Long non-coding RNA expression levels modulate cell-type specific splicing patterns by altering their interaction landscape with RNA-binding proteins

Felipe Wendt Porto[1], Swapna Vidhur Daulatabad[1], Sarath Chandra Janga[1, 2, 3, *]

Affiliations:

[1]Department of BioHealth Informatics, School of Informatics and Computing, IUPUI, Indianapolis, IN, USA.

[2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA.

[3]Centre for Computational Biology and Bioinformatics Indiana University School of Medicine, Indianapolis, IN, USA.

Corresponding author*

Email addresses:

FWP: fporto@iu.edu

 SVD: swapdaul@iupui.edu

*SCJ: scjanga@iupui.edu

Abstract

Background: Recent developments in our understanding of the interactions between long non-coding RNAs (lncRNAs) and cellular components have improved treatment approaches for various human diseases including cancer, vascular diseases, and neurological diseases. Although investigation of specific lncRNAs revealed their role in the metabolism of cellular RNA, our understanding of their contribution to post-transcriptional regulation is relatively limited. In this study, we explore the role of lncRNAs in modulating alternative splicing and their impact on downstream protein-RNA interaction networks.

Results: Analysis of alternative splicing events across 39 lncRNA knockdown and wildtype RNA-sequencing datasets from three human cell lines: HeLa (Cervical Cancer), K562 (Myeloid Leukemia), and U87 (Glioblastoma), resulted in high confidence (fdr < 0.01) identification of 11630 skipped exon events and 5895 retained intron events, implicating 759 genes to be impacted at post-transcriptional level due to the loss of lncRNAs. We observed that a majority of the alternatively spliced genes in a lncRNA knockdown were specific to the cell type, in tandem, the functions annotated to the genes affected by alternative splicing across each lncRNA knockdown also displayed cell type specificity. To understand the mechanism behind this cell-type specific alternative splicing patterns, we analyzed RNA binding protein (RBP)-RNA interaction profiles across the spliced regions, to observe cell type specific alternative splice event RBP binding preference.

Conclusions: Despite limited RBP binding data across cell lines, alternatively spliced events detected in lncRNA perturbation experiments were associated with RBPs binding in proximal intron-exon junctions, in a cell type specific manner. The cellular functions affected by alternative splicing were also affected in a cell type specific manner. Based on the RBP

binding profiles in HeLa and K562 cells, we hypothesize that several lncRNAs are likely to exhibit a sponge effect in disease contexts, resulting in the functional disruption of RBPs, and their downstream functions. We propose that such lncRNA sponges can extensively rewire the post-transcriptional gene regulatory networks by altering the protein-RNA interaction landscape in a cell-type specific manner.

Keywords

Long non-coding RNA, cell type specific, alternative splicing, functional enrichment, RNA-binding proteins, protein binding lncRNA sponges, secondary RNA structure, and cancer.

Introduction

One of the major challenges in the post-genomic era is to understand fundamental mechanisms of long non-coding RNAs' (lncRNAs) and their role in modulating cellular homeostasis. Despite increased awareness of their influence on alternative splicing and alterations in notable cancers: most lncRNAs are still not known to have a function and thousands of lncRNAs are believed to be a result of transcriptional noise[1]. Their length and low expression have acted as a barrier for experimental assays as well as for building computational models [2,3], resulting in the lack of approaches to confidently assess the full scope of lncRNA function and structure. However, with the help of novel genome editing technologies such as "catalytic dead" Cas9 (dCas9) silencing [4], genome-scale probing methods which study the loss of function phenotypes for lncRNAs are beginning to emerge.

By combining sequential dCas9-KRAB guides predicted via *in silico* models, effective knockdowns of whole lncRNA genes have been recently made, providing a promise for expanding genome-wide knockdown protocols[4]. With the help of such improved lncRNA knockdown experiments, one can begin to confidently analyze and associate the downstream phenotypic changes and resulting interaction networks of lncRNAs across cell types. CRISPR guided knockdown assays [5] and large scale lncRNA functional analysis [6] demonstrate the translational significance of CRISPR mediated lncRNA perturbations.

Other studies of lncRNAs have repeatedly shown that lncRNAs perturb or enhance biological functions in a cell-type or tissue-specific manner [7,8]. Putative lncRNAs like HOTAIR, NEAT1 and MALAT1 have been demonstrated to interact with specific RNA binding proteins (RBPs) like HUR and ELAVL1, in a tissue specific fashion [8,9]. However, few

4

databases have curated useful cross-linking immunoprecipitation (CLIP) [10] data with lncRNAs and annotated functional effects that match the quality of Seten, [11] CLIPdb's POSTAR2 [12], and ENCODE [13]. RBPs have been shown to influence alternative splicing and their interaction with pre-mRNAs [14], thereby impacting cellular functions and phenotypes. Yet, in a post-transcriptional regulatory context, our understanding of the association between RBPs and the non-coding transcriptome remains unclear.

To understand the influence of lncRNAs on alternative splicing, we integrate: dCAS9 mediated lncRNA knockdown RNA-sequencing data [4] across three cancer cell lines along with protein-RNA interaction maps from eCLIP[15] experiments performed as part of the ENCODE project. This study examines the role of lncRNAs in modulating the cell-type specific alternative splicing outcomes due to their physical interactions with RBPs.

Materials and Methods

To dissect the functional impact of lncRNAs on splicing, we downloaded publicly available RNA-seq data from a study (GSE85011)[4] where multiple lncRNAs were knocked down across 3 human cancer cell lines. As illustrated in the workflow (Figure 1), the RNA-seq data was aligned to the human reference genome. The aligned data was further analyzed for identifying alternative splicing events across lncRNA knockdowns and corresponding control samples. RBP binding profiles were analyzed to understand their binding preference around alternative splice events. The following sections discuss individual steps involved in our workflow in detail.

