nature biotechnology

Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites

Ying Liu[1,2,4], Zhongzheng Cao[1,2,4], Yinan Wang[1,2,4], Yu Guo[1,3,4], Ping Xu[1], Pengfei Yuan[1,3], Zhiheng Liu[1,2], Yuan He[1] & Wensheng Wei[1]

The functions of many long noncoding RNAs (lncRNAs) in the human genome remain unknown owing to the lack of scalable loss-of-function screening tools. We previously used pairs of CRISPR–Cas9 (refs. 1–3) single guide RNAs (sgRNAs) for small-scale functional screening of lncRNAs[4]. Here we demonstrate genome-wide screening of lncRNA function using sgRNAs to target splice sites and achieve exon skipping or intron retention. Splice-site targeting outperformed a conventional CRISPR library in a negative selection screen targeting 79 ribosomal genes. Using a genome-scale library of splicing-targeting sgRNAs, we performed a screen covering 10,996 lncRNAs and identified 230 that are essential for cellular growth of chronic myeloid leukemia K562 cells. Screening GM12878 lymphoblastoid cells and HeLa cells with the same library identified cell-type-specific differences in lncRNA essentiality. Extensive validation confirmed the robustness of our approach.

The CRISPR–Cas9 system has been harnessed to identify gene functions in large-scale screens[5–8]. Most commonly Cas9 perturbs gene function through frameshift mutations generated within exons. Frameshift mutations are only effective when targeting protein-coding genes, but these account for only 2% of the human genome. Increasing evidence suggests, however, that many transcripts do not encode proteins but function as noncoding RNAs[9]. Among them, lncRNAs with more than 200 nucleotides represent a large subgroup without apparent protein-coding potential[10,11]. Previous studies indicate that the total number of human lncRNAs outstrips that of protein-coding genes, as the number of identified lncRNAs continues to climb[11]. lncRNAs play critical roles in diverse cellular processes at the transcriptional or post-transcriptional level by *cis*- or *trans*-regulating gene expression[12]. However, most of them are not functionally characterized.

Because lncRNAs are generally insensitive to reading frame alterations, it is difficult to apply the CRISPR–Cas9 system in a conventional way to disrupt their expressions. We have previously developed a deletion strategy using a paired guide RNA (pgRNA) library for the loss-of-function screens of lncRNAs[4], but it is laborious to scale up. Although screens based on RNA interference[13,14] or CRISPRi[15] proved effective for the functional identifications of lncRNAs, RNAi methods suffer from potential off-target effects[16], and both approaches are limited by the effectiveness of transcript knockdown.

We noticed in our previous study that an sgRNA targeting a splice site of the *CSPG4* gene was capable of disrupting its expression without changing the gene's reading frame[17]. Targeting lncRNA splice sites with morpholino antisense oligonucleotides has also been used to successfully disrupt the maturation of individual lncRNA[18]. Here, we show that sgRNAs targeting splice sites efficiently cause exon skipping or intron retention, and that such effects could be exploited to perturb noncoding RNA function in a large-scale fashion.

It has been reported that the intronic sequences in various species are flanked by an almost invariant GT at the 5′ splice donor (SD) site and AG at the 3′ splice acceptor (SA) site[19]. Using Weblogo3 tools[20], we confirmed that about 99% of intronic regions in the human genome are flanked by GT and AG (**Fig. 1a**). Notably, AG sequences are predominantly present as the last two bases of exons just upstream of the SD sites. To verify the effectiveness of sgRNAs in producing exon skipping or intron retention, we designed sgRNAs targeting either SD or SA sites of two ribosomal genes, *RPL18* and *RPL11*, both of which are indispensable for cell growth and proliferation[21]. In HeLa cells stably expressing *Cas9* and *OCT1* genes[8,22–24], $sgRNA1_{RPL18}$ targeting an SD site and $sgRNA2_{RPL18}$ targeting an SA site successfully generated intron 3 retention and exon 4 skipping, respectively, at the *RPL18* locus in the genome, which were confirmed by both reverse transcription–PCR (RT-PCR) and Sanger sequencing (**Fig. 1b**,**c** and **Supplementary Fig. 1a**). The same results were obtained for the *RPL11* gene, in which $sgRNA3_{RPL11}$ and $sgRNA4_{RPL11}$ produced intron 2 retention and exon 4 skipping, respectively, at the *RPL11* locus (**Supplementary Fig. 1b–d**).

To further assess the power of targeting splicing in a CRISPR screen, we designed an sgRNA library targeting splice sites of 79 ribosomal genes, most of which are essential for cellular growth in various cell lines[21]. This library contained 5,788 sgRNAs, whose cutting sites are tiled within −50 bp to +75 bp surrounding every 5′ SD site and −75 bp to +50 bp surrounding every 3′ SA site of these 79 genes (**Supplementary Table 1**). The cell libraries harboring these

[1]Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [3]Present address: EdiGene Inc, Beijing, China. [4]These authors contributed equally to this work. Correspondence should be addressed to W.W. (wswei@pku.edu.cn).

Received 26 September 2017; accepted 27 August 2018; published online 5 November 2018; doi:10.1038/nbt.4283

sgRNAs were constructed through lentiviral delivery at a multiplicity of infection of <0.3 in Cas9-expressing HeLa and Huh7.5 cells[8,22]. The cells were cultured for 15 d, and the sgRNAs leading to reduced cell viability were identified by NGS analysis.

By calculating the $\log_2$ fold change of sgRNAs between 15-d experimental and control samples, we ranked all sgRNAs. The Spearman correlation between the two biologically independent samples of control and experimental in both HeLa and Huh7.5 hepatocellular carcinoma cells showed that all results were highly reproducible (**Supplementary Fig. 2**). To determine the effectiveness of gene disruption brought about by splicing targeting, we merged all SD site-targeting data and SA site-targeting data and arranged them according to their physical distances from to SD or SA sites (**Fig. 1d** and **Supplementary Table 2**). sgRNAs affecting splice sites outperformed those targeting only exonic regions in both HeLa and Huh7.5 cells. The closer the distances from sgRNAs' cutting sites to splice sites, the better their effects on gene disruption, with peak points slightly toward the exons in the case of both SD and SA sites (**Fig. 1d**). In comparison, the vast majority of sgRNAs targeting introns were rarely depleted throughout the screens, suggesting that they had little effects on gene function and consequently on cell viability. The only exceptions were those sgRNAs targeting intronic regions close to SA sites, which include branch points followed by polypyrimidine tracts, known for their involvement in RNA splicing[25,26].

