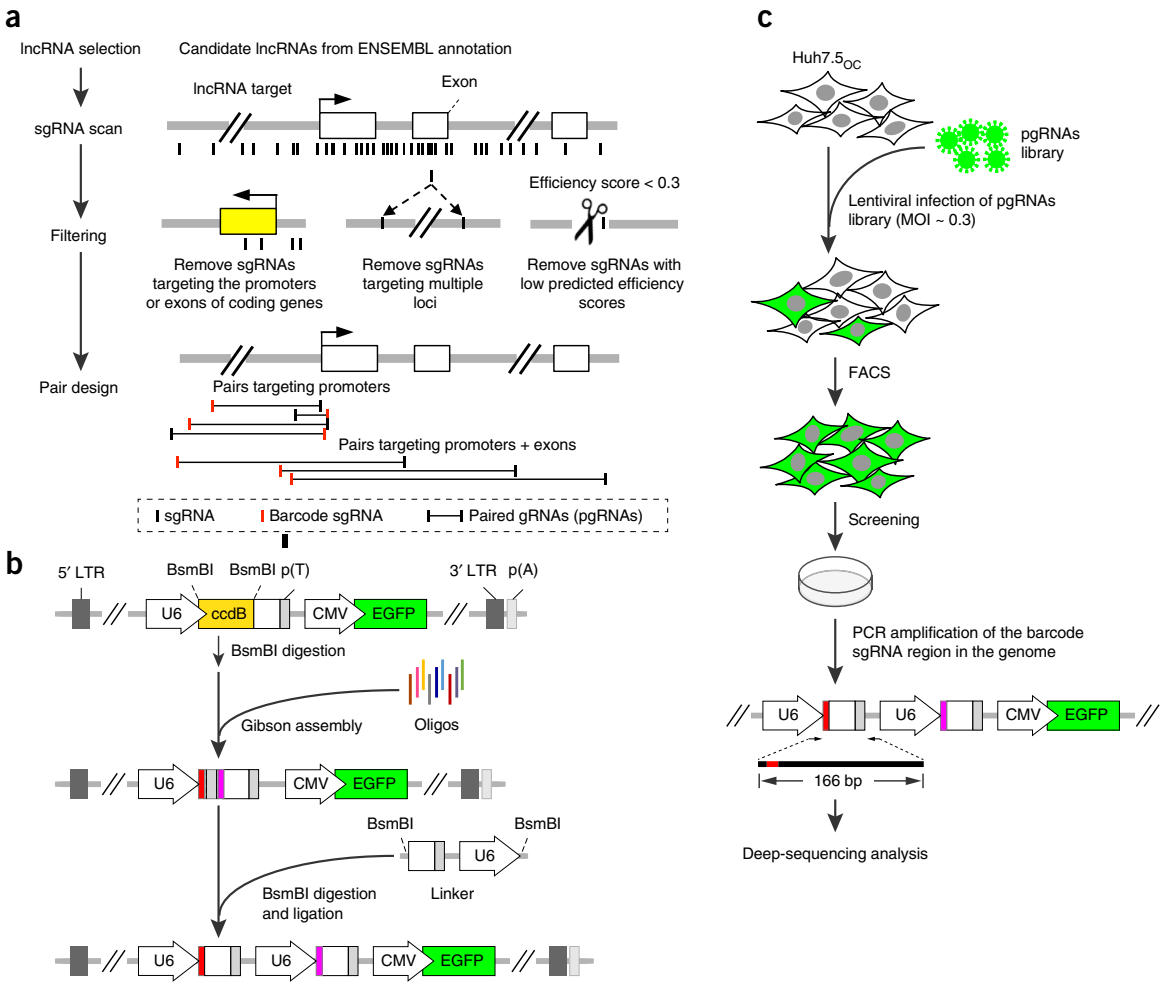

Figure 2 pgRNA library design, cloning and screening. (**a**) pgRNA library design. 671 candidate lncRNAs were identified from ENSEMBL[35]. For each candidate lncRNA, pgRNAs targeting promoter or gene bodies were designed (using the pgRNADesign algorithm) with one unique gRNA for each pair serving as decoding barcode (Online Methods and **Supplementary Code**). (**b**) pgRNA plasmid library construction. Each array-synthesized 137-nt DNA oligo contains two gRNAs (represented in red and purple). Oligos were amplified to produce double-stranded DNA (dsDNA) molecules and cloned into a lentiviral backbone using a Gibson reaction. The final constructs were obtained after the insertion of a linker segment by BsmBI digestion and ligation (Online Methods and **Supplementary Fig. 1**). (**c**) The pgRNA library was delivered into Huh7.5$_{OC}$ cells by lentiviral infection with a MOI of about 0.3. Infected cells were harvested by FACS for green fluorescence 3 d post infection. For screening, library cells were cultured for 30 d before genome DNA extraction and high-throughput sequencing analysis of the barcode gRNA regions.

We used the MAGeCK algorithm to identify the top hits by comparing samples from day 30 with day-0 controls[21]. MAGeCK evaluates the statistical significance of individual pgRNA abundance changes using a negative binomial (NB) model, and compares the ranks of pgRNAs targeting each lncRNA with a null model of uniform distribution (Online Methods). The output of MAGeCK is a set of negatively (or positively) selected lncRNAs, or lncRNAs whose knockout disrupts (or stimulates) cell proliferation. In total, MAGeCK identified 43 negatively selected and 8 positively selected lncRNAs with statistical significance (false discovery rate < 0.25; **Supplementary Table 5**). Gene set enrichment analysis (GSEA) showed that positive control pgRNAs were significantly enriched in the ranked list of negatively selected pgRNAs (**Fig. 3c**), as expected given the essential roles of their targets[22]. The top negatively selected genes include two positive control genes: *RPL18A*, a ribosomal gene, and *EZH2*, a gene that encodes a member of the Polycomb-group family that has an essential role in the proliferation of liver cancer cells[23]. pgRNAs targeting the promoters and exons of *RPL18A* and *EZH2* were consistently depleted (**Fig. 3d**,**e**). Similarly, 89% of the pgRNAs targeting top-ranked negatively selected lncRNAs were depleted, while 76% of the pgRNAs targeting positively selected lncRNAs were enriched (**Fig. 3f**,**g** and **Supplementary Fig. 4a**,**b**). In contrast, the abundances of pgRNAs with non-targeting controls and targeting the AAVS1 loci were similar between control and treatment conditions (**Supplementary Fig. 4c**). Intriguingly, 266 pgRNAs targeting 25 intronic regions of essential genes decreased cell viability (**Fig. 3d**), possibly as a result of the deletion of regulatory elements or modulation of alternative splicing of the target genes[24,25].