Sequence alignment of RNA-Seq reads using HISAT

Data from the study GSE85011 was generated by creating a guide RNA library to facilitate CRISPRi lncRNA knockdowns via precise heterochromatization[4]. Single end RNA-seq data of lncRNA knockdowns and controls in the HeLa (cervical cancer), U87 (glioblastoma), and K562 (leukemia) cell lines, was downloaded from GEO[16] in the 'fastq' format using NCBI's sratoolkit (version 2.8.2) [17]. We used a highly efficient short-read RNA-seq alignment tool, HISAT2 [18], to align the downloaded data to the human reference genome. Default parameters of HISAT2 were used to align the reads against hg38 reference genome. By utilizing SAMtools (version 1.3.1) [19] for data processing, the SAM (Sequence Alignment/Map) files from HISAT2 alignment were converted to the BAM (Binary Alignment/Map) format. Concurrently, the BAM files were transformed into the sorted-BAM format. The average alignment rate across the samples was 91%. The corresponding SRA sample run IDs (SRR IDs) and their respective alignment statistics are reported in Supplementary table 1.

Identifying differentially spliced events using rMATS

RNA-seq data enables us to observe alternative splicing events across two given conditions. In this study we deployed rMATS (version 3.2.5) [20] on samples corresponding to 39 lncRNA knockdowns and their respective controls, to identify the differential alternative splice (AS) events. rMATS takes sorted-BAM files as input and identifies a variety of splicing events, that are differential across two conditions. We used sorted-BAM files of lncRNA knockdowns, control reads, and corresponding replicates as input for rMATS. Additionally, human transcript annotation was provided with a GTF (Gene transfer format,

version 84) file from Ensembl [21]. By implementing our inputs into the pipeline with default thresholds, we could compare AS patterns in the presence and absence of a lncRNA. Thereby, rMATS enabled us to analyze the inclusion/exclusion of target exons/introns pertaining to different types of AS events: skipped exon (SE), alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), mutually exclusive exons (MXE), and retained intron (RI) across lncRNA knockdown experiments. We ran rMATS on 54 comparisons of samples pertaining to 39 lncRNA knockdowns in three cell lines; however, not all the lncRNAs had knocked down RNA-seq data available across the three cell lines.

Custom python scripts were developed to parse through rMATS output files across all comparisons to generate AS event frequency matrices. A filter of False Discovery Rate < .10 and p-value < .05 was employed over these tables, thereby enabling us to extract only the significant AS events [Supplementary table 3].

Visualizing incidence of differentially spliced genes in lncRNA knocked down cell lines

The frequency matrices generated from the rMATS output were further analyzed to gauge how often each gene was alternatively spliced when a specific lncRNA is knocked down in a given cell line. To optimize data processing and visualization, the matrix was further filtered to remove instances which were not populated with data. LncRNA's genome wide influence over splicing was visualized by generating a heatmap that depicts the incidence of a gene being alternatively spliced in a lncRNA knock down by using Morpheus[22], a heatmap generation tool from broad institute. Both, rows and columns were clustered hierarchically based on the "one minus spearman rank correlation" distance option. The spearman distance option was the most relevant method to discriminate the

genes' cell type specificity, as it was a non-parametric test and gave the most consistent groupings after trials of similar distance metrics. We used Venny (v2.1) [23], to generate Venn Diagrams to depict the differentially spliced gene overlaps across cell lines.

Functional enrichment analysis of alternatively spliced genes

To understand the functional impact of lncRNAs on cellular functions, we performed functional enrichment analysis using ClueGO (v2.50) [24] over alternatively spliced gene sets from each lncRNA knockdown. ClueGO is a Cytoscape[25] plug-in which enables users to perform comprehensive functional interpretation of gene sets. ClueGO can analyze gene sets and visualize the respective functionally grouped annotations based on pathway enrichments recognized in the corresponding gene sets. We extracted the gene sets that were exclusively alternatively spliced in a given cell line and plugged them into ClueGO, thereby extracting functional contributions of lncRNAs. We used all the GO terms irrespective of their level in the GO hierarchy for identifying enrichment terms after searching for term annotations from databases like KEGG[26], GO Cellular Components[27,28], GO Biological Process[27,28], and REACTOME Pathways[29]. Only significantly enriched terms (p-value < .05) were used for the downstream analysis. The functional enrichment of the smaller gene sets affected by alternative splicing across cell types revealed no significant (pvalue < .05) pathways or functional annotations.

Extraction of RBP binding profiles across AS events

Experimental strategies like CLIP[10], iCLIP[30], and eCLIP[15] have significantly improved our understanding of the RNA bound proteome. By observing variations of RBP binding preferences around AS events in the presence and absence of a given lncRNA, we gained a new perspective on the functional dynamics and crosstalk between lncRNAs and RNA binding proteins (RBPs). Publicly available eCLIP experimental data in the HeLa and K562 cell lines were obtained from the ENCODE project [13]. The eCLIP data was then processed to extract the genome level RBP binding profiles in both cell lines.

Additionally, we extracted the genomic loci where the AS events occurred from the rMATS output data. Potential genomic loci involved in AS can be captured by examining adjacent upstream and downstream regions in the exon start and end loci. A total of six types of proximal coordinates were extracted: upstream exon start and end, AS event specific exon start and end, and the downstream exon start and end. Each of the six coordinates were given a distance allowance of 500 base pairs on either side of the sites to confirm locations which overlap with the binding sites of RBPs. Both, RBP binding profiles and AS event sites were processed and converted into the BED (Browser Extensible Data) format. Bedtools (version 2.18.1) [31] is one of the most efficient tools to analyze genomic loci based information. We used the Bedtools-intersect option to extract the overlaps across RBP binding sites and AS events and proximal loci, to understand which RBPs are potentially associated with a specific AS event [see Supplementary materials 11]. However, the eCLIP data on ENCODE did not host high confidence U87 cell line experiments, therefore no RBP binding profiles from U87 were extracted.

Enrichment of RBP binding over alternative splicing events

Although we have identified RBP and AS event interactions based on eCLIP data, we wanted to identify only the most likely interactions for further analysis. A hyper geometric statistical test was performed to observe the enrichment of a RBP's binding preference with respect to each lncRNA that was knocked down, thereby, determining significant relative RBP binding preference.

