Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing

Teng Fei[a,b,c,d,1,2], Yiwen Chen[e,1], Tengfei Xiao[b,c,d], Wei Li[d], Laura Cato[b,c], Peng Zhang[e], Maura B. Cotter[f,g], Michaela Bowden[f,g], Rosina T. Lis[f,g], Shuang G. Zhao[h,i], Qiu Wu[j], Felix Y. Feng[h,i], Massimo Loda[f,g,k,l], Housheng Hansen He[m,n], X. Shirley Liu[d,2], and Myles Brown[b,c,2]

[a]College of Life and Health Sciences, Northeastern University, Shenyang 110819, People's Republic of China; [b]Division of Molecular and Cellular Oncology, Department of Medical Oncology, Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215; [c]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215; [d]Department of Biostatistics and Computational Biology, Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02215; [e]Department of Bioinformatics and Computational Biology, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX 77030; [f]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215; [g]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215; [h]Michigan Center for Translational Pathology, Comprehensive Cancer Center, University of Michigan, Ann Arbor, MI 48109; [i]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109; [j]Department of Bioinformatics, School of Life Sciences and Technology, Tongji University, Shanghai, 200092, China; [k]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215; [l]Broad Institute of Harvard and MIT, Cambridge, MA 02142; [m]Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada M5G1L7; and [n]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada M5G1L7

Contributed by Myles Brown, May 14, 2017 (sent for review November 9, 2016; reviewed by Mariano Garcia-Blanco and Jindan Yu)

Alternative RNA splicing plays an important role in cancer. To determine which factors involved in RNA processing are essential in prostate cancer, we performed a genome-wide CRISPR/Cas9 knockout screen to identify the genes that are required for prostate cancer growth. Functional annotation defined a set of essential spliceosome and RNA binding protein (RBP) genes, including most notably heterogeneous nuclear ribonucleoprotein L (HNRNPL). We defined the HNRNPL-bound RNA landscape by RNA immunoprecipitation coupled with next-generation sequencing and linked these RBP–RNA interactions to changes in RNA processing. HNRNPL directly regulates the alternative splicing of a set of RNAs, including those encoding the androgen receptor, the key lineage-specific prostate cancer oncogene. HNRNPL also regulates circular RNA formation via back splicing. Importantly, both HNRNPL and its RNA targets are aberrantly expressed in human prostate tumors, supporting their clinical relevance. Collectively, our data reveal HNRNPL and its RNA clients as players in prostate cancer growth and potential therapeutic targets.

CRISPR screen | HNRNPL | prostate cancer | RNA binding protein | alternative splicing

Prostate cancer is among the most prevalent adult malignancies in developed countries. The principal treatment for prostate cancer once it is no longer amenable to surgery or radiation treatment is androgen deprivation therapy, which targets androgen or androgen receptor (AR) signaling. However, resistance to androgen deprivation therapy often develops and leads to a state termed "castration-resistant prostate cancer," which still lacks an effective cure (1–3). Therefore, significant efforts have been devoted to better understand the mechanism of oncogenesis and to develop additional effective therapeutics targeting pivotal oncogenes, cancer-related signal transduction pathways, and epigenetic regulators (4, 5).

Alternative RNA splicing is a fundamental cellular process by which a single gene can give rise to multiple different transcripts and proteins. This process is tightly regulated by core spliceosomes and other splicing factors, such as the serine/arginine-rich family of proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) (6, 7). Multiple studies indicate that deregulation of alternative splicing is implicated in cancer progression and that the splicing machinery may be targeted therapeutically (8–10). In addition to RNA splicing, the physical interactions between RNAs and RNA binding proteins (RBPs) underlie multiple RNA processing steps, such as capping, polyadenylation, transport, localization, modification, and translation, thereby regulating many aspects of RNA fate (11). Which RBPs and their related RNA processing steps are functionally important, especially in prostate cancer, remains elusive.

The recent implementation of the clustered regularly interspaced short palindromic repeats (CRISPR)-associated nuclease Cas9 genome editing system has proved effective in high-throughput loss-of-function screens (12–14). Compared with RNA interfering (RNAi)-based gene knockdown, CRISPR/Cas9-based gene knockout confers a more thorough deletion of target gene expression and has less off-target effects when the guide RNA is appropriately designed (15, 16). The power of the pooled CRISPR screen has been demonstrated by several studies investigating the genes involved in drug and toxin resistance (13, 17), cancer metastasis (18), and immune response (19).

Here, by a genome-wide CRISPR screen and in depth mechanistic studies, we sought to systematically identify functional RBPs or RNA processing factors that are essential for prostate cancer growth and underlie pivotal cancer-related RNA processing steps, especially RNA splicing, thus potentially enabling the development of novel cancer therapeutics targeting RBPs or RBP–RNA interactions.

Significance

Alternative RNA splicing and the spliceosome machinery have been implicated in cancer progression. A genome-wide CRISPR screen identified the RNA processing factor heterogeneous nuclear ribonucleoprotein L (HNRNPL) as required for prostate cancer growth by regulating alternative RNA splicing and circular RNA formation. HNRNPL and its RNA clients are overexpressed during prostate cancer progression, supporting their potential role as therapeutic targets.

Author contributions: T.F., X.S.L., and M. Brown designed research; T.F., Y.C., T.X., W.L., P.Z., M.B.C., S.G.Z., and Q.W. performed research; T.F., Y.C., T.X., W.L., L.C., P.Z., M.B.C., M. Bowden, R.T.L., S.G.Z., Q.W., F.Y.F., M.L., H.H.H., X.S.L., and M. Brown analyzed data; and T.F., Y.C., X.S.L., and M. Brown wrote the paper.

Reviewers: M.G.-B., University of Texas Medical Branch; and J.Y., Northwestern University.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE72844).

[1]T.F. and Y.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: feiteng@mail.neu.edu.cn, xsliu@jimmy.harvard.edu, or myles_brown@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617467114/-/DCSupplemental.

Results

Genome-Wide CRISPR/Cas9 Knockout Screen in Prostate Cancer Cells. To systematically identify the functional RBPs or RNA processing factors in prostate cancer, we performed a pooled genome-wide CRISPR/Cas9 knockout screen in prostate cancer LNCaP cells using the GeCKO v2 lentiviral single guide RNA (sgRNA) library (20) (Fig. 1*A*). This library contains 123,411 unique sgRNAs targeting 19,050 genes, with 6 sgRNAs per gene. After selection with puromycin, half of the infected cells were harvested as a day_0 control sample and the rest of the cells were continuously cultured for an additional two weeks as the day_14 sample. The sgRNAs incorporated in the cells were then amplified from genomic DNA and subjected to massively parallel sequencing for quantification. We defined the negatively selected essential genes as those with significantly depleted sgRNAs at day 14 and positively selected genes as those with significantly enriched sgRNAs at day 14 (Dataset S1). MAGeCK and MAGeCK-VISPR are statistical algorithms we previously developed for CRISPR screen analyses (21, 22). Whereas MAGeCK was developed to call genes from a single CRISPR screen experiment, MAGeCK-VISPR was developed to compare the gRNA abundance across many conditions. The magnitude of negative selection for the top 1,000 essential genes in LNCaP cells is comparable to that of known essential ribosomal protein genes (Fig. S1*A*). We thus chose these top 1,000 genes as the LNCaP essential gene set for the downstream analysis. By comparing this list of essential genes to publically available data for 10 other cell lines, we found that ∼one-third of the LNCaP essential genes are broadly essential across multiple cell types (as pan-essential genes), whereas the remaining genes were either LNCaP-specific or essential only in a subset of cell types (Fig. 1*B*). The most enriched functional categories for the pan-essential genes fall into protein biosysthesis, ribonucleoprotein, ribosome, proteasome, and cell cycle that are all core functional components or processes in the cell (Fig. S1*B*). In terms of enriched pathways, LNCaP essential genes are enriched for AR and PI3K pathways that are known to play pivotal signaling roles in LNCaP cells, supporting the relevance of our screen (Fig. S1*C*). In addition, many known functional genes in prostate cancer were under significant negative selection, including STAT3, PIAS1, PIK3C2A, AR, and MYC (Fig. 1*C*) (23–25). Overall, these top 1,000 essential genes established an interconnected network highlighting multiple key signaling molecules, complexes, or processes, including spliceosome and ribonucleoproteins (Fig. 1*D*). Furthermore, we investigated the gene copy number effects on CRISPR screen performance and confirmed that genes in genomic amplified regions tend to exhibit more negative selection as revealed by recent studies (16, 26) (Fig. S1*D* and Dataset S1). These potential