Validation of selected lncRNA candidates

From the positively or negatively selected lncRNAs with statistical significance, we obtained top-ranked hits whose corresponding pgRNAs were consistently depleted (for negative selection) or enriched (for positive selection) in three independent experimental replicates (**Fig. 3f**,**g** and **Supplementary Fig. 5**). To validate the functions of some of these lncRNAs, we chose two pairs of gRNAs that were present in the original

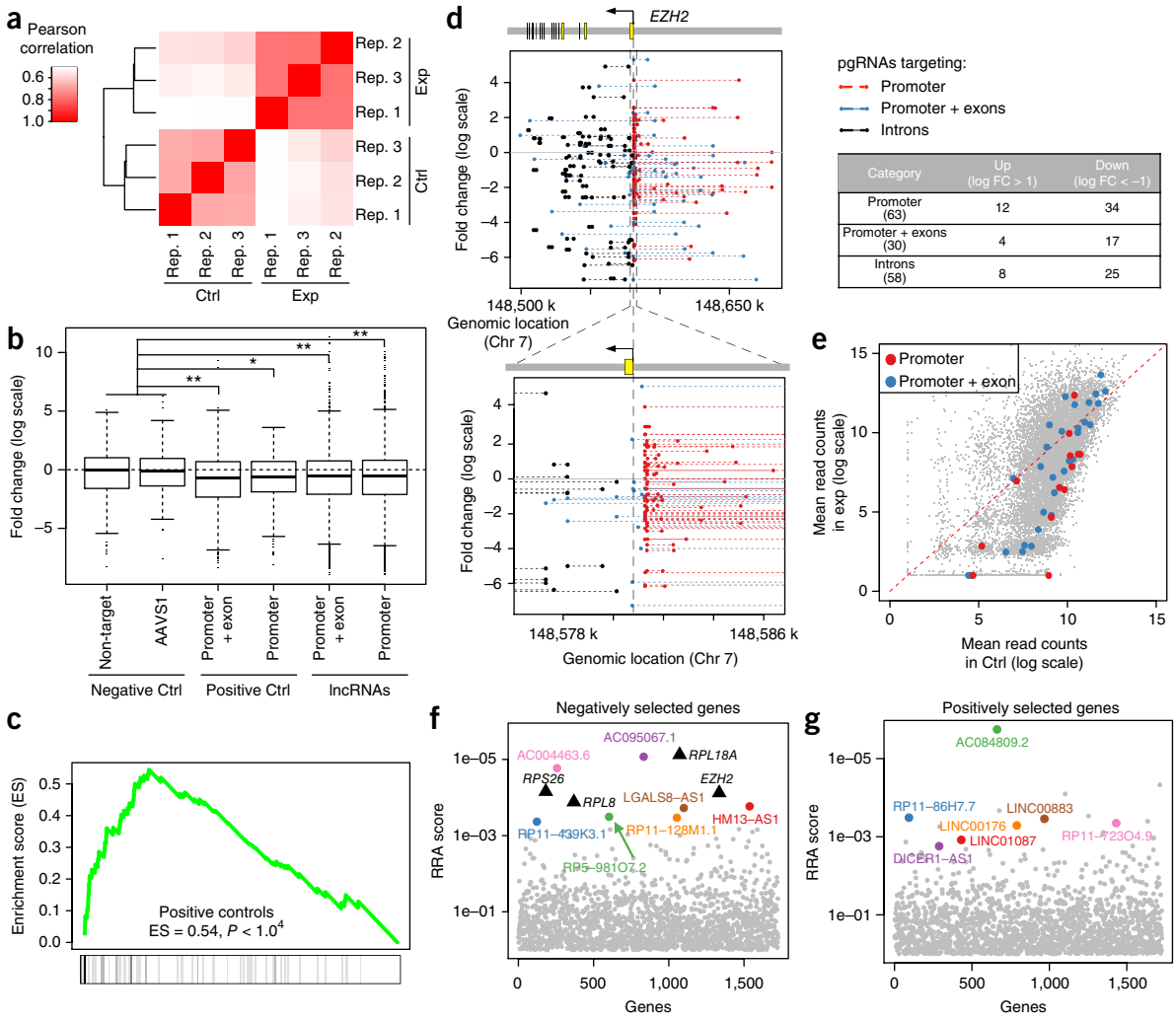

Figure 3 Identification of negatively and positively selected lncRNAs. (**a**) The Pearson correlation coefficient between the replicates (Rep.) of control samples (Ctrl) and 30-d enrichment samples (Exp). (**b**) The log fold-change distribution of pgRNAs targeting negative controls, positive controls and lncRNAs. $*P < 0.05$; $**P < 0.01$; Wilcoxon rank–sum test. Center lines represent median values; box limits represent the interquartile range; whiskers each extend 1.5 times the interquartile range; dots represent outliers. (**c**) Essential genes served as positive controls are enriched in negative selections 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 GSEA. (**d**) The log fold-change values and genomic locations of all pgRNAs targeting *EZH2*, a positive control gene. A zoom-in view of pgRNAs near the promoter of *EZH2* is also shown. The majority of pgRNAs are depleted, including pgRNAs targeting the promoter, promoter + exons and introns of *EZH2* (table at right). (**e**) The mean read counts of pgRNAs targeting the promoter and promoter + exon of a positive control ribosomal gene, *RPL18A*. (**f**,**g**) The robust rank aggregation (RRA) scores of top ranked negatively selected lncRNAs (**f**) and positively selected lncRNAs (**g**) calculated by MAGeCK. Some positive control genes that are negatively selected are also shown as black triangles. A smaller RRA score indicates a stronger selection of the corresponding lncRNAs.

screening library and designed up to three additional new pgRNAs for each gene. In addition, three pairs of gRNAs were designed to target the AAVS1 loci to serve as negative controls (**Supplementary Table 6**). All pgRNAs were transduced afresh into Huh7.5$_{OC}$ cells using a lentiviral backbone carrying CMV–EGFP, and proliferation of cells was quantified based on the percentage change of EGFP-positive cells. Deletion of the promoter of *RPL18A*, one ribosomal gene that ranked on top of the negative selection list from the screen, strongly decreased cell proliferation, while deletions of the AAVS1 loci had negligible effect on cell growth (**Fig. 4a**).

