

UC San Diego

UC San Diego Previously Published Works

Title

Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression

Permalink

<https://escholarship.org/uc/item/3j35h9gq>

Journal

Genome Medicine, 12(1)

ISSN

1756-994X

Authors

Okholm, Trine Line Hauge
Sathe, Shashank
Park, Samuel S
[et al.](#)

Publication Date

2020-12-01

DOI

10.1186/s13073-020-00812-8


Peer reviewed

RESEARCH

Open Access



Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression

Trine Line Hauge Okholm^{1,2*} , Shashank Sathe³, Samuel S. Park³, Andreas Bjerregaard Kamstrup⁴, Asta Mannstaedt Rasmussen^{1,2}, Archana Shankar³, Zong Ming Chua³, Niels Fristrup⁵, Morten Muhlig Nielsen^{1,2}, Søren Vang^{1,2}, Lars Dyrskjøl^{1,2}, Stefan Aigner³, Christian Kroun Damgaard⁴, Gene W. Yeo³ and Jakob Skou Pedersen^{1,2,6*}

Abstract

Background: Circular RNAs (circRNAs) are stable, often highly expressed RNA transcripts with potential to modulate other regulatory RNAs. A few circRNAs have been shown to bind RNA-binding proteins (RBPs); however, little is known about the prevalence and distribution of these interactions in different biological contexts.

Methods: We conduct an extensive screen of circRNA-RBP interactions in the ENCODE cell lines HepG2 and K562. We profile circRNAs in deep-sequenced total RNA samples and analyze circRNA-RBP interactions using a large set of eCLIP data with binding sites of 150 RBPs. We validate interactions for select circRNAs and RBPs by performing RNA immunoprecipitation and functionally characterize our most interesting candidates by conducting knockdown studies followed by RNA-Seq.

Results: We generate a comprehensive catalog of circRNA-RBP interactions in HepG2 and K562 cells. We show that KHSRP binding sites are enriched in flanking introns of circRNAs and that KHSRP depletion affects circRNA biogenesis. We identify circRNAs that are highly covered by RBP binding sites and experimentally validate individual circRNA-RBP interactions. We show that circCDYL, a highly expressed circRNA with clinical and functional implications in bladder cancer, is almost completely covered with GRWD1 binding sites in HepG2 cells, and that circCDYL depletion counteracts the effect of GRWD1 depletion. Furthermore, we confirm interactions between circCDYL and RBPs in bladder cancer cells and demonstrate that circCDYL depletion affects hallmarks of cancer and perturbs the expression of key cancer genes, e.g., *TP53*. Finally, we show that elevated levels of circCDYL are associated with overall survival of bladder cancer patients.

Conclusions: Our study demonstrates transcriptome-wide and cell-type-specific circRNA-RBP interactions that could play important regulatory roles in tumorigenesis.

Keywords: Circular RNAs, RNA-binding proteins, RBP sponges, circCDYL, Bladder cancer, Tumorigenesis

* Correspondence: trine.okholm@clin.au.dk; jakob.skou@clin.au.dk

¹Department of Molecular Medicine (MOMA), Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Circular RNAs (circRNAs) are covalently closed RNA molecules often derived from precursor mRNA (pre-mRNA) through a backsplicing event, in which a downstream 5' splice donor backsplices to an upstream 3' splice acceptor [1]. First identified in the early 1990s, eukaryotic circRNAs were thought to be rare and a result of erroneous splicing events [2]. Twenty years later, the advent of high-throughput sequencing of non-polyadenylated transcriptomes and bioinformatic analyses have made it possible to detect thousands of circRNAs, many of which are highly abundant and conserved across species [3–5]. Accumulating evidence links circRNAs to development and progression of different diseases (reviewed in [6]) and several recent studies have shown that circRNAs are involved in tumorigenesis [7, 8]. Due to their structural stability [4], tissue specificity [3], and relatively high expression levels in exosomes [9], blood [10], and plasma [11], circRNAs have been suggested as a new class of biomarkers and potential therapeutic targets.

RNA-binding proteins (RBPs) are proteins that bind to double- or single-stranded RNA. Some RBPs contain well-established RNA-binding domains, including the RNA recognition motif (RRM) or the K-homology (KH) domain and bind to well-defined motifs. However, many RBPs rely on contextual features as well, e.g., secondary structure, flanking nucleotide composition, or short non-sequential motifs, complicating RNA target predictions from sequence alone [12]. RBPs play crucial roles in all aspects of RNA biology, e.g., RNA transcription, pre-mRNA splicing, and polyadenylation as well as modification, stabilization, localization, and translation of RNA (reviewed in [13]). Recent studies have shown that RBPs also affect all phases of the circRNA lifecycle (reviewed in [14]). Some RBPs are involved in circRNA biogenesis as has been shown for Quaking (QKI) [15], FUS [16], HNRNPL [17], RBM20 [18], and Muscleblind [19], which bind to specific intronic RBP motifs and promote formation of some circRNAs in certain biological settings. Besides RBP-binding motifs, complementary sequences in both flanking introns, like *Alu* elements, facilitate circRNA production by RNA pairing [4]. The RBP immune factors NF90/NF11 promote circRNA formation by directly binding to inverted repeated *Alus* (*IRAlus*) [20], while ADAR1 [21] and DHX9 [22] reduce circularization by destabilizing *IRAlu*-mediated RNA pairing.

The functional role of most circRNAs is still unknown. A number of circRNAs, e.g., ciRS-7 and circSRY, have been reported to function as miRNA sponges by binding a large number of microRNAs (miRNAs) and thereby regulating miRNA target genes [3, 23]. More recent evidence using a ciRS-7 knockout mouse suggests that

ciRS-7 is important for normal brain function and for maintaining proper miR-7 levels [24], which in the absence of ciRS-7, becomes efficiently destroyed through target RNA-directed miRNA degradation (TDMD), promoted by the long non-coding RNA (lncRNA) Cyrano [25]. Other specific circRNAs have been shown to modulate host gene expression by interacting with RBPs, such as circMbl, which regulates the expression of its parent gene in a negative feedback loop between MBL and circMbl production [19]. CircRNAs may also regulate translation efficiency, as reported for the PABPN1 mRNA, whose translation is inhibited by the encoded circPABPN1 that efficiently “sponges” a translation stimulator RBP, HuR [26]. Additionally, circFoxo3 was found to interact with CDK2 and p21 to repress cell cycle progression in cancer cell lines [27] and to promote cardiac senescence by interacting with ID-1, E2F1, FAK, and HIF1 α in the mammalian heart [28]. Another study showed that a specific RBP, IGF2BP3, associates with several circRNAs [29].

Despite few examples of circRNA-RBP interplay, little is known about the overall ability of circRNAs to interact with RBPs. Based on binding sequence motifs of 38 RBPs and nucleotide sequence alone, You et al. found that neuronal circRNAs are not enriched with RBP binding sites compared to mRNAs [30]. However, a comprehensive understanding of circRNA-RBP interactions on a global scale is missing. Through extensive analysis of high-throughput data sets of experimentally defined RBP binding sites combined with circRNA profiling, we can screen the entire genome for circRNA-RBP interactions and study regulatory dependencies. Since RBPs are essential to maintain normal function of the cells, defects in the expression or localization of RBPs can cause diseases [31]. The abundance, high stability, and general lack of protein translation, which normally would displace most bound RBPs, make circRNAs ideal binding platforms for more than transient RBP interactions. Thus, binding and deregulation of RBPs through circRNA-RBP interactions could likely have long-term cellular effects.

Here, we evaluate the overall potential of circRNAs to interact with RBPs. Based on eCLIP data profiling the binding sites of 150 RBPs in HepG2 and K562 [32, 33] and deep total RNA-sequencing (RNA-Seq) samples allowing circRNA quantification [34], we comprehensively study the ability of circRNAs to interact with RBPs as well as the capability of RBPs to influence circRNA formation (Fig. 1a, b). We show that KHSRP binding sites are enriched in intronic regions flanking circRNAs compared to non-circularizing exons and that KHSRP depletion diminishes circRNA expression. Additionally, we find that circularizing exons are enriched with RBP binding sites compared to non-circularizing exons

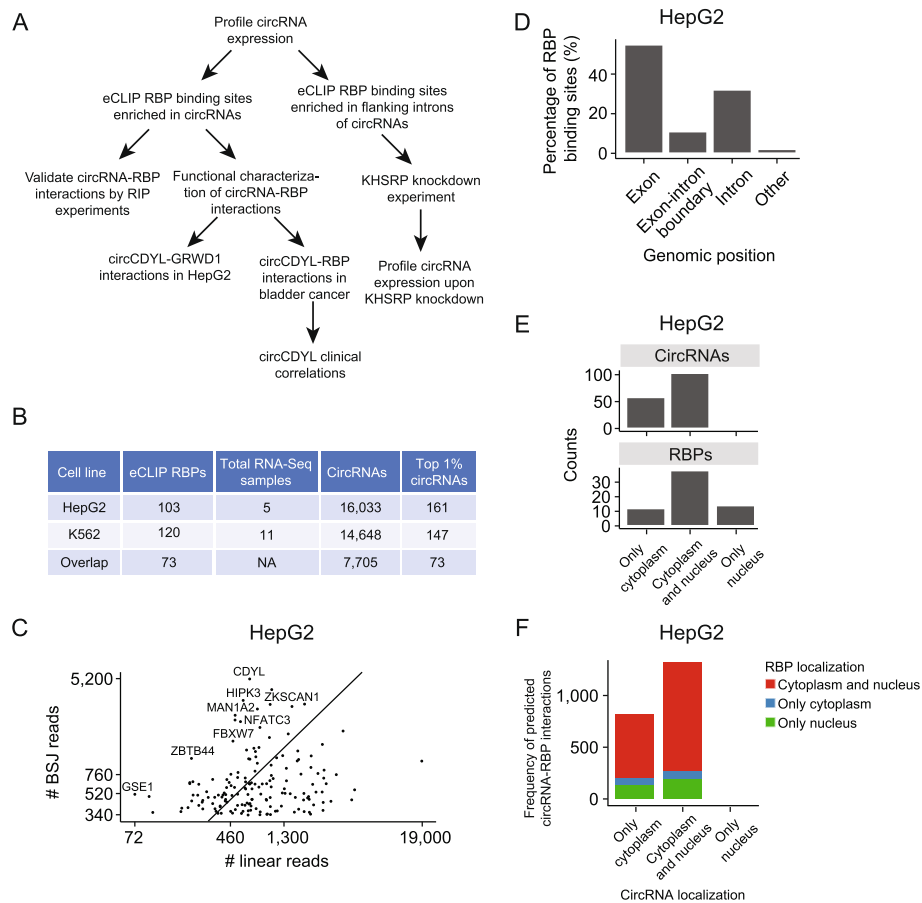


Fig. 1 Circular RNAs are highly expressed in HepG2 and K562 and generally colocalizes with RBPs. **a** Flowchart overview of analyses. The initial analyses and characterizations are based on the ENCODE cell lines HepG2 and K562. Follow-up experiments are performed in HepG2 cells and bladder cancer cell lines to evaluate the relevance of findings in tumorigenesis. **b** Summary of input data sets and inferred sets of circRNAs. The sets of high-confidence, highly expressed circRNAs (top 1% circRNAs) are used in most of the downstream analyses. **c** Number of backsplice junction (BSJ) reads supporting the 1% highest expressed circRNAs ($n = 161$), and the number of reads spanning canonical splice sites in the corresponding linear transcripts in HepG2. X- and Y-axes are plotted on a logarithmic scale (log10) showing actual counts. Some circRNAs are depicted with their host gene name. The number of reads represents the summarized values across all five HepG2 data sets. **d** Genomic location of RBP binding sites in the top 1% circRNA loci in HepG2. Most RBP binding sites are found within exonic parts of the circRNAs. Exonic parts span 0.119 Mb, while introns span 1.836 Mb. **e** Localization of top 1% circRNAs ($n = 161$; top) and RBPs ($n = 64$; bottom) across cellular compartments of HepG2. For circRNAs, all are found in the cytoplasm ($n = 161$), with a large subset that is also found in the nucleus ($n = 103$). No circRNAs are exclusively expressed in the nucleus. For RBPs, most are located in both cellular compartments ($n = 38$), while some RBPs are expressed exclusively in the cytoplasm ($n = 12$) or the nucleus ($n = 14$). **f** Localization of top 1% circRNAs and RBPs that are predicted to interact from eCLIP data in HepG2. CircRNA localization is shown on the x-axis, while RBP localization is indicated by color. For 94% of the predicted circRNA-RBP interactions, the circRNA and RBP are expressed in the same subcellular compartment

indicating regulatory potency of circRNA-RBP interactions. We investigate the potential of individual circRNAs to function as RBP sponges and show experimentally that circRNAs interact with RBPs in a cell-type-specific manner. Specifically, we demonstrate that the highly expressed circCDYL is almost completely covered with GRWD1 binding sites in HepG2 and that circCDYL depletion in HepG2 cells counteracts the effect of GRWD1 depletion on target genes. In bladder cancer cell lines, we show that circCDYL interacts with IGF2BP1 and IGF2BP2 and that depletion of circCDYL or these RBPs disturb several hallmarks of cancer.

Specifically, we show that some key tumor genes, including *TP53* and *MYC*, are affected by circCDYL knockdown. Finally, we show that the expression of circRNAs highly covered with RBP binding sites, including circCDYL, is positively correlated with overall survival of bladder cancer patients.

Methods

HepG2 and K562 encode cell lines

We downloaded all total and fractionated samples for the cell lines HepG2 ($n = 5$) and K562 ($n = 11$) from ENCODE (<https://www.encodeproject.org/>), generated by

the labs of Brenton Graveley, UConn; Thomas Gingeras, CSHL; and Éric Lécuyer, IRCM (Additional files 1+2: File S1 + S2) [34]. Before profiling gene and circRNA expression, we combined the fastq files from biological replicates of the same sample.

Detection of circRNAs in HepG2 and K562

We used the CIRI2 (v2.0.6) [35] and CIRCexplorer (v1.1.10) [36] pipelines to detect circRNAs in all samples. Before running the CIRI2 pipeline, we trimmed reads with Trim Galore and cutadapt (v0.4.1 and v1.9). We aligned the reads to the human genome (hg19) using bwa (v0.7.15) and samtools (v1.3). The CIRI2 pipeline was run with a gene transfer format (GTF) file (hg19) to annotate the overlapping gene of the circRNAs. The pipeline does not take strand into consideration so afterwards we named the overlapping gene of the circRNAs according to strand information. For CIRCexplorer, raw fastq files were mapped against the human genome (hg19) using Tophat (v2.0.13) and Bowtie (v2.2.8.0) and the gencode v19 gene model. Tophat was run with the flags `-a 6 --microexon-search -m 2`. Unmapped reads were investigated using tophat fusion in conjunction with bowtie (v1.1.1), before the output was processed using CIRCexplorer. We only reported circRNAs that are supported by at least two reads spanning the backsplice junction by both pipelines. For all analysis, we used the CIRI2 output for circRNA expression. The expression of the linear counterpart of circRNAs is quantified by the CIRI2 pipeline as the number of linearly spliced reads mapping across the backsplice junction of circRNAs. As described for CIRI2, relative ratios are calculated as $(2 \times \text{circRNA}) / (2 \times \text{circRNA} + \text{linear RNA})$ because a BSJ read is generated from two ends of a circular junction but only counted once while reads mapping to the corresponding linear RNA are generated from one end.