As the numbers of sgRNAs designed for different loci were not equal, we compared the percentages of highly efficient (over fourfold dropout) sgRNAs at every locus for a fair comparison. With such normalization, we further confirmed that both SD- and SA-targeting sgRNAs were substantially superior to those targeting only exonic regions (**Supplementary Fig. 3a**). To better quantify our results, we classified all sgRNAs into three categories: intron-targeting (with cutting sites of sgRNAs within introns and at least 30 bp away from SD or SA sites), exon-targeting (with cutting sites of sgRNAs within exons and at least 30 bp away from SD or SA sites), and splicing-targeting (with cutting sites of sgRNAs between −10 bp and +10 bp flanking SD or SA sites; − and + refer to the intronic and exonic direction, respectively). In both HeLa and Huh7.5 cells, the percentages of sgRNAs leading to over two- or fourfold dropouts were 2–25 times higher in splicing-targeting than the other two categories (**Supplementary Fig. 3b**,**c**). We also compared the performance of sgRNAs targeting the 3′ splice sites and failed to observe any correlations between the gene disruption effects and whether the sizes of the immediate downstream exons were integral multiples of 3 (**Supplementary Fig. 4a**). As the average lengths of exons that are integral multiple of 3 in these ribosomal genes are more than 100 bp (**Supplementary Fig. 4b**), we assume that such deletions in essential ribosomal gene are likely sufficient to cause cell death or decrease cell proliferation, at least in most cases. In addition, it appeared more effective to target splice sites toward N termini than C termini, similarly to the performance of sgRNAs targeting protein-coding regions (**Supplementary Fig. 4c**). Culturing the cells for longer times did not change this picture, as no differences between 22 and 15 d were observed for either 5′ SD- or 3′ SA-site targeting, in HeLa or Huh7.5 cell lines (**Supplementary Fig. 5**).

As RNA splicing is a conserved mechanism for both coding genes and noncoding RNAs, we constructed a splice-site-targeting sgRNA library for genome-scale functional screening of lncRNAs. Among the 14,470 lncRNAs retrieved from GENCODE dataset V20, we first filtered out 2,477 lacking splice sites. We also imposed several other rules: for instance, all sgRNAs' cutting sites should be within −10 bp to +10 bp surrounding splice sites, and sgRNAs should be predicted to have high cleavage activity[21,27,28] without off-target effect on any known essential gene. We ultimately generated a library containing 126,773 sgRNAs targeting 10,996 unique lncRNAs. Together with 500 nontargeting control sgRNAs and 350 sgRNAs targeting essential ribosomal genes, we constructed the cell library in K562 cells engineered to stably express Cas9 protein (**Fig. 2a** and **Supplementary Table 3**) by lentiviral transduction at a low multiplicity of infection of <0.3. We continued to culture the library cells for 30 d after infection to screen for lncRNAs affecting cell growth and proliferation. sgRNA dropout was subsequently detected by next-generation sequencing[4,8].

The read distributions of the two biologically independent samples of the control library showed a high level of correlation (**Supplementary Fig. 6a**). After 30 d of culture, sgRNAs targeting lncRNAs and essential genes were both depleted compared with the nontargeting sgRNAs (**Fig. 2b**,**c**), indicating an effects on cell viability or proliferation. For each lncRNA, we computed the fold changes of sgRNAs and obtained their *P* values by comparing with nontargeting sgRNAs through a Wilcoxon test. We randomly sampled nontargeting sgRNAs to generate 'negative control genes', thus correcting the lncRNA genes' *P* values by their distribution. For each lncRNA, a screen score was computed by combining the mean fold change and corrected *P* values (Online Methods, **Supplementary Fig. 6b** and **Supplementary Tables 4**–**6**). A total of 230 lncRNA candidates whose depletion would lead to cell growth inhibition or cell death in the K562 line were selected with a threshold of the screen score of 2 (**Fig. 2d**). All 36 essential control genes were significantly enriched in the ranking list of negatively selected genes, confirming the reliability of the screening approach and the data analysis method.

From the negatively selected lncRNAs whose corresponding sgRNAs were consistently depleted in two replicates (for example, **Supplementary Fig. 7**), we chose the 35 top-ranked lncRNA genes for further validation. For each candidate, we cloned the two top-ranked sgRNAs obtained from the library screen into a lentiviral backbone with an EGFP selection marker. A nontargeting sgRNA and a sgRNA targeting the nonfunctional adeno-associated virus integration site 1 (*AAVS1*) locus were chosen as negative controls, and an sgRNA targeting the ribosomal gene *RPL18* was also included as the positive control (**Fig. 3a**). Each sgRNA was transduced into K562 cells, and cell proliferation was quantified through the percentage change in EGFP-positive cells. All sgRNAs targeting the 35 top-ranked lncRNA loci effectively led to the inhibition of cell proliferation in K562 cells (**Fig. 3b**,**c** and **Supplementary Figs. 8** and **9**). Furthermore, we chose the pgRNA-mediated deletion method[4] to independently investigate the roles of lncRNA hits from our screen. We selected 8 lncRNAs from the validated 35 hits and another 6 candidates from the top hits that were not included in the above validation. Four pair of guide RNAs were designed for each lncRNA to target its promoter and first exon (Online Methods). The *AAVS1* locus or ribosomal genes *RPL19* and *RPL23A* were chosen for pgRNA targeting as negative control or positive controls, respectively (**Fig. 3d**). Through the cell proliferation assay, the essentiality of all 14 lncRNAs was validated by deletion (**Fig. 3e** and **Supplementary Figs. 10** and **11**). Validation results from targeting splicing correlated well with those from the deletion strategy (correlation coefficient 0.93, $P = 0.002$) (**Fig. 3f**). This indicates that targeting splicing is a reliable and robust approach for lncRNA gene disruption, as all 41 lncRNA hits chosen for further validation were confirmed to be critically important for K562 cell growth and proliferation.

The cell type specificity of lncRNA function has been previously reported[1,15]. To further explore the difference in lncRNA functions between cancer and normal cells, we performed the splicing-targeting

screen in the lymphoblastoid cell line GM12878 as well as the HeLa cell line. GM12878 cells have a relatively normal karyotype and belong to tier 1 ENCODE cell lines, as do K562 cells[29,30]. Two hundred twenty lncRNA candidates were negatively selected in GM12878 and 115 were negatively selected in HeLa (**Fig. 4a**,**b** and **Supplementary Fig. 12a–d**). Only 20 lncRNAs affected cell growth and proliferation