Using the same method, we selected lncRNAs without any overlap with coding genes from the pgRNA library screening for validation. From the initial screen, we chose five negatively selected lncRNAs (AC004463.6, AC095067.1, HM13-AS1, RP11-128M1.1 and RP11-439K3.1) and four positively selected lncRNAs (LINC00176, LINC01087, LINC00882 and LINC00883). We designed pgRNAs to target the promoters or exons of these lncRNAs. For the divergently transcribed pair LINC00882 and LINC00883, which share the same promoter, we designed three additional pgRNAs to target their exons. All five negative-selected lncRNAs were found upon individual deletion to be essential for cell proliferation, and all four positive-selected lncRNAs were confirmed to negatively regulate cell proliferation (**Fig. 4b**,**c** and **Supplementary Fig. 6a**,**b**). We further introduced a cDNA clone of LINC00882 into two groups of LINC00882-deleted Huh7.5$_{OC}$ cells and demonstrated that the ectopic expression of LINC00882 could inhibit cell proliferation (**Supplementary Fig. 6c**,**d**). Some pgRNAs, such as RP11-439K3.1_p3 and RP11-439K3.1_p4, did not produce distinct phenotypes (**Fig. 4b**) because of their failure to generate

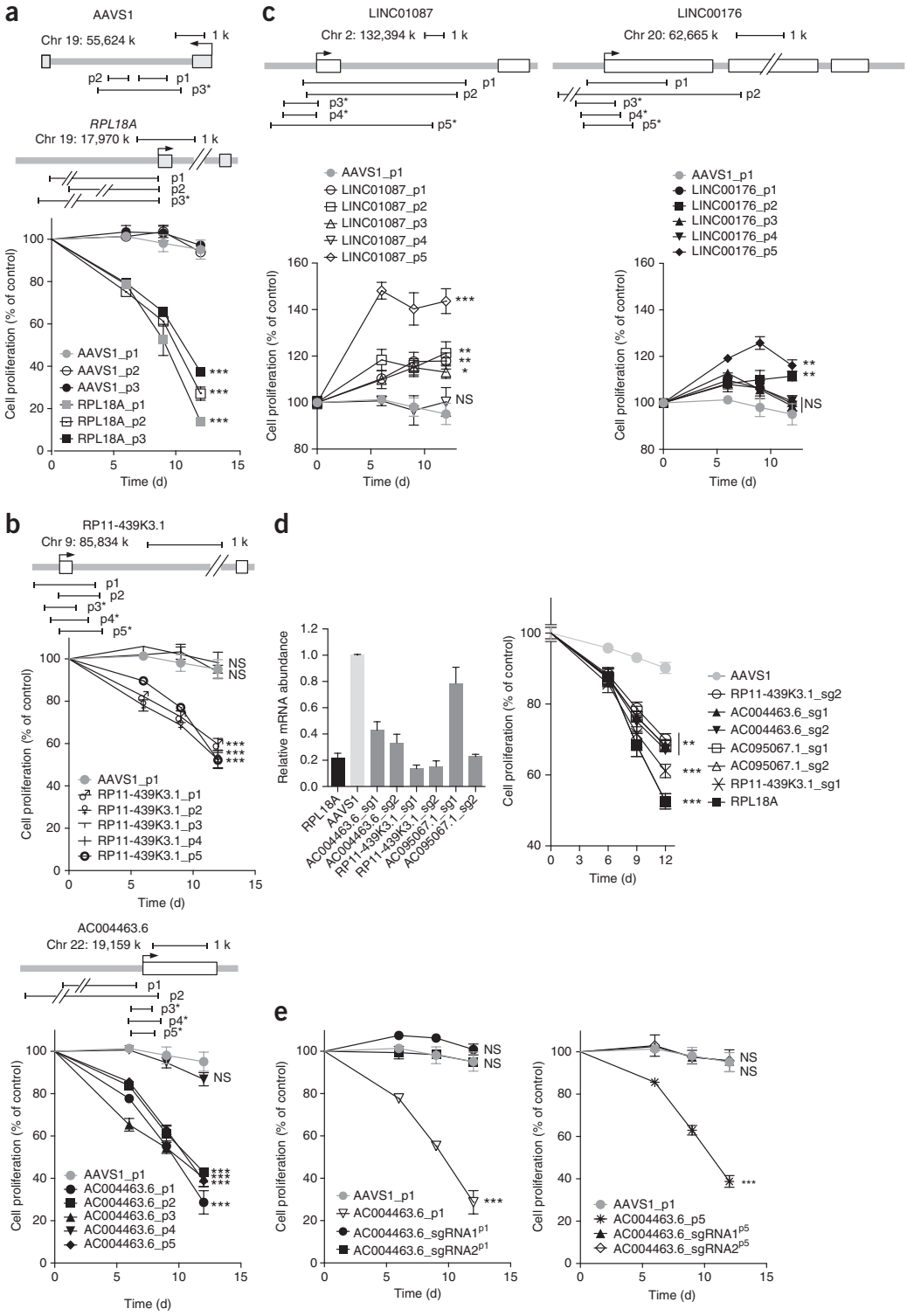

Figure 4 Validation of candidate lncRNAs. (**a**–**c**) Effects of large-fragment deletions of *RPL18A* (**a**), negatively selected lncRNAs (**b**) and positively selected lncRNAs (**c**) on cell proliferation in Huh7.5$_{OC}$ cells. ~3–5 pairs of gRNAs of each lncRNA targeting promoter or promoter + exon were chosen for validation. The pgRNAs, expressing from a backbone carrying CMV promoter-driven EGFP, were delivered into cells by lentiviral infection, and the percentage of EGFP$^+$ cells was quantified by FACS. The first quantification started from 3 d post viral infection, labeled as day 0 in this and the remaining figures. Cell proliferation was determined by normalizing EGFP$^+$ percentages at indicated time points with control (day 0). Newly designed pgRNAs different to those used in the original library are marked with an asterisk (*). The arrows indicate the transcriptional start sites. Open and shaded boxes refer to exons of non-coding and coding genes, respectively. (**d**) Effects of transcriptional repression of negatively selected lncRNAs on cell proliferation. RPL18A, AC004463.6, RP11-439K3.1 and AC095067.1 mRNA levels (normalized to GAPDH) were quantified. All primers used for quantitative PCR are listed in **Supplementary Table 12**. (**e**) Effects of sgRNA and pgRNAs targeting negatively-selected lncRNAs on cell proliferation in Huh7.5 cells. Data are presented as the mean ± s.d. ($n = 3$). *P* values were calculated using Student's *t*-test and corrected for multiple comparison using Benjamini–Hochberg procedure, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; NS, not significant.

