SCIENTIFIC REP&RTS

OPEN

Expression profiling and *in situ* screening of circular RNAs in human tissues

Ammar Zaghlool[1], Adam Ameur[1], Chenglin Wu[2], Jakub Orzechowski Westholm[2], Adnan Niazi[1], Manimozhi Manivannan[3], Kelli Bramlett[3], Mats Nilsson[1,2] & Lars Feuk[1]

Circular RNAs (circRNAs) were recently discovered as a class of widely expressed noncoding RNA and have been implicated in regulation of gene expression. However, the function of the majority of circRNAs remains unknown. Studies of circRNAs have been hampered by a lack of essential approaches for detection, quantification and visualization. We therefore developed a target-enrichment sequencing method suitable for screening of circRNAs and their linear counterparts in large number of samples. We also applied padlock probes and *in situ* sequencing to visualize and determine circRNA localization in human brain tissue at subcellular levels. We measured circRNA abundance across different human samples and tissues. Our results highlight the potential of this RNA class to act as a specific diagnostic marker in blood and serum, by detection of circRNAs from genes exclusively expressed in the brain. The powerful and scalable tools we present will enable studies of circRNA function and facilitate screening of circRNA as diagnostic biomarkers.

Circular RNAs (circRNAs) were recently discovered as a novel class of noncoding RNA[1–3]. It is believed that most circRNAs are produced from precursor mRNA through back-splicing of exons, where a downstream (3′) splice donor site is spliced together with an upstream (5′) splice acceptor site. This process mainly occurs at annotated exon boundaries, forming covalently closed single-stranded circular transcripts[4]. Although circRNAs were originally reported more than two decades ago, only few examples were found, and circRNAs were mostly considered to be splicing noise[5,6]. However, with the recent advances in RNA sequencing technologies and newly developed bioinformatics approaches, several reports revealed that circRNA expression is widespread in human tissues and appears to be a general feature in all eukaryotes[1–3,7–11]. It is estimated that as much as 10% of expressed genes in examined cells and tissues produce circRNAs, often times in a cell or tissue and developmental-stage specific manner[1,3,7,11–14]. Even though circRNA are often expressed at low levels, there are at least 100 genes that express circRNAs as the predominant transcript isoform[10].

Although the function of the vast majority of circRNAs is still unknown, several studies demonstrated that circRNAs could function as competing endogenous RNAs or microRNA sponges by sequestering microRNA and influencing their ability to act on their target genes. The best-studied examples are *CDR1as* and *SRY* circR-NAs[8,15,16]. Moreover, emerging evidence suggests potential roles for circRNAs in important physiological functions, for example, in the nervous system[11,17–20]. In addition to their role as microRNA sponges, circRNAs are believed to regulate RNA polymerase II transcription, and compete with the canonical splicing machinery[14,19–21]. Recently, several reports provided *in vivo* and *in vitro* evidence in multiple metazoan species, including human, that a subset of circRNAs are translated into proteins by a cap-independent translation, thereby adding a new layer to the complexity of eukaryotic gene and protein expression[22–24].

CircRNAs are extraordinarily enriched in the mammalian brain, particularly in the synapses, and are upregulated during neuronal differentiation[11,25], and have been demonstrated to be involved in the function of the brain[25]. In addition, circRNAs have been shown to accumulate with age in Drosophila[9] and mouse brains[26]. Recent data suggest that circRNAs may play crucial roles in the development of various diseases, such as neurological disorders and various cancers mainly through differential expression analysis between healthy and disease

[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [2]National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. [3]Clinical Sequencing Division, Life Science Solutions Group, Thermo Fisher Scientific, San Francisco, CA, USA. Correspondence and requests for materials should be addressed to A.Z. (email: ammar.zaghlool@igp.uu.se) or L.F. (email: lars.feuk@igp.uu.se)
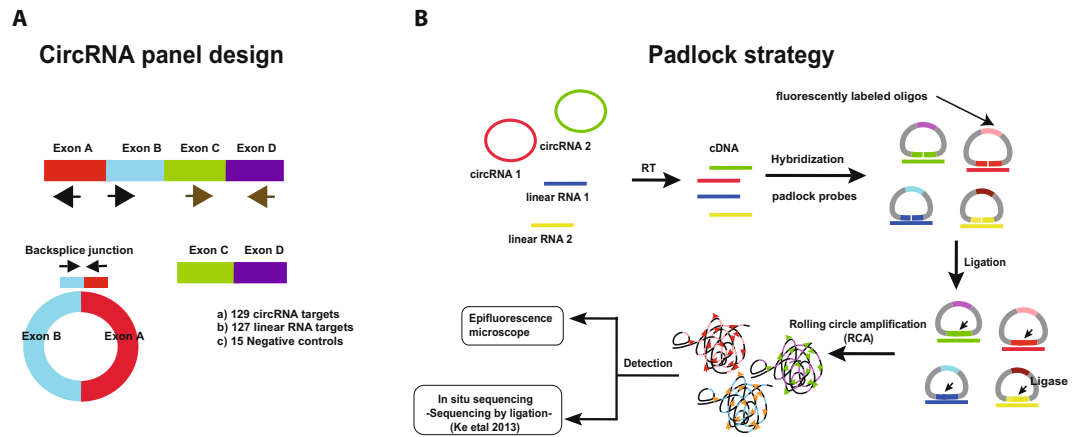
Figure 1. (**A**) A schematic illustration for the design of the primers used to enrich for circRNA and their corresponding mRNA in the AmpliSeq panel. Target specific primers for sequencing-library preparation included: 129 circRNAs (primers facing outwards), 127 linear RNAs (primers facing inwards) and 15 negative controls (primer facing outwards in regions where no circRNAs are annotated). CircRNA targets were selected based on human brain RNA-seq data. (**B**) A schematic illustration for the detection of circRNA using padlock probes and RCA. RNA is first reverse transcribed followed by hybridization of the padlock probes. Upon perfect hybridization to the target sequence, the 5′ and 3′ of the padlock probe come in proximity and circularized by target-dependent ligation. Circularized padlock probes can be amplified by RCA *in situ* and visualized using fluorescence microscopy or screened using *in situ* sequencing.