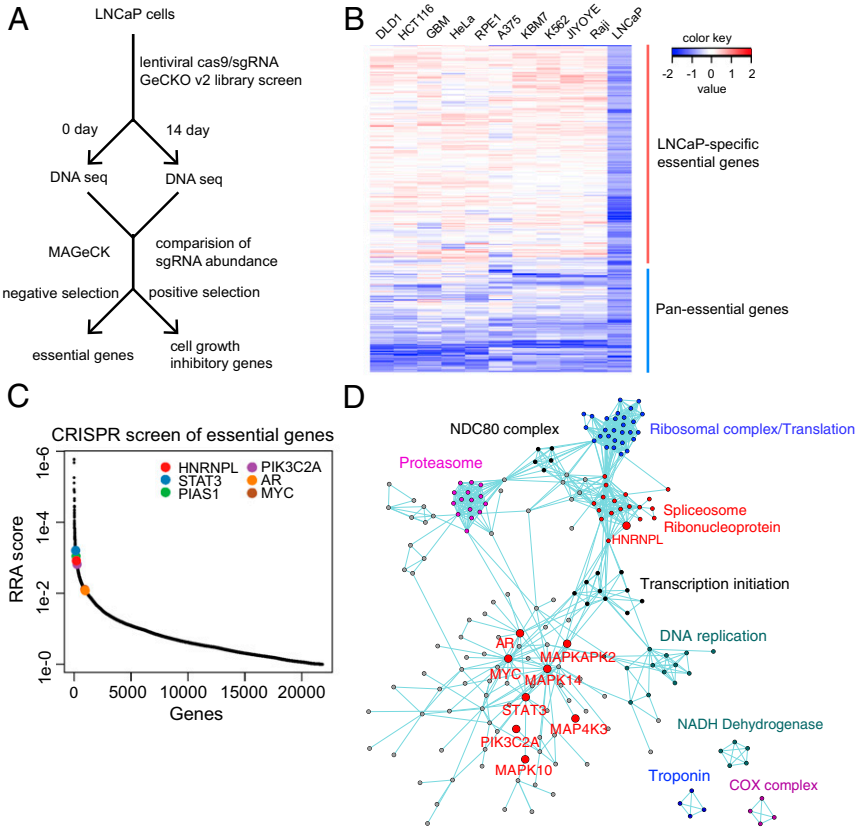

Fig. 1. Genome-wide CRISPR/Cas9 screen in prostate cancer. (*A*) Workflow of the CRISPR/Cas9 screen in LNCaP cells cultured in the presence of 10 nM dihydrotestosterone (DHT) with charcoal/dextran-treated FBS-containing medium. (*B*) Multiple CRISPR screening datasets (12, 15, 49) reveal the functions of LNCaP essential genes across different cell types. The β-essentiality scores of top 1,000 LNCaP essential genes, calculated by MAGeCK-VISPR, are shown. Negative (or positive) β-score indicates the corresponding gene is undergone negative (or positive) selection. The less the β-score, the more essential the corresponding gene. These genes consist of two categories: pan-essential genes (genes that are also essential in more than half of the other cell types) and LNCaP-specific essential genes. (*C*) CRISPR screen of essential genes in LNCaP cells. Genes were rank-ordered by robust rank aggregation (RRA) scores calculated by MAGeCK (21), where a smaller RRA score indicates more essentiality. HNRNPL and known LNCaP functional genes are highlighted. (*D*) A network view of the top 1,000 LNCaP essential genes. Nodes represent genes, and an edge connecting two genes if both are in the same pathway. AR and some known LNCaP oncogenes are highlighted, and some major gene clusters are also marked using different colors. The pathway information is extracted from GeneMANIA database (50).

false-positive hits from copy number-increased regions were removed from the downstream functional analysis.

HNRNPL Is Required for Prostate Cancer Cell Growth. To explore which factors might play a role in RNA processing, we examined the RBPs and RNA splicing factors that were essential in the screen. Interestingly, several HNRNP family members, such as HNRNPL, HNRNPC, and FUS, are among the top essential RBPs (Dataset S1). Notably, HNRNPL was the top HNRNP essential gene in LNCaP cells and comparable to the known LNCaP essential genes identified by CRISPR screen, such as AR (Fig. 1*C* and Dataset S1). To cross-validate and more carefully examine the HNRNP family genes, we additionally used a targeted short interfering RNA (siRNA) approach to knock down each HNRNP gene with two independent and effective siRNAs, and examined the effects on cell growth after 6 d of siRNA exposure. Among all of the HNRNP genes tested, HNRNPK and HNRNPL showed the most dramatic and consistent cell growth reduction upon knockdown, followed by HNRNPC (Fig. 2*A* and Fig. S2*A*). HNRNPK was not identified as essential in the CRISPR screen, which might be because of differences between the CRISPR and RNAi technologies, such as inefficient sgRNAs or siRNA off targets.

To further minimize the potential siRNAs off-target effects for HNRNPK and HNRNPL, we used two additional independent siRNAs targeting either HNRNPK or HNRNPL, and examined the cell-growth effect in several prostate cancer cell line models. In both the androgen-dependent LNCaP cells and the castration-resistant CWR22RV1 cells, as well as the AR-negative prostate cancer DU145 and PC3 cell lines, knockdown of either HNRNPL or HNRNPK could diminish cell growth (Fig. 2*B* and Fig. S2 *B* and *C*). Interestingly, in a normal-like prostate epithelial cell line, RWPE-1, HNRNPL is not required for cell growth, whereas HNRNPK remains essential (Fig. 2*B*), indicating that HNRNPL, but not HNRNPK, may have cancer-specific functions. Moreover, recent studies suggested that superenhancer-associated genes tend to be functionally important in the control of cell identity and disease (27). HNRNPL and PTBP1 were the only members of the HNRNP family categorized as superenhancer-associated genes in LNCaP cells (Fig. S2*D*). Taken together, our CRISPR screen, siRNA validation, and superenhancer analysis identified HNRNPL, among the other RBPs, as worthy of further study.