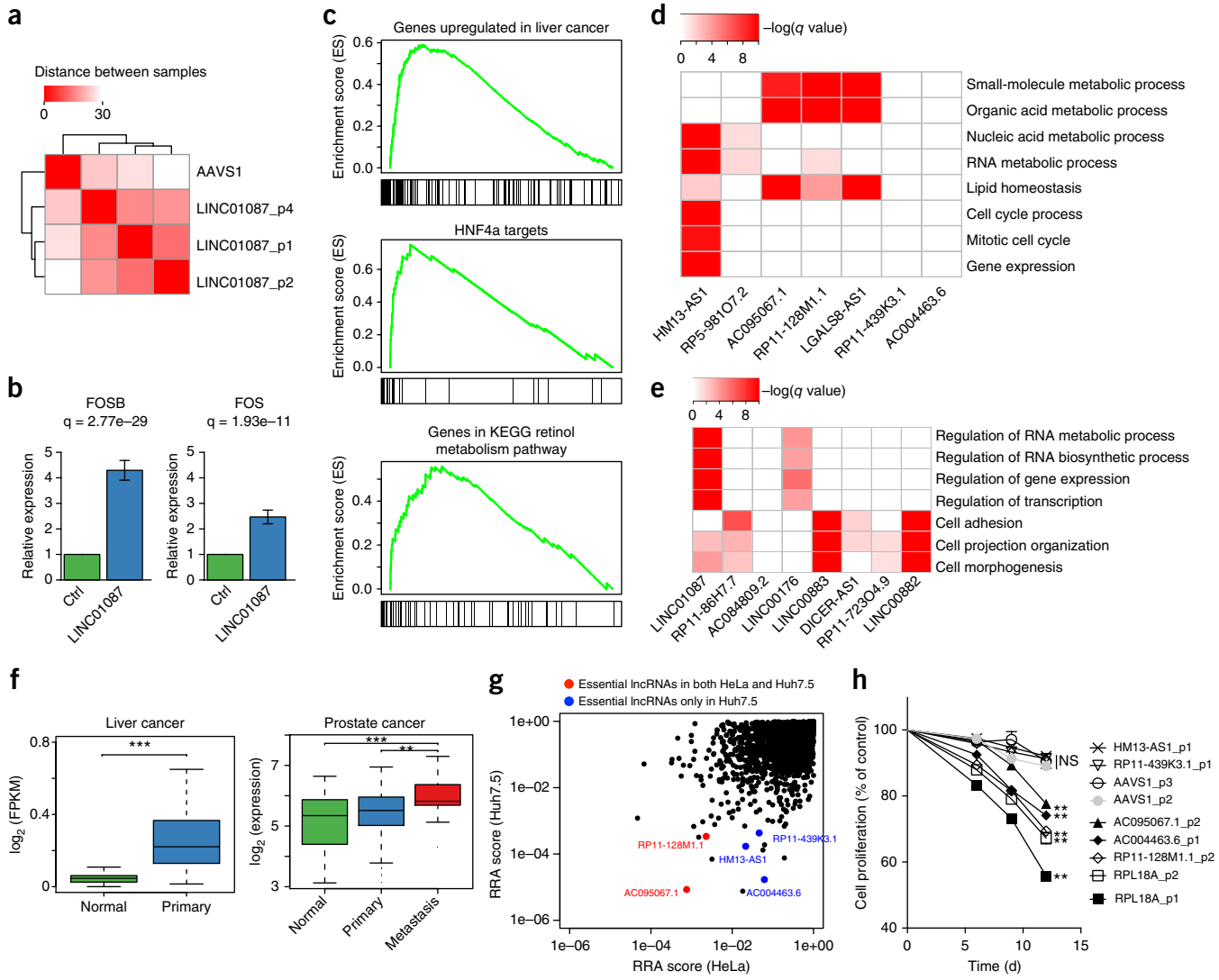

Figure 5 Functional expression analysis of selected selection of identified lncRNAs. (**a**–**c**) RNA-seq profiling of cells that have a LINC01087 (positively selected) knockout. (**a**) RNA-seq gene expression clustering of cells with a LINC01087 knockout using three different sets of pgRNAs, and control cells in which the AAVS1 locus is deleted. Clustering is based on the Euclid distances between samples, using the expressions of all genes. (**b**) Two Fos family genes, FOS and FOSB, were upregulated in the LINC01087 knockout. (**c**) Enriched signatures of genes that are upregulated upon deletion of LINC01087 using GSEA analysis. These signatures include genes that are upregulated in liver cancer (MSigDB ID: LEE_LIVER_CANCER_SURVIVAL_UP), HNF4α target genes (SUMI_HNF4A_TARGETS), and genes in the KEGG retinol metabolism pathway (KEGG_RETINOL_METABOLISM). All signatures are statistically significant ($q = 0$, permutation test). (**d**,**e**) The gene ontology (GO) terms of genes that are co-expressed with top negatively (**d**) or positively (**e**) selected lncRNAs in liver cancer. (**f**) One negatively selected lncRNA, AC004463.6, is significantly overexpressed in liver cancer tumors and metastatic prostate tumors. (**g**) Comparison of lncRNA screens in Huh7.5 and HeLa cells (a cervical cancer cell line). Two lncRNAs are essential in both cell lines (red dots), and three lncRNAs are essential only in Huh7.5 cells (blue dots). Similarly to what is seen **Figure 3f**,**g**, a smaller RRA score indicates stronger selection of the corresponding lncRNAs. (**h**) Effects of deletion of selected lncRNAs on HeLa cell proliferation. Data are presented as the mean ± s.d. ($n = 2$). *P* values were calculated using Student's *t*-test and corrected for multiple comparisons using the Benjamini–Hochberg procedure. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; NS, not significant.

genomic deletions (**Supplementary Fig. 6e**). To further validate candidate genes, we used a CRISPR-inhibitor (CRISPRi) method[26] that can reduce the transcription of the targeted gene. Out of the five negatively selected lncRNAs, we were able to successfully decrease the expression of three (AC004463.6, RP11-439K3.1 and AC095067.1) using CRISPRi, and all significantly decreased cell proliferation (**Fig. 4d**). We also carried out cell lethality assays on lines with deletion on the five negatively selected lncRNAs and on the CRISPRi lines (transcription of lncRNAs repressed), and found all five lncRNAs to be essential for cell viability (**Supplementary Fig. 7** and **Supplementary Table 7**). For positively selected gene candidates LINC01087 and LINC00882, we used a CRISPR-activator (CRISPRa) method[27] to upregulate their transcription, and found both lncRNAs to be lethal when overexpressed (**Supplementary Fig. 8** and **Supplementary Table 7**). Therefore, the CRISPR–Cas9 screen strategy from genomic deletion works well for both negatively and positively selected lncRNAs with high efficiency and reliability.