tissues[27–29]. CircRNAs are highly stable as compared to linear RNA transcripts due to their closed structure that prevents their cleavage and degradation by exonucleases, and can be detected in saliva, exosomes and blood[30–33]. They are therefore considered as a promising new class of biomarkers[31,33,34].

Although thousands of circRNAs have been identified using total RNA-seq data, the detection and characterization of circRNAs remain challenging. CircRNAs represent a small fraction of the total RNA pool and the majority of circRNAs are expressed at low levels. In addition, circRNAs are made up of the same sequence as their linear counterparts and can only be differentiated by the sequence reads spanning the back-splice junctions. Such back-splice junction reads are rare in RNA-seq data, making up less than 1/1000 reads in a typical sequencing library. A plethora of bioinformatics approaches have been developed to predict circRNAs and measure their expression from RNA-seq data[21,35,36], yet the efficiency of circRNA calls and reproducibility across different approaches remain relatively low. RNAse R treatment, a process that digests all linear RNAs but preserves circRNAs, followed by PCR or RNA-seq has been instrumental for validating circRNAs identified from total RNA-seq. However, this strategy hampers the ability to quantify the abundance of circRNAs in relation to their cognate linear RNA. Although some studies attempted to define the localization of circRNAs in cells and tissues using available *in situ* methods such as FISH[12,37], it is still challenging to define the localization of circRNAs at the subcellular level. This is mainly due to their generally low expression levels compared to mRNA, relatively short sequence and the sequence similarity with their cognate RNA. Therefore, developing tools to study subcellular localization of circRNAs would provide novel insights to their function. For instance, if circRNAs are enriched in a particular subcellular compartment different from their cognate RNA, the spatial discrepancy between the two molecules could indicate independent functions and regulatory pathways.

Although circRNAs are currently under intensive investigation, there is a lack of approaches to study their function, including high-throughput methods to quantify their abundance in relation to the expression of their cognate linear mRNA. Also there is a lack of methods to define the spatio-temporal expression of circRNAs across tissues, within tissues, and at the subcellular level. To address these methodological issues, we developed a robust target-enrichment sequencing approach to simultaneously measure the abundance of circRNA expression and their cognate linear transcripts across many samples. Furthermore, we adapted the padlock probes to visualize circRNA and their correlated linear transcript *in situ*, and to determine their subcellular localization in cells and tissues.

Results

In this study, we developed an efficient and cost effective sequencing strategy to accurately measure circRNA expression from various input materials in large numbers of samples. This strategy, here called the circRNA AmpliSeq panel, is based on target enrichment of circRNAs and their cognate linear isoforms using the Ion AmpliSeq™ target-enrichment sequencing panel (Fig. 1A). We also report improvements for the padlock method combined with *in situ* sequencing to define the spatio-temporal expression of circRNAs at the subcellular level (Fig. 1B).