The relationship between RBP binding frequencies was highly influenced by many missing values, which might not be representative of the whole relationship of binding. The importance of a reported binding had no numerical value attached, so a hypergeometric probability was needed. A majority of the lncRNAs had a higher proportion of binding rather than non-binding, and the fishers exact odds ration test evaluated the representation of binding. A fisher's exact test ran for each of the bound and unbound genes to a motif created by a lncRNA knockdown and all the genes from each lncRNA. The p-values from each fisher's test were then adjusted in reference to all the other fisher's exact tests.

Table 1 shows the contingency table of bound and unbound protein frequencies used to conduct the fishers exact test. Each square in the heatmaps shown in Figure 4A and 4B demonstrate the significance (-log(corrected pvalue)) from each of these tests. The maps were clustered hierarchically to see if there was a relationship between RBP and lncRNAs.

Correlation analysis of lncRNA specific RBP binding activity and AS events

A correlation analysis was performed across the number of skipped events with the amount of RBPs relative and each lncRNA specifically to determine a relationship between RBP binding frequency and alternative splicing events. However, a significant correlation was not observed given the small amount of data points. Supplementary table 4 has the attempts and tables used to show the connection between the lncRNAs and RBP.

Results

Framework for studying alternative splicing outcomes in lncRNA knockdown RNA-sequencing experiments

In this study, we investigated alternative splicing events, functions of alternatively spliced gene sets, and RBP-lncRNA binding patterns across 39 lncRNA knockdowns in HeLa (Cervical Cancer), K562 (Myeloid Leukemia), and U87 (Glioblastoma) cell lines. An overview of the three-step analysis is illustrated in Figure 1 (see Materials and Methods). First, we collected and processed the raw RNA-Seq data for multiple lncRNA knockdown samples in human cancer cell lines and identified splicing events using the replicate Multivariate Analysis of Transcript Splicing (rMATS) [20] pipeline. Secondly, alternative splicing summaries were analyzed, and functional enrichment analysis of the effected gene sets was performed. Third, we analyzed the binding profiles of RNA binding proteins (RBPs) with respect to the skipped exon (SE) and retained intron (RI) events across 39 lncRNA knockdowns: finding that the RBPs binding profiles around SE and RI events were also unique to each cell line. Additionally, we observed RBPs' substantial binding preference for

11

knocked down lncRNAs and proposed that certain lncRNAs demonstrate a sponge-like RBP binding activity.

Skipped Exon events followed by Retained Intron events are the most prominent alternative splicing events occurring due to the loss of lncRNAs across multiple human cell lines

We collected the RNA sequencing data from dCAS9 based lncRNA knockdown study [4] which contained 39 lncRNA knockdowns in U87, HeLa, and K562 cell lines (see Materials and Methods). LncRNA knockdown sequence replicates and control samples were aligned onto the human reference genome (hg38) using HISAT2[18]. The overall percentage of alignment is highlighted in Supplementary table 1 demonstrates a high fraction of read alignment to the reference genome (average alignment rate ≥ 91%). The alternative splicing (AS) event identification analysis was performed by deploying the rMATS pipeline over the three cell lines. A filter of fdr < 0.1 and p-value < 0.05 was employed to extract the most likely alternative splicing events (see Materials and Methods). A total of 26167 high confidence AS events were extracted from the rMATS analysis output across three cell lines. Skipped exon (SE) events were the most predominant events, constituting for 44.4% of the total events, followed by retained intron (RI) events at 22.5%. Sample wise distributions of splice events are provided as Supplementary table 2. The output from the rMATS analysis was parsed and summarized into a total of 17525 unique, statistically significant splice events (11630 SE and 5895 RI) and their AS event frequency across 39 lncRNA knockdowns in the three cell lines is available as Supplementary table 3.

Hierarchically clustered heatmaps of AS event frequency in lncRNA knockdowns reveal cell type specific alternative splicing signatures

Annotated splicing summaries were compiled to generate matrices which depicted the frequency of alternative splicing events occurring in lncRNA knockdowns across three different cell lines. In order to visualize the AS events from the matrix, heatmaps were generated for SE and RI events. A legible resolution of the SE and RI heatmaps were obtained after missing data was filtered out. Across RI and SE events, we observed that the alternatively spliced gene clusters were predominantly unique across each cell line [Supplementary figures 4 and 5]. Consequently, even though the same lncRNA was knocked out across two cell lines, the genes that were alternatively spliced are cell type specific. SE events had the most pronounced cell type specificity, in comparison to other AS events. In tandem, we also observed that the cellular and phenotypic functions affected by genes alternatively spliced across SE and RI events were specific to each cell line.

Among the lncRNA knockdown dataset described in the materials and methods section, we found only 12 lncRNAs to be common across three cell lines. Amongst all the AS events in the 3 different cell lines, not a single gene was affected in the same way by a knockdown. The resulting alternatively spliced genes from the 12 lncRNA knockdowns which were present in at least 2 cell types were observed to be cell type specific. Figure 2 illustrates how the cell lines which had a lncRNA knocked down affected the frequency of SE events in specific genes. The genes affected by RI events were altered in a cell type specific manner as well. However, the intensity of the signal was relatively lower in RI events when compared to SE events.

In the AS event frequency heatmap, the SE events had very defined clusters and continued to be clustered in a distinct cell type specific manner [see Supplementary figure 5]. The genes which were annotated with SE events, were rarely shared across two cell lines. Out of the total 759 alternatively spliced genes, only 7.3% (37 genes) were alternatively spliced across at least 2 cell lines. When enrichment analysis was performed over the few SE affected genes which were shared across different cell lines, there was too little information for a statistically significant functional annotation to emerge. The number of SE event affected genes shared across cell lines can be seen in the Venn Diagram in Figure 2. Similarly, a heatmap of RI events frequencies was generated, akin to the observations from SE event analysis, the gene sets with RI events were mostly cell type specific [see Supplementary figure 5].

The gene sets affected by RI events were cell type specific, but these gene sets overlapped slightly more among the cell lines, relative to SE distributions. Given the small number of genes alternatively spliced and RI events shared (maximum of 30 shared events), the scope for significant functional annotation was low. The 162 genes affected by RI events in the K562 cell line had a distinct signal from the other cell lines, further emphasizing their cell type specificity.

lncRNA knockdowns not only enriched for cell type specific functions but occasionally favor similar functions via varied alternatively spliced gene sets

In an attempt to understand the downstream functions being influenced by lncRNAs, functional enrichment analysis was performed over 759 alternatively spliced genes corresponding to 39 lncRNA knockdowns in three cell lines (see Materials and Methods).