For both CRISPR screening and candidate validation, we introduced paired gRNAs into cells. It is possible that the phenotypic changes we observed were due to the effect of one gRNA-mediated

double-strand break (DSB) instead of pgRNAs-mediated genomic deletion. To exclude this possibility, we compared the effects of pgRNAs targeting AC004463.6 and AC095067.1 with the effects of introducing only one of their corresponding gRNAs. Only pgRNAs significantly affected cell proliferation in both cases, while none of the single gRNAs targeting introns or exons altered cell survival (**Fig. 4e** and **Supplementary Fig. 6f**). This suggests that, at least for these two lncRNAs, pgRNA-mediated genomic deletion is required to generate functional knockout, an effect unlikely to be achieved through indels created by single gRNAs.

Functional analysis of validated lncRNAs

We next sought to investigate the potential functions of LINC01087, one of the top positively selected lncRNAs in the screen. We knocked out LINC01087 with three different pgRNAs and observed similar changes in gene expression patterns from RNA-seq (**Fig. 5a** and **Supplementary Fig. 9a–c**). Knocking out LINC01087 did not affect the expression of neighboring protein-coding genes (**Supplementary Fig. 9d**), but instead upregulated a set of genes associated with liver cancer. The upregulated genes included *FOS* and *FOS*B (**Fig. 5b**), which encode members of the FOS gene family and AP-1 transcription factor complex[28]; liver-cancer-upregulated genes; targets of the hepatocellular oncogenic transcription factor HNF4α[29]; and genes involved in retinol metabolism (**Fig. 5c**).

We also evaluated the predicted functions of our top 15 lncRNA hits using "guilt by association," a computational approach for inferring lncRNA function from the enriched functions of co-expressed coding genes[30]. We analyzed the expressions of genes and lncRNAs in five different cancers (liver, prostate, ovarian, lung cancer and glioblastoma multiforme) using data sets from the Cancer Cell Line Encyclopedia (CCLE) and The Cancer Genome Atlas (TCGA), respectively[18,31] (Online Methods). Many of the genes that were co-expressed with negatively selected lncRNAs were enriched in essential processes such as RNA metabolism and cell cycle, whereas genes correlated with positively selected lncRNAs were enriched in the negative regulation of these essential processes (**Fig. 5d,e**; **Supplementary Fig. 10** and **Supplementary Table 8**). This is consistent with the finding that knockout of these negatively (or positively) selected lncRNAs disrupts (or enhances) cell proliferation and viability. One of the negatively selected lncRNAs, AC004463.6, is significantly overexpressed in liver cancer and metastatic prostate cancer (**Fig. 5f**), and five out of the seven negatively selected lncRNAs in Huh7.5 are significantly overexpressed in metastatic prostate cancers (**Supplementary Fig. 11**). In addition, two of the five validated negatively selected lncRNAs, AC004463.6 and RP11-439K3.1, were confirmed to be essential in 22RV1, a relapsed prostate cancer cell line[32] (**Supplementary Fig. 12**). These results suggest that the lncRNAs selected in liver cancer cell lines may also function in other cancer types.

lncRNA screening in HeLa cells

To assess the functions of lncRNAs in a different cell type, we screened HeLa cells using our lncRNA library (**Supplementary Table 9**). Positive control pgRNAs and genes were negatively selected, an indication that our screen works well in HeLa cells as well (**Fig. 5g** and **Supplementary Fig. 13**). A further comparison of the screens done in Huh7.5 and HeLa cells revealed different roles for distinct lncRNAs in these two cell types (**Fig. 5g**). Of the top negatively selected and validated lncRNAs in Huh7.5, we tested five lncRNAs in HeLa, including two that seemed to be essential (AC095067.1 and RP11-128M1.1) and three that appeared to be non-essential (HM13-AS1, AC004463.6 and RP11-439K3.1, **Fig. 5h**) in the HeLa cell screen. Indeed, knocking out two essential lncRNAs reduced cell proliferation, and knocking out two of the three non-essential lncRNAs had no effect on cell proliferation. Our screen missed AC004463.6, which was found to be essential in HeLa through individual validation, an indication that the current lncRNA pgRNA library still has space for improvement.

DISCUSSION

The vast majority of all mammalian genomes comprise non-coding regions, many of which have important regulatory roles. Functional analyses of non-coding regions have been challenging, and an effective screening strategy for non-coding regions based on genomic deletion was, until now, lacking. We have established a genome deletion screening method using CRISPR–Cas9 screens with paired gRNAs in mammalian cells. Using this method, we screened approximately 700 human lncRNAs and identified lncRNAs that have oncogenic or tumor suppressor activities in cancer cells. Orthogonal validation of top hits using individual CRISPR–Cas9 knockout, CRISPR inhibition and activation, gene expression profiling and expression correlation analysis confirmed the findings of our screens and showed that our method has a high level of fidelity and specificity.

There are potential limitations to our lncRNA screening strategy. First, deleting lncRNAs may also affect other proximal functional elements, including enhancers and microRNAs. It is desirable to avoid designing pgRNAs that overlap with other functional elements where possible, to examine hits for potential enhancer function and to validate screening results using orthogonal technologies. Our screening approach cannot reveal mechanisms of lncRNA action[30], so a detailed investigation is needed to further understand the functions of identified lncRNAs. More than 30% of the lncRNAs are located in the introns of other coding genes with diverse biological functions[33]. Further characterization of these lncRNAs is challenging, as disrupting introns may perturb splicing or other regulatory elements and have deleterious effects on cell proliferation (for example, the intron-targeting pgRNAs in **Fig. 3d**). Finally, pgRNA orientation seems to have negligible effects on the knockout phenotype, but the number of pgRNAs per lncRNA is crucial to reduce the false negative rate of the screens (**Supplementary Figs. 14** and **15a**). Not all of the positive controls were identified in our screen, which shows that the sensitivity of the screen needs improvement. As the deletion frequencies for pgRNAs vary (**Fig. 1d** and **Supplementary Fig. 1**), a sufficient number of pgRNAs (preferably >20) targeting each lncRNA is desirable to reduce the false negative rate. Although our CRISPR pgRNA library might theoretically cause incorrect pgRNA assembly due to paired gRNA recombination in the lentiviral packaging and integration step (owing to the sequence similarity of two U6 promoters and two repeats of gRNA scaffold sequences), we found that our screen was unaffected, a result of a low recombination rate. We could further reduce the potential lentiviral recombination rate by using different types of U6 promoters (of human and murine origins)[34] and alternative sgRNA scaffold sequences.