CircRNA AmpliSeq panel design and validation.
In order to identify targets to include in the AmpliSeq panel enriching for selected circular and linear RNAs, we first performed standard RNA-seq on total RNA extracted from human fetal frontal cortex, fetal liver, fetal heart, and SHSY-5Y cells. To identify circRNAs expressed in these tissues, we then called sequence reads mapping to back-splice junctions from known exon
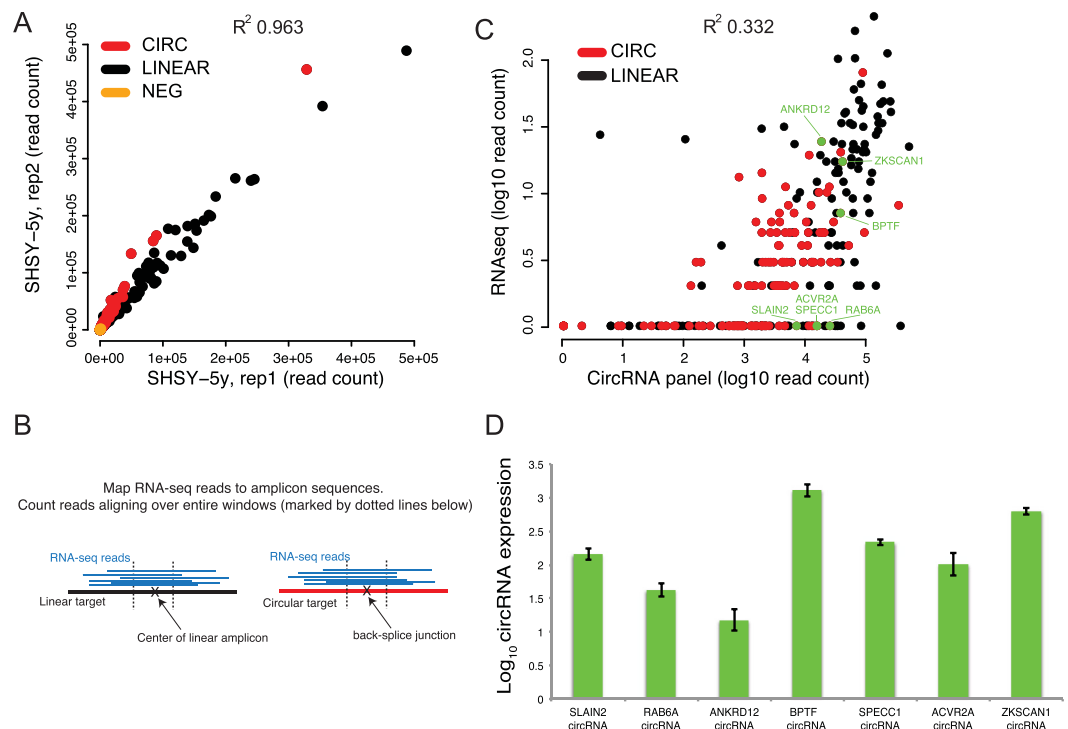
Figure 2. (**A**) Correlation of expression levels (raw read counts) obtained from the circRNA AmpliSeq panel, for two replicate RNA samples from the SHSY-5Y cell line. The black dots correspond to linear RNA targets, red dots correspond to circular RNA targets, and yellow dots correspond to negative control targets. (**B**) To make the RNA sequencing data comparable to results obtained from the circRNA panel, we aligned RNA-seq reads to the reference sequences obtained from the AmpliSeq panel design and required at least 50 bases to be mapped over the middle of each amplicon in order to call a match (indicated by the dotted lines). This implies that the RNA-seq read covers at least 25 bp at each side of the back-splice junction for circRNAs, and 50 consecutive bases for linear RNAs. (**C**) Quantification of circRNA expression levels from total RNA-seq (y-axis) and the AmpliSeq circRNA panel (x-axis) in SHSY-5Y RNA. The axis scales are logarithmic and the value 1 was added to each target, implying that non-expressed have a value of 0 in the plot. Each dot represents a target amplicon in the circRNA panel, either a circular RNA (red) or a linear RNA (black). (**D**) Validation of circRNA expression in SHSY-5Y cells. The expression level of circRNAs from *SLAIN2*, *RAB6A*, *SPECC1*, *ACVR2A*, *ZKSCAN1*, *ANKRD12 and BPTF* (represented by the green dots highlighted in 2B), and, a Neg control (negative control target from the circRNA panel, as in (**A**)) were measured using qrtPCR. All expression values were first normalized to the level of B-actin and then circRNA expression was calculated as a fold of the expression levels from the Neg control using the $\Delta\Delta Ct$ method. Expression levels represent mean values of the technical replicates and error bars are $\pm$ SD. All expression levels are presented as $\log_{10}$.

junctions using the circRNA_finder pipeline described in[9]. For initial testing, we focused on brain tissues, as circRNAs are highly abundant and diverse in brain compared to other tissues. From the circRNAs identified in the frontal cortex, we selected 129 circRNAs expressed at high levels as well as 127 linear RNA targets from the same genes for the design of the circRNA AmpliSeq panel. Additionally, we included 15 negative control targets, i.e. back-splice junctions with no support from the RNA-seq data and with no circRNA reported in literature (see Supplementary materials for the targets and their primer sequences). A schematic representation of the target enrichment approach for the circRNA and linear RNA is described in Fig. 1A. To test if our targeted approach successfully captures circRNAs, we performed a pilot sequencing experiment using the circRNA panel on total RNA extracted from human adult frontal cortex tissues. We also used the panel to assay technical replicates of total RNA extracted from the human neuroblastoma SHSY-5Y cell line.

The sequencing generated approximately 8 million reads per sample, with 97% of the reads mapping to the panel targets. Based on sequencing read counts for all targets in the panel, we found high correlation in circRNA and linear RNA expression between technical replicates ($R^2 = 0.963$). We also found the circRNA to linear RNA ratio to be highly correlated between these replicates, highlighting the reproducibility of the circRNA AmpliSeq panel (Fig. 2A and Supplementary Figure 1). To evaluate the efficiency of the circRNA panel as compared with circRNA detection from total RNA-seq data, we compared the number of sequence reads for all individual targets in the circRNA panel to the number of reads mapping to the same targets in the standard total RNA-seq data from the same samples. For this analysis, we aligned the RNA-seq reads to the exact same target sequences as defined by the AmpliSeq panel and thereby obtained an RNA-seq count for each target. Then, we directly compared the resulting RNA-seq counts to the corresponding counts obtained from the circRNA panel (Fig. 2B). Our results
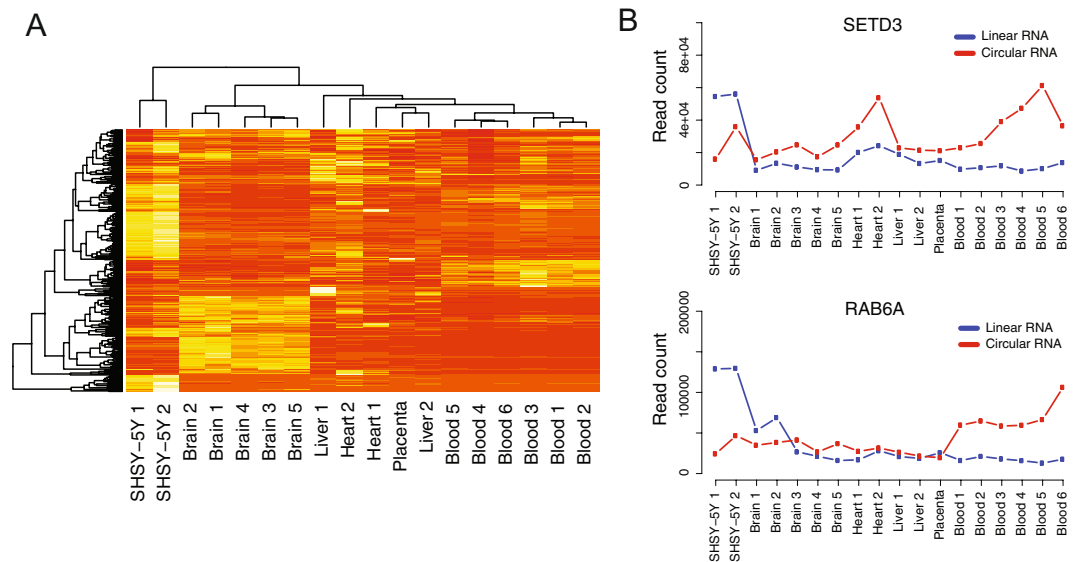
Figure 3. (**A**) Hierarchical clustering different RNA samples, based on raw read counts obtained from the AmpliSeq circRNA panel. Data from all targets (circular, linear and negative control) was used for the clustering analysis. (**B**) Read counts for the circular and linear isoforms of the genes *SETD3* and *RAB6A* in the SHSY-5Y cell line, blood and several tissue samples.