Gene expression profiling in HepG2 and K562

Illumina paired-end reads were stripped of library adapters (Trim Galore and cutadapt as above) and mapped to the human genome (hg19) using TopHat2 (version 2.1.1) [37] and Bowtie2 (version 2.2.8.0) [38], and Cufflinks (v2.1.1) [39] and HTSeq (v0.6.1p1) [40] were used to estimate the transcript abundance using transcript information from GENCODE v19. Samtools (v1.3) [41] and Picard (v2.0.1) were used for quality control and statistics.

To compare the expression of circCDYL and the mRNAs of its interacting RBPs, we evaluated the expression of the entire exon that constitutes circCDYL in FPKM (chr6:4891946-4892613) using Cufflinks as described above. Since reads that align to the circCDYL exon could come from both the circular and the linear

form of the exon, we took the relative ratio between circular-to-linear reads based on CIRI2 to obtain an estimate of circCDYL abundance in FPKM. Same approach was used to profile gene and circCDYL expression in the bladder cancer cell lines J82 and UMUC3.

eCLIP of HepG2 and K562

The eCLIP data of RNA-binding protein (RBP) targets from HepG2 and K562 was obtained from a previous study [32, 33]. Briefly, RNA-RBP interactions were covalently linked by UV, followed by RNA fragmentation and immunoprecipitation (IP) with a specific antibody against the RBP of interest. RNA from RNA-RBP complexes were prepared into paired-end high-throughput sequencing libraries and sequenced. All experiments were performed with a size-matched input control (SMInput), which controls for nonspecific background signals. RBP targets were determined using the CLIPper algorithm [42]. To report RBP target sites, IP fold enrichment was calculated based on the number of reads overlapping peaks identified by CLIPper in both the IP and SMInput samples. A cut-off of $\log_2(\text{foldchange}) \geq 3$ was applied, which means an 8-fold enrichment in the IP. Additionally, we only considered RBP binding sites supported by at least 10 reads in the IP.

eCLIP reads spanning BSJ

We generated a reference set of all possible exonic BSJ events within transcripts by extracting 30 bp from each side of all exons and pasting them together in reverse order. Unmapped eCLIP reads (~20–25 bp) from both IP and input experiments [32] were mapped against the BSJ reference set, allowing two mismatches. We removed PCR duplicates based on the barcode and position of reads and merged barcode replicates. Before counting mapped reads spanning backsplice junctions, we removed read 1 as this was only used to identify PCR duplicates. Only reads spanning the BSJ site by ≥ 5 bp were considered. To identify specific RBP binding sites, we used the same cut-off as above: only RBP binding sites supported by ≥ 10 reads in both IP replicates and with a $\log_2(\text{foldchange}) \geq 3$ between IP and input were reported. A pseudo count of 1 was added before counting the foldchange.

Subcellular localization of RBPs in HepG2

Localization of RBPs by immunofluorescence was performed for another study [43]. We only considered cellular fractions where we had both circRNA and RBP data, e.g., the nuclear and cytoplasmic fractions of HepG2.

Genomic annotations of exons

We extracted all exons from known and novel protein-coding transcripts (hg19). Exons from different transcript isoforms were merged using bedtools merge (v2.25.0) to obtain one genomic interval per exon per gene. We used bedtools intersect to obtain exonic regions of the circRNAs. Exons in circRNAs were annotated as backsplice junction (BSJ) circ-exons if they are involved in the backsplicing event and circ-exons if they are internal exons in the circRNAs. An exon might be an internal circ-exon in one circRNA but a BSJ circ-exon in another circRNA. In that case, it will be annotated as a BSJ circ-exon. Intronic regions were achieved in the same way.

RBP binding sites in circRNAs

We used bedtools intersect to annotate the overlap between the exonic parts of circRNAs and RBP binding sites. We disregarded RBPs with a binding site of < 4 bp or a circRNA-RBP overlap of < 4 bp long. To get that fraction of circRNAs covered with RBP binding sites, we merged the RBP binding sites overlapping circRNAs and intersected them with merged exons in circRNAs.

RBP binding sites transcriptome-wide

We extracted all merged exon positions (of known and novel protein-coding transcripts from above) for expressed genes in each cell line and intersected them with RBP binding sites. Again, we disregarded RBPs with a binding site of < 4 bp or an exon-RBP overlap of < 4 bp. We obtained the fraction of exons covered with RBP binding sites by merging these RBP binding sites and intersecting them with merged exon positions from expressed genes.

To compare RBP coverage of circRNAs to comparable genes sets, the genes were divided into 100 bins based on expression, and gene expression was considered a proxy for exon expression. We randomly drew internal non-circ-exons from genes not producing circRNAs, matching the gene expression distribution of genes producing highly expressed circRNAs, with one hundred repetitions. The same approach was conducted to extract exons of highly expressed genes (percentile 90–95 and 96–100).

To evaluate the enrichment of circRNA-RBP coverage to non-circ-exons in host genes, we added a pseudo count of 1 to the RBP coverage of circRNAs and non-circ-exons to include circRNAs and non-circ-exons with no RBP binding sites in the analysis. We divided circRNA-RBP coverage by the mean RBP coverage of internal non-circ-exons within host genes. CircRNAs from genes with zero internal non-circ-exons are not depicted in the figure and are denoted with NA in Additional file 3: Table S1.

RBP binding sites in introns flanking circRNAs and non-circularizing exons

RBP binding sites in intronic regions were evaluated in the same way as exonic RBP binding sites (above). Intronic regions overlapping exonic regions in a transcript isoform were omitted. To identify RBPs with binding sites in both flanking introns of circRNAs, flanking regions (10,000 bp to each side of the circRNAs) were intersected with RBP binding sites in intronic regions. For non-circularizing exons, the first and last exons from expressed genes were removed from this analysis as they are not flanked by introns on both sides. Additionally, exons found in circRNAs were disregarded. As for the circRNAs, flanking regions of internal non-circularizing exons (10,000 bp to each side) were intersected with RBP binding sites in intronic regions.

Alu repeats in introns

Positions of *Alu* repeats were obtained from the UCSC Browser RepeatMasker track as described in a previous study [44]. *Alu* repeat positions were intersected with flanking regions of circRNAs and non-circularizing exons from above.

Gene Ontology enrichment analysis

Gene Ontology (GO) term data was extracted from BioMart [45]. Two subsets of the eCLIP data were generated: RBPs that bind relatively more to BSJ circ-exons than non-circ-exons and a negative set containing all other RBPs for each of the two cell lines. In K562, the subset of RBPs that are more prone to bind to BSJ circ-exons ($n = 10$) contains NOLC1, IGF2BP1, IGF2BP2, YBX3, U2AF2, SND1, PUM1, RBM15, GRWD1, and ZNF800. In HepG2, the subset ($n = 7$) contains UCHL5, ZNF800, BCLAF1, GRWD1, RBM15, TRA2A, and IGF2BP3. Only GO terms represented by at least two RBPs were considered in the analysis for each cell line. For the GO terms in each of the GO domains, Biological process (BP), Cellular component (CC), and Molecular function (MF), a one-tailed Fisher's exact test was performed.

Pathway analyses and distribution of oncogenes

We used the R package *gage* [46] for gene set enrichment and pathway analyses. We used the gene sets provided by the package for KEGG pathway analyses. We obtained the 50 hallmarks of cancer gene sets from The Molecular Signatures Database (MSigDB) [47]. We downloaded the list of all cancer census genes from The Cosmic Cancer Gene Census [48]. Some genes were classified into several categories, so we restricted our analyses to genes that were only classified as oncogenes.

RIP assay for HepG2 and K562 cells

Five RBPs and four circRNAs were chosen for validation. We designed primers against the unique backsplice junction of specific circRNAs to validate their expression in HepG2 and K562. Antibodies were obtained for the following five RBPs: GRDW1 (Bethyl A301-576A Lot 1), UCHL5 (Bethyl A304-099A Lot 1), YBX3 (A303-070A Lot 1), IGF2BP1 (MBL RN007P Lot 004), and IGF2BP2 (MBL RN008P Lot 005). Rabbit IgG Isotype Control (Invitrogen Cat# 02-6102) was used as negative control.

K562 cells (American Type Culture Collection (ATCC[®]), CCL-243[™]) were grown in RPMI 1640 media (Gibco[™], Life Technologies, 11875119) with 10% FBS (Gibco[™] Life Technologies, 26140079). HepG2 cells (American Type Culture Collection (ATCC[®]), HB-8065) were grown in DMEM (Gibco[™], Thermo Fisher Scientific) with 10% FBS (Gibco[™] Life Technologies, 26140079). 20×10^6 snap frozen cells were lysed in 1 ml of iCLIP Lysis Buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% NP-40 (Igepal CA630), 0.1% SDS, 0.5% sodium deoxycholate) with 5.5 μ l Protease Inhibitor Cocktail Set III EDTA Free (EMD Millipore Corp. 539134-1ML) and 11 μ l Murine RNase Inhibitor (New England BioLabs Inc.[®] M0314L) for 15 min and were then centrifuged at 20,000g for 20 min at 4 °C. The supernatant was placed in a solution containing specific primary (10 μ g)/secondary (1.25 g) (anti-Rabbit magnetic DynaBeads, Invitrogen, 11204) antibody-antibody (incubated on a rotator at 25 °C for 45 min) to immunoprecipitate overnight on a rotator at 4 °C. The RNA-RBP pull-down was then purified by stringently washing with NET-2 wash buffer (5 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton X-100). Isolation of RNA from the RNA-RBP complexes was accomplished with the addition of TRIzol[™] Reagent (Invitrogen[™], 15596018) followed by the Direct-zol[™] RNA MiniPrep (Zymo Research Cat No. R2052). Isolated RNA was reverse transcribed with the SuperScript[™] III First Strand Synthesis System (Invitrogen[™] 18080051) using Random Hexamers (Thermo Fisher Scientific N8080127), and circRNAs were amplified using GoTaq[®] DNA Polymerase (Promega, M3005); 4 μ l of 1:2.5 diluted cDNA, 1 μ l of each primer for 34 cycles at the following conditions: strand separation at 95 °C for 30 s, primer hybridization at 55 °C for 30 s, and elongation at 72 °C for 20 s followed by a final elongation step at 72 °C for 5 min. Amplicons were run on a 3% agarose gel at 135 V for 35 min at 4 °C alongside a 50-bp ladder marker. CircRNAs bound by RBPs were identified on the gel based on amplicon size.

Validation of RBP pull-down in IP

Approximately 5×10^6 cells from the overnight incubation with the primary/secondary antibody-antibody complex were stringently washed with NET-2 buffer then

denatured in a DL-Dithiothreitol (DTT) (Sigma-Aldrich D9779), NuPAGE[™] LDS Sample Buffer (4 \times) (Invitrogen[™] NP0008) mixture at 70 °C for 10 min in a Thermomixer at 1200 rpm. Denatured protein samples were subject to SDS-PAGE and western blotting with Anti-Rabbit IgG HRP (Rockland Inc. 18-8816-33).

RBP knockdown in HepG2

RBP knockdown samples were originally generated for the ENCODE project by Brenton Graveley's Lab, UConn [34]. RNA-Seq data of polyadenylated transcripts were obtained from ENCODE (<https://www.encodeproject.org/>).

siRNA-mediated knockdown

3×10^5 cells (UMUC3, J82 or HepG2) were seeded per well in 6-well plates (2 ml per well). Then, 24 h later cells were transfected with siRNA (Additional file 4: Table S2) using SilentFect (Bio-Rad) in biological triplicates: 100 μ l OptiMEM (Gibco) + 2 μ l SilentFect (Bio-Rad) was incubated at RT for 5 min before mixing with 2 μ l siRNA (20 μ M stock) pre-diluted in 100 μ l OptiMEM. After gentle mixing by pipetting, the solution was incubated for 20 min at RT prior to dropwise addition to cells (final concentration of 20 nM). Cells were incubated for 56 h and harvested by addition of 1 ml Trizol (Thermo Fisher Scientific) to each well followed by mixing. RNA was purified according to the manufacturer's protocol with an additional chloroform extraction to increase quality. A parallel set of identically treated samples were used to harvest cells for protein lysates by addition of 300 μ l 2 \times SDS sample buffer [4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue, and 0.125 M Tris-HCl, pH 6.8] per well followed by incubation at 90 °C for 8–10 min until all cell material is dissolved. Knockdown efficiency was assessed by either western blotting or qRT-PCR.

For qRT-PCR, RNA samples were treated with DNase I (Thermo Fisher Scientific) according to the manufacturer's protocol. First-strand cDNA synthesis was carried out using the Maxima First Strand cDNA synthesis Kit for qPCR (Thermo Fisher Scientific) according to the manufacturer's protocol. qPCR reactions were prepared using gene-specific primers and Platinum SYBR Green qPCR Supermix-UDG (Thermo Fisher Scientific) according to the manufacturer's protocol. An AriaMx Real-time PCR System (Agilent Technologies) was used for quantification of RNA levels and the X_0 method was used for calculations of relative RNA levels [49] normalized to *GAPDH* mRNA.

For western blotting, cell lysates dissolved in SDS load buffer were heated at 90 °C for 3 min and separated on a Novex WedgeWell 4–12% Tris-Glycine Gel (Invitrogen). Proteins were transferred to a PVDF Transfer Membrane (Thermo Scientific) using standard procedures.

The membranes were blocked in 5% skimmed milk powder in PBS for 1 h at room temperature. The membranes were incubated at 4 °C overnight with primary antibodies diluted as indicated in 5% skimmed milk powder in PBS. After three times wash in 13 ml PBS, the membranes were incubated with goat polyclonal HRP-conjugated secondary antibodies (Dako) diluted 1:20,000 in 5% skimmed milk powder in PBS. After 1 h of incubation at room temperature, the membranes were washed three times in 13 ml PBS and the bound antibodies were detected using the SuperSignal West Femto maximum sensitivity substrate (Thermo Scientific) according to the manufacturer's protocol and the LI-COR Odyssey system (LI-COR Biosciences).