14

Only 28 genes were shared between the SE and RI events, while 481 genes were unique to SE events, and 250 genes were unique to RI events. Figure 3 illustrates the significant ($p < 0.05$) potential functions associated to genes alternatively spliced in each cell line resulting from a lncRNA knockdown. As expected, the functions affected within each cell line, were also unique to each cell line. On the rare occasion of a common enriched function between cell lines, gene set composition leading to that function varied drastically.

Genes alternatively spliced via SE events in U87, HeLa, and K562 reveal cell type specific functional enrichment

SE events affected a total of 509 genes, and the respective functional annotation terms were enriched from four different pathway databases (see Materials and Methods). A unique gene set of 138 genes corresponding to 14 terms were enriched in the U87 cell line. The highest number of genes that were alternatively spliced in samples were induced by the knockdown of lncRNA families XLOC and RP11. The U87 cell type specific gene cluster has areas of higher magnitude of splicing events and shared only 25 genes with the K562 cell line. Of the 25 shared genes no significant pathways were annotated. The top functions that are affected by the alternative splicing induced by the lncRNA knockdowns in the U87 cell line correspond with DNA endonuclease repair (37% of genes), and with ribosomal complex formation (16% of genes) [see Supplementary figure 6]. These functional observations are in coherence with literature, as it is known that endonuclease repair activity is a function targeted by cancerous U87 [32].

The 141 genes affected in the K562 cell line were enriched for a total of 17 functional terms. The top functions affected by SE events in the K652 cell line were: cell microfiber construction (45% of genes), phosphotransferase activity (23% of genes), and p53 signaling pathway (12% of genes) [see Supplementary figure 7]. Functional annotations like cell growth and cell death have been identified as key checkpoints in various cancers, including leukemia (K562) [33]. However, this study appears to provide a novel functional annotation of 'cell microfiber construction' being affected in leukemia (K562). Thereby, our genome wide functional analysis is not only able to predict novel functions based on genes affected by AS events, but also support functions previously annotated to these cell lines.

The HeLa clusters affected genes in a large gene cluster, with 193 unique genes over represented with 28 functional terms. The top functions affected by SE events in the HeLa cell line were: deoxyribonucleoside monophosphate metabolic processes (29% of genes), and vesicle formation and vesicle movement (15% of genes) [see Supplementary figure 8]. Vesicle formation and movement have been observed to be attenuated in the HeLa cell line, in a gene (SPIN90) dependent manner, which our analysis was able to capture [34]. No significant pathways were annotated for the 12 genes the Hela cell line shared with the K562 cell line.

Genes alternatively spliced via RI events in HeLa, and K562 reveal cell type specific functional enrichment

RI events affected a total of 278 genes across three cell lines. The respective functional annotation terms were enriched from four different pathway databases (see Materials and Methods). While the clustered gene set in the HeLa cell line yielded no significant (p < 0.05)

16

pathways, the knockdown of RP5-1148A21.3 in the HeLa cell line was able to alternatively splice all 278 genes [see Supplementary figure 5]. Thus, the 278 genes affected in the HeLa cell line by RP5-1148A21.3, enriched 53 functional terms: responses to endoplasmic reticulum stress (19% of genes), holiday junction resolvase complex (9% of genes), and RNA splicing (8% of genes). The few genes affected within the U87 cell line did not enrich any significant pathways.

The 162 genes affected in the K562 cell line were enriched for 18 functional terms composed of concepts such as 'ribosomal construction' and 'negative regulation of autophagy'. 'Cytosolic large ribosomal subunit' constituted for 50% of the genes alternatively spliced in the K562 gene set [see Supplementary figure 9]. Other prominent functions associated with the K562 were ribosomal RNA processing in the nucleus and cytosol (7% of genes), and negative regulation of autophagy. The functional annotation of the term 'targeted rRNA processing', onto K562, is in coherence with literature; as 'targeted rRNA processing' has been identified to play a key role in K562 cells [35].

Analysis of lncRNA AS events proximal to RBP binding sites reveals cell type specific interactions and supports a lncRNA-RBP sponge model

As a conduit to understand lncRNA's role in alternative splicing: lncRNA's interactions with RBPs were extracted. RBP binding profiles for 22 lncRNAs were obtained from documented eCLIP experiments from the ENCODE database. Supplementary material 11 highlights the binding profiles which overlapped with proximal (±500bp) alternative splice site locations revealing 4261897 RBP binding locations for 148 RBPs on 22 lncRNAs. The significance of relative frequencies of bound and unbound RBP sites was gauged by

deploying the fisher's exact test on each lncRNA's RBP binding preferences. Figure 4 showcases the intensity of each RBPs' interactions to their respective 11 lncRNAs in the K562 cell line (Figure 4A) and 14 lncRNAs in the HeLa cell line (Figure 4B). Additionally, we observed lncRNAs which acted as an RBP sequestering sponge, which is illustrated in Figure 4C, based on their extensive interactions with RBPs. Figure 4D demonstrates how sponging lncRNAs like LINC00909 have many interactions with a variety of RBPs.

The lncRNA – RBP binding profile-based clustering analysis across both cell lines was not very informative. However, an interesting behavior was revealed where certain lncRNAs, like LINC00909, LINC00263, and LINC00910, had extensive number of binding events across many RBPs. Therefore, the downstream alternative splicing caused by the loss of a lncRNA is induced by the absence of an RBP binding sponge. As highlighted in the model shown in Figure 4C, in cancer cells a lncRNA might bind to many RBPs where its expression level could facilitate extensive number of RBP interactions. However, in the event of a lncRNA knockdown or due to the loss of function of a lncRNA, an abundance of RBPs interact with pre-mRNA targets illustrated in Figure 4C, thereby inducing alternative splicing.