show that the target enrichment leads to a read count several orders of magnitude higher in the circRNA panel compared with RNA-seq (Fig. 2C and Supplementary Figure S2), but the results of the two approaches are still correlated with R^2 values of 0.332 and ranging between 0.114 and 0.495 for the rest of the samples.

The targeted enrichment and subsequent higher read coverage provides an increased sensitivity for detecting circRNAs compared to RNA-seq. Since many circRNAs are expressed at low levels, this increased sensitivity provides an advantage for detection and quantification of circRNA, which would otherwise require deep total RNA-sequencing or RNase R treatment for detection. This is relevant when screening for targets of interest in new tissues or samples. It also makes the panel an attractive choice for screening large cohorts after an initial circRNA discovery effort based on total RNA-seq. Our results also indicate that several circRNAs that were not detected or detected at very low levels using standard RNA-seq are readily captured using the panel, and these may explain the relatively low correlation values between the two methods. To validate these findings using an independent method, we used both the circRNAs panel and RNA-seq data from SHSY-5Y cells and selected four circRNA targets (*SLAIN2, RAB6A, SPECC1 and ACVR2A*) with no or low coverage in our total RNA-seq and in previously published data[11], but that were detected at high levels with the circRNA panel (Fig. 2C). In addition, we included three targets (*ZKSCAN1, ANKRD12 and BPTF*), which had coverage in both total RNA-seq and in the circRNA panel. We used qrtPCR to measure their actual expression levels in the original RNA sample, and found the qrtPCR results to be in agreement and correlate better with the circRNA panel data compared to RNA-seq (Fig. 2D and Supplementary Figure S3).

CircRNA expression in human tissues. To explore circRNA expression in a large range of human tissues, we applied the circRNA panel to RNA extracted from a set of human adult and fetal tissues including frontal lobe, liver, heart, placenta and blood (n = 12, Supplementary Materials). To obtain an overview of circRNA expression patterns in the different tissues, we performed hierarchical clustering of the expression counts obtained from the circRNA panel for all the samples (Fig. 3A). In agreement with previously published data[8,13], we found that circRNAs are expressed in all tissues with distinguishable expression patterns among the different tissues. Samples cluster by tissue type, with the exception for liver where adult and fetal samples do not cluster together. We also used the circRNA panel to measure the ratio between circRNA and their cognate linear RNA and found that circRNA/linear RNA ratios vary between samples. For some genes and in some tissue types, circRNAs were expressed at higher levels than their corresponding linear RNA. In Fig. 3B, we show two genes, *SETD3* and *RAB6A*, where both the absolute expression of circRNA and the ratio between circRNAs and their linear isoforms differ between the tissues. The complete table containing expression counts for linear and circular RNAs in all samples is available as a Supplementary Data file.