Our approach to screening for function of the non-coding genome could be applied to investigate phenotypic changes of interest other than growth by incorporating a reporter system. Finally, our paired-guide RNA screening strategy could be more broadly applied to study other non-coding sequences including microRNAs, *cis* elements and other currently uncategorized elements.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Accession codes. CRISPR screening results for Huh7.5 cells can be accessed in NCBI Short Read Archive (SRA) with the accession number SRX2148757 and SRX2148759. Screening results for HeLa cells can be accessed in SRA with the accession number SRX2149095. RNA-seq reads can be accessed in SRA with the accession number SRX2152480. Source codes to design pgRNAs are available in **Supplementary Code**, as well as in Bitbucket repository (https://bitbucket.org/liulab/pgrnadesign).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We acknowledge the staff of the BIOPIC sequencing facility (Peking University) for their assistance, and National Center for Protein Sciences Beijing (Peking University) for help in Fluorescence Activated Cell Sorting. The project was supported by funds from the National Science Foundation of China (NSFC31430025, NSFC31170126, NSFC81471909), Beijing Advanced Innovation Center for Genomics at Peking University, and the Peking-Tsinghua Center for Life Sciences (to W.W.), the NIH grant U01 CA180980 (to X.S.L.), R01 HG008728 (to M.B. and X.S.L.), and the Claudia Adams Barr Award in Innovative Basic Cancer Research from the Dana-Farber Cancer Institute.

AUTHOR CONTRIBUTIONS

X.S.L. and W.W. conceived and supervised the project. W.W., S.Z., J.P., and P.Y. designed the experiments. S.Z., J.L., P.X. and Z.C. performed the experiments with help from W.L., T.X. and M.B. and T.X. W.L., C.-H.C. and H.X. designed the oligos used for pgRNA library construction. W.L. performed the data analysis, with the help of Q.L. on the functional expression analysis of candidate lncRNAs. S.Z., W.L., X.S.L. and W.W. wrote the manuscript with the help of all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
3. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
4. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
5. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
6. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
7. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, Mdel C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
8. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
9. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
10. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
11. Shalem, O., Sanjana, N.E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
12. Peng, J., Zhou, Y., Zhu, S. & Wei, W. High-throughput screens in mammalian cells using the CRISPR-Cas9 system. *FEBS J.* **282**, 2089–2096 (2015).
13. Canver, M.C. *et al.* BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197 (2015).
14. Han, J. *et al.* Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol.* **11**, 829–835 (2014).
15. Yin, Y. *et al.* Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16**, 504–516 (2015).
16. Ren, Q. *et al.* A dual-reporter system for real-time monitoring and high-throughput CRISPR/Cas9 library screening of the hepatitis C virus. *Sci. Rep.* **5**, 8865 (2015).
17. Zheng, Q. *et al.* Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques* **57**, 115–124 (2014).
18. Du, Z. *et al.* Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* **20**, 908–913 (2013).
19. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
20. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157 (2015).
21. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
22. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
23. Cheng, A.S. *et al.* EZH2-mediated concordant repression of Wnt antagonists promotes β-catenin-dependent hepatocarcinogenesis. *Cancer Res.* **71**, 4028–4039 (2011).
24. Gillies, S.D., Morrison, S.L., Oi, V.T. & Tonegawa, S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**, 717–728 (1983).
25. Xiao, X. *et al.* Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat. Struct. Mol. Biol.* **16**, 1094–1100 (2009).
26. Gilbert, L.A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
27. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
28. Eferl, R. & Wagner, E.F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3**, 859–868 (2003).
29. Hatziapostolou, M. *et al.* An HNF4α-miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis. *Cell* **147**, 1233–1247 (2011).
30. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
31. Barretina, J. *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
32. Sramkoski, R.M. *et al.* A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell. Dev. Biol. Anim.* **35**, 403–409 (1999).
33. Louro, R., Smirnova, A.S. & Verjovski-Almeida, S. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* **93**, 291–298 (2009).
34. Vidigal, J.A. & Ventura, A. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nat. Commun.* **6**, 8083 (2015).
35. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).

ONLINE METHODS

Cells and reagents. Huh7.5 cells were from S. Cohen's laboratory (Stanford University School of Medicine) and maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) with MEM nonessential amino acids (NEAA; Gibco), 22RV1 cells were from M. Brown's laboratory and maintained in RPMI1640 medium (Gibco) and HeLa cells were from Z. Jiang's laboratory (Peking University) and were maintained in DMEM (Gibco), all supplemented with 10% FBS (CellMax) with 5% CO_2 at 37 °C. All cells were checked to ensure they are free of mycoplasma contamination.

Plasmid construction. The lentiviral pgRNA-expressing vector was constructed by cloning the human U6 promoter, ccdB cassette and gRNA scaffold into pLL3.7 (Addgene, Inc.) by replacing its original U6 promoter[8]. The scaffold-linker-U6 fragment was cloned into pEASY-Blunt plasmid (TransGen Biotech).

lncRNA selection. *lncRNA targets in cancer.* lncRNA targets consist of known cancer-related lncRNAs and lncRNAs that are differentially expressed in tumors. We used lncRNA expression estimation from a recent study repurposing exon array probes to lncRNAs[18] and used the Limma algorithm[36] to identify overexpressed lncRNAs in cancer. In total, 671 lncRNAs were selected and up to 20 pgRNAs were designed for each target. Among the 20 pairs, 10 target the promoter regions, and the other 10 target promoters plus exons.

Positive controls. Positive control genes consist of 20 genes, including 17 ribosomal genes and 3 cancer-related genes, *FOXA1*, *HOXB13* and *EZH2*. We designed 100 pairs for each positive control gene, including 20 targeting promoters (the distance between two gRNAs in each pair is between 200 bp and 5 kb), and 80 targeting promoters plus exons. Among the 80 pairs targeting promoters plus gene bodies, 60 were designed such that their gRNA orientations are consistent with gene orientations. This is because gRNAs with the same orientation of their targeting genes have a better knockout effect than gRNAs with a distinct orientation[37]. The rest 20 pairs were designed to have at least one different orientation with the targeting gene.