As highlighted in Figure 4A, most lncRNAs in the K562 cell line bound generally to RBPs like: KHSRP, CSTF2T, YBX3, ZNF622, SAFB2, SRSF1, and QKI. LncRNAs LINC00910, LINC00680, RP11-392P7.6, and LINC00909 showed a very high number of RBP interactions in the K562 cell line [see Supplementary figure 4] (Figure 4D) and exemplify the proposed RBP-sponge binding model. Other interesting patterns of lncRNA-RBP binding included distinct RBP binding preferences for lncRNAs from the same family, namely XLOC_042889 and XLOC_038702.

Despite only having 10 RBPs binding within the Hela cell line proximal AS events, lncRNA LINC00909's RBP interactions further reinforce our proposed lncRNA-RBP sponge model. As illustrated in Figure 4B, RBPs ELAVL1 and HNRNPU were observed to have many nan values across lncRNA knockdown samples and one RBP (HNRNPC) had many significant binding associations across all lncRNA knockdowns. The parameters of the fisher's exact test require bound and unbound frequencies which can be observed in the contingency table (Table 1), thus any RBP which is not reported to be unbound to a lncRNA will yield a nan result. Both ELAVL1 and HNRNPU were manually checked across all lncRNAs inside of the HeLa cell line, and they only have values for binding across lncRNAs. Thus, the binding preference of ELAV1 and HNRNPU is very indifferent and will not be considered as a contributor to the RNA binding protein sponge model; however, binding specificity of these RBPs can be revealed as more interaction data is collected across other lncRNAs.

Discussion

In this study, we investigated the splicing alterations in lncRNA knockdown experiments, and depicted a molecular mechanism of RBP's influence on alternative splicing. The analysis of the alternative splicing heat maps indicated that transcriptional networks were perturbed in a cell type specific pattern. The observed cell type specific pattern of perturbation is in line with previous profiling of expression in lncRNA knockdowns [4]. Unique to this experiment, alternative splicing displays characteristics of being cell type specific over many lncRNAs, with extreme examples depicted in the SE experiments.

Based on our RBP binding analysis, lncRNAs display RBP sponge-like behavior and hint at other methods of inducing alternative splicing. This study also tries to implicate that lncRNAs like LINC00909 and LINC00910 act as sponges; however, there could be other means of inducing splicing. For instance, lncRNAs like RP5-1148A21.3 seem to participate in a different narrative, which is depicted in its high involvement in RI events with only 18 RBPs reported to be bound. Other studies implicate lncRNAs' potential interaction with themselves to make silencing structures, crystalline structures, or molecular machinery [9,36]. Additionally, subnuclear bodies have shown promise in understanding lncRNA's potential in creating other cellular machinery [37], and lncRNAs which bind to many RBPs could be recruiting those proteins to form a subnuclear body. Groups invested in mechanisms of post-transcriptional regulation may begin to examine RBP binding data, and multiple sequence alignment of lncRNAs in order to understand intramolecular interactions in subnuclear bodies and secondary lncRNA structure [38,39]. Hence, further attempts to understand non-coding RNA structure, should account for interactions of lncRNAs and RBPs inside of the nucleus and in the cytoplasm in order to reveal other means of lncRNAs' involvement in splicing.

While it is known that lncRNAs bind to microRNAs through a binding sponge mechanism [40], lncRNA-RBP binding activity is not often described as a sponge titration. By sponging microRNAs, lncRNAs revealed many significant instances of their influence on post-transcriptional regulation in a variety of cancers [41,42]. Their influence is controlled through mechanisms like the competitive endogenous model (ceRNA) [43] which will ultimately affect gene expression based on the binding activity of microRNAs. Despite established investigations of miRNA and lncRNA interactions, lncRNA-RBP binding activity

is not well explored in a context of post-transcriptional regulation. Other non-coding RNAs like circular RNAs have been reported to act as protein binding sponges [44] and lncRNAs have been reported to act as a binding sponge for the Hur protein [45]. However, this study identifies many RBPs binding to a handful of lncRNAs which act as a binding sponge. Thereby, a direct relationship between all the alternative splicing and RBP sponge activity is not clear, but further analysis of RBP and lncRNA binding activity can evaluate the strength of the proposed sponge model.

Conclusion

Alternative splicing induced by lncRNA knockdowns has been shown to be cell type specific within the human cancer cell lines: HeLa, K562, and U87. The downstream cellular functions of 759 genes were significantly affected by 11630 skipped exon events and 5895 retained intron events, specifically within each cell line. LncRNA and RBP interactions were also shown to be cell type specific. We hypothesize that several lncRNAs like LINC00909, LINC00910, and LINC00263 are likely to titrate RBPs in cancer, resulting in the functional disruption of RBPs and their downstream functions. We propose that such lncRNA sponges can extensively rewire the post-transcriptional gene regulatory networks by altering the protein-RNA interaction landscape in a cell-type specific manner. Our study is one of the first to implicate lncRNAs as an RBP sponge, and it reveals more diverse RBP sponge activity than previously observed [39].

However, the sponge model was not able to represent all lncRNAs which played a significant role in splicing. LncRNAs, like RP5-1148A21.3, which affected splicing very significantly and had few reported RBPs bound could have its structure researched. Investigations of lncRNA secondary structure and tertiary structure have revealed methods of epigenetic and post-transcriptomic regulation [38,46] and may further implicate lncRNA's influence over alternative splicing. Thereby, we propose that lncRNAs like RP5-1148A21.3 may interact intramolecularly and alter post-transcriptional gene regulatory networks.

Reports on the interactions between lncRNAs in unique human cell lines are becoming more prevalent as well curated databases are emerging to improve the quality of lncRNA annotation [6,47]. As more lncRNA knockdown procedures and RBP binding data are released to the public, pipelines like the one conducted in this study can be utilized to investigate the mechanisms of lncRNAs in alternative splicing.

Availability of data and materials

The RNA sequencing dataset studied within this experiment was from GSE85011. The RBP binding information was obtained from the ENCODE project. All pertinent generated data is included within this article. Additional preliminary datasets generated by the current study are available from the corresponding author on request.

Competing Interests

The authors declare that they have no competing interests.