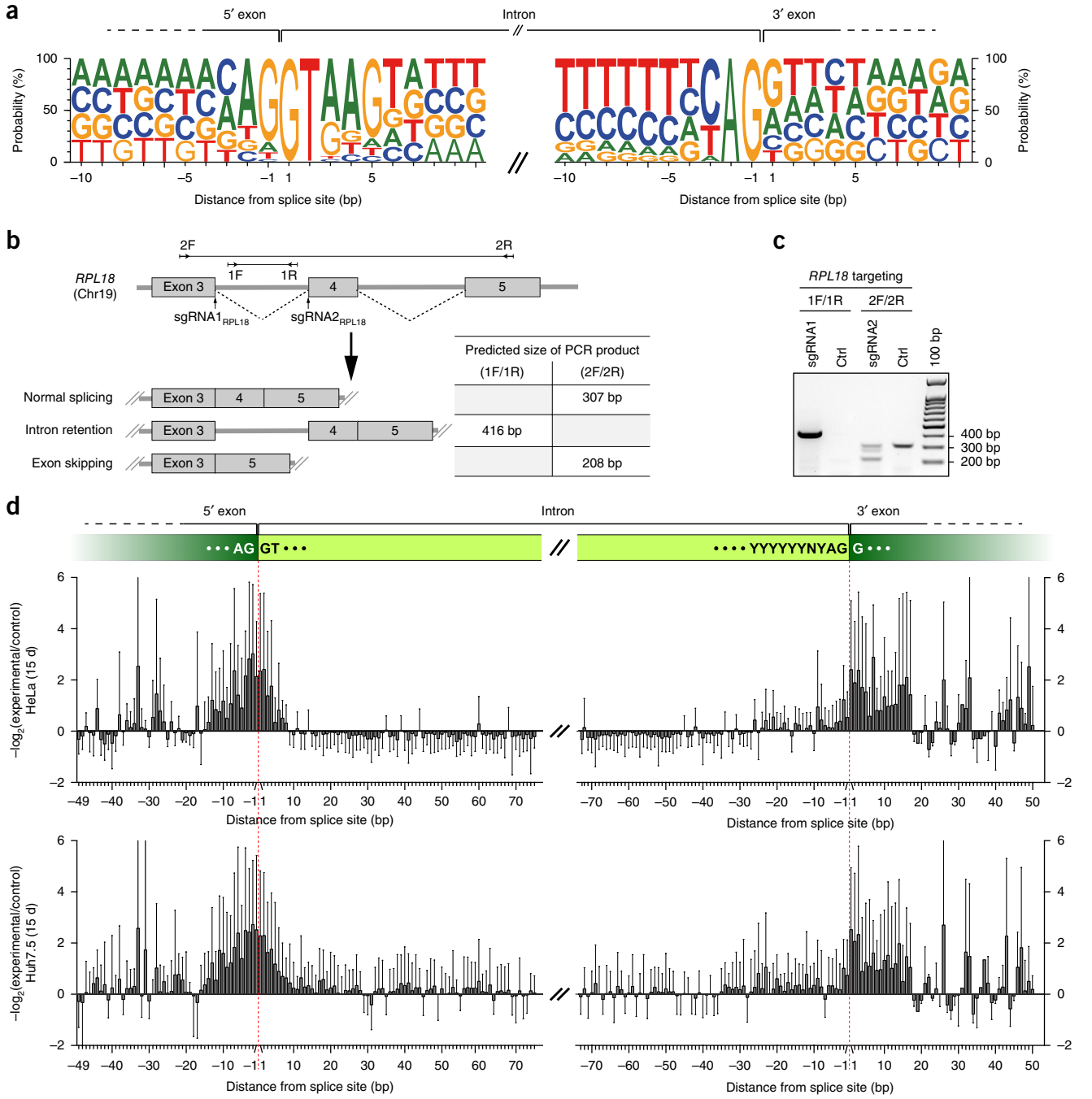

Figure 1 Lentivirally delivered sgRNAs generate intron retention or exon skipping by disrupting splice sites. (**a**) Genomic sequence features and base specificity of splice sites in eukaryotes *Homo sapiens*. ~270,000 intronic sequences covering all protein coding genes of *H. sapiens* from the hg38 reference genome were analyzed. The vertical axis indicates the probability of bases at each locus. (**b**) Schematic of intron retention or exon skipping induced by sgRNAs targeting the SD or SA site of *RPL18* in HeLa cells. (**c**) RT-PCR analysis of intron retention or exon skipping induced by lentivirally delivered sgRNAs in *RPL18*. Primer pairs L1/R1 and L2/R2 were chosen for RT-PCR amplification at the indicated locus, as labeled in **b**. All infected HeLa cells targeted by different sgRNAs were sorted by fluorescence-activated cell sorting 72 h after infection (Online Methods). Wild-type HeLa cells without sgRNA transduction were set as the control (Ctrl). The experiments were performed twice with similar results. (**d**) Deep sequencing analysis of CRISPR screen of the sgRNA library targeting ribosomal genes in HeLa and Huh7.5 cell lines. The sgRNA saturation mutagenesis library was designed to target regions from –50 bp to +75 bp surrounding 5′ SD sites and from –75 bp to +50 bp surrounding 3′ SA sites of 79 ribosomal genes. The pooled plasmid library was lentivirally transduced into HeLa and Huh7.5 cells expressing Cas9 protein. The dropouts of all sgRNAs at indicated loci were calculated and averaged as $-\log_2$(experimental/control) of the normalized read counts ($n = 2$ biologically independent experiments). The black bars represent the mean fold changes of all sgRNAs at their loci. The red dotted lines indicate the positions of splice sites. Data are presented as mean ± s.d. (n is all sgRNAs at their indicated loci).