RNA Binding Landscape of HNRNPL by RNA Immunoprecipitation Coupled with Next-Generation Sequencing. As a classic RBP, HNRNPL functions through protein–RNA interaction. To characterize the genome-wide landscape of HNRNPL-associated RNAs, we used RNA immunoprecipitation coupled with next-generation sequencing (RIP-seq) to map the HNRNPL–RNA interactome in LNCaP cells. Similar to ChIP for analyzing protein–DNA interactions, the RIP approach uses a specific antibody against the

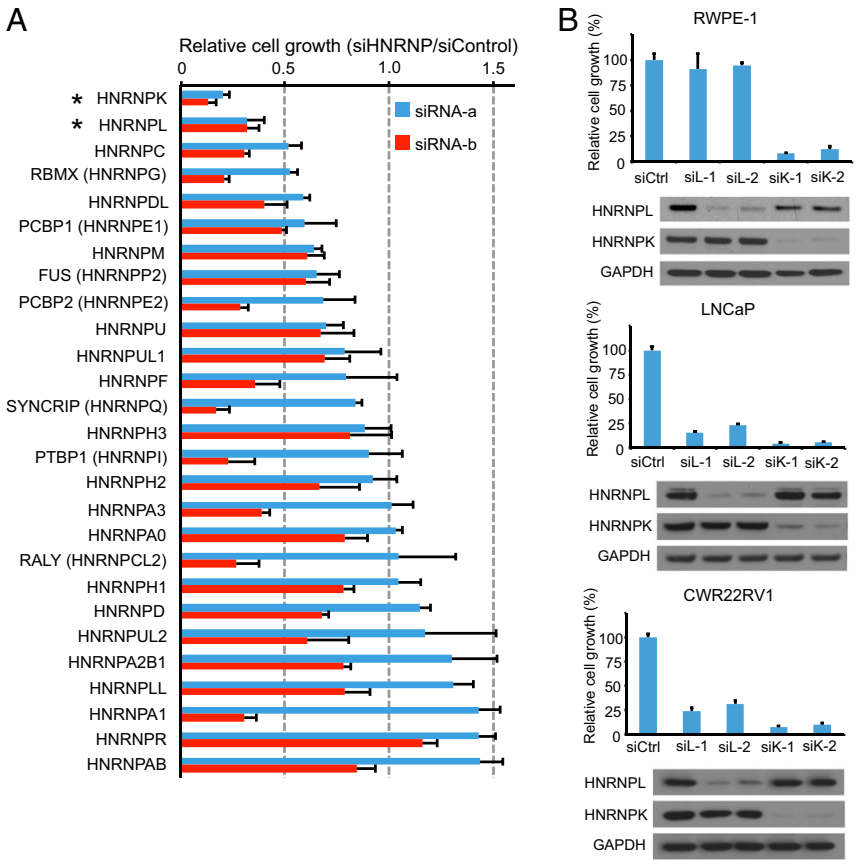

Fig. 2. HNRNPL is a functional RBP in prostate cancer. (*A*) Validation of essentiality of HNRNPs using siRNA oligos in LNCaP cells. LNCaP cells cultured in full medium were transfected with corresponding siRNAs and cell growth was determined by cell counting after 6 d of transfection. Each HNRNP was targeted by two independent and effective siRNA oligos (siRNA-a and siRNA-b). Data are shown as relative cell growth by comparison with control knockdown cells, mean ± SD, $n = 3$. Functional HNRNPs, denoted with an asterisk, are defined as exhibiting less than 0.5 of relative cell growth by two independent knockdowns. (*B*) Cell growth validation upon HNRNPL and HNRNPK knockdown using an additional two independent siRNA oligos (siRNA-1 and siRNA-2) in three cell lines: normal-like RWPE-1 prostate cells, LNCaP androgen-dependent prostate cancer cells, and CWR22Rv1 castration-resistant prostate cancer cells. Cells were counted after 6 d of siRNA transfection and data are shown as mean ± SD, $n = 3$. Knockdown efficiency was determined by Western blot analysis.

protein of interest to pull down formaldehyde cross-linked RNAs. Biological replicates of HNRNPL RIP-seq identified 6,892 consensus binding sites across 1,549 genes (Fig. 3*A* and Dataset S2). The reproducibility of our RIP-seq data was evidenced by the correlation of read count across 10-kb windows between replicates ($r = 0.98$) (Fig. 3*B*). We further categorized the distribution of the binding sites across different genomic elements and found that HNRNPL-associated RNA peaks are mostly enriched in introns and 3′UTRs (Fig. 3*C* and Fig. S3*A*). De novo motif analysis showed that the most enriched binding motif in LNCaP cells is a CA-repeat or CA-enriched pattern (Fig. 3*D*), consistent with previous studies of HNRNPL–RNA interactions in other cell types (28, 29).

To further validate the RIP-seq peaks, we used two independent monoclonal HNRNPL antibodies (4D11 and D5) to perform RIP-qPCR analysis on several representative RIP-seq peaks with differential binding strengths. Although HNRNPL does not bind to the negative control RPS28 transcript, it has significant but differential binding for CTBP1, ROR2, and STX3 transcripts (Fig. 3 *E* and *F*). In addition, we performed RNA pulldown assays to validate these peak regions. Biotin-labeled peak-region RNA was in vitro transcribed and incubated with