Funding

Author's Contributions

FWP, SVD, and SCJ conceived the study. SCJ supervised the study. SVD assisted in data mining and provided excellent guidance and reviews. SVD designed a representation of the sponge model used in Figure 4, and FWP designed and generated all other figures. FWP performed the analysis and worked with SVD to write the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Figure & Table Legends

Figure 1. Depiction of the workflow and analysis conducted in this study. Automated scripts were developed to download data, align reads to the corresponding reference, and predict the alternative splicing events across samples. From further processing of the predicted events, both alternative splicing event information and proximal RBP binding locations were extracted. Alternative splicing events were then analyzed by generating heatmaps and by performing functional enrichment analysis of the alternatively spliced genes. The overlapping proximal RBP binding locations were extracted by analyzing RBP binding profiles obtained from ENCODE.

Figure 2. Summary of cell type specific Skipped Exon events. A hierarchically clustered heatmap demonstrates that the genes which are spliced via skipping exon: isolate themselves in a mostly cell type specific manner. The numbers in the Venn diagram show how many alternatively spliced genes are shared across cell lines. Across all the alternatively spliced events in the 3 different cell lines, not a single gene displayed the same pattern.

Figure 3. Summary of cell type specific GO functional annotation in Skipped Exon events. Functional annotation of cell type specific gene sets that were affected by SE events. The lines connecting the corresponding gene sets across each cell type to respective functional annotation. The functions enriched in each cell type are exclusive.

Figure 4. RNA-Binding Protein eCLIP based binding profile analysis reveals possible lncRNAs acting as an RBP sponge. The intensities are log scaled FDRs of RBP protein interaction with genes alternatively spliced in corresponding lncRNA knockdown. Figures 4A and 4B showcase the lncRNAs knocked down in respective cell lines A) K562 and B) HeLa and the resulting proximal RBP interaction intensity. The grey values represent NA values

within the heatmaps. C) An illustration of the proposed RBP sponge model, portraying lncRNA – RBP proposed behavior, in the presence and absence of a lncRNA. C(I) demonstrates when the lncRNA is present, and C(II) demonstrates the proliferated RBP interactions with mRNAs, when the corresponding lncRNA is absent or less abundant. D) Shows the extent of RBPs interacting with lncRNA LINC00909 (data from ENCODE), and supports the lncRNA- RBP sponge model.

Table 1. Representation of Fisher's exact test input contingency table for each iteration of RBP binding data. Conservative hypergeometric tests were performed in order to evaluate the importance of reported proximal RBP binding activity to a lncRNA. Each iteration of the fisher's exact test required this input space, which evaluates if a proximal binding with RBP is occurring in significant frequencies or by chance.

	Bound	Unbound
Regions where RBP's may bind across a lncRNA in a cell line	X	Y
All proximal splicing regions generated from a lncRNA knock down in a cell line	A	B

Supplementary materials

Supplementary Table 1: Table of alignment percentages of control reads to appropriate experimental reads. "Supplementary materials 1.xlsx"

The table contains the SRR read id from and its respective HISAT2 RNA-seq alignment percentages with control sequences and human genome: hg38. The overall alignment had a median of 77%, and a mean of 91%. These initial percentages indicated a good quality of alignment so that the next step of analysis could continue.

Supplementary Table 2: RMATs event summaries for each lncRNA combination. "Supplementary materials 2.xlsx"

A table that neatly summarizes the alternative splicing predictions from RMATs. After looking at the quantity of different splicing events and the difficulty of predicting different methods of splicing, the relative frequencies of alternative splicing events indicated that skipped exon and retained intron events would be the most statistically sound events to investigate.

Supplementary Tables 3: Annotated alternative splicing summaries and correlation analysis of RBP binding activity with frequency of splicing events. "Supplementary materials 3.xlsx"

The first two tables contain the annotated skipped exon and retained intron splicing events from each lncRNA respectively. Attempts to correlate the amount of RBP binding to number of splicing events are contained in the third table with respective graphs. The plots within the correlation analysis reveal that there is no linear correlation between bound RBP frequencies and alternative splicing event frequencies.

27

Supplementary Figure 4: Heatmap of skipped exon events induced by the respective knockdowns of lncRNAs in three cell lines. "Supplementary materials 4.pdf"

The entire heatmap of skipped exon splicing alternative splicing across the 39 lncRNAs in their own respective cell line. The columns of the heatmap are lncRNAs which have been knocked down, and annotated with the name of their originating cell line. The rows represent the genes which have been alternatively spliced via skipped exon events. The hierarchical clustering of both rows and columns reveals cell type specific influences of splicing.

Supplementary Figure 5: Heatmap of retained intron events induced by the respective knockdowns of lncRNAs in three cell lines. "Supplementary materials 5.pdf"

The entire heatmap of retained intron splicing alternative splicing across the 39 lncRNAs in their own respective cell line. The columns of the heatmap are lncRNAs which have been knocked down, and annotated with the name of their originating cell line. The rows represent the genes which have been alternatively spliced via retained intron events. The hierarchical clustering of both rows and columns reveals cell type specific influences of splicing.

Supplementary Figure 6: Functional enrichment of genes affected by skipped exon splicing in U87. "Supplementary materials 6.png"

A pie chart of functionally enriched terms affected by skipped exon splicing due to the knockdown of the lncRNAs within the u87 cell line. The terms annotated to the many genes which were affected related to tRNA processing formation, formation of the 43S ternary complex, and protein localization to ciliary transition zone.

28

Supplementary Figure 7: Functional enrichment of genes affected by skipped exon splicing in K562. "Supplementary materials 7.png"

A pie chart of functionally enriched terms affected by skipped exon splicing due to the knockdown of the lncRNAs within the K562 cell line. The affected terms were annotated to many genes relating to microtubular organization, phosphotransferase activity, and the p53 signaling pathway. The found terms are indicative of cancer proliferation.

Supplementary Figure 8: Functional enrichment of genes affected by skipped exon splicing in HeLa. "Supplementary materials 8.png"

A pie chart of functionally enriched terms affected by skipped exon splicing due to the knockdown of the lncRNAs within the HeLa cell line. The top terms were annotated to many genes relating to vacuole organization, interstrand cross-link repair, and monophosphate metabolic processes.