We quantified mRNA expression using QuantSeq [50]. Sequencing libraries were generated using the QuantSeq 3'mRNA Library Prep Kit Protocol (Lexogen). Input was 500 ng RNA and 12 PCR cycles were applied. Library concentrations were measured on Qubit 3.0 (Invitrogen), and the average size of the library was measured on TapeStation. The libraries were 1 × 75 bp sequenced on a NextSeq500 system (Illumina). QPCR Universal Human Total Reference RNA (UHR) (Agilent, Cat no: 750500) was included in all batches to assess batch effect.

The raw reads were converted to fastq format and demultiplexed using Illumina's bcl2fastq v2.20.0.422, and library adapters were removed from the read pairs (trim_galore v0.4.1). Reads were mapped to the human genome (hg19) using TopHat2 (version 2.1.1) [37] and Bowtie2 (version 2.1.0.0) [38], and Cufflinks (v2.1.1) [39] and HTSeq (v0.11.2) [40] were used to estimate the transcript abundance using transcript information from GENCODE v19. Samtools (v1.3) [41] and Picard (v2.0.1) were used for quality control and statistics.

RIP assay for FL3 cells

1.65×10^6 FL3 cells were seeded in a P10 dish and transfected 24 h later with 12 µg total DNA (2 µg Twin-Strep-Tag-RBP expression vector and 10 µg pcDNA3 PL). Forty-eight hours later, cells were washed with PBS and placed on ice. One milliliter cold lysis buffer (50 mM TRIS-HCl pH 7.5, 10 mM NaCl, 5 mM MgCl₂, 0.5% Triton X-100, 2% Hexane-1,6-diol, 1 pill Complete pr. 10 mL) was added per plate, and cells were scraped off and transferred to an Eppendorf tube. Samples were mixed and spun (13,000 RPM, 15 min at 4 °C), and 50 µL supernatant was transferred to 50 µL SDS load buffer while 100 µL supernatant was transferred to 0.5 mL Trizol (INPUT). Eight hundred microliters supernatant was incubated with pre-equilibrated MagStrep type 3 XT beads (IBA Life sciences) and rotated for ≥ 2 h at 4 °C. Five hundred microliters supernatant was collected (FT) and samples were washed 4× with 1.5 mL WASH1 buffer

(10 mM TRIS-HCl pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 0.1% Triton X-100) and 2× with WASH2 buffer (10 mM TRIS-HCl pH 7.5, 150 mM NaCl, 5 mM MgCl₂). During the last wash, sample beads were divided 1:10 (RNA:protein) and added 40 µL SDS load buffer or 0.5 mL Trizol (IP). INPUT, FT, and IP samples were analyzed with western blotting and RT-qPCR.

Total RNA-Seq of KHSRP KD in HepG2 and K562

Total RNA from KHSRP knockdown and control samples were obtained from Brenton Graveley's Lab, UConn [34]. Total RNA-sequencing libraries were prepared using the KAPA RNA HyperPrep Kit with RiboErase for Illumina Platforms (KAPA Biosystems). Briefly, DNA oligos were hybridized to rRNA and digested using RNase H treatment followed by a 2.2× bead-based cleanup. Next, RNAs hybridized to rRNA targeting oligos were removed from the samples via RNase H digestion, followed by a 2.2× bead-based cleanup. The rRNA-depleted RNA was then fragmented to an insert size of 200–300 bp at 94 °C for 6 min in the presence of Mg²⁺. First-strand cDNA was synthesized using random primers followed by 2nd-strand dscDNA synthesis marked by dUTP and A-tailing with dAMP at the 3'-end. 3'-dTMP adapters were ligated to the 3'-dAMP library fragments. A 0.63× bead-based cleanup followed by a 0.7× bead-based cleanup were performed, and the purified, adapter-ligated library DNA was amplified with 12 amplification cycles followed by a 1× bead-based cleanup. Post-capture libraries were barcoded and pooled for sequencing. Libraries were sequenced on an Illumina HiSeq 4000 instrument to a depth of approximately 40 million 100 bp paired-end reads. CircRNA expression was profiled using CIRI2 as described for HepG2 and K562 with the following tool versions: Trim Galore (v0.4.1), cutadapt (v1.15), bwa (v0.7.17), and samtools (v1.9).

Bladder cancer patient cohort

RNA was paired-end sequenced using an Illumina NextSeq 500 instrument. Reads were demultiplexed using bcl2fastq v2.18.0.12 trimmed for traces of adapters using Trim Galore v0.4.1 and mapped to the hg19 genome build using tophat v2.1.1. Gene expression was estimated using cufflinks v2.1.1 and HTseq v0.6.1. CircRNA expression was quantified using the CIRI2 pipeline as described above for HepG2 and K562.

Statistical analyses

All statistical tests were performed in R [51, 52]. The non-parametric Wilcoxon rank sum test was utilized to evaluate the enrichment of RBP binding sites in circularizing exons compared to non-circularizing exons from different groups. It was also applied to compare the

enrichment of RBP binding sites in circularizing exons to non-circularizing exons for individual RBPs. The chi-square test was assessed to evaluate enrichment of individual RBPs in flanking introns of circRNAs compared to non-circularizing exons. The empirical p value was calculated to evaluate the significance of a specific observation compared to the distribution of all observations. *DESeq2* [53] was used for differential expression analyses of circCDYL and RBP knockdown data. The R packages *Survival* [54, 55] and *Survminer* [56] were used to produce Kaplan-Meier plots and curves were compared statistically by the log-rank test. For multiple testing corrections, we applied Benjamini-Hochberg correction and statistical differences were declared significant at $FDR < 0.1$. When multiple testing was not applied, statistical differences were declared significant at $P < 0.05$. Most of the plots were produced with the R package *ggplot2* [57].

Results

Circular RNAs are highly expressed in HepG2 and K562 and generally colocalize with RBPs

To identify and quantify the expression of circRNAs in HepG2 and K562 cells, we utilized all available ENCODE whole transcriptome RNA-Seq data sets for these cell lines (HepG2; $n = 5$ and K562; $n = 11$) (Additional file 1+2: File S1 + S2, Additional file 5: Table S3) [34]. Employing the CIRI2 [35] and CIRCexplorer [36] pipelines, we detected 16,033 and 14,648 unique circRNAs supported by at least two reads spanning the backsplice junction (BSJ) by both pipelines in HepG2 and K562, respectively, of which 7705 circRNAs were identified in both cell lines (Fig. 1b, Additional file 6+7: Table S4 + S5). Of all identified circRNAs, 55% (HepG2) and 61% (K562) are described in circBase, a large compendium of circRNAs from different studies [58]. As many circRNAs are lowly expressed and hard to distinguish from artifacts [59] (Additional file 8: Fig. S1A), we generated two sets of high-confidence circRNAs by considering only the 1% highest expressed circRNAs in each cell line (Fig. 1b). These sets contain 161 circRNAs in HepG2, each supported by at least 338 reads, and 147 circRNAs in K562, each supported by at least 270 reads, and were used in most downstream analyses. The vast majority of these highly expressed circRNAs (top 1%) are found in circBase (97% in HepG2 and 98% in K562). We compared the expression of the top 1% circRNAs to their corresponding linear transcript (Fig. 1c, Additional file 8: Fig. S1B (K562)). Taking the relative ratios between circular and linear reads (“Methods”), we found that 72% of circRNAs in HepG2 are more expressed than the linear counterpart, among these common circles like circHIPK3 and circCDYL. Corroborating previous findings, we observed that most highly expressed circRNAs are

comprised of 5 or less exons [3] (Additional file 8: Fig. S1C) and that exons giving rise to 1-exon-circRNAs are significantly longer than exons giving rise to multiple-exon-circRNAs [4] ($P < 4.8e-07$, Wilcoxon rank sum test, Additional file 8: Fig. S1D).

Generally, circRNAs are only comprised of exonic sequences as introns are spliced out in the circularization process [4]. Based on eCLIP data and a high-confidence set of peak calls representing binding sites of RBPs in HepG2 ($n = 103$) and K562 ($n = 120$) (“Methods”, Additional file 9+10: File S3 + S4), we evaluated the genomic positions of RBP targets in the top 1% circRNA loci (Additional file 11: Table S6). Among protein-coding genes, 54% of the RBP binding sites are within exonic parts of the circRNAs (span 0.119 Mb), while 35% of the binding sites are found strictly in introns (span 1.836 Mb) and 10% at exon-intron boundaries in HepG2 (Fig. 1d, Additional file 8: Fig. S1E (K562)). The remaining 1% are found in other regions of the genome than protein-coding genes. As introns are usually spliced out and RBPs that bind to exon-intron boundaries are likely involved in normal pre-mRNA splicing [60], we focused initially on RBPs with binding sites in exonic regions of circRNAs.

To be able to interact, circRNAs and RBPs should be present in the same cellular compartments of the cell. Based on ENCODE subcellular fractionated RNA-Seq expression data and immunofluorescence imaging of RBP occupancy in HepG2 [43], we evaluated the localizations of circRNAs and RBPs in HepG2. We found that all circRNAs are located in the cytoplasm as ordinarily the case [4, 5, 61] (Fig. 1e). Though some circRNAs are also found in the nucleus, no circRNAs are more highly expressed in the nucleus than in the cytoplasm (Additional file 8: Fig. S1F). For the RBPs ($n = 64$), most are found in both compartments ($n = 38$), while some are found exclusively in the cytoplasm ($n = 12$) and some in the nucleus ($n = 14$) (Additional file 12: Table S7). For 94% of the circRNA-RBP interactions inferred from eCLIP data, we found that the involved circRNA and RBP were co-located in the same subcellular compartments, indicating potential to interact (Fig. 1f).

KHSRP binding is enriched in introns flanking circRNAs and affects biogenesis

To comprehensively analyze if RBPs influence circRNA formation in HepG2 and K562 cells, we evaluated whether binding sites for individual RBPs are enriched in introns flanking circRNAs compared to non-circularizing exons. We divided the genomic regions into different sets (Fig. 2a): (1) the subset of highly expressed, high-confidence circRNAs; (2) all other circRNAs; and (3) non-circularizing exons (non-circ-exons) from genes producing circRNAs. For the latter, we excluded the first

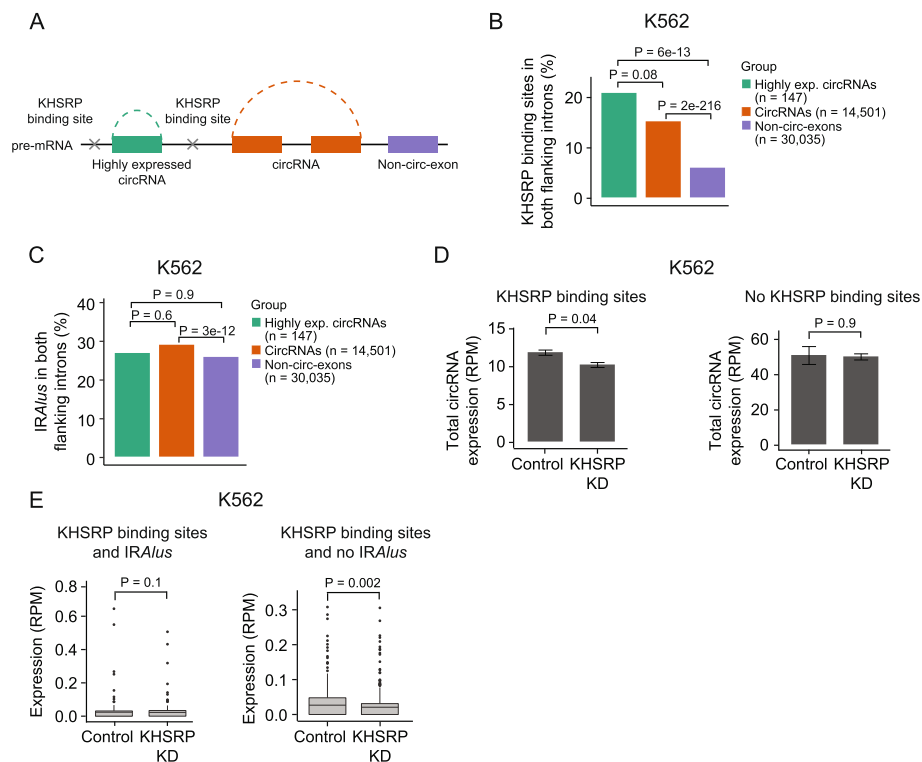


Fig. 2 KHSRP binding is enriched in introns flanking circRNAs and affects biogenesis. **a** Illustration of circRNA and exon categories and intronic location of KHSRP binding sites. The categories are defined as the top 1% highest expressed circRNAs (green), all other circRNAs (orange), and non-circularizing exons from genes producing circRNAs (purple). First and last exons in each gene are disregarded for the analyses. **b** Percentage of circRNAs and non-circ-exons in genes producing circRNAs with KHSRP binding sites in both flanking introns in K562. *P* values obtained by chi-square test. **c** Percentage of circRNAs and non-circ-exons in genes producing circRNAs with inverted repeated *Alu* elements (*IRAlus*) in both flanking introns in K562. *P* values obtained by chi-square test. **d** Total expression of circRNAs with (left, $n = 297$) and without (right, $n = 1537$) KHSRP binding sites in both flanking introns in KHSRP knockdown (KD) and control samples in K562. *P* values obtained by *T*-test. RPM = reads per million. **e** Expression of circRNAs with KHSRP binding sites and with (left, $n = 95$) or without (right, $n = 202$) *IRAlus* in both flanking introns in KHSRP knockdown (KD) and control samples in K562. *P* values obtained by *T*-test

and last exons in each gene as they are not surrounded by introns on both sides.

For all individual RBPs, we evaluated the presence of binding sites in both intronic regions (10 kb to each side) of circRNAs and non-circ-exons (Additional file 8: Fig. S2A + B). Noteworthy, we found that > 20% of the highly expressed circRNAs ($n = 141$) and 15% of all other circRNAs ($n = 14,501$) in K562 possess KHSRP binding sites in both flanking introns, which is 3.4 and 2.5 times more than non-circ-exons ($n = 30,035$), respectively (all $P < 0.001$ for all comparisons between circRNAs and non-circ-exons, chi-square test, Fig. 2b). Similar significant results were observed in HepG2, although fewer circRNAs are surrounded by KHSRP binding sites (all $P < 0.001$, chi-square test, Additional file 8: Fig. S2C).

Since *IRAlus* can also facilitate circularization, we evaluated the existence of *IRAlus* in both flanking introns of circRNAs and non-circ-exons. Highly expressed circRNAs showed no enrichment of *IRAlus* ($1.0\times$, $P = 0.9$, chi-square test) while all other circRNAs were only

weakly enriched ($1.1\times$, $P = 3e-12$, chi-square test) compared to non-circ-exons in K562 (Fig. 2c). The same pattern is observed for HepG2 cells (Additional file 8: Fig. S2D). This indicates that *IRAlus* are not a major driver of circularization in K562 and HepG2 cells. Specifically, flanking introns of circRNAs with KHSRP binding sites were not enriched with *IRAlus* (Additional file 8: Fig. S2E).