Negative controls. We designed 500 pgRNAs of negative controls with three different types. The first type of negative control (100 pairs) consists of pgRNAs that do not target any loci in the human genome. These pgRNAs will be constructed directly from existing non-target control gRNAs from GeCKO v2 library[38]. The second type of control (100 pairs) consists of pgRNAs targeting the AAVS1 region, which is a non-essential region in genome and is frequently used in CRISPR studies for efficiency test. The third type of negative control (300 pairs) consists of pgRNAs targeting the introns of positive control genes.

sgRNA filtering and design. *Target regions*. For positive control genes and lncRNA, the target regions are their promoters and the whole gene bodies. For promoter regions, 5 kb upstream and 200 bp downstream loci of each transcription start site (TSS) were selected as the target regions.

gRNA scanning and filtering. After the regions were selected, we identified all possible gRNAs by searching the PAM motif in the genome sequence. We only kept the gRNAs if (1) their sequences are uniquely mapped to the intended loci, (2) have at least 2 mismatches to any other loci of the genome, and (3) their predicted efficiency scores are above 0.3. The efficiency score prediction was calculated from our recently published machine-learning model[20]. For gRNA pairs targeting lncRNAs, we further required (4) the GC content to be between 0.2 and 0.9, and (5) that the gRNAs not include the UUUU/TTTT polymer. This is because gRNAs with extreme GC content or with UUUU/TTTT sequence have been shown to have lower cleavage efficiency[26,37].

pgRNA design. For all sgRNAs targeting each lncRNA or positive control gene, we first enumerated all possible pgRNAs and then kept pairs that satisfy all of the following conditions:

1. Include one sgRNA before TSS and one after TSS;
2. Do not overlap with any exons of coding regions (for lncRNA targets);
3. Have the same sgRNA orientation as target lncRNA or gene;
4. Are at least 5 kb away from the promoters of coding regions (for lncRNA targets);
5. Are at least 50 bp away from the exon-intron boundary of coding genes (for lncRNAs located inside the introns of another coding gene).

For each lncRNA or gene, if there are not enough pgRNAs, we also included pgRNAs that (1) do not cross over TSS or (2) have different orientation compared with targeted lncRNA or gene. For all pgRNAs that pass the filter mentioned above, we next sought to identify desired number of pgRNAs with barcode (**Fig. 2**), and require the barcode gRNA is used only once in the library. Note that randomly assigning one of the two gRNAs as the barcode may result in some pgRNAs with no available barcode (see **Supplementary Text 1**). Alternatively, we designed an iterative "greedy" algorithm to identify possible pgRNAs and their barcodes, and proved that this algorithm can identify the optimal number of pgRNAs with barcodes (see **Supplementary Text 1**).

The pgRNA design algorithm, "pgRNADesign", is open-source and freely available at https://bitbucket.org/liulab/pgrnadesign. Besides the pgRNA design and barcode assignment, pgRNADesign further allows users to specify a list of "blackout" regions. Once specified, pgRNADesign will avoid designing pgRNAs that overlap with these blackout regions.

Construction of the CRISPR–Cas9 pgRNA library. We created a library targeting 671 lncRNAs with 12,472 pairs of gRNAs as mentioned above (**Supplementary Table 3**). The 137-nt oligonucleotides containing each pairs of pgRNA-coding sequences were designed (**Supplementary Table 10**) and synthesized (CustomArray, Inc.). Then, primers targeting the flanking sequences of oligonucleotides were used for the amplification to create 60-bp homologies with BsmBI digested pgRNA-expressing backbone. The amplified DNA products were ligated into the lentiviral vector using Gibson cloning method[39] and were transformed into Trans1-T1 competent cells (TransGen, Biotech) to obtain the plasmids. Plasmids were then digested by BsmBI and ligated with BsmBI-digested scaffold-linker-U6 fragment (**Supplementary Fig. 2b**), and the ligation mixture was transformed into Trans1-T1 competent cells (TransGen, Biotech) to obtain the final library plasmids (see **Supplementary Text 2** for sequences). The lentivirus of the pgRNA library was produced by co-transfection of library plasmids with two viral packaging plasmids pVSVG and pR8.74 (Addgene, Inc.) into HEK293T cells using the X-tremeGENE HP DNA transfection reagent (Roche). Huh7.5$_{OC}$ cell library was constructed through transduction of low MOI (~0.3) virus, followed by FACS for EGFP$^+$ cells, 72 h after infection.

Recombination rate calculation. The recombination rates were calculated in both plasmid constructs and chromosomal integrations in cells after transduction. For plasmid, we amplified the entire pgRNA sequence from the library plasmid as the template. For chromosomal integrations in cells, the pgRNA sequence was amplified from the genome of library cells as the template. The PCR products were then cloned into vectors for sequencing analysis. 80 and 120 clones were randomly selected from the plasmids and the cell libraries for sequencing, respectively.

CRISPR–Cas9 pgRNA library screening. A total of 1.2×10^7 pgRNA library cells were plated onto 150 mm Petri dishes and three replicates were arranged. The library cells of control group were collected for genomic DNA extraction and that of experimental group were incubated for one month. Then genomic DNA of experimental group was also extracted, followed by PCR amplification of the barcode gRNA-coding regions and deep-sequencing analysis.

Identification of candidate pgRNA sequences and data analysis. The genomic DNA of every replicate was isolated from 4×10^6 cells using the DNeasy Blood and Tissue kit (Qiagen). gRNA-coding regions integrated into the chromosomes were then PCR amplified (TransTaq DNA Polymerase High Fidelity; TransGen) with 28 cycles of reaction using primers targeting U6 promoter and the linker between two gRNAs of each pair (**Supplementary Fig. 2** and **Supplementary Table 11**). In every tube, 0.6 μg of genomic DNA was used as the template and 20 PCR reactions were performed for each replicate. The PCR products of each replicate were pooled together and purified with DNA Clean & Concentrator-25 (Zymo Research Corporation), followed by deep-sequencing analysis (Illumina HiSeq 2500).

The computational analysis of screens. We used the latest version of MAGeCK (0.5.3) we previously developed to analyze the screening data[21]. We used the MAGeCK "count" command to generate read counts of all samples. Briefly, the qualities of fastq files are evaluated using fastqc. If the fastq files are of high quality, then all reads are mapped to the screening library without tolerating any mismatches, and the raw read counts of all pgRNAs of all samples are merged into a count matrix. The distribution of read counts is reported in **Supplementary Figure 3c**, and the correlations between samples are reported in **Supplementary Figure 3a**,**b**.