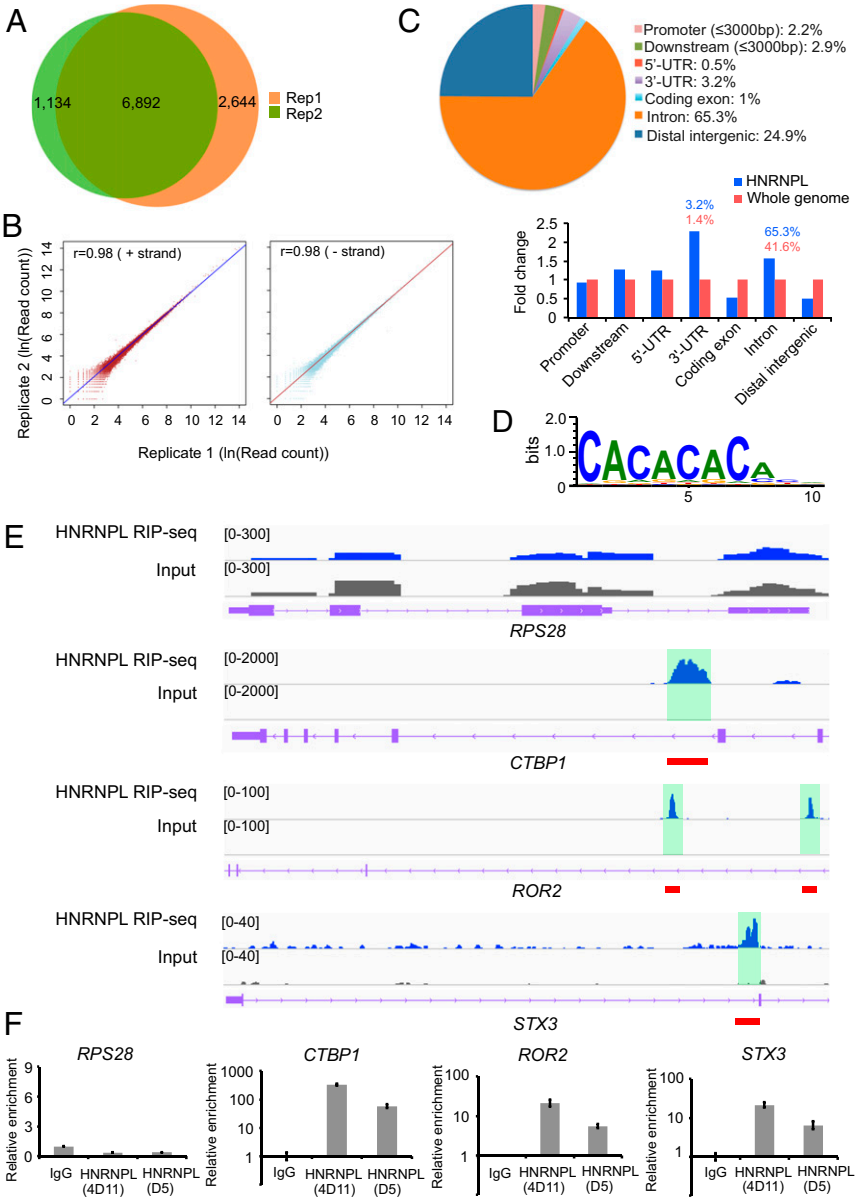

Fig. 3. The genome-wide landscape of HNRNPL binding sites on RNA. (*A*) Venn diagram of HNRNPL RIP-seq peaks from two biological replicates. LNCaP cells cultured in charcoal/dextran-treated FBS were treated with 10 nM androgen DHT for 4 h and harvested for RIP-seq analysis. The peaks from two biological replicates were compared. (*B*) Scatter plot and correlation coefficience of the RIP-seq read count within a 10-kb window between two replicates. (*C*) HNRNPL RIP-seq peaks were enriched in introns and 3′UTR region. All RIP-seq peaks were categorized according to the distribution on different genomic elements and compared with the genomic background. (*D*) De novo motif analysis identifying CA-repeat motif as the only enriched motif within the top HNRNPL RIP-seq peaks. (*E*) Representative HNRNPL RIP-seq peaks were shown as track signals in an integrative genomic viewer. Peak area is highlighted in light green and peak region is denoted as red rectangle. RPS28 served as a negative control for HNRNPL enrichment. (*F*) RIP-qPCR analysis of HNRNPL binding sites as shown in *E*. The left side peak of ROR2 in *E* was chosen for RIP-qPCR validation. IgG RIP served as negative RIP control. RPS28 served as negative site for RIP enrichment. Data were shown as mean ± SEM, $n = 3$.

cell lysate. Streptavidin beads then captured the RNA–protein complex and interacting proteins were resolved by Western blot analysis. Only the sense strand of peak RNAs, but not the antisense strand, could associate with HNRNPL, further proving the sequence specificity of HNRNPL–RNA association (Fig. S3*B*). Taking these data together, using RIP-seq we identified a specific set of HNRNPL-associated RNAs in prostate cancer with high confidence and precision, suggesting that these may underlie the mechanism of HNRNPL function.

HNRNPL Regulates Alternative Splicing in Prostate Cancer Cells. The significant intron-binding pattern implied that HNRNPL is primarily bound to pre-mRNA and might directly regulate alternative splicing, consistent with a previously reported function of HNRNPL (29, 30). To gain a comprehensive view of HNRNPL-dependent alternative splicing events in prostate cancer, we transfected either control or HNRNPL-targeting siRNAs in LNCaP cells and applied paired-end sequencing on the polyadenylated transcriptome on three biological replicates. Alternative splicing events were analyzed by the rMATS algorithm (31). In total, 206 alternative splicing events were identified as HNRNPL-dependent, with the majority falling in the skipped exon category (Fig. 4*A* and Dataset S3). HNRNPL binding is significantly associated with these splicing changes (Fig. 4*B*), supporting a direct effect of HNRNPL–RNA interaction on alternative splicing.

Interestingly, the AR pre-mRNA is significantly bound by HNRNPL and an HNRNPL-dependent alternative splicing event was identified in the AR transcript (Fig. 4*C*, Fig. S3*C*, and Dataset S3). A cryptic exon 2b was identified as alternatively spliced between constitutive exon 2 and exon 3 in the AR transcript (Fig. 4*D*). Using primers that discriminate the two alternatively spliced isoforms, we found that the constitutive isoform connecting exon 2 and 3 decreased, whereas the alternative isoform connecting exon 2 and 2b increased upon HNRNPL knockdown (Fig. 4*D*). The pre-mRNA level of AR was not affected (Fig. 4*D*), confirming that HNRNPL regulates AR isoform expression through alternative splicing but not direct transcription. Moreover, RIP-qPCR analysis confirmed that HNRNPL binds to AR intron 3, contiguous with exon 3 that is alternatively spliced (Fig. 4 *C* and *E*). The regulation of the protein level of the constitutive AR isoform by HNRNPL was further confirmed by Western blot analysis (Fig. 4*F*). In addition to AR, we also validated another HNRNPL bound and regulated alternative splicing target MYH10 (Fig. S4 *A*–*C* and Dataset S3), where HNRNPL knockdown significantly reduced

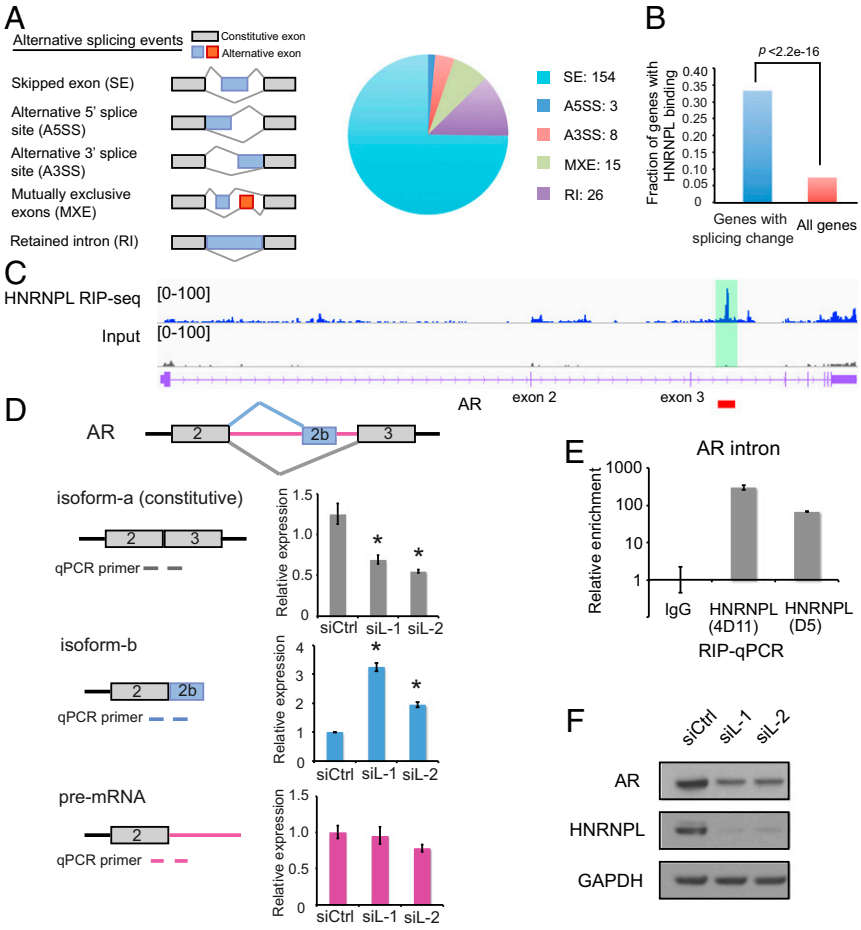

Fig. 4. HNRNPL directly regulates alternative splicing in prostate cancer. (*A*) Pie chart showing different classes of HNRNPL-dependent alternative splicing events. (*B*) HNRNPL binding is significantly associated with HNRNPL-dependent alternative splicing. Fraction of genes with HNRNPL binding from differential splicing genes was compared with that from all genes. (*C*) Genome browser representation of HNRNPL binding over AR loci from RIP-seq data. (*D*) RT-qPCR confirmation of HNRNPL-dependent alternative splicing of AR. Both two alternatively spliced isoforms and pre-mRNA of AR were examined using corresponding PCR primers by RT-qPCR analysis upon HNRNPL knockdown. Data were shown as mean ± SEM, $n = 3$, $^*P < 0.05$. (*E*) RIP-qPCR confirmation of HNRNPL binding on AR RNA. Data were shown as mean ± SEM, $n = 3$. (*F*) Western blot analysis of full-length AR protein level upon HNRNPL knockdown. GAPDH served as loading control.