KHSRP binds to single-stranded RNA and exerts diverse functions in RNA metabolism, e.g., by promoting mRNA decay, inducing miRNA maturation, and affecting alternative RNA splicing [62, 63]. Based on immunofluorescence imaging in HepG2, we see that KHSRP is only detectable in the nuclear fraction of the cell (Additional file 12: Table S7). We hypothesize that KHSRP binds to flanking introns of circRNAs in the pre-mRNA transcript in the nucleus and thereby promotes circRNA biogenesis.

To test our hypothesis, we profiled the expression of circRNAs using CIRI2 upon KHSRP knockdown (KD) in

K562 and HepG2 cells (“Methods”) [64]. Overall, we found no difference in total circRNA expression (supported by at least 2 reads) between K562 control and KHSRP KD samples ($n = 1834$) ($P = 0.6$, T -test, Additional file 8: Fig. S2F). However, for circRNAs with KHSRP binding sites in both flanking introns (KHSRP-circRNAs, $n = 297$), we found a 15% decrease in circRNA expression upon KHSRP KD ($P = 0.04$, T -test, Fig. 2d), while there was no effect on expression levels for circRNAs without KHSRP binding sites in both flanking introns in K562 ($n = 1537$) ($P = 0.9$, T -test, Fig. 2d). For KHSRP-circRNAs, we evaluated the presence of *IRAlus* in both flanking introns. We found that the expression of KHSRP-circRNAs without *IRAlus* in flanking introns ($n = 202$) is more affected by KHSRP depletion ($P = 0.002$, Wilcoxon rank sum test, Fig. 2e) than KHSRP-circRNAs surrounded by *IRAlus* ($n = 95$, $P = 0.1$, Wilcoxon rank sum test, Fig. 2e), supporting the role of KHSRP in the biogenesis of a subset of circRNAs.

Since our observations could be explained by overall splicing perturbations in KHSRP KD samples, we evaluated the expression of the corresponding linear RNA of circRNAs. We found no difference in linear RNA expression between KHSRP KD and control samples for circRNAs with ($P = 0.9$, T -test, Additional file 8: Fig. S2G) or without KHSRP binding sites ($P = 0.5$, T -test, Additional file 8: Fig. S2G), indicating that KHSRP specifically affect the expression of circRNAs.

We observed no effect of KHSRP KD on circRNA expression levels of circRNAs with ($P = 0.5$, T -test, Additional file 8: Fig. S2H) or without ($P = 0.1$, T -test, Additional file 8: Fig. S2H) KHSRP binding sites in both flanking introns in HepG2. This could be explained by lower expression levels of KHSRP (1.92 FPKM in the nucleus of HepG2 cells compared to 2.98 FPKM in the nucleus of K562 cells) and a generally lower fraction of circRNAs with KHSRP binding sites in flanking introns (Additional file 8: Fig. S2C).

Taken together, our results identify KHSRP as an RBP that appears to be involved in the biogenesis of a subset of circRNAs with KHSRP binding sites in flanking introns in K562 cells.

Exons comprising circRNAs are enriched with RBP binding sites

Most circRNAs consist of protein-coding exons. If RBP binding is enriched in circRNAs, it suggests a regulatory layer in addition to their protein-coding capacity.

To evaluate the enrichment of RBP binding sites in circRNAs, several considerations had to be taken into account. Since circRNAs share sequence with their cognate linear transcript, RBP targets from eCLIP data do not directly distinguish between interactions with

circular or linear RNA transcripts. Additionally, gene expression levels influence the ability to detect RNA-RBP interactions. By dividing all genes into expression deciles, we found that the fraction of exons covered with RBP binding sites (RBP coverage) increases with transcript abundance as expected ($R > 0.4$, $P < 2.2e-16$, Pearson’s product-moment correlation, Additional file 8: Fig. S3A).

To evaluate whether circRNAs show more RBP enrichment than expected simply from the expression level of their parent gene, we compared the RBP coverage of exons comprising circRNAs to linear exons that are not involved in circularization within different comparable gene sets: (A) circRNA host genes, (B) genes of the same expression level, and (C) the most highly expressed genes. We divided exons into three different groups: (1) backsplice junction circularizing exons (BSJ circ-exons), which are directly involved in backsplicing; (2) circularizing exons (circ-exons) that are potentially part of circRNAs but not supported by BSJ reads; (3) and non-circ-exons, which are not part of circRNAs (Fig. 3a). Since many RBPs regulate transcription and translation by binding to the 5′- and 3′-UTRs of mRNAs [65] and since they give rise to few circRNAs, we disregarded the first and last exons in each gene (Additional file 8: Fig. S3B).

Interestingly, for the high-confidence, highly expressed circRNAs (top 1%), we found that BSJ circ-exons were significantly more covered with RBP binding sites than internal non-circ-exons within the same genes for both K562 ($P = 5.1e-08$, Wilcoxon rank sum test, Fig. 3b) and HepG2 ($P = 0.027$, Wilcoxon rank sum test, Additional file 8: Fig. S3C). The difference was also significant when examining the complete set of circRNAs and their associated non-circ-exons (all $P < 2.2e-16$, Wilcoxon rank sum test, Additional file 8: Fig. S3D).

To compare RBP coverage of BSJ circ-exons to genes of the same expression levels as genes producing highly expressed circRNAs (circ-genes), we divided all genes into expression level percentiles. We randomly drew non-circ-exons, while ensuring the same expression level distribution as circ-genes, which are distributed across percentiles ~ 20 –100 (Additional file 8: Fig. S3E). We found that BSJ circ-exons are significantly more covered with RBP binding sites than non-circ-exons of the same expression level in both cell lines ($P < 0.01$, empirical p value, Fig. 3c (K562), Additional file 8: Fig. S3F (HepG2)).

Even when BSJ circ-exons were compared against the highest expressed genes, their RBP coverage were significantly higher ($P < 0.01$, percentiles 90–95) or similar (P not significant, percentiles 96–100), indicating specific circRNA-RBP interactions (Fig. 3c (K562), Additional file 8: Fig. S3F (HepG2)).

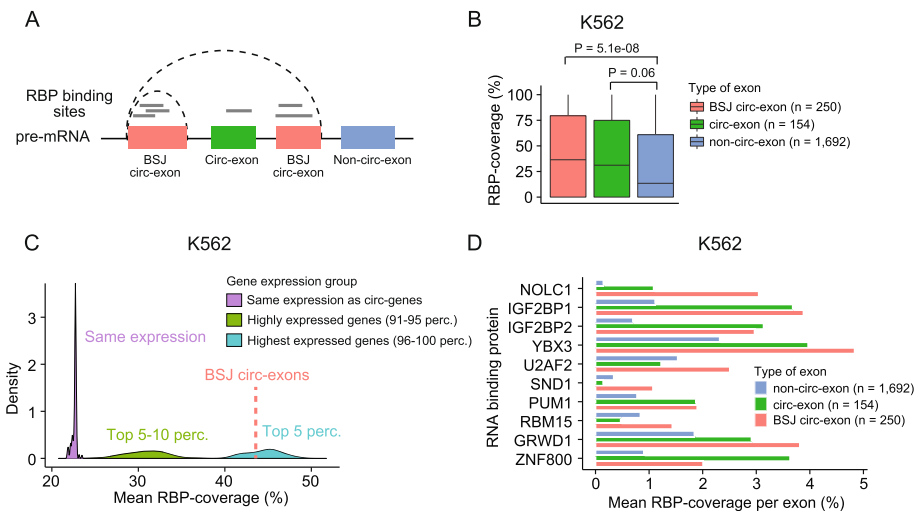


Fig. 3 Exons comprising circRNAs are enriched with RBP binding sites. **a** Illustration of exon categories in genes forming circRNAs. Backsplice junction circularizing exons (BSJ circ-exons, red) are involved in the backsplicing event. Circularizing exons (circ-exons, green) are potentially part of circRNAs, if not spliced out. Non-circularizing exons (non-circ-exons, blue) are not part of circRNAs. Gray lines represent RBP binding sites. To evaluate the fraction of exons covered by RBP binding sites overall, the individual overlapping RBP binding sites were merged. **b** RBP binding site coverage for each category of exons in genes that produce highly expressed (top 1%) circRNAs in K562 ($n = 124$). P values obtained by Wilcoxon rank sum test. **c** Comparison of RBP coverage between BSJ circ-exons from top 1% circRNAs and exons in groups of genes of different expression levels (K562). The mean RBP coverage of the highly expressed BSJ circ-exons ($n = 250$) is 44% (red punctuated line; median = 36%). Exons randomly sampled from genes while ensuring the same expression profile as genes producing highly expressed circRNAs (circ-genes) have a much lower mean coverage of 23% (purple; median = 0%). Exons of highly expressed genes (top 5–10 percentiles) also showed a lower mean coverage of 31% (green; median = 3%), while the most highly expressed genes (top 5 percentiles) had a slightly higher mean of 45% (blue; median = 43%). The random sampling procedures were repeated with 100 iterations. Empirical P values for BSJ circ-exons vs. same expression, $P < 0.01$; BSJ circ-exons vs. Top5–10, $P < 0.01$; BSJ circ-exons vs. Top5, $P = 0.67$. **d** Mean RBP coverage per exon for individual RBPs (K562). All RBPs shown here have significantly more target sites in highly expressed BSJ circ-exons than in non-circ-exons of the same genes (FDR < 0.1 , Wilcoxon rank sum test). Only RBPs with at least 20 distinct binding sites in total are considered

We evaluated whether any individual RBPs preferentially bind to circRNA exons. Several RBPs have significantly higher RBP coverage on average in BSJ circ-exons compared to non-circ-exons in circ-genes (false discovery rate (FDR) < 0.1 for all shown RBPs, Wilcoxon rank sum test, Fig. 3d (K562), Additional file 8: Fig. S3G (HepG2)). We found that some RBPs are more prone to bind to BSJ circ-exons (and to circ-exons in some cases) in both cell lines, e.g., GRWD1 and ZNF800, while others seem to be cell line specific, like NOLC1, IGF2BP1, and YBX3 in K562. We performed GO enrichment analysis to identify GO terms that are over-represented in the subset of RBPs that bind more to BSJ circ-exons (K562; $n = 10$ and HepG2; $n = 7$). Although we observed no significant results, we found that a larger fraction of RBPs that are more prone to bind to BSJ circ-exons are associated with regulation of mRNA stability (3.8 fold), translation (3.7 fold), and binding to the mRNA 5'-UTR (10.1 fold) and 3'-UTR (3.4 fold) compared to the rest of the RBPs in K562 (Additional file 13: Table S8). These results suggest that circRNAs could influence transcript expression and function by interacting with RBPs that are involved in mRNA maturation.

Finally, we explored the ability of RBPs to interact with the unique BSJ of circRNAs. For this, we aligned eCLIP reads to a constructed reference set of all possible backsplicing events based on annotated splice sites (“Methods”, Additional file 8: Fig. S3H). Some potential backsplicing events were covered by eCLIP reads in HepG2 ($n = 111$) and K562 ($n = 133$) (Additional file 8: Fig. S3I); however, none of these correspond to a BSJ of the circRNAs we called.

Overall, our results show that exons comprising circRNAs are enriched with RBP binding sites and that some RBPs preferentially bind to circRNAs. Though RBP binding sites overlapping BSJ circ-exons could simply stem from binding to the linear form of the exons, circularizing exons generally had much higher RBP coverage than comparable linear exons, supporting that they specifically interact with RBPs.

CircRNAs interact with RBPs in a cell-type-specific manner

Next, we wanted to identify specific circRNAs enriched with RBP binding sites that could potentially function as RBP sponges or in other ways interact with RBPs to deregulate the expression, function, or localization of RBPs. We evaluated the coverage of RBP binding sites in the

exonic parts of the top 1% circRNAs (Additional file 3: Table S1). Many known and highly expressed circRNAs are completely or almost completely covered by RBP binding sites in one or both of the cell lines, e.g., circRNAs arising from the genes *RBM39*, *GSE1*, *SMARCA5*, *RBM33*, *ZKSCAN1*, and *CDYL* (Fig. 4a). Overall, RBP coverage is not significantly associated with the number of exons comprising the circRNAs (Additional file 8: Fig. S4A). We compared the RBP coverage in highly expressed circRNAs to the RBP coverage of

internal non-circularizing exons transcriptome-wide and found that several of the abovementioned circRNAs fall in the 5% tail (as indicated by stars) in one, e.g., circGSE1 and circSMARCA5, or both cell lines, e.g., circCDYL and circRBM33 (Fig. 4b, Additional file 3: Table S1).