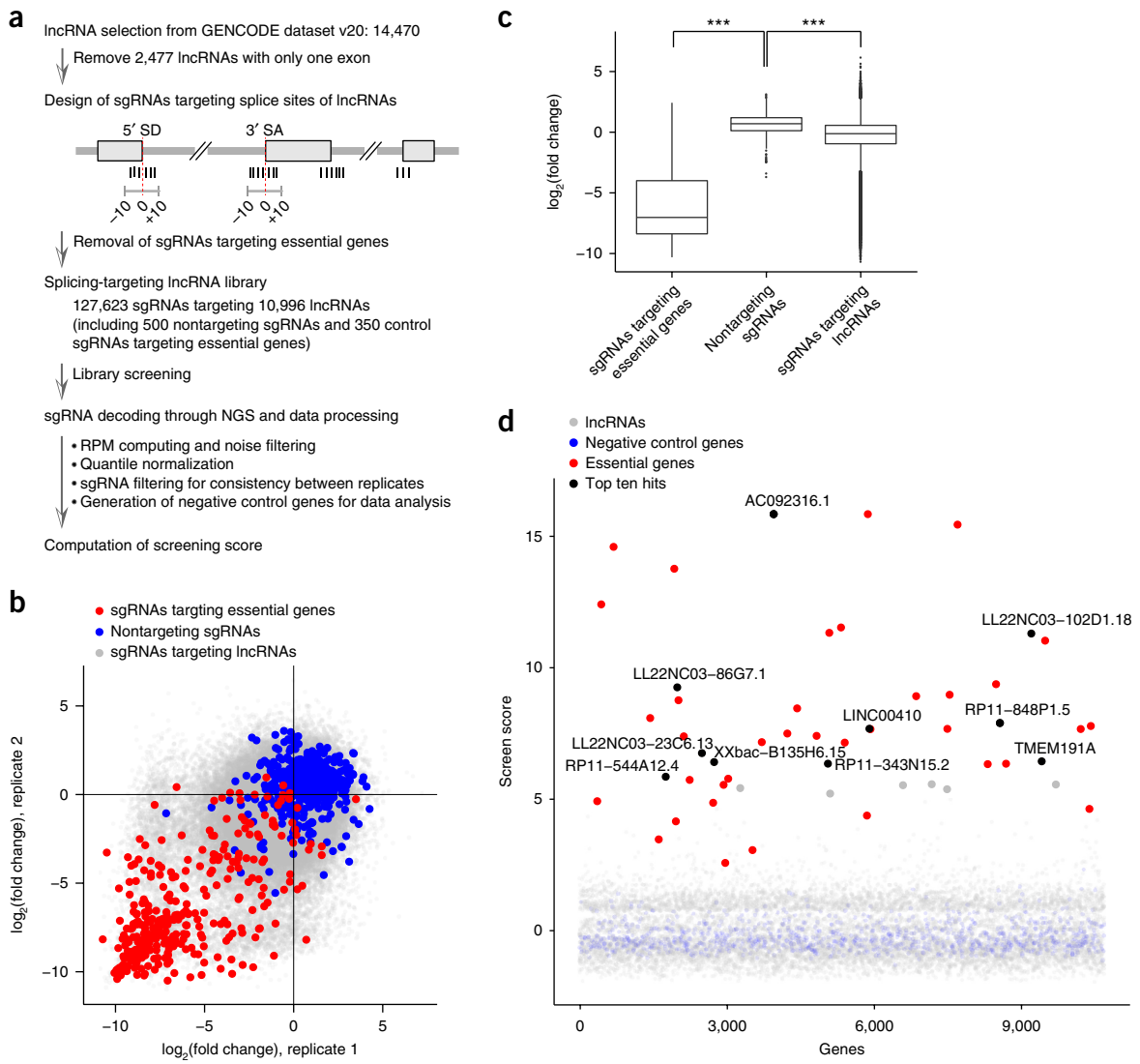

Figure 2 Splicing-targeting enables genome-scale screening for the identification of lncRNAs essential for cell growth and proliferation. (**a**) The workflow of splicing-targeting sgRNA library construction, screening and data analysis. (**b**) Scatter plot of sgRNA fold change between two biologically independent experiments. (**c**) The $\log_2$(fold change) distribution of nontargeting sgRNAs, sgRNAs targeting essential genes and lncRNAs. The fold change of each group was respectively compared with that of nontargeting sgRNAs by two-sided Student *t*-test ($n = 2$ biologically independent experiments). $^{***}P < 0.001$. In the box plot, center lines represent median values, box limits represent the interquartile range, whiskers extend 1.5 times the interquartile range and dots represent outliers. (**d**) Screen scores of negatively selected lncRNAs by splicing-targeting CRISPR screening. For each lncRNA, the fold changes of all targeting sgRNAs were compared with negative control sgRNAs by Wilcox test and the generated *P*-value was further corrected by the null distribution of negative control genes (blue), which were obtained by randomly sampling negative control sgRNAs. The screen score was calculated from the mean fold change and corrected *P*-value ($n = 2$ biologically independent experiments) (Online Methods). The top ten lncRNA hits and negatively selected essential genes are labeled in green and red, respectively.

in all three cell lines (**Fig. 4c**). Among these 20, 6 were further tested and successfully validated in both K562 and GM12878 cells. Of the lncRNAs previously validated in K562 cells, 18 appeared essential for the growth of GM12878 cells as well (**Supplementary Fig. 8**), while 6 and 11 showed weak or no detectable effects, respectively, on cell viability in GM12878 (**Fig. 4d**,**e** and **Supplementary Fig. 9**).

We further analyzed the correlation between expression levels and the screen scores of those lncRNAs in each cell line and found that lncRNAs with higher expression tended to have higher screen scores (**Supplementary Fig. 12e**–**g**). Nevertheless, some lncRNAs showing low expression levels in RNA sequencing (RNA-seq) analysis (between 0 and 10 transcripts per million) had high screen scores (**Supplementary Fig. 12e**–**g**). It remains to be determined whether these lncRNAs are indeed poorly expressed[31] or possess no poly(A) tails[32,33].

To better understand the mechanisms leading to these varied phenotypes in K562 and GM12878 cells, we further explored the functions of lncRNA *BMS1P20*, which was essential for cell viability only in K562 but not in GM12878 (**Fig. 4d**). We performed RNA-seq analysis of both K562 and GM12878 cells, with and without *BMS1P20* knockout with two validated sgRNAs targeting its splice sites (**Fig. 4e**). The expression levels of the top 500 genes showing variance between control and sgRNA-targeting samples in each cell line were evaluated,