We next used MAGeCK "test" command to identify the top negatively and positively selected lncRNAs. The MAGeCK algorithm consists of four steps: normalization, pgRNA mean-variance modeling, pgRNA ranking and lncRNA ranking. In the normalization step, MAGeCK adjusts the effect of sequencing depth of all samples by calculating a size factor for each sample. The factor is estimated from the "median ratio normalization" approach described before[40]. Instead of calculating the size factor from all pgRNAs (the default normalization method for MAGeCK), we estimated the size factor from all AAVS1 targeting pgRNAs ("AAVS1 normalization"), since AAVS1 normalization provides a more realistic estimation about the log fold-change distribution of the negative control pgRNAs (**Supplementary Fig. 15b**). In the mean-variance modeling step, MAGeCK estimates the mean and variance of every pgRNA across independent experimental replicates, and fits a linear regression model to better estimate variances based on the mean of pgRNA counts. In the sgRNA-ranking step, MAGeCK estimates the *P* value of every pgRNA based on the negative binomial (NB) model of read counts. The parameters of the NB distribution are estimated from the mean-variance model built in previous step. In the final lncRNA ranking step, MAGeCK estimates the level of negative (or negative) selection of each lncRNA by comparing the rankings of all pgRNAs targeting that lncRNA with a null model (where all pgRNAs are distributed uniformly in the ranked list). MAGeCK uses a α-Robust Rank Aggregation (α-RRA) algorithm to calculate the "RRA score" of each lncRNA, a score to describe the degree of negative (or positive) selection. The *P* value of the RRA score is calculated by permuting all pgRNAs, and the adjusted *P* values are obtained from the Benjamini–Hochberg method. To increase the statistical power, we filter lncRNAs that have fewer than 2 statistical significant pgRNAs, and only perform multiple comparison *P*-value correction on the remaining lncRNAs. A detailed description of the algorithm can be found in the original study[21].

Cell proliferation assay. All the pgRNAs targeting the positive control gene and lncRNAs to be validated were cloned into a lentiviral expressing backbone carrying CMV promoter-driven EGFP, and were delivered into cells through transduction. The percentage of $EGFP^+$ cells was quantified by FACS. The first quantification started from three days post viral infection, labeled as day 0, serving as control for normalization. Cell viability was determined by normalizing $EGFP^+$ percentages at indicated time points with day 0 control.

Cell lethality assay. All the pgRNAs targeting negatively selected lncRNAs were delivered into $Huh7.5_{OC}$ cells through lentiviral infection and all the sgRNAs that were designed to repress or activate the transcription level of lncRNAs were delivered into Huh7.5 cells through transient transfection. The cells were conducted with FACS enrichment 72 h after infection or transfection, and the LDH lethality assay were performed from one day to three days post FACS. LDH staining and detection were performed as described in the product instruction (CytTox96; Promega). The death signal represented by the amount of LDH release was normalized to the wells based on the maximum LDH activity of the total lysed cells. Each data point and related error bar shown in the figures represent the average results from three replicates.

CRISPR inhibitor and CRISPR activator. For CRISPRi, the KRAB–dCas9-P2A-mCherry (Addgene; #60954) plasmid was delivered into Huh7.5 cells through lentivirus infection. And the mCherry-positive cells were enriched by FACS 3 d after infection. Then, the sgRNAs targeting the negatively selected lncRNAs were delivered into cells with stable expressing of dCas9–KRAB by lentivirus infection followed by cell proliferation assay and cell lethality assay. For CRISPRa, the three plasmids dCAS–VP64_Blast (Addgene; # 61425), MS2–P65–HSF1_Hygro (Addgene # 61426) and sgRNAs carrying EGFP for each positively selected lncRNAs were delivered into cells through transient transfection. Then the EGFP-positive cells were enriched by FACS 3 days after transfection followed by cell lethality assay.

Real-time PCR. RNA of cultured cells was extracted using RNAprep Pure Micro kit (TIANGEN; DP420), and the cDNA was synthesized using QuantScript RT kit (TIANGEN; KR103-03). Real-time PCR was performed with SYBR Premix Ex TaqII (TaKaRa; RR820A) on LightCycler96 qPCR system. And GAPDH transcript levels were measured as normalized controls.

RNA sequencing and data analysis. LINC01087 targeting pgRNAs (LINC01087_p1, LINC01087_p2 and LINC01087_p4) were delivered into Huh7.5 cells through lentivirus infection. The EGFP-positive cells were enriched by FACS three days after infection and cultured for another nine days. All the samples were harvested using RNAprep Pure Micro kit (TIANGEN; DP420) and deep sequenced on the Illumina Hiseq 4000 platform. RNA-seq reads are mapped to the human reference genome (hg19) using Tophat2 (ref. 41). The read counts of genes are collected using HTSeq[42], and the differential expression analysis is performed using DESeq2 (ref. 43).

Functional analysis of lncRNAs. We collected the expression data of genes and lncRNAs from five different cancer types: liver, prostate, lung, ovarian and brain. The expression levels from prostate, lung, ovarian and brain cancers were downloaded from our previous study[18]. In this study, gene expressions were measured from human exon arrays, and lncRNA expressions were measured by repurposing some of the probes to lncRNAs. The expression profiles include 150 tumor samples from MSKCC Prostate Oncogenome Project[44], 451 samples from glioblastoma multiforme (GBM)[45], 585 samples from ovarian cancer[46], and 113 samples from lung squamous cell carcinoma from The Cancer Genome Atlas Research (TCGA) project[47]. We downloaded the RNA-seq expression profiles of liver cancer patients in TCGA from the TANRIC database, an integrative platform to explore the lncRNA functions[48]. For the expression profiles of liver cancer cell lines, we downloaded the RNA-seq data of 32 liver cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE)[31,49]. RNA-seq reads were obtained from UCSC Cancer Genomics Hub (http://cghub.ucsc.edu) and mapped to the human reference genome (hg19) using Tophat2 (ref. 41), and the expressions of genes and lncRNAs were calculated using Cufflinks[50]. For gene ontology (GO) analysis, we calculated the expression correlations of all coding genes for each lncRNA, chose genes with top 10% highest positive correlation, and used the topGO R package to estimate the statistical significance of enriched GO terms[51].

36. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
37. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
38. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
39. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361 (2011).
40. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
41. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
42. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
43. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
44. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
45. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
46. Partensky, F. & Garczarek, L. Microbiology: arms race in a drop of sea water. *Nature* **474**, 582–583 (2011).
47. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
48. Li, J. *et al.* TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res.* **75**, 3728–3737 (2015).
49. Wilks, C. *et al.* The cancer genomics hub (CGHub): overcoming cancer through the power of torrential data. *Database (Oxford)* **2014**, bau093 (2014).
50. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
51. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

doi:10.1038/nbt.3715

Corrigendum: Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library

Shiyou Zhu, Wei Li, Jingze Liu, Chen-Hao Chen, Qi Liao, Ping Xu, Han Xu, Tengfei Xiao, Zhongzheng Cao, Jingyu Peng, Pengfei Yuan, Myles Brown, Xiaole Shirley Liu & Wensheng Wei

***Nat. Biotechnol.*; doi:10.1038/nbt.3715; corrected online 9 November 2016**

In the version of this article initially published online, genomic deletion values (listed in base pairs) in Figure 1e were incorrect: in the second line, the value given as 3,476 should have been 3,511; third line, 3,479 should have been 3,514; fourth line, 3,475 should have been 3,510; fifth line, 3,484 should have been 3,519. The errors have been corrected for the print, PDF and HTML versions of this article.