To ensure our observations are not simply explained by host gene expression levels, we compared the RBP coverage of highly expressed circRNAs to non-circularizing exons within the parent gene (“Methods”,

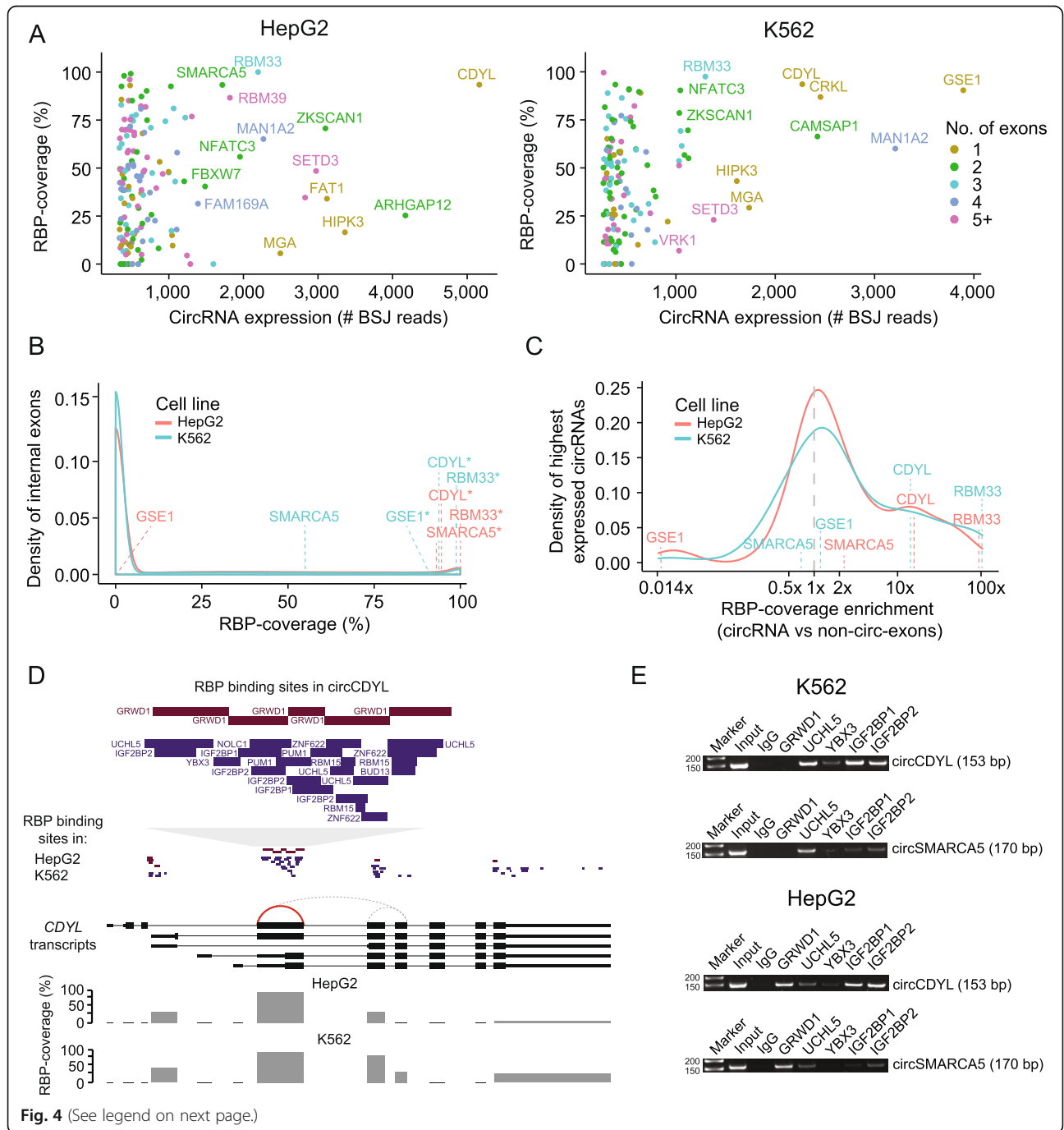


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 CircRNAs interact with RBPs in a cell-type-specific manner. **a** Percentage of the exonic part of circRNAs that is covered with RBP binding sites in HepG2 (right) and K562 (left). Only the top 1% highest expressed circRNAs are shown. Some circRNAs are depicted with their host gene name. Some circRNAs are highly covered with RBP binding sites in both cell lines, e.g., circCDYL and circRBM33. Colors indicate the number of exons constituting each circRNA. We only consider circRNAs with annotated exons. **b** Comparison of RBP coverage for individual highly expressed circRNAs and all internal non-circ-exons in expressed genes in HepG2 (red) and K562 (blue). Common circRNAs that are highly covered with RBP binding sites in one or both cell lines are depicted with their host gene name. Stars indicate an RBP coverage in the 5% tail. **c** RBP coverage enrichment for the top 1% highest expressed circRNAs. The RBP coverage enrichment was calculated by dividing circRNA-RBP coverage with the mean RBP coverage of internal non-circ-exons of the host gene in HepG2 (red) and K562 (blue). Most circRNAs contain an equal amount or slightly more RBP binding sites than non-circ-exons in their parent gene ($\geq 1\times$ enrichment), while some circRNAs are highly enriched with RBP binding sites, e.g., circCDYL ($> 10\times$) and circRBM33 ($100\times$). **d** RBP binding sites in exons of the *CDYL* gene. Exons of the five main *CDYL* transcripts (thick part indicates protein-coding regions) and all circRNAs (stipulated lines) supported by at least two reads by both pipelines are shown. The highly expressed circCDYL (chr6:4891946-4892613; supported by 5163 (HepG2) and 2272 (K562) reads based on CIRI2) is indicated by a red line. This exon is highly covered by RBP binding sites in both cells. Zoom-in at the top shows the binding sites and names of the individual RBPs in both cell lines. All other *CDYL*-circRNAs are supported by less than 22 reads. Bars show the percentage of each exon covered with RBP binding sites. **e** Validation of circRNA-RBP interactions from RNA immunoprecipitation. All circRNA-RBP interactions, except for one, were verified. Here, RNA immunoprecipitation (RIP) experiments confirmed that circCDYL interacts with several RBPs in K562, e.g., UCHL5, IGF2BP1, IGF2BP2, but not with GRWD1. In HepG2, RIP experiments confirmed strong circCDYL-GRWD1 interactions. Predicted interactions between circSMARCA5 and UCHL5 in K562 and GRWD1 in HepG2 were also validated. 50-bp markers were used

Additional file 3: Table S1). We found that most circRNAs (71%) have the same or higher RBP coverage ($\geq 1\times$ enrichment) than non-circularizing exons in the same gene (Fig. 4c). Of these, 24% are highly enriched ($\geq 10\times$). Our results indicate that gene expression cannot explain the high RBP coverage of specific circRNAs, as some circRNAs differ in RBP coverage between cell lines despite similar circRNA and host gene expression levels. For instance, circGSE1 as well as non-circ-exons in *GSE1* are highly covered with RBP binding sites in K562 ($1.2\times$), while there are no RBPs overlapping circGSE1 in HepG2 but many RBP binding sites in *GSE1* non-circularizing exons (ratio = 0.02). In both cell lines, circGSE1 is among the 1% highest expressed circRNAs and the parent gene *GSE1* is found in similar gene expression percentiles in K562 (percentile 81) and HepG2 (percentile 83).

Interestingly, circCDYL ($\sim 11\times$) and circRBM33 ($\sim 100\times$) are highly covered with RBP binding sites compared to non-circ-exons in their parent gene in both cell lines (Fig. 4c). These enrichments are significantly higher than for other genes with the same number of non-circ-exons ($P < 0.05$, empirical p value). The *RBM33* gene is highly expressed in both K562 (percentile 73) and HepG2 (percentile 91). The high enrichment for circRBM33 is due to $\sim 100\%$ RBP coverage of circRBM33 and no RBP binding sites in the single internal non-circ-exon in the gene. The *CDYL* host gene, which gives rise to the highest expressed circRNA, is itself modestly expressed in both K562 (percentile 48) and HepG2 (percentile 62). CircCDYL is 93% covered with RBP binding sites in both cell lines while the non-circ-exons in *CDYL* have a mean RBP coverage of only 7% (Fig. 4d). Remarkably, based on the analyzed eCLIP data, the RBPs binding to the circCDYL exon differ between the cell lines. In HepG2,

only one RBP, GRWD1, is binding across almost the entire sequence, while nine different RBPs have at least one binding site in K562. We observe that cell-type-specific circRNA-RBP interactions are a general phenomenon. Between 73 circRNAs that are highly expressed in both cell lines (Additional file 8: Fig. S4B) and 28 RBPs evaluated in both cell lines (Additional file 8: Fig. S4C), only 11% of circRNA-RBP interactions ($n = 634$) are shared between the cell lines (Additional file 8: Fig. S4D).

To evaluate the validity of the predicted circRNA-RBP interactions based on the eCLIP data, we performed RNA immunoprecipitation (RIP) of five RBPs, IGF2BP1, IGF2BP2, GRWD1, YBX3, and UCHL5, that are predicted to bind to one or several of four highly expressed and highly RBP-covered circRNAs, circCDYL, circRBM33, circZKSCAN1, and circSMARCA5 (Additional file 8: Fig. S4E). RBP antibody specificity was validated using western blots (Additional file 8: Fig. S4F). We designed divergent primers against the unique backsplice junction of the circRNAs to specifically verify circRNA pull-down in RIP experiments by RT-PCR (Additional file 14: Table S9). We evaluated all possible interactions between the RBPs and circRNAs and all, except for one, were verified. In K562, the RIP experiments confirmed that circCDYL interacts with UCHL5, YBX3, IGF2BP1, and IGF2BP2 and that circSMARCA5, circZKSCAN1, and circRBM33 binds UCHL5 (Fig. 4e, Additional file 8: Fig. S4G). In HepG2, the RIP experiments confirmed interactions between GRWD1 and all four circRNAs (Fig. 4e, Additional file 8: Fig. S4G). There were no eCLIP data of IGF2BP2 in HepG2 but strong bands in the IGF2BP2 immunoprecipitation (IP) experiment showed that IGF2BP2 interacts with circZKSCAN1 and circCDYL. The only interaction that could not be confirmed was between GRWD1 and

circZKSCAN1 in K562 (Additional file 8: Fig. S4G). One likely explanation is that GRWD1 only interacts with the linear form of *ZKSCAN1* or with circZKSCAN1 under certain conditions. Faint bands indicate that the circRNAs interact with other RBPs than the ones identified from the eCLIP data with lower affinity. The lack of low-affinity and nonspecific RBP binding sites in the eCLIP data could be explained by the stringent cut-off used, of 8-fold enrichment in IP compared to input (“Methods”). In accordance with the eCLIP data, our results show that circRNAs interact with RBPs in a highly cell-type-specific manner.

Functional studies of circCDYL-RBP interactions in HepG2

There are several ways in which circRNAs could interact with and regulate RBPs. CircRNAs could function as protein decoys and retain certain RBPs to specific cellular compartments, or as scaffolds to facilitate contact between two or more RBPs, or as a unit in larger functional complexes.

To understand the regulatory potential of circRNA-RBP interactions in more depth, we focused our attention on circCDYL’s interaction with RBPs. CircCDYL is one of the highest expressed circRNAs across tissues in both humans and mice [68] and is deregulated in diseases, including cancer and myotonic dystrophy [9, 44, 69]. However, little is known about its regulatory functions and its potential to interact with RBPs remains unexplored.

Here, we found that circCDYL is one of the highest expressed circRNAs in both cell lines and almost entirely covered with RBP binding sites (Fig. 4a). As seen for most circRNAs, circCDYL levels are highest in the cytosol and the circle is more abundant than its linear counterpart in both HepG2 (Fig. 5a) and K562 (Additional file 8: Fig. S5A). All RBPs confirmed to interact with circCDYL are expressed in the cytoplasm as well (Additional file 12: Table S7). Therefore, we hypothesized that circCDYL interacts with RBPs in the cytoplasm and regulates associated RBP target genes.

To functionally characterize circCDYL, we conducted siRNA-mediated knockdown (KD) in HepG2 with an siRNA specifically recognizing the backsplice junction of circCDYL (“Methods”, Additional file 4: Table S2). We validated efficient and specific circCDYL KD by qRT-PCR ($P = 7.6e-05$, T -test, Additional file 8: Fig. S5B, Additional file 15: Table S10) and observed no downregulation of the parent gene ($P = 0.061$, T -test, Additional file 8: Fig. S5B, Additional file 15: Table S10). We harvested RNA in triplicates upon circCDYL KD and quantified mRNA expression using QuantSeq [50, 70]. We identified 2233 genes that are differentially expressed

(DE) upon circCDYL KD, of which 1255 are upregulated and 978 are downregulated (Fig. 5b). KEGG pathway analyses showed that pathways in cancer and specifically the MAPK signaling pathway are activated upon circCDYL KD (FDR < 0.1, Fig. 5c).

Our analyses showed that circCDYL interacts strongly and specifically with GRWD1 in HepG2. RIP experiments showed that circCDYL also interacts with IGF2BP1 and IGF2BP2. If circCDYL regulates mRNA abundance through its interaction with RBPs, we would expect an enrichment of RBP binding sites in genes affected by circCDYL KD. Based on eCLIP data, we evaluated the presence of GRWD1 and IGF2BP1 binding sites in all genes and found that GRWD1 binding sites are enriched in downregulated genes ($P = 8e-14$, chi-square test), while IGF2BP1 binding sites are enriched in upregulated genes ($P = 5e-12$, chi-square test, Fig. 5d). For interaction to occur, circCDYL has to be present in comparable numbers as its interacting RBPs. Importantly, we found that circCDYL is much higher expressed than the mRNAs of *GRWD1* (> 16×) and *IGF2BP1* (3.8×) in the cytosol of HepG2 cells (“Methods”, Additional file 8: Fig. S5C).

GRWD1 is a multifunctional protein (reviewed in [71]), which is overexpressed in cancer cells [72]. Specifically, elevated GRWD1 expression has been shown to negatively regulate *TP53* and promote tumorigenesis [73]. Since circCDYL is highly covered with GRWD1 binding sites, it could potentially function as a sponge for GRWD1 and lower its binding to targets, incl. *TP53*. We assessed whether circCDYL and GRWD1 regulate the same genes by evaluating the overlap of genes affected by KD. We obtained mRNA expression upon GRWD1 KD in HepG2 from ENCODE [34] and found a significant overlap between altered genes upon circCDYL KD and GRWD1 KD ($P < 2.2e-16$, Fisher’s exact test, Fig. 5e). Consistent with an RBP sponge hypothesis, we observed that most downregulated genes upon circCDYL KD are upregulated upon GRWD1 KD (Fig. 5f). Although not significantly, we found a tendency to decreased *TP53* levels upon circCDYL depletion in HepG2 cells ($P = 0.07$, Wald test, Additional file 8: Fig. S5D), while the opposite was observed upon GRWD1 KD ($P = 0.2$, Wald test, Additional file 8: Fig. S5E).