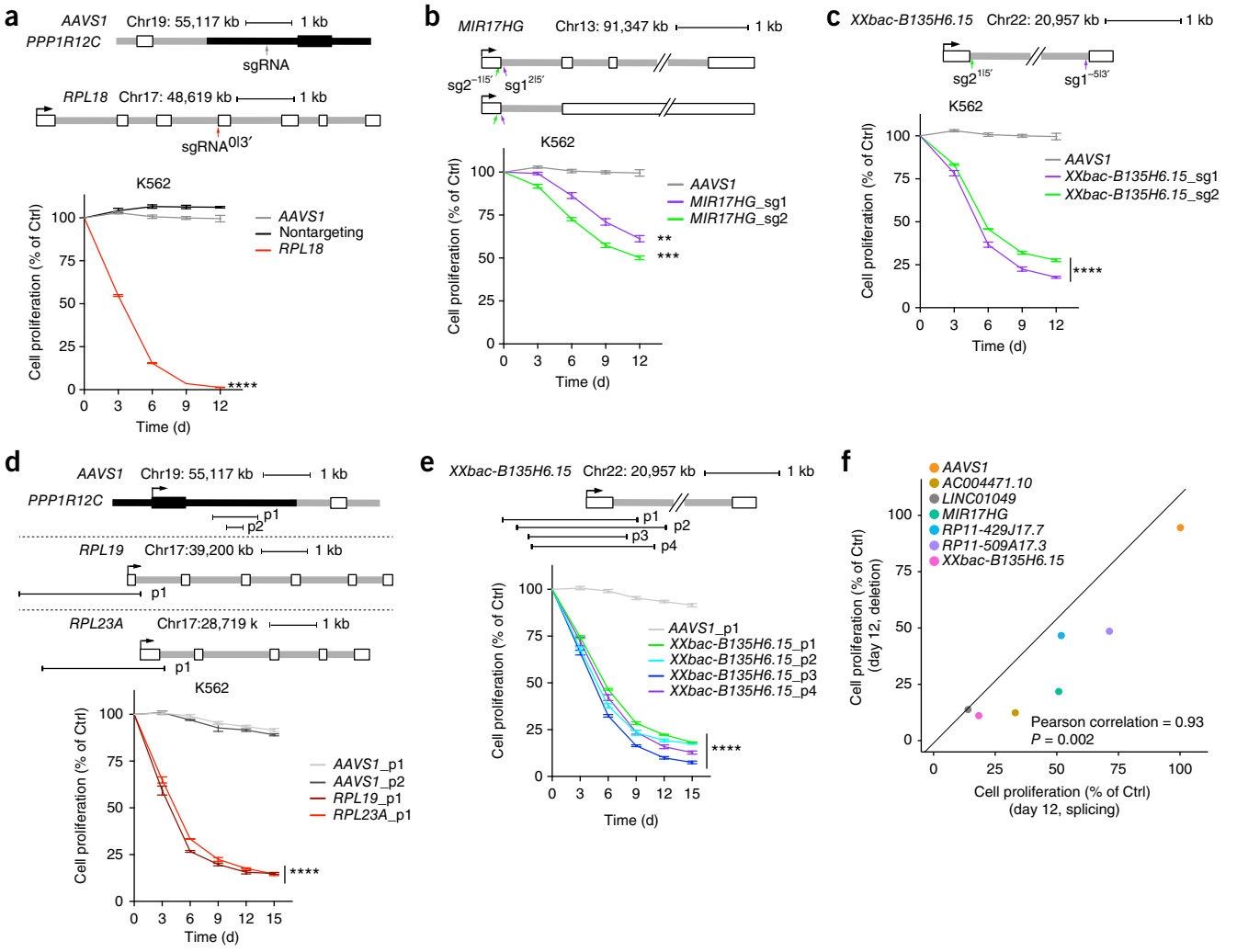

Figure 3 Validation of candidate lncRNAs. (**a**–**c**) Effects of indicated sgRNAs on cell proliferation in K562 cells, which include three kinds of control sgRNAs, nontargeting sgRNA, sgRNA targeting *AAVS1*, sgRNA targeting the splice site of *RPL18*—an essential gene for cell growth (**a**)—and two negatively selected lncRNAs (**b**,**c**). The notation sgRNA$^{0|3'}$ indicates that this sgRNA targets the 3′ splice acceptor site and the distance between the sgRNA's cutting site and the corresponding splice site is 0 bp. Each lentivirus of the sgRNA expression vector harboring a CMV-promoter-driven EGFP marker was transduced into K562 cells. The percentage of EGFP-positive cells was measured every 3 d by fluorescence-activated cell sorting, indicating the fraction of sgRNA-infected cells. The first cell sorting analysis started 3 d after infection (labeled as day 0), and then the pooled cells were passaged for 12 d. Cell proliferation of each sample was determined by dividing the percentages of EGFP-positive cells at the indicated time points by that at day 0. Data are presented as the mean and s.d. of three biologically independent experiments. *P* values represent comparisons with sgRNA targeting AAVS1 at the assay end point (day 12), calculated using a two-tailed Student's *t*-test and adjusted using the Benjamini–Hochberg method. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. (**d**,**e**) Cell proliferation assay performed by large-fragment deletions of the *AAVS1* locus, essential genes *RPL19* and *RPL23A* (**d**), and lncRNA *XXbac-B135H6.15* (**e**) in K562 cells. Two pgRNAs were designed for *AAVS1*, and one pair was designed for each essential gene to delete the promoter and the first exon. For *XXbac-B135H6.15*, four pgRNAs were designed to delete its promoter and first exon. The pgRNA for each locus was labeled p1, p2, etc. The pgRNAs were expressed from the backbone containing the EGFP marker, and the cell proliferation assay was performed as in **a**–**c** and as previously described[4]. Data are presented as the mean and s.d. of three biologically independent experiments. Asterisks represent *P* values compared with AAVS1_p1 at day 15, calculated using two-tailed Student's *t*-test and adjusted using the Benjamini–Hochberg method: $^{****}P < 0.0001$. (**f**) The correlations of knockout effects on top lncRNA candidates between splicing-targeting and pgRNA-mediated deletion methods. The knockout effect in each condition was shown as the normalized percentage of sgRNA/pgRNA-infected cells at day 12, which is the mean effect of the two splicing-targeting sgRNAs or four pgRNAs for each lncRNA. The *P*-value was computed by two-sided Student's *t*-test. All sgRNA and pgRNA sequences used for individual validation are listed in **Supplementary Table 7**.

and different expression patterns were observed in the two lines after knocking out the lncRNA (**Fig. 4f**). The two sgRNAs targeting the same splice site resulted in similar changes in expression patterns (**Supplementary Fig. 13a**,**b**). In the K562 cell line, changing the splicing pattern of *BMS1P20* downregulated 178 known essential genes[21] ($P = 0.05$, **Fig. 4g**), suggesting possible mechanisms by which this lncRNA affects the growth of K562 cells. These essential genes were enriched in several essential pathways, such as regulation of translational initiation, cell division and DNA repair (**Supplementary Fig. 13c**). We found that disruption of *BMS1P20* up- or downregulated the expression of a series of protein-coding genes in both K562 and GM12878 cells (**Supplementary Fig. 13d**,**e**). We further