the exon 6-containing isoform but slightly increased the exon 6-skipping isoform.

HNRNPL Regulates Circular RNA Formation. Circular RNA (circRNA), generated by nonsequential back splicing of pre-mRNA (Fig. 5*A*), has drawn increasing attention in the RNA field (32–34). Increasing length and depth of high-throughput sequencing allowed increasing number of circRNAs to be discovered and characterized. Rather than splicing noise, circRNA has been shown to be tightly regulated and functional (35). Recent advances in the understanding of circRNA biogenesis suggested splicing factors as regulators involved in circRNA formation (32, 36, 37).

To determine whether HNRNPL is involved in circRNA regulation, we profiled the circRNA transcriptome in LNCaP cells by RNase R enrichment and RNA-seq profiling in either control or HNRNPL knockdown conditions, and determined

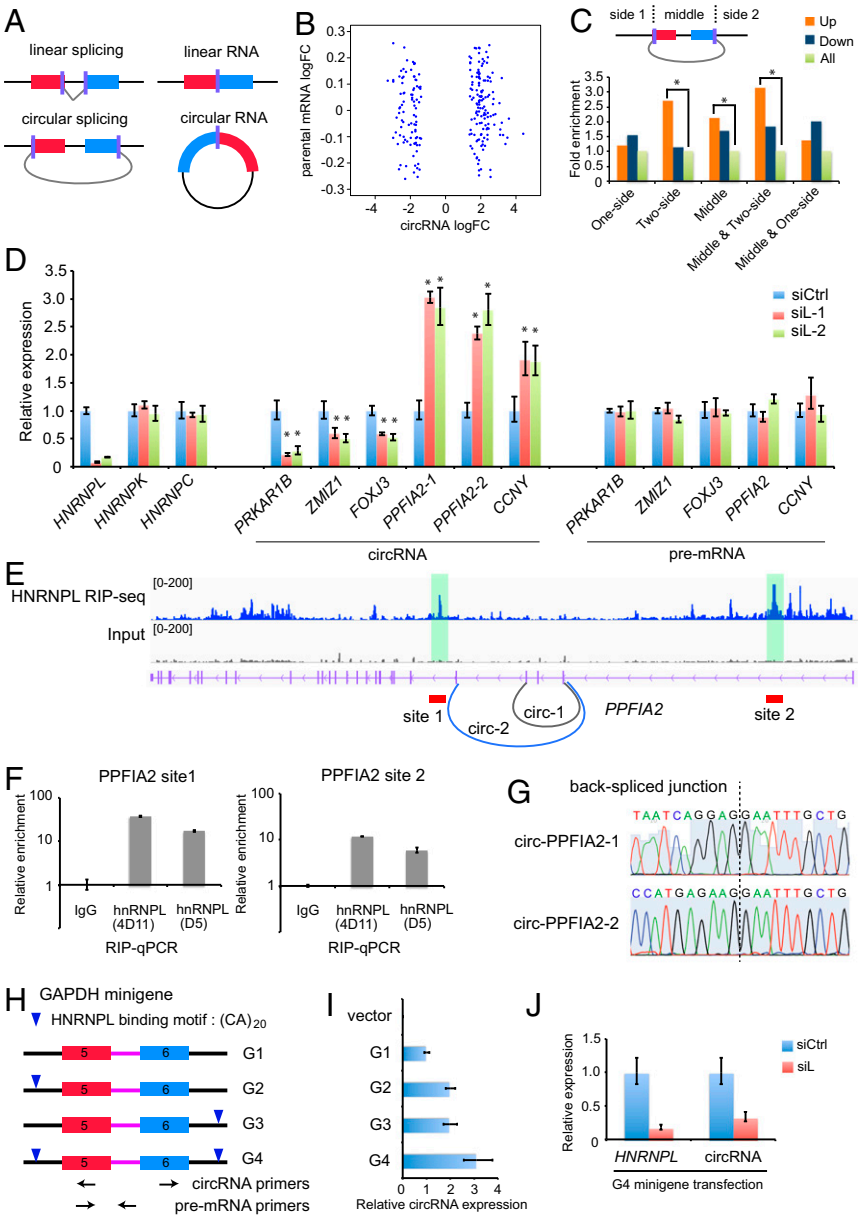

Fig. 5. HNRNPL is involved in CircRNA regulation. (*A*) Schematic depiction of circRNA formation via back splicing of linearly aligned exons. (*B*) Correlation of fold-change in abundance of HNRNPL-regulated circRNA (*y* axis) and its cognate parental mRNA (*x* axis). Note the scale difference between the two axes. (*C*) Association of HNRNPL binding status with circRNA expression regulation. The fold-enrichment represents the fraction for each category over all circRNA genes. (*$P < 0.05$). (*D*) RT-qPCR analysis of several HNRNPL-regulated circRNAs and their cognate pre-mRNAs. Validated circRNAs are selected as more than ~twofold change upon HNRNPL knockdown and with HNRNPL binding in any region of parental genes from Dataset S4. Data are shown as mean ± SEM, $n = 3$, *$P < 0.05$. (*E*) Genome browser representation of HNRNPL binding over PPFIA2 loci as determined by RIP-seq. The back-spliced exons that generate two different circRNAs (circ-1 and circ-2) are denoted. The two flanking HNRNPL-binding peaks are highlighted. (*F*) RIP-qPCR validation of two HNRNPL-binding peaks over PPFIA2 transcript highlighted in *E*. Data are shown as mean ± SEM, $n = 3$. (*G*) Sanger sequencing traces for confirmation of back-spliced exon junctions of the two separate PPFIA2 circRNA. (*H*) Schematic description showing the design of GAPDH minigene assay. (*I*) RT-qPCR analysis of relative circRNA formation normalized by pre-mRNA levels for different GAPDH minigenes transfected into HEK293T cells. Data are shown as mean ± SEM, $n = 3$. (*J*) CircRNA formation of the GAPDH minigene is dependent on HNRNPL. RT-qPCR analysis was performed in G4 minigene-transfected samples upon control or HNRNPL knockdown. Data are shown as mean ± SEM, $n = 3$.

differential expression of circRNAs (Dataset S4). Among 4,389 detected circRNAs in the LNCaP transcriptome, we identified 139 significantly up-regulated and 93 down-regulated circRNAs upon HNRNPL knockdown, whose parental mRNA gene expression remains unchanged (Fig. 5*B*). HNRNPL binding significantly associates with up-regulated circRNAs and to a lesser extent with down-regulated circRNAs upon HNRNPL knockdown when the binding events occur at both flanks or within the circRNAs (Fig. 5*C*). Using divergent PCR primers, we validated the expression change of several HNRNPL-regulated circRNAs upon HNRNPL knockdown, but their parental pre-mRNAs remain unchanged (Fig. 5*D*). The circularity and back splicing of these circRNAs were confirmed by RNase R treatment and Sanger sequencing (Fig. S5 *A* and *B*). Interestingly, one parental transcript can give rise to multiple circRNAs, as exemplified by PPFIA2 (Fig. 5 *D*–*G*). Between the two highlighted HNRNPL binding peaks in Fig. 5*E*, two HNRNPL-repressed circRNAs (denoted as circ-1 and circ-2) were present, indicating the in vivo complexity of circRNA regulation by HNRNPL.