Supplementary Figure 9: Functional enrichment of genes affected by retained intron splicing in K562. "Supplementary materials 9.png"

A pie chart of functionally enriched terms affected by retained intron splicing due to the knockdown of the lncRNAs within the K562 cell line. The top terms were annotated to many genes relating to cytosolic large ribosomal subunits and large ribosomal subunit biogenesis. Other functions include negative regulation of autophagy and nucleic sugar biosynthesis.

Supplementary Figure 10: Functional enrichment of genes affected by retained intron splicing in HeLa from the knockdown of RP5-1148A21.4. "Supplementary materials 10.png"

A pie chart of functionally enriched terms affected by retained intron splicing due to the knockdown of RP5-1148A21.4. The top terms annotated to the many genes which the

lncRNA affected were translation, response to endoplasmic reticulum stress, holiday junction resolvase complex, and RNA splicing.

Supplementary materials 11: Zip file which contains the bedfiles of the proximal RBP splicing locations which overlapped with RBP binding locations noted in ENCODE "Supplementary materials 11.zip"

Inside each bedfile, there will be columns from the lncRNA which was overlapped "lncRNA name|cell line|ENSG id|location relative to splice site" and from RBP locations "RBP name|cell line". Take note that there are regions in one cell line that have reported RBP data in other cell lines.

References

1. Kopp, F.; Mendell, J.T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **2018**, *172*, 393-407, doi:10.1016/j.cell.2018.01.011.
2. Leone, S.; Santoro, R. Challenges in the analysis of long noncoding RNA functionality. *FEBS Lett* **2016**, *590*, 2342-2353, doi:10.1002/1873-3468.12308.
3. Shukla, C.J.; McCorkindale, A.L.; Gerhardinger, C.; Korthauer, K.D.; Cabili, M.N.; Shechner, D.M.; Irizarry, R.A.; Maass, P.G.; Rinn, J.L. High-throughput identification of RNA nuclear enrichment sequences. *EMBO J* **2018**, *37*, doi:10.15252/embj.201798452.
4. Liu, S.J.; Horlbeck, M.A.; Cho, S.W.; Birk, H.S.; Malatesta, M.; He, D.; Attenello, F.J.; Villalta, J.E.; Cho, M.Y.; Chen, Y., et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **2017**, *355*, doi:10.1126/science.aah7111.
5. Zare, K.; Shademan, M.; Ghahramani Seno, M.M.; Dehghani, H. CRISPR/Cas9 Knockout Strategies to Ablate CCAT1 lncRNA Gene in Cancer Cells. *Biol Proced Online* **2018**, *20*, 21, doi:10.1186/s12575-018-0086-5.
6. Xu, J.; Shi, A.; Long, Z.; Xu, L.; Liao, G.; Deng, C.; Yan, M.; Xie, A.; Luo, T.; Huang, J., et al. Capturing functional long non-coding RNAs through integrating large-scale causal relations from gene perturbation experiments. *EBioMedicine* **2018**, *35*, 369-380, doi:10.1016/j.ebiom.2018.08.050.
7. Gilbert, L.A.; Larson, M.H.; Morsut, L.; Liu, Z.; Brar, G.A.; Torres, S.E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E.H.; Doudna, J.A., et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **2013**, *154*, 442-451, doi:10.1016/j.cell.2013.06.044.
8. Yu, D.; Zhang, C.; Gui, J. RNA-binding protein HuR promotes bladder cancer progression by competitively binding to the long noncoding HOTAIR with miR-1. *Onco Targets Ther* **2017**, *10*, 2609-2619, doi:10.2147/OTT.S132728.
9. Hirose, T.; Virnicchi, G.; Tanigawa, A.; Naganuma, T.; Li, R.; Kimura, H.; Yokoi, T.; Nakagawa, S.; Benard, M.; Fox, A.H., et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* **2014**, *25*, 169-183, doi:10.1091/mbc.E13-09-0558.
10. Ule, J.; Jensen, K.; Mele, A.; Darnell, R.B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **2005**, *37*, 376-386, doi:10.1016/j.ymeth.2005.07.018.
11. Budak, G.; Srivastava, R.; Janga, S.C. Seten: a tool for systematic identification and comparison of processes, phenotypes, and diseases associated with RNA-binding proteins from condition-specific CLIP-seq profiles. *RNA* **2017**, *23*, 836-846, doi:10.1261/rna.059089.116.
12. Zhu, Y.; Xu, G.; Yang, Y.T.; Xu, Z.; Chen, X.; Shi, B.; Xie, D.; Lu, Z.J.; Wang, P. POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res* **2019**, *47*, D203-D211, doi:10.1093/nar/gky830.
13. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57-74, doi:10.1038/nature11247.
14. Neelamraju, Y.; Gonzalez-Perez, A.; Bhat-Nakshatri, P.; Nakshatri, H.; Janga, S.C. Mutational landscape of RNA-binding proteins in human cancers. *RNA Biol* **2018**, *15*, 115-129, doi:10.1080/15476286.2017.1391436.