IGF2BP1 is a post-transcriptional regulator that affects the stability, translatability, and localization of essential mRNAs involved in tumor cell proliferation, growth, and invasion (reviewed in [74]). IGF2BP1 is generally assigned oncogenic roles and has been shown to interact with and regulate the expression of genes involved in the MAPK signal transduction pathway [75–77], providing a link to the activation of this pathway upon circCDYL KD. The gene *PLA2G2A* is an extreme outlier upon circCDYL KD with

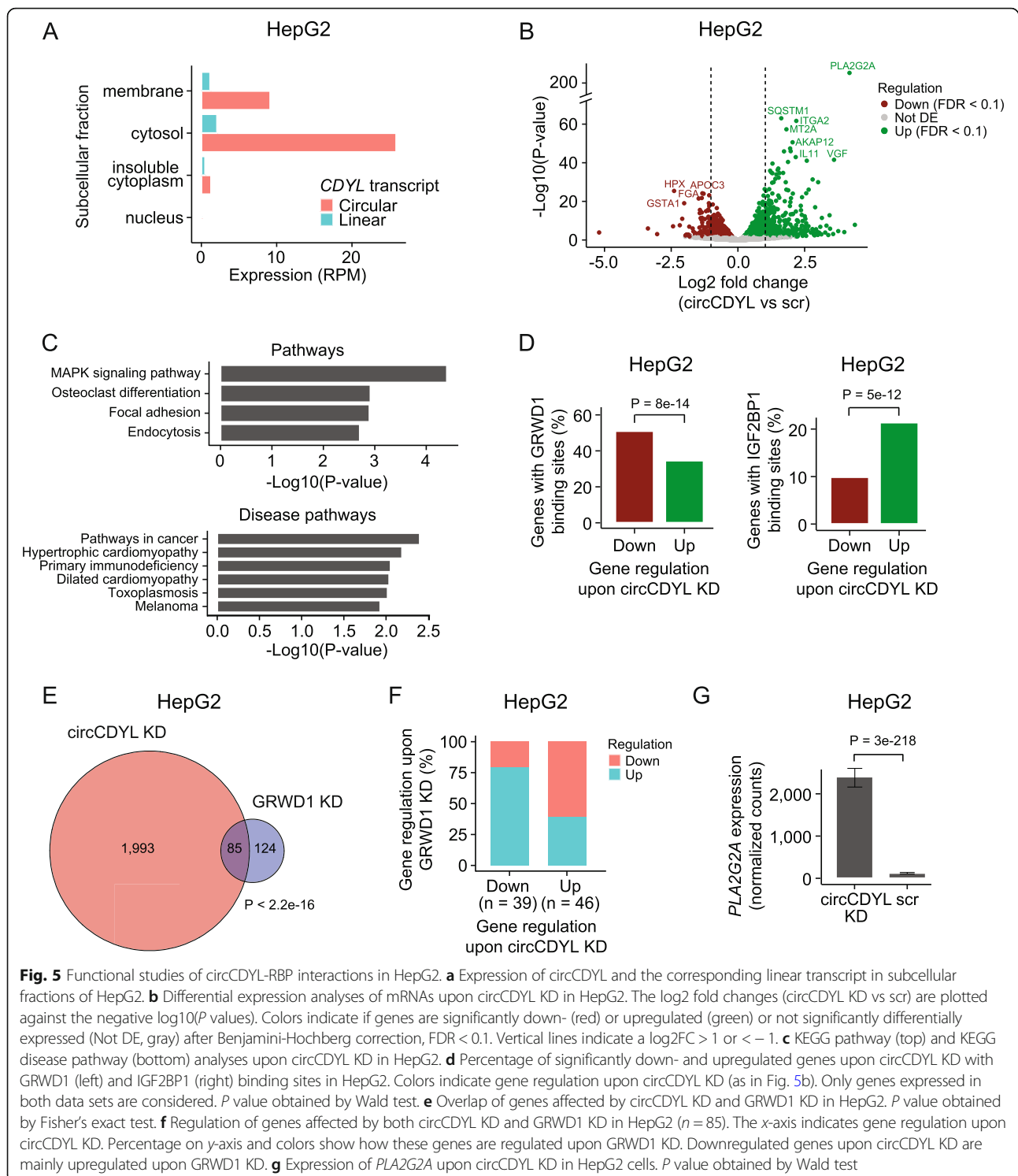


Fig. 5 Functional studies of circCDYL-RBP interactions in HepG2. **a** Expression of circCDYL and the corresponding linear transcript in subcellular fractions of HepG2. **b** Differential expression analyses of mRNAs upon circCDYL KD in HepG2. The \log_2 fold changes (circCDYL KD vs scr) are plotted against the negative $\log_{10}(\text{P values})$. Colors indicate if genes are significantly down- (red) or upregulated (green) or not significantly differentially expressed (Not DE, gray) after Benjamini-Hochberg correction, FDR < 0.1. Vertical lines indicate a $\log_2\text{FC} > 1$ or < -1 . **c** KEGG pathway (top) and KEGG disease pathway (bottom) analyses upon circCDYL KD in HepG2. **d** Percentage of significantly down- and upregulated genes upon circCDYL KD with GRWD1 (left) and IGF2BP1 (right) binding sites in HepG2. Colors indicate gene regulation upon circCDYL KD (as in Fig. 5b). Only genes expressed in both data sets are considered. P value obtained by Wald test. **e** Overlap of genes affected by circCDYL KD and GRWD1 KD in HepG2. P value obtained by Fisher's exact test. **f** Regulation of genes affected by both circCDYL KD and GRWD1 KD in HepG2 ($n = 85$). The x-axis indicates gene regulation upon circCDYL KD. Percentage on y-axis and colors show how these genes are regulated upon GRWD1 KD. Downregulated genes upon circCDYL KD are mainly upregulated upon GRWD1 KD. **g** Expression of *PLA2G2A* upon circCDYL KD in HepG2 cells. P value obtained by Wald test

abundance increasing more than 19 times ($P < 3 \times 10^{-218}$, Wald test, Fig. 5g). *PLA2G2A* is part of the MAPK pathway, is upregulated in colorectal cancer [78], and is associated with poor clinical outcomes [79, 80]. Contrary, *PLA2G2A* is downregulated upon IGF2BP1 KD ($P < 3 \times 10^{-53}$, Wald test, Additional file 8:

Fig. S5F), consistent with an oncogenic role of IGF2BP1 and MAPK pathway activation.

Our results suggest that circCDYL interacts with GRWD1 in HepG2 to regulate GRWD1 target genes overall consistent with an RBP sponge hypothesis. Besides, circCDYL could be involved in other regulatory

mechanisms to regulate classes of genes and individual transcripts in cancer-related pathways.

CircCDYL interacts with RBPs in bladder cancer and abundance is associated with overall survival

In a previous study, we found that circCDYL is highly expressed in patients with non-muscle invasive bladder cancer (NMIBC) and correlates positively with good prognosis independently of the parent gene [44]. Based on our findings here, we hypothesized that circCDYL possess regulatory functions in bladder cancer (BC) by binding RBPs.

First, to investigate circCDYL-RBP interactions in BC, we performed RIP experiments using the three RBPs confirmed to interact with circCDYL in HepG2 and/or K562: IGF2BP1, IGF2BP2, and GRWD1, in BC FL3 cells. The IPs were validated with western blotting (Additional file 8: Fig. S6A), and pull-down of circCDYL (Additional file 8: Fig. S6B) and the *CDYL* mRNA (Additional file 8: Fig. S6C) was quantified by qRT-PCR (Additional file 15: Table S10). The RIP experiments showed that circCDYL interacts with all three RBPs in BC with higher binding affinity than the host gene (Fig. 6a). Contrasting our findings in HepG2, circCDYL pull-down was highly enriched (10×) for IGF2BP1 and IGF2BP2, indicating strong and specific circCDYL-IGF2BPs interactions in BC cells.

Next, to evaluate the functional role of circCDYL-RBP interactions in BC, we conducted siRNA-mediated KD of circCDYL or the three RBPs [70] (Additional file 4: Table S2). To choose which BC cell lines to use in the experiments, we evaluated the expression of circCDYL using CIRI2 [35] in total RNA-Seq data from eleven BC cell lines [44, 81], and chose J82 and UMUC3 based on cell line stability and high circCDYL expression levels (Additional file 8: Fig. S6D). Efficient RBP and circCDYL KD were validated with western blotting (Additional file 8: Fig. S6E) and qRT-PCR ($P < 0.005$, *T*-test, Additional file 8: Fig. S6F).

From differential expression analysis, we identified 2624 and 3014 DE genes upon circCDYL KD in UMUC3 and J82, respectively (Additional file 8: Fig. S6G), with a significant overlap of 976 genes ($P < 2.2e-16$, Fisher's exact test, Additional file 8: Fig. S6H). Of these, 86% show the same direction of deregulation in both cell lines, supporting circCDYL-specific alterations across BC cell lines (Fig. 6b). When we looked at the genes affected by GRWD1 KD, we found a significant overlap between UMUC3 and J82 ($P < 2.2e-16$) but not between the BC cells and HepG2 ($P > 0.4$), consistent with cell-type-specific interactions and functions of circRNAs and RBPs (Additional file 8: Fig. S6I). In both BC cells, we found a significant overlap of perturbed genes upon circCDYL KD and depletion of IGF2BP1 and IGF2BP2,

respectively ($P < 2.2e-16$ for all comparisons, Fig. 6c, Additional file 8: Fig. S6J).

CircCDYL has been suggested to suppress cell growth by inhibiting MYC protein expression levels in bladder cancer through an unknown mechanism [66]. IGF2BP1 and IGF2BP2 possess oncogenic roles and positively regulate the expression of various oncogenes (reviewed in [74, 82]). We evaluated how circCDYL and RBP depletion affect 50 hallmarks of cancer [47]. We found that several proliferation pathways, e.g., E2F targets, G2M checkpoint, and MYC targets, are upregulated upon circCDYL KD in UMUC3 cells, while genes affecting immune processes are downregulated (FDR < 0.1 , Fig. 6d). Consistent with a role for circCDYL as a sponge for IGF2BP1 and IGF2BP2, proliferation pathways are downregulated upon IGF2BP1 KD in both UMUC3 (FDR < 0.1 , Fig. 6e) and J82 cells (FDR < 0.1 , Additional file 8: Fig. S6K), and immune pathways are upregulated upon IGF2BP2 KD (FDR < 0.1 , Additional file 8: Fig. S6K). Consistent with our observations in HepG2, circCDYL is much higher expressed than the mRNAs of its interacting RBPs, *IGF2BP1* ($> 180\times$) and *IGF2BP2* ($> 5\times$), in both bladder cancer cell lines, supporting circCDYL's potential to act as a sponge (Additional file 8: Fig. S6L).

We further examined the distribution of oncogenes [48] among differentially expressed genes upon circCDYL KD and found twice as many oncogenes among upregulated genes than among downregulated genes in all cell lines (Fig. 6f). Accordingly, we found that *MYC* expression is upregulated upon circCDYL KD in UMUC3 cells ($P < 0.005$, Wald test, Fig. 6g) and in J82 cells, though not significantly ($P = 0.2$, Wald test, Additional file 8: Fig. S6M). Consistent with the tendencies in HepG2 cells, we observed that *TP53* levels are significantly downregulated upon circCDYL depletion in both UMUC3 and J82 cells ($P < 8e-04$, Wald test, Fig. 6h, Additional file 8: Fig. S6N).

Finally, we analyzed the expression and clinical correlations of circCDYL in an additional, local cohort of patients with NMIBC ($n = 56$). Corresponding with our previous findings, high expression levels of circCDYL are positively correlated with overall survival ($P = 0.011$, log-rank test, Fig. 6i) and recurrence-free survival ($P < 0.014$, log-rank test, Additional file 8: Fig. S6O). Expression levels of the *CDYL* host gene are not associated with clinical outcomes ($P > 0.4$, log-rank test, Fig. 6i, Additional file 8: Fig. S6O). Additionally, clinical correlations were observed for other circRNAs highly covered with RBP binding sites, e.g., circZKSCAN1 and circRBM33 ($P < 0.05$, log-rank test, Additional file 8: Fig. S6P) and for overall circRNA expression ($P = 0.026$, log-rank test, Additional file 8: Fig. S6Q), independent of gene expressions.

Our results show that circCDYL interacts with the oncogenic RBPs IGF2BP1 and IGF2BP2 in BC cells and

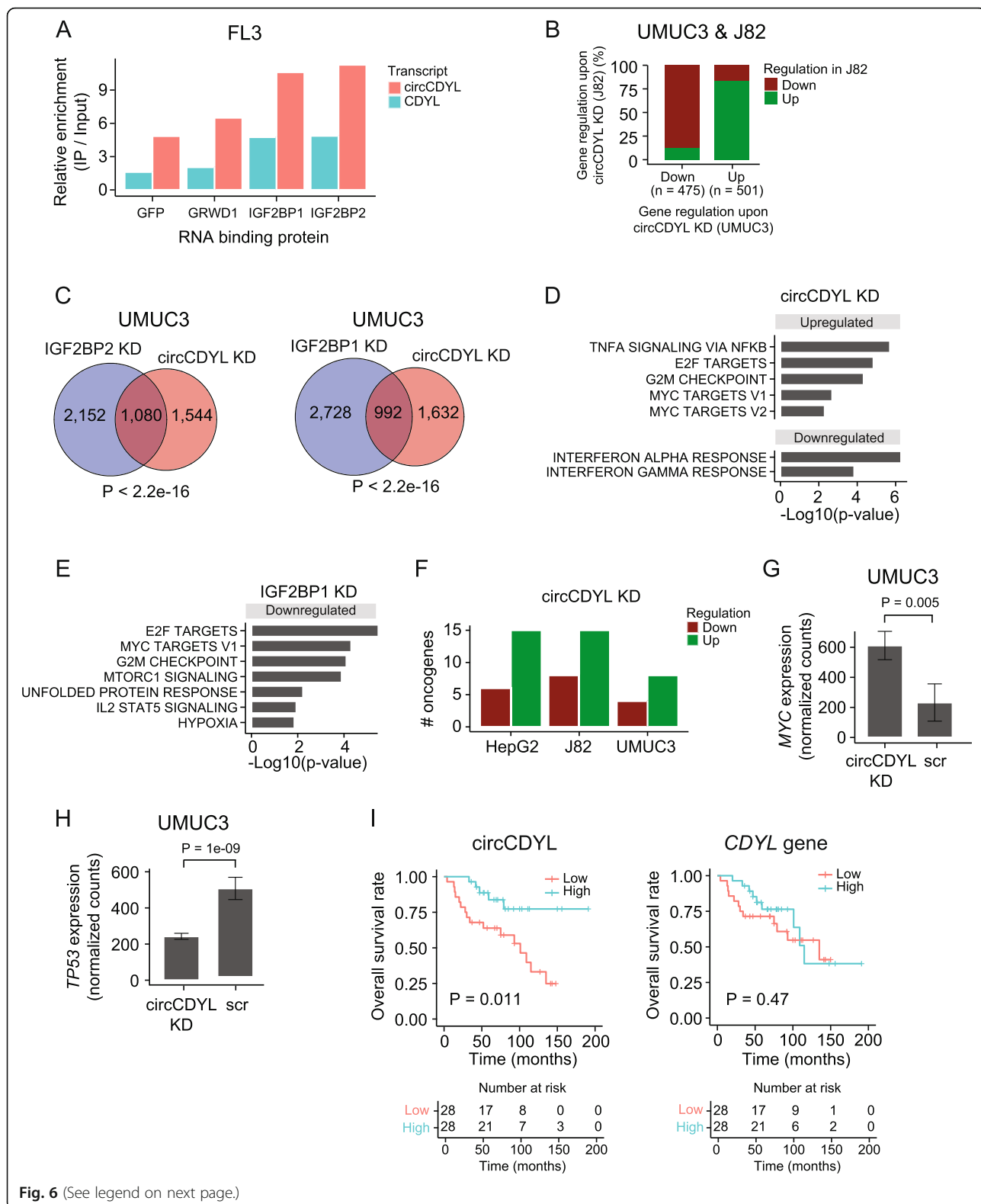


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 CircCDYL interacts with RBPs in bladder cancer and abundance is associated with overall survival. **a** Relative enrichment of circCDYL and CDYL host gene expression levels between immunoprecipitation (IP) and input in the bladder cancer cell line FL3. GFP was used as control. **b** Regulation of genes that are differentially expressed upon circCDYL KD in both J82 and UMUC3 cells. 86% of the genes show the same direction of perturbation in both cell lines. **c** Overlap of genes affected by circCDYL KD, and IGF2BP1 and IGF2BP2 KD, respectively, in UMUC3. *P* values obtained by Fisher's exact test. **d, e** Gene set enrichment analysis of 50 hallmarks of cancer upon circCDYL KD (**d**) and IGF2BP1 KD (**e**) in UMUC3 cells. **f** Distribution of oncogenes among up- and downregulated genes upon circCDYL KD in HepG2, J82, and UMUC3 cells. **g, h** Expression of *MYC* (**g**) and *TP53* (**h**) upon circCDYL KD in UMUC3. *P* values obtained by Wald test. **i** Kaplan-Meier overall survival plots for circCDYL and the host gene in patients with non-muscle invasive bladder cancer. Median expression used as cut-off; circCDYL = 0.125 RPM and CDYL = 21.4 FPKM. *P* values obtained by log-rank test

that there is a large overlap of altered genes upon circCDYL and RBP depletions. Consistent with a suggested tumor suppressive role, circCDYL depletion activates proliferation processes and the expression of oncogenes. In line with this, we find that elevated expression of circCDYL and other circRNAs identified here to interact with RBPs are associated with good clinical outcomes.