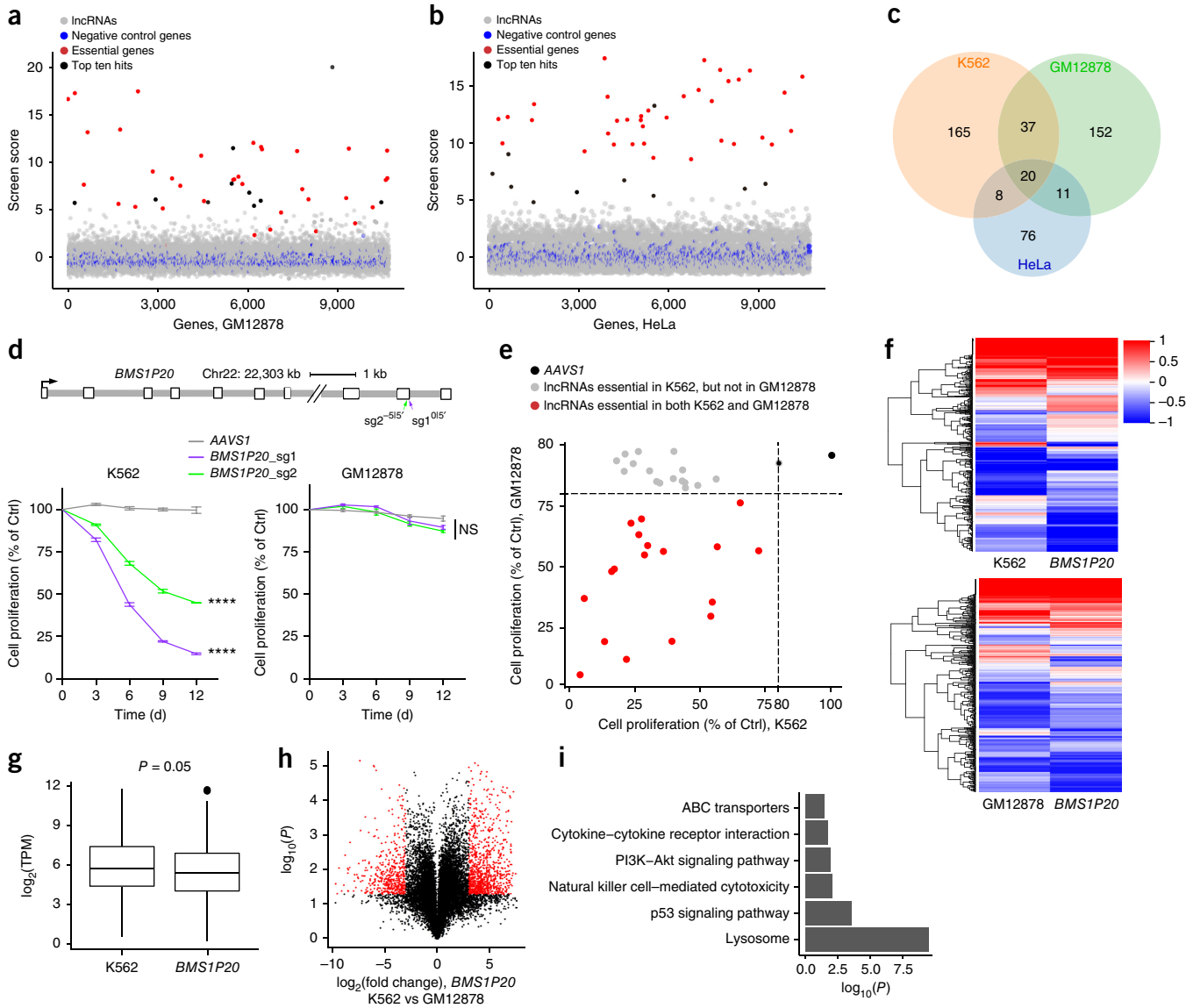

Figure 4 Cell type specificity of lncRNA function across multiple cell lines. (**a**,**b**) Screen scores of negatively selected lncRNAs by splicing-targeting CRISPR screening in GM12878 (**a**) and HeLa (**b**) cell lines. The analysis was performed as in **Figure 2d**. The top ten lncRNA hits and negatively selected essential genes are labeled in green and red, respectively. (**c**) Venn diagram of negatively selected lncRNA candidates in K562, GM12878 and HeLa cell lines. (**d**) Effects of the indicated sgRNAs on cell proliferation in K562 and GM12878 cells. Cell proliferation assay and data analysis as in **Figure 3a–c**. $^{****}P < 0.0001$; NS, not significant. The sgRNA sequences used for individual validation are listed in **Supplementary Table 7**. (**e**) Cell proliferation of 35 top candidate lncRNAs in K562 cells compared with that in GM12878 cells by splicing-targeting strategy. The threshold was set at 80%, which was calculated as the normalized percentage of sgRNA-infected cells at day 12. Gray dots indicate lncRNAs essential only in K562 cells and red dots indicate those exhibiting growth phenotypes in both K562 and GM12878 cells. (**f**) Expression patterns of the top 500 genes showing the highest variance across *BMS1P20* knockout cells and their corresponding controls. (**g**) The expression levels of downregulated essential genes in *BMS1P20* knockout cells compared with the wild-type K562 cells, expressed in $\log_2$ of transcripts per million (TPM). The *P*-value is computed by two-sided Student's *t*-test. In the box plot, center lines represent median values, box limits represent the interquartile range, whiskers extend 1.5 times the interquartile range and dots represent outliers. (**h**) Volcano plots for differential expression following infection of splicing-targeting sgRNAs for *BMS1P20* in K562 cells compared with GM12878 cells. Black and red dots represent all genes and differentially expressed genes, respectively. The *P*-value is computed by two-sided Student's *t*-test. (**i**) Gene Ontology (GO) terms and KEGG annotations of genes that were downregulated (red dots in **h**) in K562 cells. The *P*-value is computed by one-sided Fisher's exact test. In **g**–**i**, the expression levels were calculated from two biologically independent samples of each condition for RNA-seq analysis.

investigated the differentially expressed genes after knocking out this lncRNA in K562 versus in GM12878 (**Fig. 4h**). These downregulated genes in K562 were enriched in processes such as p53 signaling pathway and PI3K–Akt signaling pathway, which might affect cell growth and proliferation (**Fig. 4i**). There were also upregulated genes (**Supplementary Fig. 13f**), and these differentially expressed genes all contributed to the phenotypic difference of *BMS1P20* knockouts in cell growth between these two cell lines.

In sum, splicing-targeting provides an alternative to generating frameshift mutations in protein-coding genes and is applicable to all

transcripts that undergo splicing. This feature is essential for knocking out reading-frame-insensitive noncoding RNAs. In addition, this strategy could be also useful when it is difficult to design sgRNAs targeting genes with conserved coding sequences.

Previously, pgRNA-mediated deletion[4] and CRISPRi[15] have been applied to identify functional lncRNAs. Although it is technically easier to scale up using a CRISPRi strategy than pgRNA-mediated genomic deletion, CRISPRi as well as CRISPR activation (CRISPRa) methods generally act within a 1-kb window around the targeted transcriptional start site[15,34]; by this criterion, one would risk affecting expression of neighboring genes inadvertently for nearly 60% of lncRNA loci[35]. In the CRISPRi screening by Liu *et al.*[15], of the 144 lncRNAs identified in K562 cells, 79 neighbored essential coding genes[36], making it difficult to determine whether the observed phenotypes were due to lncRNA knockdown or the inhibition of neighboring genes. We also investigated the remaining 65 lncRNA hits identified by the CRISPRi study[15] in K562, only 31 of which were included in our library owing to difference of database and lncRNA annotations. Two out of these 31 hits appeared essential for cell viability in our splicing-targeting screening. To better understand the marked difference in results between these two approaches, we selected 5 lncRNAs from the remaining 29 hits that were only identified by CRISPRi[15] for further verification using the pgRNA deletion strategy and/or CRISPRi strategy. These lncRNAs have no evident overlap with any other genes. Among them, only *LINC00910*, which was high-ranked in our splicing-based screen, was validated to be essential for cell growth and proliferation in K562 (**Supplementary Fig. 14a–c**). Owing to the stringent cutoff, it was not included in our final essential lncRNA list. It remains to be determined whether such a discrepancy is due to cell line variation. We further selected four lncRNAs that were only identified in our screen and had been validated by both splicing and deletion approaches (**Fig. 3c**,**e** and **Supplementary Figs. 8–10**), and used the CRISPRi strategy for verification. The essentiality of only lncRNA *MIR17HG* was confirmed in K562 by CRISPRi (**Supplementary Fig. 14d**). The above results suggest that CRISPRi merely decreases gene expression instead of completely knocking out the target locus, leaving room for false-negative results. In addition, the splicing-targeting strategy could effectively avoid cutting regions close to neighboring genes, thereby decreasing the false positive rate.

Nevertheless, there are some limitations for splicing-targeting strategy. In comparison with the CRISPRi strategy, which interferes with both *trans*- and *cis*-acting lncRNAs, this method is restricted to study *trans*-acting lncRNAs that carry out functions by regulating their targeted genes at the post-transcriptional level, but is not suitable to study *cis*-acting lncRNAs that regulate expression of nearby genes[37]. However, it is possible to interrupt the function of some *cis*-acting lncRNAs by targeting their 5′ splice sites adjacent to promoters, which have been shown to be important for local regulation of certain lncRNAs[37]. Despite these limitations, this new strategy has demonstrated advantages in CRISPR screening of coding genes complementary to conventional exon-targeting methods, and enables large-scale loss-of-function screening of noncoding genes using a sgRNA-CRISPR library.

METHODS

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

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 High-Throughput Sequencing Center (Peking University) for their assistance in next-generation sequencing analysis, the National Center for Protein Sciences Beijing (Peking University), and the core facilities at School of Life Sciences (Peking University) for help in fluorescence-activated cell sorting. We also acknowledge the High-performance Computing Platform of Peking University. This project was supported by funds from the National Science Foundation of China (NSFC31430025), the Beijing Advanced Innovation Center for Genomics at Peking University, and the Peking-Tsinghua Center for Life Sciences (W.W.).

AUTHOR CONTRIBUTIONS

W.W. conceived and supervised the project. W.W., Y.L. and Z.C. designed the experiments. Y.L., Z.C., P.X. and Y.H. performed the experiments. Y.G. designed the oligonucleotides used for ribosomal gene mutagenesis and genome-wide lncRNA library, and Z.L. designed the pgRNAs used for individual validation. Y.W., Y.G. and P.Y. performed the bioinformatics analysis. Y.L., Z.C., Y.W. and W.W. wrote the manuscript with the help of all other authors.

COMPETING INTERESTS

A patent has been filed relating to the data presented. W.W. is a founder and scientific advisor for EdiGene.

1. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
2. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
5. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
6. 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).
7. Koike-Yusa, H., Li, Y., Tan, E.P., Del Castillo Velasco-Herrera, M. & 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. Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866–5878 (2014).
10. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
11. Quinn, J.J. & Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62 (2016).
12. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231–235 (2013).
13. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
14. Lin, N. *et al.* An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol. Cell* **53**, 1005–1019 (2014).
15. Liu, S.J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, eaah7111 (2017).
16. Adamson, B., Smogorzewska, A., Sigoillot, F.D., King, R.W. & Elledge, S.J. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.* **14**, 318–328 (2012).
17. Yuan, P. *et al.* Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. *Cell Res.* **25**, 157–168 (2015).
18. Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H. & Bartel, D.P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).
19. Lim, L.P. & Burge, C.B. A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci. USA* **98**, 11193–11198 (2001).
20. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
21. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
22. 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).
23. 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).
24. Zhu, S., Zhou, Y. & Wei, W. Genome-wide CRISPR/Cas9 screening for high-throughput functional genomics in human cells. *Methods Mol. Biol.* **1656**, 175–181 (2017).

25. Matlin, A.J., Clark, F. & Smith, C.W. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
26. Taggart, A.J., DeSimone, A.M., Shih, J.S., Filloux, M.E. & Fairbrother, W.G. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat. Struct. Mol. Biol.* **19**, 719–721 (2012).
27. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
28. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157 (2015).
29. Heidari, N. *et al.* Genome-wide map of regulatory interactions in the human genome. *Genome Res.* **24**, 1905–1917 (2014).
30. Muller, R.Y., Hammond, M.C., Rio, D.C. & Lee, Y.J. An Efficient method for electroporation of small interfering RNAs into ENCODE project tier 1 GM12878 and K562 cell lines. *J. Biomol. Tech.* **26**, 142–149 (2015).
31. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
32. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
33. Fang, Y. & Fullwood, M.J. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinformatics* **14**, 42–54 (2016).
34. Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* **548**, 343–346 (2017).
35. Goyal, A. *et al.* Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res.* **45**, e12 (2017).
36. Gilbert, L.A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
37. Engreitz, J.M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).

ONLINE METHODS

Cells and reagents. The HeLa cell line was from Z. Jiang's laboratory (Peking University) and cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco C11995500BT). The Huh7.5 cell line from S. Cohen's laboratory (Stanford University School of Medicine) was cultured in DMEM (Gibco) supplemented with 1% MEM nonessential amino acids (NEAA, Gibco 1140-050). K562 cell from H. Wu's laboratory (Peking University) and GM12878 cell from Coriell Cell Repositories were cultured in RPMI1640 medium (Gibco 11875-093). All cells were supplemented with 10% FBS (CellMax BL102-02) with 1% penicillin/streptomycin, cultured with 5% CO_2 at 37 °C.

Reverse transcription PCR (RT-PCR) for testing intron retention or exon skipping. The sgRNAs were cloned into a lentiviral expression vector carrying a CMV-promoter-driven mCherry marker, then transduced into HeLa cells[5–8] through viral infection at an multiplicity of infection (MOI) <1. The mCherry-positive cells were FACS-sorted 72 h after infection, and the total RNA of each sample was extracted using an RNAprep Pure Cell/Bacteria kit (Tiangen DP430). The cDNAs were synthesized from 2 μg of total RNA using a Quantscript RT kit (Tiangen KR103-04), and RT-PCR reactions were performed with TransTaq HiFi DNA polymerase (TransGen AP131-13).

Sequences of sgRNAs targeting the *RPL18* or *RPL11* gene. $sgRNA1_{RPL18}$: 5′-GGACCAGCCACTCACCATCC

$sgRNA2_{RPL18}$: 5′-AGCTTCATCTTCCGGATCTT

$sgRNA3_{RPL11}$: 5′-TCCTTGTGACTACTCACCTT

$sgRNA4_{RPL11}$: 5′-AACTCATACTCCCGCACCTG

Primers used for RT-PCR:

1F: 5′-CTGGGTCTTGTCTGTCTGGAA; 1R: 5′-CTGGTGTTTACATTCAGCCCC;

2F: 5′-GGCCAGAAGAACCAACTCCA; 2R: 5′-GACAGTGCCACAGCCCTTAG;

3F: 5′-TCAAGATGGCGTGTGGGATT; 3R: 5′-GACCAGCAAATGGTGAAGCC;

4F: 5′-GATCCTTTGGCATCCGGAGA; 4R: 5′-GCTGATTCTGTGTTTGGCCC.

Construction and screening of splicing-targeting sgRNA library on essential ribosomal genes. The annotations of 79 ribosomal genes were retrieved from NCBI. We scanned all potential sgRNAs targeting −50 bp to +75 bp surrounding every 5′ SD site and −75 bp to +50 bp surrounding every 3′ SA site of these 79 genes (**Supplementary Table 1**). We ensured that all sgRNAs had at least two mismatches to any other locus of the human genome and that the GC content was between 20% and 80%. A total of 5,788 sgRNAs targeting 79 ribosomal genes were synthesized using a CustmoArray 12K array chip (CustmoArray, Inc.) (**Supplementary Table 2**) , and construction of the plasmid library was the same as described before[4,8].

The cell library harboring these sgRNAs were constructed through lentiviral delivery at an MOI of <0.3 in Cas9-expressing HeLa and Huh7.5 cells[22], with a minimum coverage of 400×. 72 h after viral infection, the cells were sorted by FACS (BD) for mCherry expression. The control cells (2.4×10^6) of each library were collected for genomic DNA extraction using the DNeasy Blood and Tissue kit (Qiagen 69506), and the experimental cells were continuously cultured for 15 d before genomic DNA extraction. For each replicate, the lentivirally integrated sgRNA-coding regions were PCR-amplified by TransTaq HiFi DNA polymerase (TransGen AP131-13) and further purified with DNA Clean & Concentrator-25 (Zymo Research Corporation D4034) as previously described[4,8]. The resulting libraries were prepared for high-throughput sequencing analysis (Illumina HiSeq2500) using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB E7370L).

Design and construction of the genome-scale human lncRNA library. lncRNA annotations were retrieved from GENCODE dataset V20, which contains 14,470 lncRNAs. In this dataset, 2,477 lncRNAs without splice sites were removed in the first filtering process. For the rest of the lncRNAs, all potential 20-nt sgRNAs targeting −10 bp to +10 bp regions surrounding every 5′ SD site and 3′ SA site were designed. To ensure cleavage efficiency, we only kept sgRNAs whose GC content was between 20% and 80%, and removed those sgRNAs that contained ≥4-bp homopolymeric stretch of T nucleotides. To achieve the best coverage, certain sgRNAs with 1-bp or 0-bp mismatches to other loci were retained as long as they did not target any essential genes conserved in four cell lines, including K562 (ref. 21), and the total number of mismatched sites was no more than two. A total of 126,773 sgRNAs targeting 10,996 lncRNAs were ultimately synthesized. In the library, we also included 500 nontargeting sgRNAs that had at least two mismatches to any locus in human genome as negative controls and 350 sgRNAs targeting 36 essential ribosomal genes as positive controls. The oligonucleotides were synthesized using the CustmoArray 90K array chips (CustmoArray, Inc.), and library construction was the same as described above.

Genome-scale lncRNA screening. A total of 5×10^8 K562 or GM12878 cells were plated onto 175-cm^2 flasks (Corning 431080), and 2.6×10^8 HeLa cells were plated onto 15-cm plates. Each cell line was arranged in duplicate. Cells were infected with sgRNA library lentiviruses at an MOI of less than 0.3 (1,000× coverage for K562 and GM12878, 500× coverage for HeLa) in 24 h. The library cells were subjected to puromycin treatment (3 μg/ml for K562 and GM12878, 1 μg/ml for HeLa; Solarbio P8230) for 2 d. For each replicate, a total of 1.3×10^8 K562 or GM12878 cells and 6.5×10^7 HeLa cells were collected as the day-0 control samples for genome extraction. Each cell line was passaged every 2 d and cultured for 22 d (GM12878, HeLa) or 30 d (K562). Experimental cells at a minimum coverage of 1,000× (K562, GM12878) or 500× (HeLa) were isolated at the endpoint for genome extraction and NGS analysis[4,8].