CircRNA expression in FFPE and serum samples. We then tested if we are able to enrich for and detect circRNA from specimens with low RNA quality such as formalin-fixed paraffin embedded (FFPE) tissues or low RNA content such as in serum. Total RNA was first extracted from 6 FFPE tissues and 1 serum sample (Supplementary Data file) and then sequenced using the circRNA panel. In this experiment, we also performed RNA sequencing using the circRNA panel following RNAse R treatment from one of the adult frontal cortex (Brain 5) as a control for the panel specificity. The relative circRNA expression (circRNA reads/(LinOut reads + circRNA reads)) in each sample is shown in Fig. 4A. As expected, the RNAse R treatment resulted in the loss of the sequencing signal over the linear RNA targets without affecting the sequence coverage over the
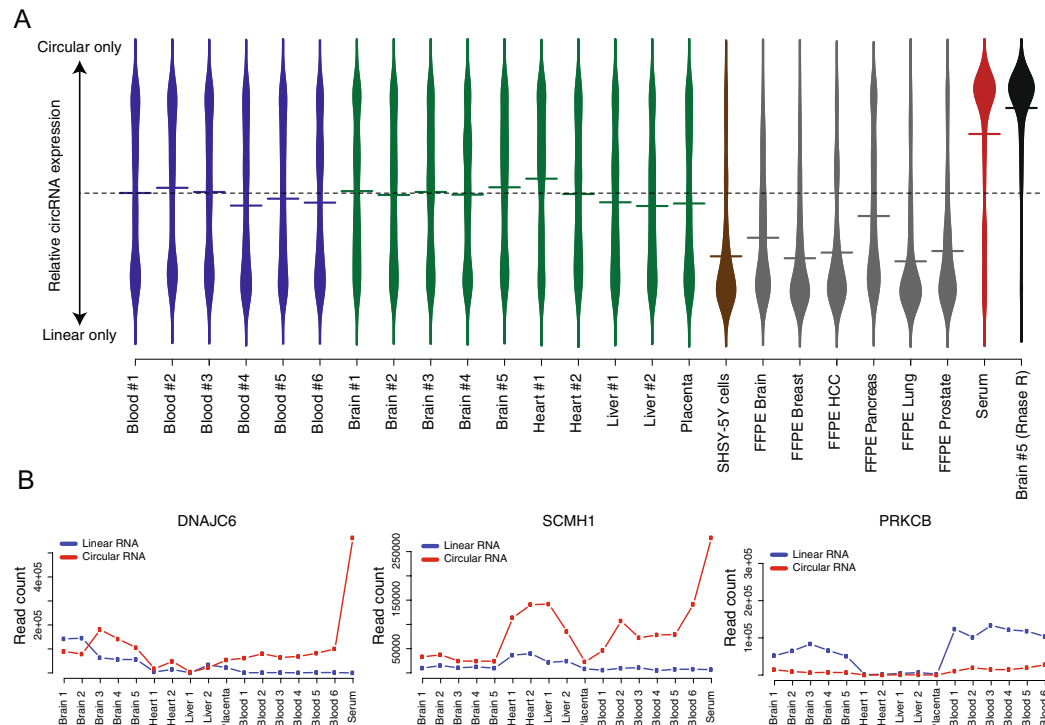
A

B

Figure 4. (**A**) Violin plot showing the relative circRNA expression for all genes in the AmpliSeq panel (i.e. ratio circRNA/(circRNA + linearRNA). The average circRNA expression values are represented by horizontal lines. The dashed line represents the value where circRNA and linear RNA are equally expressed. The RNAse R treated sample was used as a control to evaluate the specificity of the circRNA panel (**B**) Read counts for the circular and linear isoforms of the genes *DNAJC6*, *SCMH1* and *PRKCB* in tissue samples, blood and serum. All three genes show a dramatic increase of relative circular RNA expression in serum, indicating that the linear isoforms of *DNAJC6*, *SCMH1* and *PRKCB* are virtually absent from that sample.

circRNA targets. This highlights the ability of the panel to enrich for circRNAs with high specificity. The results from FFPE samples showed substantial reduction of the number of circRNAs detected, as well as reduced expression relative to their linear counterparts, when compared with results from all other samples. We speculate that the reduction of the number of circRNAs detected in the FFPE samples is caused by the strong crosslinking and fragmentation caused by the tissue fixation process. More adjustment and optimization for the RNA extraction from FFPE could improve circRNA detection.

Strikingly, we find several circRNAs to be highly expressed in the serum sample. Three examples of such circRNAs are shown in Fig. 4B. Expression of *DNAJC6* is of particular interest, since only the circRNA isoform is expressed in blood and serum. *DNAJC6* encodes the neuronal protein Auxilin, and it is known have a role in clathrin-mediated endocytosis and to be implicated in the pathogenesis of Parkinson disease and intellectual disability[38,39]. Previous data from The Genotype-Tissue Expression (GTEx) project shows that *DNAJC6* is exclusively expressed in the brain[40,41]. The presence of circRNA in serum reflects the high stability of circRNA as compared to linear RNA transcripts. *DNAJC6* represent an example of circRNAs that could be detected in blood and serum. Including larger number of serum samples and thorough validation experiments could pave the way to unravel the possibility of using circRNA as biomarkers for a particular biological or pathological condition.