To further test whether HNRNPL directly regulates circRNA formation via HNRNPL–RNA interaction, we designed a minigene system in which a fragment of GAPDH pre-mRNA with exon 5 and 6, as well as their flanking introns, was cloned into an expression plasmid. We then created different minigene constructs by introducing an HNRNPL binding site $(CA)_{20}$ to either or both of the flanking introns to evaluate the effect of HNRNPL binding on circRNA formation. These minigene constructs were transfected into HEK293T cells and relative circRNA levels were determined and normalized against pre-mRNA expression (Fig. 5*H*). This approach has several advantages in the evaluation of *cis* element or *trans* factor effects on circRNA formation, as there is no endogenous circRNA formation in the GAPDH locus and little endogenous pre-mRNA expression compared with the transfected minigene. HNRNPL binding on the flanking introns significantly enhances circRNA formation, especially when present at both sides (Fig. 5*I*), and HNRNPL knockdown significantly reduces the circRNA expression (Fig. 5*J*). Sanger sequencing confirmed the back-splicing exon junction of this GAPDH circRNA (Fig. S5*C*). Consistent with this, we found that one or more HNRNPL binding sites correlates with an increased likelihood to form circRNAs in our LNCaP cell profiling data (Fig. S5*D*). Because this may reflect the co-occurrence of HNRNPL with other key circRNA-promoting factors, integrating data on the other factors might be needed to better elucidate the role of HNRNPL in enhancing or repressing circRNA expression in vivo.

HNRNPL and Its RNA Targets are Clinically Relevant in Human Prostate Cancer. To evaluate the clinical relevance of HNRNPL and its RNA targets, we analyzed the mRNA expression of HNRNPL in several independent prostate cancer cohorts (38–41). HNRNPL expression is consistently higher in prostate tumors versus normal prostate (Fig. 6*A*). Moreover, we confirmed its higher protein expression in cancer by immunohistochemical analysis of tissue microarrays with 79 pairs of matched prostate tumor and benign samples (Fig. 6 *B* and *C* and Fig. S6 *A* and *B*).

Using Oncomine Concepts Map analysis (42), we found that both HNRNPL-regulated alternatively spliced genes and HNRNPL-regulated circRNA genes are significantly associated with over-expressed gene signatures in prostate cancer versus normal prostate in several independent cohorts (Fig. 6 *D* and *E*). Furthermore, HNRNPL-regulated circRNA genes are significantly enriched among the LNCaP essential gene set (Fig. S6*C*), indicating a functional link between HNRNPL and its downstream RNA targets. The association of both HNRNPL expression and its target RNA expression with prostate cancer progression suggests that HNRNPL may regulate prostate cancer growth through these downstream RNA targets. In addition to increased expression in prostate cancer, HNRNPL also exhibits frequent genomic amplification in many other tumor types (Fig. S6*D*), indicating that its importance may not be limited to prostate cancer.

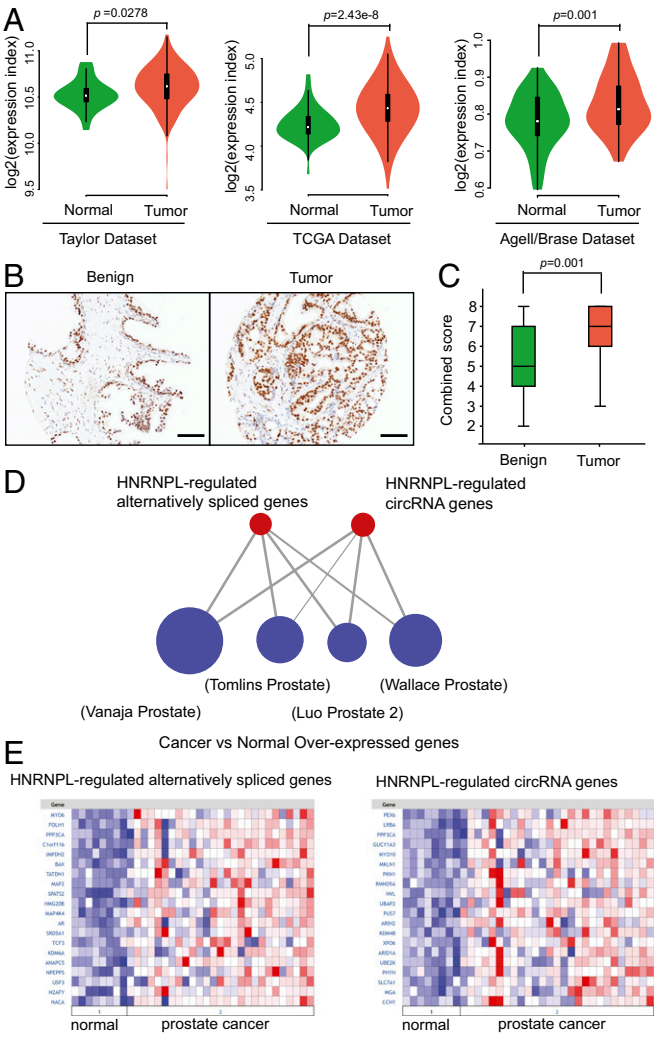

Fig. 6. HNRNPL and its RNA targets are clinically relevant. (*A*) Violin plot showing the HNRNPL expression difference between normal prostate and prostate tumors using the gene-expression data from three independent cohorts. TCGA dataset were from prostate adenocarcinoma data in The Cancer Genome Atlas (TCGA). The Agell/Brase datasets were combined because of the limited sample number in each individual dataset. (*B*) Representative images of tissue microarray analysis at 20× magnification staining HNRNPL in both tumor and matched benign prostate tissues. (Scale bar, 200 μm.) (*C*) Statistics of tissue microarray analysis on HNRNPL protein expression. Seventy-nine pairs of tumor and matched normal biopsies were analyzed. Combined score adapted from the Allred scoring system taking into account both staining intensity and percentage was used to quantify the protein expression. (*D*) Gene-signature association network between HNRNPL-regulated RNA targets (either alternatively spliced genes from Dataset S3 or circRNA genes from Dataset S4) and published gene signatures deposited in Oncomine database. Each node represents a gene set signature and each edge represents a significant association between two connected gene sets ($P \leq 4.91E\text{-}4$). The size of node is proportional to the gene numbers within gene sets. The thickness of the edge is proportional to the significance of the association with thicker edge meaning more significant association. (*E*) Representative gene heatmap showing the HNRNPL-regulated alternatively spliced genes or circRNA genes are correlated with overexpressed gene signatures in prostate cancer versus normal prostate in Vanaja Prostate cohorts in *D*.