Discussion

By comprehensive analyses of a large atlas of eCLIP RBP binding sites and circRNA expression in the ENCODE cell lines HepG2 and K562, we showed that KHSRP binding sites are enriched in introns flanking circRNAs and that KHSRP depletion affects circRNA expression. Additionally, we found that exons comprising circRNAs generally contain more RBP binding sites than non-circularizing exons and that some RBPs preferentially bind to circRNAs. Furthermore, we examined the potential of individual circRNAs to function as RBP sponges and showed experimentally that circRNAs interact with RBPs in a cell-type-specific manner. We specifically investigated the function of circCDYL, which is highly covered with RBP binding sites in both cell lines. We found that circCDYL interacts with GRWD1 in HepG2 cells and that circCDYL depletion has the opposite effect of knocking down GRWD1. Finally, we showed that circCDYL, which is positively correlated with good clinical outcomes in BC, interacts with IGF2BP1 and IGF2BP2 in BC cell lines and that circCDYL and RBP KD perturb hallmarks of cancer gene sets and, specifically, that circCDYL KD affects the expression of key tumor genes, e.g., *TP53* and *MYC*.

First, we evaluated the overall potential of circRNAs as a group to interact with RBPs. No previous studies have comprehensively cataloged circRNA-RBP interactions using experimental data. In a previous study, You et al. found that circRNAs are not more prone to bind RBPs than linear mRNAs based on computational nucleotide sequence prediction of RBP binding sites [30]. However, due to contextual RBP binding preferences and the unique structures of circRNAs, nucleotide sequences are likely not adequate to predict circRNA-RBP interactions. Here, we defined RBP binding sites transcriptome-wide using experimental eCLIP data and found that RBP

coverage generally increases with transcript abundance. To ensure that enrichment of RBP binding sites in circRNAs is not simply explained by transcript abundance, we compared the RBP coverage of circRNAs to comparable gene sets. Contrasting the findings by You et al., our analyses showed that circRNAs are highly enriched with RBP binding sites compared to linear exons in host genes and genes of the same expression. Even compared to linear exons in the most highly expressed genes, circRNAs were enriched with or covered by an equal amount of RBP binding sites. These observations indicate specific circRNA-RBP interactions independent of transcript abundance and that non- or inefficiently translated circRNAs are ideal binding platforms for RBPs, since they are not in direct competition with elongating ribosomes.

Since the eCLIP data set was generated for linear transcripts, we evaluated the ability of circRNAs to interact with RBPs across the BSJ by mapping unmapped eCLIP reads to a reference set of BSJ sequences. We identified potential backsplicing events covered by eCLIP reads fulfilling our cut-off ("Methods"), of which a few conform to circRNAs identified in circBase. Nevertheless, none of them correspond to the BSJ of a circRNA expressed in HepG2 or K562. The lack of eCLIP reads spanning circRNA BSJs could be explained by short eCLIP sequencing reads around 20–25 bp long. This is considerably shorter than reads from most total RNA-Seq data sets, which typically use paired-end libraries and obtain longer read lengths (≥ 100 bp). Hence, these are not ideal for robust mapping across the BSJs of circRNAs [67]. Apparent backsplicing sequences supported by eCLIP reads could be artifacts or stem from other mechanisms such as template switching by reverse transcriptase, tandem duplication, structural variation between individuals, and RNA trans-splicing [83]. They could potentially also be real circRNAs not detected by CIRI2.

RBP binding sites in circRNA loci could reflect RBPs binding to the circRNA, the corresponding linear transcript, or both. Additionally, many factors could influence RBP binding capacity, like expression of cofactors and secondary structures of circRNAs. To confirm the validity of predicting circRNA-RBP interactions from

eCLIP data, we experimentally validated a set of circRNA-RBP interactions by RIP. All predicted circRNA-RBP interactions were validated except for one. Negative results would be expected if the RBP only interacts with the linear transcript and not the circRNA or if the circRNA-RBP interaction only occurs under certain circumstances.

By analyzing the RBP coverage of abundant circRNAs, we found that circCDYL is almost completely covered with RBP binding sites in both cell lines. In HepG2, GRWD1 binding sites cover 93% of the exonic circCDYL sequence, which is comparable to the efficient miRNA sponge, ciRS-7, which is ~90% covered by motifs comprising miR-7 binding sites [84]. RIP experiments showed that circCDYL interacts with GRWD1, IGF2BP1, and IGF2BP2 in both HepG2 and BC cells. From eCLIP data, we observed no IGF2BP1 binding sites in circCDYL in HepG2; however, we found that circCDYL was pulled down more efficiently by both of the IGF2BPs than by GRWD1 in BC cells. CircCDYL was also slightly enriched for GFP, which is likely due to the huge amount of GFP in the IP (Additional file 8: Fig. S6A). Importantly, even though there is extensively more GFP, we see a significantly larger enrichment for IGF2BP1 and IGF2BP2 (>10×) than for GFP (<5×), indicating a higher binding affinity to cellular proteins. Different binding affinities, competition for binding sites, dynamic interactions, and expression of RBPs and cofactors could influence interactions and explain the different observations in HepG2 and BC cells.

CircCDYL is clinically interesting, because it is deregulated in diseases [9, 44, 69], correlated to clinical outcomes [44], and highly expressed in patient plasma samples [69] and exosomes [9], indicating potential as a non-invasive biomarker. Sun et al. suggested that circCDYL overexpression inhibits MYC at the protein level, but found no effect on MYC mRNA levels [66], contrasting our findings here. Since circCDYL is almost completely covered with RBP binding sites, we speculate that circCDYL regulates important tumor genes and hallmarks of cancer through RBP interactions. Although circCDYL/mRNA ratios show that circCDYL is relatively abundant as required for sponging to take place, several factors affect final stoichiometries, including repeated translation of mRNAs and protein stability. Additionally, it is currently unknown how RBP multimerization or phase separation, which is a common characteristic of many RBPs, may influence the sponging capacity of circRNAs, leaving precisely measured stoichiometries between RNAs and interacting proteins somewhat ambiguous. Further experiments should address the regulatory potency of circCDYL on tumorigenesis and assess its clinical value as a new biomarker. Despite clinical correlations, circRNA perturbation in cancer might

not be causative but a consequence of underlying biological mechanisms.

The interplay between circRNAs and RBPs is highly complex as both factors have been shown to modulate the function and expression of the other. Zhang et al. implied that circRNA localization might be facilitated by RBP-mediated transportation in HepG2 [61] and Huang et al. provided evidence that the RBPs URH49 and UAP56 control nuclear export of short and long circRNAs, respectively, in HeLa cells [85]. Many of the RBPs we identified with enriched binding sites in circRNAs occupy both the nuclear and cytoplasmic fraction of the cell (Additional file 12: Table S7) and could be involved in circRNA nuclear export or otherwise in circRNA localization. For instance, the IGF2BPs, which we found on average have more target sites in circularizing exons, are regulators of essential mRNAs in tumorigenesis and are important players of mRNA stability and transportation to subcellular compartments [82]. Accordingly, IGF2BP3 was found to bind to several circRNAs [29]. Additionally, RBPs can regulate circRNA biogenesis by binding to intronic regions under certain circumstances. Here, we found that KHSRP depletion affects the expression of a subset of circRNAs with KHSRP binding sites in flanking intronic regions in K562 cells. Even though we observed an enrichment of KHSRP binding sites in flanking introns of circRNAs in HepG2, our experiments suggest that KHSRP is not essential to circRNA formation in HepG2, which could be explained by different binding activity in the two cell lines. Finally, RBPs binding to circRNAs could mark the circRNAs as “self” to prevent innate immune responses as suggested by Chen et al. [86].

CircRNAs can also regulate RBP localization and function. Binding of specific RBPs across multiple circRNAs and overall clinical correlations might reflect large-scale regulatory roles of circRNAs as a group. Additionally, circRNAs could act as dynamic scaffolds that bring together regulatory complexes. CircRNAs enriched with RBP binding sites, e.g., circCDYL, could bind and sequester a large number of RBPs to certain subcellular compartments and abrogate their normal function similarly to the role of the miR-7 sponge, ciRS-7 [24, 84]. Our results showed that circRNAs interact with RBPs in a highly cell-type-specific manner, consistent with findings for circFoxo3, which bind diverse RBPs in different biological settings [27, 28, 87]. CircRNAs might obtain different tertiary structures in various tissues and cellular conditions [88], partly explaining diverse functions and dynamic circRNA-RBP interactions as observed for circCDYL.

The list of potential functions of circRNA-RBP interactions is long, and their regulatory dependencies are complicated. How circRNAs mechanistically

control RBP localization, activity, and homeostasis awaits further investigation, but here we have provided strong comprehensive and global-scale evidence to support such roles.

Conclusions

In conclusion, we provided global-scale analyses of circRNA-RBP interactions in the main ENCODE cell lines. We showed that KHSRP affects circRNA biogenesis and that circRNAs interact with RBPs in a cell-type-specific manner. Specifically, we demonstrated that circCDYL is highly covered with RBP binding sites and that circCDYL and RBP depletion affect hallmarks of cancer gene sets, potentially playing roles in bladder cancer pathogenesis.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13073-020-00812-8>.

Additional file 1: File S1. Bash script containing the code to extract all HepG2 total RNA-Seq samples from ENCODE. The following samples were obtained for HepG2: total cell ($n = 1$), nucleus ($n = 1$), membrane ($n = 1$), cytosol ($n = 1$), and insoluble cytoplasmic fraction ($n = 1$).

Additional file 2: File S2. Bash script containing the code to extract all K562 total RNA-Seq samples from ENCODE. The following samples were obtained for K562: total cell ($n = 4$), nucleus ($n = 1$), nucleoplasm ($n = 1$), nucleolus ($n = 1$), chromatin ($n = 1$), membrane ($n = 1$), cytosol ($n = 1$), and insoluble cytoplasmic fraction ($n = 1$).

Additional file 3: Table S1. Genomic and statistical information of the highly expressed circRNAs in K562 and HepG2. First six columns provide circRNA_ID, genomic positions of circRNAs, and information of circRNAs found in circBase. BSJ_reads and linear_reads give the number of reads supporting the circRNA and its corresponding linear transcript based on the CIRI2 pipeline. BSJ_reads_CIRCexplorer holds the number of reads supporting the circRNA by the CIRCexplorer pipeline. Exonic_length and N_exons hold the exonic length of the circRNA and the number of exons comprising the circRNA. N_RBPs give the number of RBPs with binding sites in the exonic part of the circRNA. RBP_overlap_bp is the number of bp overlapped by RBP binding sites (merged) in the exonic part of the circRNA. RBP_coverage is the fraction of the circRNA (exonic part) covered by RBP binding sites ($\text{RBP_overlap_bp/exonic_length} * 100$). N_non_circ_exon_host_gene indicates the number of non-circularizing exons in the host gene. The mean RBP-coverage of non-circularizing exons in the host gene is provided in the column mean_RBP_coverage_non_circ_exons and was added a pseudo count of 1 to calculate the ratio between circRNA RBP-coverage and RBP-coverage of non-circ-exons in the host gene. This ratio is provided in Ratio_RBP_coverage_circRNA_non_circ_exons ($(\text{RBP_coverage} + 1) / (\text{mean_RBP_coverage_non_circ_exons} + 1)$).

Additional file 4: Table S2. Target sequences for siRNA-mediated knockdown of circCDYL, IGF2BP1, IGF2BP2, and GRWD1 in HepG2 and/or UMUC3 and J82.

Additional file 5: Table S3. Sample information for HepG2 and K562.

Additional file 6: Table S4. Count matrix of backsplice junction reads supporting circRNAs in HepG2 (txt).

Additional file 7: Table S5. Count matrix of backsplice junction reads supporting circRNAs in K562.

Additional file 8: Fig. S1-S6. Pdf file with all supplementary figures (Fig. S1-S6) with corresponding figure legends.

Additional file 9: File S3. Name of the RNA binding proteins with eCLIP data in HepG2.

Additional file 10: File S4. Name of the RNA binding proteins with eCLIP data in K562.

Additional file 11: Table S6. Genomic overlap of circRNAs and RBP binding sites in HepG2 and K562.

Additional file 12: Table S7. Subcellular localization of RBPs in HepG2.

Additional file 13: Table S8. GO enrichment analysis of RBPs that bind relatively more to BSJ circ-exons than non-circ-exons.

Additional file 14: Table S9. Primers for validating circRNA expression.

Additional file 15: Table S10. qPCR primers to detect circCDYL and CDYL host gene expression levels.

Acknowledgements

We thank Brenton Graveley for providing total RNA from KHSRP knockdown samples. We thank Thomas Birkballe Hansen for help with running CIRCexplorer.

Authors' contributions

TLHO and JSP conceived the project. TLHO performed the data analyses and generated all figures. SP and ZMC performed RNA immunoprecipitation in HepG2 and K562 cells, while CKD and ABK performed RNA immunoprecipitation in FL3 cells. CKD and ABK also conducted siRNA-mediated knockdown of circCDYL in HepG2, J82, and UMUC3 cells and of RBPs in J82 and UMUC3 cells. AS prepared total RNA-Seq libraries from KHSRP knockdown samples. SS generated the BSJ reference set and contributed initial analyses of eCLIP data. SV profiled circRNA and gene expression in RNA-Seq data sets. AMR performed GO enrichment analysis. MMN contributed insights into expression analyses. NF and LD generated and contributed clinical cohorts. SA and GWY contributed eCLIP data and supervised analyses. JSP supervised the project. TLHO and JSP wrote the manuscript with input from the other authors. All authors read and approved the final manuscript.