CircRNA detection using padlock probes and RCA. In order to validate our findings from the circRNA panel and to evaluate the possibility of visualizing circRNAs *in situ*, we adapted a protocol for padlock probes in combination with rolling circle amplification (RCA)[42]. Initially, we tested this method in human fixed cells. We performed multiplex detection of three circRNAs (*CDRas1*, *HIPK3* and *MAN1A2*) using padlock probes targeting sequences surrounding the back-splice junctions followed by RCA (Fig. 1B). These targets were selected due to their high expression as seen in our sequencing data, and all of them were successfully detected in SHSY-5Y cell line (Fig. 5A). Although, the detection efficiency of the padlock in the cell line is less sensitive than circRNA panel, the relative abundance of the RCA signal for the circRNA targets correlated well with the expression pattern obtained from the circRNA panel (Fig. 5B). To test the method's ability for dual detection of circRNAs and their corresponding linear RNA, we designed additional padlock probes for the *HIPK3* and *MAN1A2* linear RNA transcripts. Both the circRNA and linear RNA were simultaneously detected for both genes (Fig. 5C). In accordance with our sequencing data, the padlock experiment showed higher expression for circRNAs from both genes as compared to their corresponding linear isoforms. Raw files for the original microscope images are presented in Supplementary Figure S4.
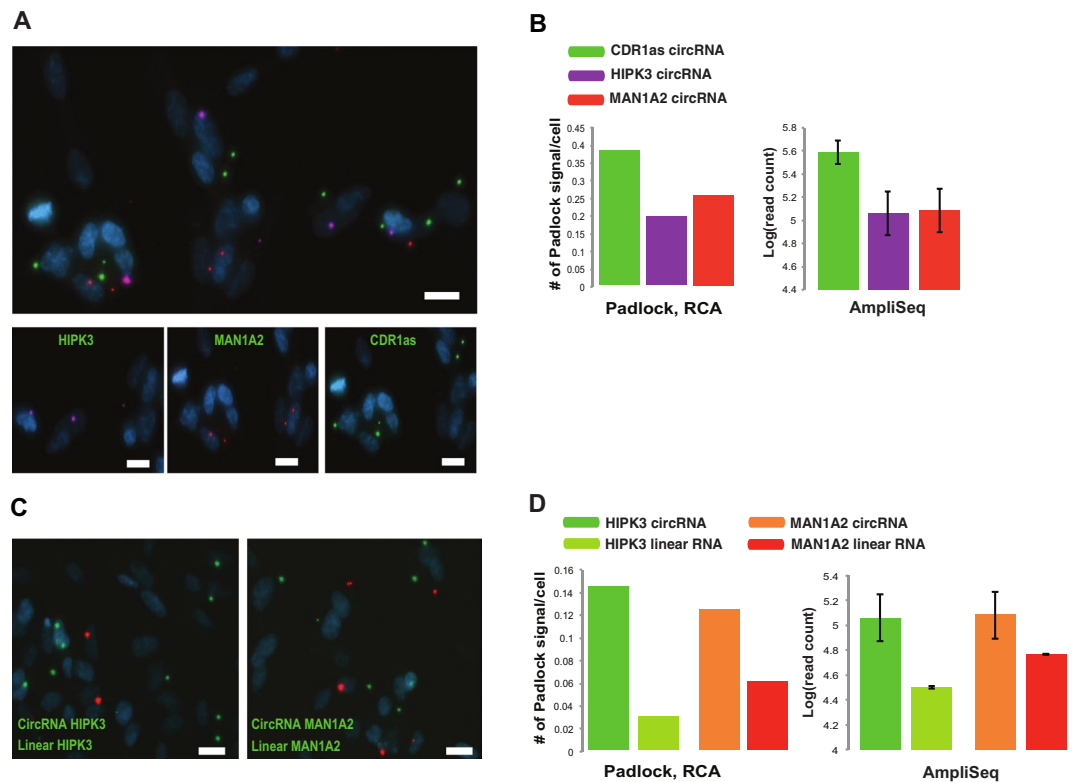
Figure 5. Detection of circRNA and linear RNA in SHSY-5Y cell lines using padlock probes and RCA (**A**) Multiplex detection of three circRNAs from *HIPK3*, *MAN1A2* and *CDR1as* using padlock probes complementary to the backs-plice junction (Supplementary Materials). RCA products (RCPs) from each of the circRNAs is represented with different dyes, *HIPK3;* purple, *MAN1A2*; red and, *CDR1as*; green. DAPI-stained nuclei are shown in blue. Top; is a merged signal from the three circRNAs. Bottom; shows the signal from each of the circRNAs. (**B**) A comparison between the padlock and RCA signal number, and circRNAs expression from circRNA panel. Left panel shows the average RCPs/cell for circRNAs from *HIPK3, MAN1A2 and, CDR1as*, while the right panel shows $\log_{10}$ read count identified using circRNA panel for the same circRNAs. Similar expression patterns were observed in both approaches highlighting the reproducibility of our data. (**C**) Dual detection of circRNA and linear RNAs form *MAN1A2* and *HIPK3* in SHSY-5Y cells. RCPs from circRNA and linear RNA are labeled in red and green fluorescence respectively. (**D**) A comparison between the padlock and RCA signal number, and circRNAs expression from circRNA panel. Left panel shows the average RCPs/cell for circRNA and linear RNA from *HIPK3* and *MAN1A2*, while the right panel shows $\log_{10}$ read count identified using circRNA panel for the same circRNAs.

Subcellular localization of circRNA in the brain. While it is believed that the majority of circRNAs are cytoplasmic, the detection of circRNAs in the nucleus is also reported[19]. Since current reports suggest that circR-NAs could display functions independent from the function of their host gene, we motivate that it is interesting to investigate if some circRNAs show different localization than their cognate linear RNA molecules. Therefore, we extended the use of the padlock approach and RCA to study the subcellular localization of circRNA *in vivo*.