Discussion

Posttranscriptional events, including alternative splicing and circular RNA formation, are increasingly appreciated to play important roles in cancer progression. Factors regulating RNA biogenesis including RBPs have been implicated in multiple fundamental biological processes, both physiologically and pathologically. However, which and how RBPs are functionally important in the context of prostate cancer still remains largely unknown. To this end, we systematically identified the genes that are essential for prostate cancer growth by a genome-wide CRISPR screen and found HNRNPL to be a top essential RBP.

To determine the landscape of HNRNPL-associated RNAs, we performed HNRNPL RIP-seq and found that HNRNPL binds preferentially to CA-repeats or CA-enriched RNA motifs in introns and 3′UTR regions. Although these observations agree with two recent studies of HNRNPL–RNA interactions by cross-linking immunoprecipitation (CLIP)-seq in T cells and iCLIP-seq in HeLa cells, respectively (43, 44), we identified different RNA targets. This result may be because of context-specific HNRNPL–RNA interactions, as our RIP-seq data are aimed to understand HNRNPL function in prostate cancer.

Several hnRNP family proteins are known to function in alternative splicing (45). HNRNPL has been shown previously to enhance and repress alternative splicing of different targets (30, 46). Here we systematically profiled HNRNPL-dependent alternative splicing events in prostate cancer cells and integrated this with RIP-seq data to obtain a comprehensive view of direct HNRNPL-regulated alternative splicing. The identification of AR as an HNRNPL-regulated splicing target is interesting, in light of the central role of AR in prostate cancer growth. However, it is unlikely that HNRNPL loss-of-function decreases cell growth uniquely through the regulation of AR, for the following reasons. First, overexpression of the full-length AR only partially rescues the cell-growth phenotype caused by HNRNPL knockdown (Fig. S3*D*). Second, the magnitude of full-length AR reduction is not sufficient to significantly affect androgen/AR-regulated target gene expression, nor does androgen affect genome-wide HNRNPL–RNA interactions (Fig. S4 *D–F*). Third, despite lower expression of the full-length AR upon HNRNPL knockdown, the increase of the short AR isoform, corresponding to a constitutively active variant that could bind to DNA and execute transcription without androgen stimulation, would be expected to compensate for the reduction of full-length AR in regulating gene expression and cell growth (47). Fourth, HNRNPL is required in AR-null prostate cancer cell lines (Fig. S2*C*), indicating that AR is not the only key target of HNRNPL. Consistently, the HNRNPL-regulated splicing targets are collectively associated with the overexpressed gene signatures in prostate cancer, implying that these targets may together contribute to mediate HNRNPL function (Fig. 6 *D* and *E*).

A very limited number of factors have been shown to regulate circRNA formation (32, 36, 37). Our study added HNRNPL to this list and suggested that a broader range of RBPs, especially splicing factors, might be involved in circRNA regulation. Minigene assays of HNRNPL confirmed the previous model that *trans* factor binding or direct complementary sequence pairing at two flanking introns could promote circRNA formation by bringing back spliced exon ends into close proximity (32, 36, 37). However, the pattern of circRNA regulation in vivo may prove to be more complicated because HNRNPL could either positively or negatively interfere with the splicing machinery that is responsible for circRNA generation. The diverse association pattern of HNRNPL binding and circRNA regulation also implies that *trans* factors do not have to bind only at the flanking introns to affect circRNA expression, broadening the way that circRNAs are regulated. It is worth noting that the HNRNPL-regulated circRNA genes are relevant to prostate cancer progression, implying the potential function of the circRNAs generated from these genes.

Considering the many other proposed functions of HNRNPL, such as affecting RNA half-life or translation efficiency through 3′UTR binding (28, 48), our investigation here may not be exhaustive. We found that HNRNPL regulates its RNA targets through either linearly alternative splicing or back-spliced circRNA formation, which may collectively contribute to its essentiality in prostate cancer. Our study not only identified a cohort of prostate cancer dependency genes, but also revealed how HNRNPL orchestrates RNA biology, suggesting that targeting RBPs or RBP–RNA interactions may be a novel therapeutic approach in prostate cancer.

Materials and Methods

Cell Culture. LNCaP, CWR22Rv1, DU145, and PC3 cells were cultured in RPMI medium 1640 supplemented with 10% FBS. RWPE-1 cells were cultured in Keratinocyte Serum Free Medium (K-SFM; Kit Cat. no. 17005-042) with kit-supplied bovine pituitary extract and human recombinant epidermal growth factor. The use of human prostate samples for tissue microarray analysis has been approved by The Gelb Center Committee at Dana-Farber Cancer Institute.

CRISPR Screen. For the pooled genome-wide CRISPR screen, 1×10^8 LNCaP cells were infected with the pooled lentiviral GeCKO v2 library at a multiplicity of infection of 0.5. After 3 d of puromycin selection, half of the surviving cells were stored as 0-d control samples, and the rest of cells were cultured for an additional 2 wk. PCR was performed on genomic DNA to construct the sequencing library. Each library was sequenced at 30∼40 million reads to achieve ∼300× average coverage over the CRISPR library and data were analyzed by MAGeCK and MAGeCK-VISPR (21, 22).

siRNA Knockdown. LNCaP cells were seeded in 24-well plates and transfected with 20 nM siRNA oligos using RNAiMax reagent (Life Technology). Knockdown efficiency was determined after 72 h of transfection. Cell counting was performed after 6 d of transfection. Detailed siRNA target sequences can be found in *SI Materials and Methods*.

RNA Isolation and qRT-PCR. RNA was isolated using the RNeasy Mini Kit (Qiagen). Reverse-transcriptase (Invitrogen) was used for random-primed first-strand cDNA synthesis. Real-time PCR was carried out on ABI Prism 7300 detection system using SYBR Green PCR master mix. The ΔΔCt method was used to comparatively quantify the amount of mRNA level. RPS28 gene expression served as the internal control. Primer sequences can be found in *SI Materials and Methods*.

RIP. Cells were cross-linked with 0.3% formaldehyde for 10 min at room temperature and lysed with RIPA lysis buffer for 10 min on ice before sonication. The supernatant was collected and incubated with RIP antibodies for 4–6 h. After washing with RIPA buffer, RNA was eluted from the beads with 100 μL $NaHCO_3$ and 1% SDS in the presence of proteinase K and RNase inhibitor at room temperature for 10 min with occasional vortex. The eluted material was decross-linked at 65 °C for 45–60 min before purification of RNA using TRIzol LS reagent (Life Technology). DNase I treatment was performed to remove any residual DNA. RIP RNA was used for either library preparation or direct qPCR assay.

Next-Generation Sequencing. All RIP-seq and RNA-seq libraries were constructed with TruSeq Stranded mRNA Library Prep Kit from Illumina according to the manufacturer's manual. RIP-seq libraries were sequenced as single end reads in duplicates, whereas RNA-seq libraries were sequenced as paired end reads in triplicates (splicing analysis) or duplicates (circRNA analysis) at ∼50 million reads per library using either HiSEq. 2000 or NextSEq. 500 sequencing platforms.