The computational analysis of screens. Sequencing reads were mapped to hg38 reference genome and decoded with in-house scripts. sgRNA counts from two replicates were quantile normalized, and then average counts and fold changes between experimental and control groups were calculated. Noisy sgRNAs were then filtered with the following criteria: if a sgRNA's fold change was lower than mean fold change of positive control sgRNAs in one replicate and higher than mean fold change of negative control sgRNAs in another replicate, the sgRNA was regarded as a noisy sgRNA and excluded in the subsequent analysis. 1,000 negative control genes were generated by randomly sampling 10 negative control sgRNAs with replacement per gene. We compared the fold changes of the 10 sgRNAs targeting the 'virtual gene' to all the nontargeting sgRNAs by Wilcox test to get the *P* value. As a result, we obtained 1,000 *P* values and constructed an empirical distribution of these negative control genes' *P* values. Then, for each lncRNA after noise filtering, we also acquired a *P* value and further compared the *P* value to the empirical distribution to calculate the corrected *P* value. We ultimately defined the screen score by screen score = $\text{scale}[-\log_{10}(\text{adjusted } P)] + |\text{scale}[\log_2(\text{sgRNA fold change})]|$ (**Supplementary Table 5**). We designated those hits with screen score higher than 2 as essential lncRNAs. We also calculated the MIT score (http://crispor.tefor.net) of each sgRNA in the lncRNA library to evaluate its off-target potential, filtering out 1,357 sgRNAs with MIT score <10. Following the above procedure, the remaining sgRNAs were processed to generate the screen score for each lncRNA (**Supplementary Table 6**).

Validation of lncRNA hits. The two top-ranked sgRNAs for validation by splicing strategy were selected from library, which had at least two mismatches to any other locus in the genome. For the pgRNA deletion strategy, pgRNAs were designed to delete the promoter and the first exon of each lncRNA. We designed guide RNA pairs according to the following criteria: (1) one sgRNA targets the 2.5- to 3.5-kb regions upstream of the transcription start site (TSS) and the other one targets the 0.2- to 1.5-kb regions downstream of the TSS; (2) there should be no overlap with any exons or promoters of coding or noncoding genes. For each sgRNA of the pairs, we further ensured that (1) the GC content was between 45% and 70%, (2) the sgRNA did not include a ≥4-bp homopolymer stretch, and (3) the sgRNA contained more than two mismatches to any other locus in human genome. We included some sgRNAs with two mismatches to other loci, but the number of off-target sites was less than two. For the lncRNAs that were only identified by CRISPRi screen[15], two top-ranked sgRNAs from the CRISPRi study were selected to perform the individual validation in K562 cells stably expressing dCas9-KRAB protein by CRISPRi strategy. To avoid affecting the neighboring genes, the pgRNAs for knocking out these CRISPRi-unique lncRNA hits were designed to delete the

promoter, or the promoter and the first exon, or the gene body of the lncRNAs. Except for the deleted regions, the other criteria for designing these pgRNAs were the same as described above.

All the sgRNAs or pgRNAs targeting the selected lncRNAs to be validated were individually cloned into the lentiviral vector with a CMV-promoter-driven EGFP marker. After viral packaging, the sgRNA or pgRNA lentivirus was transduced into K562 or GM12878 cells at an MOI of <1. The cell proliferation assay was previously described[1].

RNA sequencing and data analysis. Two sgRNAs targeting the splice sites of lncRNA *MIR17HG* and *BMS1P20* were individually cloned into the lentiviral vector with an EGFP marker. The sgRNAs were delivered into K562 or GM12878 cells by lentiviral infection at an MOI of <1.0. EGFP-positive K562 or GM12878 cells (2×10^6) were sorted by FACS 5 d after infection. Total RNA of each sample was extracted using RNeasy Mini Kit (QIAGEN 79254), and the RNA-seq libraries were prepared following the NEBNext PolyA mRNA Magnetic Isolation Module (NEB E7490S), NEBNext RNA First Strand Synthesis Module (NEB E7525S), NEBNext mRNA Second Strand Synthesis Module (NEB E6111S) and NEBNext Ultra DNA Library Prep Kit for Illumina (NEB E7370L). All samples were subjected to NGS analysis using the Illumina HiSeq X Ten platform (Genetron Health; Beijing, China). Deep sequencing reads were mapped to the hg38 reference genome and gene expression was quantified by RSEM v1.2.25 (ref. 38). Differential expression analysis was conducted by EBSeq version 1.10.0 (ref. 39), and differentially expressed genes were selected from those that had adjusted $P < 0.05$ and absolute $\log_2$(fold change) > 3. Gene Ontology and KEGG analysis was conducted by DAVID 6.8 (ref. 40).

Code availability. Source code for the computational analysis of lncRNA screen described in this paper is available in **Supplementary Code**.

Reporting Summary. Further information on research design is available in the **Nature Research Reporting Summary** linked to this article.

Data availability. Sequencing data from CRISPR screen for identifying essential lncRNAs in each cell line and RNA-seq of each sample can be accessed in NCBI Sequence Read Archive (SRA) with accession code SRP157958 under BioProject ID PRJNA486076.

38. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
39. Leng, N. *et al.* EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043 (2013).
40. Jiao, X. *et al.* DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805–1806 (2012).

doi:10.1038/nbt.4283

nature research

Corresponding author(s): Wensheng Wei

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

In this study, two independent replicates were performed for all library screening. During the process of validation, three biological replicates were performed for all experimental groups.

2. Data exclusions

Describe any data exclusions.

In the computational analysis of genome-scale human lncRNA library, if a sgRNA's fold change was lower than mean fold change of positive control sgRNAs in one replicate and higher than mean fold change of negative control sgRNAs in another replicate, the sgRNA was regarded as a noisy sgRNA for filtering.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

During the cell library construction, all the cells for each independent replicate were randomly infected, FACS- or puro-selected and further passaged.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were blinded to group allocation during data collection and analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Nature Biotechnology: doi:10.1038/nbt.4283

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	In the analysis of RNA-seq data , gene expression was quantified by RSEM v1.2.2538 and differential expression analysis was conducted by EBSeq version 1.10.039. Gene Ontology and KEGG analysis was conducted by DAVID 6.8. Each of these is open-access.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	The plasmid of the splicing-targeting lncRNA library described in the manuscript will be deposited at Addgene for ease of distribution.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	The HeLa cell line was from Z. Jiang's laboratory (Peking University). The Huh 7.5 cell line was from S. Cohen's laboratory (Stanford University School of Medicine). The K562 cell line was from H. Wu's laboratory (Peking University) and GM12878 cell line was from Coriell Cell Repositories.
b. Describe the method of cell line authentication used.	STR analysis
c. Report whether the cell lines were tested for mycoplasma contamination.	All cell lines were tested negative for mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study did not involve human research participants.

Nature Biotechnology: doi:10.1038/nbt.4283