To define the cellular localization of circRNAs, we applied an *in situ* sequencing approach, which in principle allows the detection of hundreds of targets in one experiment, via the application of padlock probes and RCA[43]. We targeted highly expressed circRNAs detected in our brain RNA sequencing data and investigated their subcel-lular localization in a fresh/frozen brain tissue section. The tissue section used in this experiment is from sample 5 (see Supplementary Materials), which was sequenced using the circRNA panel. In this experiment, we designed padlock probes targeting 20 bp sequences overlapping and surrounding the backsplice junction of 8 circRNA targets (see Supplementary Materials for padlock probes sequences) (Fig. 1A). We also targeted their correspond-ing mRNA by targeting exons downstream of the backsplice junction. As a control for the cellular localization experiment, we included padlocks targeting the *Malat1* lncRNA, which is predominantly expressed in the nucleus[44,45]. Following RCA of the specifically-bound padlock probes, we subjected the RCPs from each target to sequencing-by-ligation[43]. Each individual padlock probe harbored a unique barcode (4 nucleotides) to distin-guish different targets[43] (Fig. 6A). We used DAPI to stain cell nuclei, and used these images to segment the data to signals that are located within the DAPI stained nuclear areas ("nuclear"), and signals that localize outside these areas ("cytoplasmic"). Since we are not using confocal microscopy, some true cytoplasmic signals may end up in the areas defined as nuclear with this approach, while true nuclear signals are unlikely to end up in the defined cytoplasmic compartment. We first plotted the number of sequence reads from the *in situ* sequencing for each tar-get and then assigned them to their subcellular localization. As expected, *in situ* sequencing results indicated that
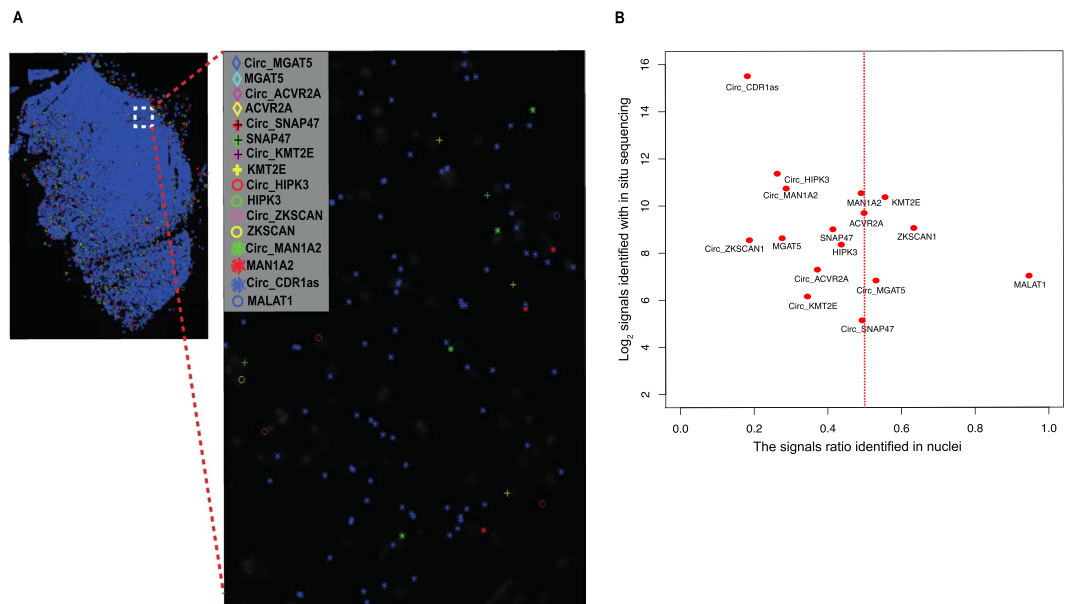
Figure 6. Multiplex detection of circRNAs and linear RNAs with padlock probes in combination with *in situ* sequencing in human brain tissue. (**A**) Padlock probes, RCA, and *in situ* sequencing were performed to detect the cellular localization of circRNAs and linear RNA in a tissue section from fresh/frozen human brain (sample 5). CircRNAs and their corresponding linear RNA from 7 genes were screened (*MGAT5, ACVR2A, SNAP47, KMT2E, HIPK3, ZKSCAN,* and *MAN1A2*). In addition, circRNA from *CDR1as* and linear RNA from *MALAT1* (known to be exclusively expressed in the nucleus, and used as a control for the localization experiment) were also included. Padlock probe sequences for each of the targets are listed in Supplementary Materials. *In situ* sequencing was performed to identify the 4-bp specific sequence barcodes in RCA products (RCPs) of each of the targets as shown in[43]. All targets were plotted on the DAPI image of the brain tissue section after *in situ* sequencing. Localization of each of the detected targets is shown as a symbol, left panel. Each symbol represents a barcode sequence that corresponds to a specific target. The nuclei are shown in grey. Scale bar represents 100 μm. Zoom in region corresponds to the region in white dash-line square. (**B**) Abundance and subcellular localization of circRNA and linear RNAs. The ratios of signals in nuclei for all targets were plotted against their barcode counts in the brain sample. Similar to the circRNA panel data, *CDR1as* was the most abundant circRNAs in the *in situ* sequencing experiment. Also, in agreement with previous data, *MALAT1* was highly expressed in the nucleus in the *in situ* sequencing experiment.

the linear *MALAT1* transcript is predominantly expressed in the nucleus (Fig. 6B, and Supplementary Figure S5), while the circRNAs *CDR1as, HIPK3* and *MAN1A2* were enriched in the cytoplasm, as previously reported[10,46,47]. Of the eight circRNA targets tested, six were found to be enriched in the cytoplasm (Fig. 6B, and Supplementary Figure S5). The circRNA target *SNAP47* showed equal distribution between nuclear and cytosolic compartments, while *MGAT5* showed predominantly nuclear localization.

Discussion

Although there is an increasing interest in unraveling the function of circRNAs and understanding their role in human disease, there is a lack of efficient and reliable tools to study their basic characteristics such as their expression levels-especially in relation to their cognate linear mRNA, and their cellular localization. In this study, we developed two approaches to allow an improved understanding of the nature and function of circRNAs. In the first approach, we aimed to develop an efficient and cost effective tool to characterize the expression patterns of circRNAs in different tissues and cells. To achieve this, we designed a pilot target-enrichment sequencing panel using the Ion AmpliSeq™ technology to enrich for circRNAs known to be expressed in human brain and their corresponding mRNA transcripts. Using this panel, we produced an average of 8 million reads/sample which resulted in a coverage over circRNAs several orders of magnitude higher than RNA-seq with 40 million reads/sample. In the second approach, we adapted the padlock technology in combination with RCA and *in situ* sequencing to detect and identify the spatio-temporal expression of circRNAs at the subcellular level in cells and fresh/frozen tissue sections.