Bioinformatics Analysis and Statistics. De novo motif analysis was performed using MDscan and SeqPos that were implemented in the Cistrome package. The RIP-seq and RNA-seq reads were aligned against the hg19 human reference genome with the University of California, Santa Cruz (UCSC) known gene transcript annotation using TopHat v2.0.9. Because the RIP-seq and RNA-seq data are strand-specific, the RIP-seq peaks were identified for "+" and "−" strands separately using macs2 v2.0.10, with a scanning window size of 100 bps but without shifting the reads [peak length ≥ 150, false-discovery rate (FDR) ≤ 0.01, fold-change ≥ 4, and RIP read count ≥ 50]. The differential

alternative splicing analysis was performed using rMATS, a computational tool that allows for the analysis of replicate RNA-seq data. The circular RNA was identified using the CIRCexplorer method (see details in *SI Materials and Methods*). The differential expression analysis of circRNAs was performed using LIMMA (FDR ≤ 0.35, fold-change ≥ 2), and the total normalized count no less than 15. Moreover, only those circRNAs, whose parental genes showed little expression change (fold-change ≤ 1.2), were considered differentially expressed. Notably, increasing sequencing depth would further enhance the power and robustness to identify or calculate the differential expression of circRNAs. Statistical analysis of tissue microarray data were performed using SPSS software, v22.0 (SPSS).

ACKNOWLEDGMENTS. We thank Dr. Albrecht Bindereif and Dr. Shaoyong Chen for sharing of reagents; and Dr. Ting Ni, Miao Han, and all members of the M. Brown and X.S.L. laboratories for helpful discussions. This project was supported by Department of Defense Prostate Cancer Research Program Postdoctoral Training Award W81XWH-14-1-0513 (to T.F.) and National Cancer Institute Grants P50 CA090381-13 (to L.C.), 5P01 CA163227 (to M. Brown), NIH R01 HG008927 (to X.S.L.), and DOD PC140817P1 (to M. Brown and X.S.L.).

1. Scher HI, Sawyers CL (2005) Biology of progressive, castration-resistant prostate cancer: Directed therapies targeting the androgen-receptor signaling axis. *J Clin Oncol* 23:8253–8261.
2. Feldman BJ, Feldman D (2001) The development of androgen-independent prostate cancer. *Nat Rev Cancer* 1:34–45.
3. Debes JD, Tindall DJ (2004) Mechanisms of androgen-refractory prostate cancer. *N Engl J Med* 351:1488–1490.
4. Shen MM, Abate-Shen C (2010) Molecular genetics of prostate cancer: New prospects for old challenges. *Genes Dev* 24:1967–2000.
5. Asangani IA, et al. (2014) Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* 510:278–282.
6. Manley JL, Tacke R (1996) SR proteins and splicing control. *Genes Dev* 10:1569–1579.
7. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG (1993) hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62:289–321.
8. Zhang J, Manley JL (2013) Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* 3:1228–1237.
9. Hsu TY, et al. (2015) The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* 525:384–388.
10. Chen J, Weiss WA (2015) Alternative splicing in cancer: Implications for biology and therapy. *Oncogene* 34:1–14.
11. Dreyfuss G, Kim VN, Kataoka N (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3:195–205.
12. Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80–84.
13. Shalem O, et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84–87.
14. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MdelC, Yusa K (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 32:267–273.
15. Hart T, et al. (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163:1515–1526.
16. Munoz DM, et al. (2016) CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov* 6:900–913.
17. Zhou Y, et al. (2014) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509:487–491.
18. Chen S, et al. (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160:1246–1260.
19. Parnas O, et al. (2015) A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162:675–686.
20. Sanjana NE, Shalem O, Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11:783–784.
21. Li W, et al. (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15:554.
22. Li W, et al. (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* 16:281.
23. DeMiguel F, et al. (2002) Stat3 enhances the growth of LNCaP human prostate cancer cells in intact and castrated male nude mice. *Prostate* 52:123–129.
24. Hoefer J, et al. (2012) PIAS1 is increased in human prostate cancer and enhances proliferation through inhibition of p21. *Am J Pathol* 180:2097–2107.
25. Robinson D, et al. (2015) Integrative clinical genomics of advanced prostate cancer. *Cell* 161:1215–1228.
26. Aguirre AJ, et al. (2016) Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov* 6:914–929.
27. Hnisz D, et al. (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155:934–947.
28. Lee DH, et al. (2009) hnRNP L binds to CA repeats in the 3′UTR of bcl-2 mRNA. *Biochem Biophys Res Commun* 382:583–587.
29. Hui J, et al. (2005) Intronic CA-repeat and CA-rich elements: A new class of regulators of mammalian alternative splicing. *EMBO J* 24:1988–1998.
30. Chiou NT, Shankarling G, Lynch KW (2013) hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. *Mol Cell* 49:972–982.
31. Shen S, et al. (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA* 111:E5593–E5601.
32. Zhang XO, et al. (2014) Complementary sequence-mediated exon circularization. *Cell* 159:134–147.
33. Memczak S, et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495:333–338.
34. Hansen TB, et al. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495:384–388.
35. Lasda E, Parker R (2014) Circular RNAs: Diversity of form and function. *RNA* 20:1829–1842.
36. Conn SJ, et al. (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160:1125–1134.
37. Ashwal-Fluss R, et al. (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 56:55–66.
38. Agell L, et al. (2012) A 12-gene expression signature is associated with aggressive histological in prostate cancer: SEC14L1 and TCEB1 genes are potential markers of progression. *Am J Pathol* 181:1585–1594.
39. Brase JC, et al. (2011) TMPRSS2-ERG-specific transcriptional modulation is associated with prostate cancer biomarkers and TGF-β signaling. *BMC Cancer* 11:507.
40. Taylor BS, et al. (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18:11–22.
41. Weinstein JN, et al.; Cancer Genome Atlas Research Network (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45:1113–1120.
42. Rhodes DR, et al. (2007) Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* 9:443–454.
43. Rossbach O, et al. (2014) Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol* 11:146–155.
44. Shankarling G, Cole BS, Mallory MJ, Lynch KW (2014) Transcriptome-wide RNA interaction profiling reveals physical and functional targets of hnRNP L in human T cells. *Mol Cell Biol* 34:71–83.
45. Han SP, Tang YH, Smith R (2010) Functional diversity of the hnRNPs: Past, present and perspectives. *Biochem J* 430:379–392.
46. Hung LH, et al. (2008) Diverse roles of hnRNP L in mammalian mRNA processing: A combined microarray and RNAi analysis. *RNA* 14:284–296.
47. Dehm SM, Tindall DJ (2011) Alternatively spliced androgen receptor variants. *Endocr Relat Cancer* 18:R183–R196.
48. Majumder M, et al. (2009) The hnRNA-binding proteins hnRNP L and PTB are required for efficient translation of the Cat-1 arginine/lysine transporter mRNA during amino acid starvation. *Mol Cell Biol* 29:2899–2912.
49. Wang T, et al. (2015) Identification and characterization of essential genes in the human genome. *Science* 350:1096–1101.
50. Warde-Farley D, et al. (2010) The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38:W214–220.
51. Liu XS, Brutlag DL, Liu JS (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* 20:835–839.
52. Liu T, et al. (2011) Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biol* 12:R83.
53. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
54. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111.
55. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.
56. Kim D, Salzberg SL (2011) TopHat-Fusion: An algorithm for discovery of novel fusion transcripts. *Genome Biol* 12:R72.
57. McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40:4288–4297.
58. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
59. Ritchie ME, et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.
60. Cerami E, et al. (2012) The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2:401–404.
61. Gao J, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6:pl1.
62. Wang Q, et al. (2009) Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* 138:245–256.