15. Van Nostrand, E.L.; Pratt, G.A.; Shishkin, A.A.; Gelboin-Burkhart, C.; Fang, M.Y.; Sundararaman, B.; Blue, S.M.; Nguyen, T.B.; Surka, C.; Elkins, K., et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **2016**, *13*, 508-514, doi:10.1038/nmeth.3810.
16. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **2002**, *30*, 207-210.
17. Leinonen, R.; Sugawara, H.; Shumway, M.; International Nucleotide Sequence Database, C. The sequence read archive. *Nucleic Acids Res* **2011**, *39*, D19-21, doi:10.1093/nar/gkq1019.
18. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **2015**, *12*, 357-360, doi:10.1038/nmeth.3317.
19. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078-2079, doi:10.1093/bioinformatics/btp352.
20. Shen, S.; Park, J.W.; Lu, Z.X.; Lin, L.; Henry, M.D.; Wu, Y.N.; Zhou, Q.; Xing, Y. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **2014**, *111*, E5593-5601, doi:10.1073/pnas.1419161111.
21. Yates, A.; Akanni, W.; Amode, M.R.; Barrell, D.; Billis, K.; Carvalho-Silva, D.; Cummins, C.; Clapham, P.; Fitzgerald, S.; Gil, L., et al. Ensembl 2016. *Nucleic Acids Res* **2016**, *44*, D710-716, doi:10.1093/nar/gkv1157.
22. Institute, B. Morpheus. Availabe online: https://software.broadinstitute.org/morpheus (accessed on
23. Oliveros, J.C. Venny. An interactive tool for comparing lists with Venn's diagrams. Availabe online: http://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed on
24. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091-1093, doi:10.1093/bioinformatics/btp101.
25. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13*, 2498-2504, doi:10.1101/gr.1239303.
26. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **1999**, *27*, 29-34.
27. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T., et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **2000**, *25*, 25-29, doi:10.1038/75556.
28. The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **2019**, *47*, D330-D338, doi:10.1093/nar/gky1055.
29. Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B., et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **2018**, *46*, D649-D655, doi:10.1093/nar/gkx1132.

30. Huppertz, I.; Attig, J.; D'Ambrogio, A.; Easton, L.E.; Sibley, C.R.; Sugimoto, Y.; Tajnik, M.; Konig, J.; Ule, J. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **2014**, *65*, 274-287, doi:10.1016/j.ymeth.2013.10.011.

31. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841-842, doi:10.1093/bioinformatics/btq033.

32. Naidu, M.D.; Mason, J.M.; Pica, R.V.; Fung, H.; Pena, L.A. Radiation resistance in glioma cells determined by DNA damage repair activity of Ape1/Ref-1. *J Radiat Res* **2010**, *51*, 393-404.

33. Van der Ven, L.T.; Roholl, P.J.; Gloudemans, T.; Van Buul-Offers, S.C.; Welters, M.J.; Bladergroen, B.A.; Faber, J.A.; Sussenbach, J.S.; Den Otter, W. Expression of insulin-like growth factors (IGFs), their receptors and IGF binding protein-3 in normal, benign and malignant smooth muscle tissues. *Br J Cancer* **1997**, *75*, 1631-1640.

34. Oh, H.; Kim, H.; Chung, K.H.; Hong, N.H.; Shin, B.; Park, W.J.; Jun, Y.; Rhee, S.; Song, W.K. SPIN90 knockdown attenuates the formation and movement of endosomal vesicles in the early stages of epidermal growth factor receptor endocytosis. *PLoS One* **2013**, *8*, e82610, doi:10.1371/journal.pone.0082610.

35. Farrar, J.E.; Quarello, P.; Fisher, R.; O'Brien, K.A.; Aspesi, A.; Parrella, S.; Henson, A.L.; Seidel, N.E.; Atsidaftos, E.; Prakash, S., et al. Exploiting pre-rRNA processing in Diamond Blackfan anemia gene discovery and diagnosis. *Am J Hematol* **2014**, *89*, 985-991, doi:10.1002/ajh.23807.

36. Novikova, I.V.; Hennelly, S.P.; Tung, C.S.; Sanbonmatsu, K.Y. Rise of the RNA machines: exploring the structure of long non-coding RNAs. *J Mol Biol* **2013**, *425*, 3731-3746, doi:10.1016/j.jmb.2013.02.030.

37. Bond, C.S.; Fox, A.H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J Cell Biol* **2009**, *186*, 637-644, doi:10.1083/jcb.200906113.

38. Lin, Y.; Schmidt, B.F.; Bruchez, M.P.; McManus, C.J. Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res* **2018**, *46*, 3742-3752, doi:10.1093/nar/gky046.

39. Rivas, E.; Clements, J.; Eddy, S.R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **2017**, *14*, 45-48, doi:10.1038/nmeth.4066.

40. Chan, J.J.; Tay, Y. Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci* **2018**, *19*, doi:10.3390/ijms19051310.

41. Du, Z.; Sun, T.; Hacisuleyman, E.; Fei, T.; Wang, X.; Brown, M.; Rinn, J.L.; Lee, M.G.; Chen, Y.; Kantoff, P.W., et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun* **2016**, *7*, 10982, doi:10.1038/ncomms10982.

42. Zhuang, M.; Zhao, S.; Jiang, Z.; Wang, S.; Sun, P.; Quan, J.; Yan, D.; Wang, X. MALAT1 sponges miR-106b-5p to promote the invasion and metastasis of colorectal cancer via SLAIN2 enhanced microtubules mobility. *EBioMedicine* **2019**, *41*, 286-298, doi:10.1016/j.ebiom.2018.12.049.

43. Long, J.; Bai, Y.; Yang, X.; Lin, J.; Yang, X.; Wang, D.; He, L.; Zheng, Y.; Zhao, H. Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma. *Cancer Cell Int* **2019**, *19*, 90, doi:10.1186/s12935-019-0817-y.

44. Du, W.W.; Zhang, C.; Yang, W.; Yong, T.; Awan, F.M.; Yang, B.B. Identifying and Characterizing circRNA-Protein Interaction. *Theranostics* **2017**, *7*, 4183-4191, doi:10.7150/thno.21299.

45. Kim, J.; Abdelmohsen, K.; Yang, X.; De, S.; Grammatikakis, I.; Noh, J.H.; Gorospe, M. LncRNA OIP5-AS1/cyrano sponges RNA-binding protein HuR. *Nucleic Acids Res* **2016**, *44*, 2378-2392, doi:10.1093/nar/gkw017.

46. Wang, C.; Wang, L.; Ding, Y.; Lu, X.; Zhang, G.; Yang, J.; Zheng, H.; Wang, H.; Jiang, Y.; Xu, L. LncRNA Structural Characteristics in Epigenetic Regulation. *Int J Mol Sci* **2017**, *18*, doi:10.3390/ijms18122659.

47. Srivastava, R.; Budak, G.; Dash, S.; Lachke, S.A.; Janga, S.C. Transcriptome analysis of developing lens reveals abundance of novel transcripts and extensive splicing alterations. *Sci Rep* **2017**, *7*, 11572, doi:10.1038/s41598-017-10615-4.