Funding

TLHO and JSP were supported by generous grants from the Lundbeck Foundation (#R191-2015-1515), the Danish Cancer Society (R124-A 7869-15-82), the Danish Council for Independent Research | Medical Sciences (FSS; DFF - 7016-00379), The Novo Nordic Foundation (NNF18OC0053222), the Harboe Foundation (19110), and Aage and Johanne Louis-Hansen Foundation (19-2B-503). ABK and CKD were supported by the Danish Cancer Society (R167-A11105). We thank the following funds for supporting TLHO's stays in GWY's lab and for supporting the experimental follow-up studies: the Harboe Foundation (18004), the Danish Cancer Society (R215-A12954), the Family Hede Nielsen's Foundation, the Lundbeck Foundation (R296-2018-2094), the Graduate School of Health at Aarhus University, and the NEYE Foundation. GWY was supported by grants from the NIH (HG004659 and HG009889). Clinical sample procurement was supported by the Danish Cancer Biobank, and LD was supported by The Danish Cancer Society (R204-A12270).

Availability of data and materials

All generated sequencing data have been deposited in NCBI's Gene Expression Omnibus. The circCDYL knockdown sequencing data in HepG2, UMUC3, and J82 cells and the GRWD1, IGF2BP1, and IGF2BP2 knockdown sequencing data in UMUC3 and J82 cells are accessible through accession number GSE146726 [70], <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146726>. Total RNA-Seq data upon KHSRP knockdown in K562 and HepG2 is accessible through accession number GSE145984 [64], <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145984>. Sensitive personal data (raw data from RNA-Seq and clinical information) cannot be shared publicly because of Danish legislation regarding sharing and processing of sensitive personal data. The publicly available data sets supporting the conclusions of this article are available in the ENCODE repository, <https://www.encodeproject.org/>, as described in the "Methods" sections, e.g., eCLIP data [32, 33], total RNA-Seq data from HepG2 and K562 [34], RBP knockdown in HepG2 [34], and subcellular localization of RBPs in HepG2 [43].

Ethics approval and consent to participate

Informed written consent to take part in future research projects was obtained from all patients, and the specific project was approved by the

National Committee on Health Research Ethics (#1708266). The research conforms to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

GWY is cofounder, member of the Board of Directors, on the SAB, equity holder, and paid consultant for Locanabio and Eclipse BioInnovations. GWY is a visiting professor at the National University of Singapore. GWY's interests have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies. The remaining authors declare no competing interests.

Author details

¹Department of Molecular Medicine (MOMA), Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark. ²Department of Clinical Medicine, Aarhus University, Palle Juul-Jensens Boulevard 82, 8200 Aarhus N, Denmark. ³Department of Cellular and Molecular Medicine, University of California San Diego, CA 92093 La Jolla, USA. ⁴Department of Molecular Biology and Genetics, Aarhus University, 8000 Aarhus C, Denmark. ⁵Department of Oncology, Aarhus University Hospital, 8200 Aarhus N, Denmark. ⁶Bioinformatics Research Center (BiRC), Aarhus University, 8000 Aarhus C, Denmark.

Received: 29 March 2020 Accepted: 12 November 2020

Published online: 07 December 2020

References

- Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung L-H, et al. Exon circularization requires canonical splice signals. *Cell Rep*. 2015;10:103–11. <https://doi.org/10.1016/j.celrep.2014.12.002>.
- Cocquerelle C, Mascres B, Hétiuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J*. 1993;7(1):155–60.
- Memczak S, Jens M, Elefantioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–8.
- Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*. 2013;19(2):141–57.
- Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *Plos One*. 2012;7(2):e30733.
- Zhang Z, Yang T, Xiao J. Circular RNAs: promising biomarkers for human diseases. *EBioMed*. 2018;34:267–74.
- Chen S, Huang V, Xu X, Livingstone J, Soares F, Jeon J, et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell*. 2019;176(4):831–43.e22.
- Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell*. 2019;176(4):869–81.e13.
- Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J, et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res*. 2015;25(8):981–4.
- Memczak S, Papavasileiou P, Peters O, Rajewsky N. Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *Plos One*. 2015;10(10):e0141214.
- Bahn JH, Zhang Q, Li F, –M. Chan T, Lin X, Kim Y, et al. The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva [Internet]. Vol. 61, *Clinical Chemistry*. 2015. p. 221–30. <https://doi.org/10.1373/clinchem.2014.230433>.
- Dominguez D, Freese P, Alexis MS, Su A, Hochman M, Palden T, et al. Sequence, structure, and context preferences of human RNA binding proteins. *Mol Cell*. 2018;70(5):854–67.e9.
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*. 2008;582:1977–86. <https://doi.org/10.1016/j.febslet.2008.03.004>.
- Zang J, Lu D, Xu A. The interaction of circRNAs and RNA binding proteins: an important part of circRNA maintenance and function. *J Neurosci Res*. 2018; <https://doi.org/10.1002/jnr.24356>.
- Conn SJ, Pillman KA, Toubia J, Conn VM, Salamanidis M, Phillips CA, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell*. 2015;160(6):1125–34.
- Erichelli L, Dini Modigliani S, Laneve P, Colantoni A, Legnini I, Caputo D, et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat Commun*. 2017;8:14741.
- Fei T, Chen Y, Xiao T, Li W, Cato L, Zhang P, et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc Natl Acad Sci*. 2017;201617467. <https://doi.org/10.1073/pnas.1617467114>.
- Khan MAF, Reckman YJ, Aufiero S, van den Hoogenhof MMG, van der Made I, Beqqali A, et al. RBM20 regulates circular RNA production from the Titin gene. *Circ Res*. 2016;119(9):996–1003.
- Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell*. 2014;56(1):55–66.
- Li X, Liu C-X, Xue W, Zhang Y, Jiang S, Yin Q-F, et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell*. 2017;67(2):214–27.e7.
- Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep*. 2015;10(2):170–7.
- Aktaş T, Avşar İlik İ, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, et al. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature*. 2017;544(7648):115–9.
- Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384–8.
- Piwecka M, Glažar P, Hernandez-Miranda LR, Memczak S, Wolf SA, Rybak-Wolf A, et al. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science*. 2017;357(6357) <https://doi.org/10.1126/science.aam8526>.
- Kleaveland B, Shi CY, Stefano J, Bartel DP. A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell*. 2018;174(2):350–62.e17.
- Abdelmohsen K, Panda AC, Munk R, Grammatikakis I, Dudekula DB, De S, et al. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol*. 2017;14(3):361–9.
- Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res*. 2016;44(6):2846–58.
- Du WW, Yang W, Chen Y, Wu Z-K, Foster FS, Yang Z, et al. Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur Heart J*. 2017;38(18):1402–12.
- Schneider T, Hung L-H, Schreiner S, Starke S, Eckhof H, Rossbach O, et al. CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs. *Sci Rep*. 2016;6:31313.
- You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci*. 2015;18(4):603–10.
- Hong S. RNA binding protein as an emerging therapeutic target for cancer prevention and treatment. *J Cancer Prev*. 2017;22(4):203–10.
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhardt C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016;13(6):508–14.
- Project Consortium ENCODE, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583(7818):699–710.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
- Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief Bioinform*. 2018;19(5):803–10.
- Zhang X-O, Wang H-B, Zhang Y, Lu X, Chen L-L, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159(1):134–47.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.

40. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9.
41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
42. Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*. 2013;20(12):1434–42.
43. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A large-scale binding and functional map of human RNA binding proteins. *bioRxiv*. 2018:179648. Available from: <https://www.biorxiv.org/content/10.1101/179648v2.abstract>. [cited 2019 Aug 21].
44. Okholm TLH, Nielsen MM, Hamilton MP, Christensen L-L, Vang S, Hedegaard J, et al. Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer. *Npj Genomic Med*. 2017;2 <https://doi.org/10.1038/s41525-017-0038-z>.
45. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46(D1):D754–61.
46. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009;10:161.
47. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1:417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
48. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019; 47(D1):D941–7.
49. Thomsen R, Sølvsten CAE, Linnet TE, Blechingberg J, Nielsen AL. Analysis of qPCR data by converting exponentially related Ct values into linearly related X0 values. *J Bioinforma Comput Biol*. 2010;8(5):885–900.
50. Moll P, Ante M, Seitz A, Reda T. QuantSeq 3' mRNA sequencing for RNA quantification. *Nat Methods*. 2014;11:i–iii. <https://doi.org/10.1038/nmeth.f.376>.
51. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019. Available from: <https://www.R-project.org/>.
52. RStudio Team. RStudio: integrated development for R. Boston: RStudio, Inc.; 2016. Available from: <http://www.rstudio.com/>.
53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
54. Therneau T. A package for survival analysis in S. Version 2.38. 2015; Available from: <https://CRAN.R-project.org/package=survival>.
55. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer; 2000.
56. Kassambara A, Kosinski M, Biecek P. survminer: drawing survival curves using “ggplot2”. R package version 0.4.6. 2019; Available from: <https://CRAN.R-project.org/package=survminer>.
57. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
58. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA*. 2014;20(11):1666–70.
59. Kristensen LS, Okholm TLH, Venø MT, Kjems J. Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. *RNA Biol*. 2018;15(2):280.
60. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136(4):701–18.
61. Zhang J, Zhang X, Li C, Yue L, Ding N, Riordan T, et al. Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells. *RNA Biol*. 2019;16(2):220–32.
62. Briata P, Bordo D, Puppo M, Gorlero F, Rossi M, Bizzozzero NP, et al. Diverse roles of the nucleic acid binding protein KHSRP in cell differentiation and disease. *Wiley Interdiscip Rev RNA*. 2016;7(2):227.
63. Min H, Turck CW, Nikolic JM, Black DL. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev*. 1997;11(8):1023–36.
64. Okholm TLH, Shankar A, Perellis M, Sathé S, Aigner S, Vang S, et al. Total RNA-Seq of KHSRP knockdown (KD) and control samples in HepG2 and K562. NCBI's Gene Expression Omnibus; Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145984>. Accessed 11 Nov 2020.
65. Singh G, Pratt G, Yeo GW, Moore MJ. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu Rev Biochem*. 2015;84:325–54.
66. Sun J, Zhang H, Tao D, Xie F, Liu F, Gu C, et al. CircCDYL inhibits the expression of C-MYC to suppress cell growth and migration in bladder cancer. *Artif Cells Nanomed Biotechnol*. 2019;47(1):1349–56.
67. Jakobi T, Dieterich C. Computational approaches for circular RNA analysis. *Wiley Interdiscip Rev*. 2019;10:e1528. <https://doi.org/10.1002/wrna.1528>.
68. Stagsted LV, Nielsen KM, Daugaard I, Hansen TB. Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles. *Life Sci Alliance*. 2019;2(3) <https://doi.org/10.26508/lsa.201900398>.
69. Voellenkle C, Perfetti A, Carrara M, Fuschi P, Renna LV, Longo M, et al. Dysregulation of circular RNAs in myotonic dystrophy type 1. *Int J Mol Sci*. 2019;20(8) <https://doi.org/10.3390/ijms20081938>.
70. Okholm TLH, Kamstrup AB, Steen H, Vang S, Damgaard CK, Pedersen JS. RNA-Seq of circCDYL knockdown (KD) samples in HepG2, J82, and U2OS cells and of GRWD1, IGF2BP1, and IGF2BP2 knockdown (KD) samples in J82 and U2OS cells. NCBI's Gene Expression Omnibus. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146726>. Accessed 11 Nov 2020.
71. Takafuji T, Kayama K, Sugimoto N, Fujita M. GRWD1, a new player among oncogenesis-related ribosomal/nucleolar proteins. *Cell Cycle*. 2017;16(15):1397.
72. Sugimoto N, Maehara K, Yoshida K, Yasukouchi S, Osano S, Watanabe S, et al. Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic Acids Res*. 2015;43(12):5898–911.
73. Kayama K, Watanabe S, Takafuji T, Tsuji T, Hironaka K, Matsumoto M, et al. GRWD1 negatively regulates p53 via the RPL11-MDM2 pathway and promotes tumorigenesis. *EMBO Rep*. 2017;18(1):123–37.
74. Huang X, Zhang H, Guo X, Zhu Z, Cai H, Kong X. Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) in cancer. *J Hematol Oncol*. 2018;11(1):88.
75. Stöhr N, Köhn M, Lederer M, Glass M, Reinke C, Singer RH, et al. IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes Dev*. 2012;26(2):176–89.
76. Rini J, Anbalagan M. IGF2BP1: a novel binding protein of p38 MAPK. *Mol Cell Biochem*. 2017;435(1–2):133–40.
77. Xu Y, Zheng Y, Liu H, Li T. Modulation of IGF2BP1 by long non-coding RNA HCG11 suppresses apoptosis of hepatocellular carcinoma cells via MAPK signaling transduction. *Int J Oncol*. 2017;51(3):791–800.
78. Kennedy BP, Soravia C, Moffat J, Xia L, Hiruki T, Collins S, et al. Overexpression of the nonpancreatic secretory group II PLA2 messenger RNA and protein in colorectal adenomas from familial adenomatous polyposis patients. *Cancer Res*. 1998;58(3):500–3.
79. Buhmeida A, Bendardaf R, Hilska M, Laine J, Collan Y, Laato M, et al. PLA2 (group IIA phospholipase A2) as a prognostic determinant in stage II colorectal carcinoma. *Ann Oncol*. 2009;20(7):1230–5.
80. He H-L, Lee Y-E, Shiue Y-L, Lee S-W, Lin L-C, Chen T-J, et al. PLA2G2A overexpression is associated with poor therapeutic response and inferior outcome in rectal cancer patients receiving neoadjuvant concurrent chemoradiotherapy. *Histopathology*. 2015;66(7):991–1002.
81. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, et al. Comprehensive transcriptional analysis of early-stage Urothelial carcinoma. *Cancer Cell*. 2016;30(1):27–42.
82. Cao J, Mu Q, Huang H. The roles of insulin-like growth factor 2 mRNA-binding protein 2 in cancer and cancer stem cells. *Stem Cells Int*. 2018;2018:4217259.
83. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;32(5):453–61.
84. Hansen TB, Kjems J, Damgaard CK. Circular RNA and miR-7 in Cancer. *Cancer Res*. 2013;73:5609–12. <https://doi.org/10.1158/0008-5472.can-13-1568>.
85. Huang C, Liang D, Tatomer DC, Wilusz JE. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes Dev*. 2018;32(9–10):639–44.
86. Chen YG, Kim MV, Chen X, Batista PJ, Aoyama S, Wilusz JE, et al. Sensing self and foreign circular RNAs by intron identity. *Mol Cell*. 2017;67(2):228–38.e5.
87. Du WW, Fang L, Yang W, Wu N, Awan FM, Yang Z, et al. Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death Differ*. 2017;24(2):357–70.
88. Du WW, Zhang C, Yang W, Yong T, Awan FM, Yang BB. Identifying and characterizing circRNA-protein interaction. *Theranostics*. 2017;7(17):4183–91.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.