The ability to quantify the abundance of circRNAs in various tissues and to identify how their expression is correlated with their corresponding mRNA is crucial to understand the nature and the function of circRNAs. Using the circRNA panel, we successfully enrich for the circRNAs and their corresponding mRNAs from various input materials. We find the circRNA panel is more efficient at detecting circRNAs than total RNA-seq. Using regular coverage total RNA-seq data (up to 100 million reads) the quantification of circRNA together with their linear counterparts is limited to the most abundantly expressed circles. The target enrichment approach leads to significantly higher sequence coverage over circRNA and mRNA targets as compared to total RNA-seq. This enabled the detection of circRNAs with very low (or no) read coverage in RNA-seq data, leading to improved and

more sensitive estimation of circRNA expression, especially for circRNA expressed at low levels or in specimens that contain a low amount of RNA. Several of these targets were detected in RNAse R treated samples, indicating that lack of detection in total RNA was not due to inability to align the sequence reads (data not shown). The target enrichment allowed us to screen multiple samples (n = 19) using one P1 Ion Proton sequencing chip, producing approximately 80 million reads/P1 chip. On the other hand, the total RNA-seq for the samples in our screening was run with 2 samples/chip, producing approximately 40 million reads/sample, highlighting the cost effectiveness of circRNA panel over total RNA-seq. We do note that two linear targets in the panel yielded low or no signal (Fig. 2C), indicating that primers failed. In light of this, it is also important to note that the circRNA panel designed for this study is a proof-of-concept pilot panel, created only to show the feasibility of high throughput screening. It is possible to scale this up, with proper testing of primers for both circular and linear target and including spike-in targets with known concentration, to quantify thousands of circRNA and linear RNA targets in multiple samples using a single panel.

Using the circRNA panel to analyze circRNA and mRNA expression in various tissues we found, in accordance with previous reports, that circRNAs show distinct expression patterns between different tissues. Moreover, the circRNA to mRNA expression ratio exhibited different patterns among the different samples. Detection of circRNA in FFPE samples was particularly inefficient. Although we could still detect a small number of highly expressed circRNAs in FFPE samples (Supplementary Materials), there was a clear reduction of circRNA as compared to their linear counterparts. These results indicate that circRNAs are not present in extracted total RNA from FFPE, either due to degradation or the general challenges associated with RNA purification from FFPE samples. Hence, it is important to note that standard extraction of RNA from FFPE samples will lead to erroneous measures of circRNA in relation to linear RNA, irrespective of how the circRNA screening is performed.

We found several examples where the circRNA isoform is highly expressed, while the linear mRNA isoform is not expressed or expressed at very low levels. One example is *DNAJC6*, a neuronal protein where the mRNA, in our data and in previous studies, is expressed exclusively in the brain[40]. Here we show for the first time that its circRNA isoform is expressed at high levels in blood, and also detectable in serum. These results support a possible role of circRNAs as stable tissue-specific biomarkers accessible in serum. The findings also highlight the efficiency and feasibility of our circRNA panel to facilitate the analysis of a large number of such samples to further evaluate and test the hypothesis that circRNA could represent potential biomarkers for diagnosis and treatment.

Prior to this study, defining the cellular localization of circRNA and their cognate linear mRNA in the same experiment using common approaches, such as FISH, was hampered due to the sequence similarity between circRNA and their corresponding mRNA and because circRNAs are typically composed of short sequences. Similar to other non-coding RNAs, defining the subcellular localization of circRNA could provide valuable insights into their function. For instance, cytoplasmic circRNAs are suggested to regulate gene expression through interfering with miRNA pathways. While nuclear circRNAs, which mostly represent intron-containing circRNAs, are suggested to interact with the splicing machinery and promote transcription[19]. In this study, we demonstrate that the padlock approach followed by RCA can be used to visualize circRNA and their corresponding mRNA in cells and tissue sections. The abundance of circRNA obtained from the padlock experiment from the SHSY-5Y cell line correlated well with the results from the circRNA panel. However, it is clear that the method has limited sensitivity as only a few targets per cell are detected. We believe that further optimization is possible to increase the sensitivity for detecting circRNAs expressed at moderate and low levels. When combined with *in situ* sequencing (sequencing-by-ligation), this approach allowed multiplex detection of circRNA and their corresponding mRNA in human brain at the subcellular level. These results highlight the possibility of using our *in situ* sequencing strategy to define the subcellular localization of circRNAs and how they co-localize with their corresponding mRNA isoforms. To our knowledge, this strategy is the first of its kind to allow a multiplex detection and subcellular localization data of circRNAs and their corresponding mRNAs *in situ* at the subcellular level. In addition, it offers an opportunity to resolve the spatio-temporal expression of circRNA in complex tissues such as the brain.

Both the circRNA panel and the circRNA padlock visualization approaches presented here will pave the way for an improved hypothesis-driven targeted analysis of circRNA functions in cells and tissues. The enrichment of the circRNA panel allows a cost-effective approach to perform surveys of circRNA expression profiles in large sample sets. The circRNA panel setup offers the possibility to increase the number of targets and to perform custom design for the circRNAs of interest. On the other hand, the padlock strategy offers, for the first time, an experimental approach to perform multiplex detection of circRNA cellular localization *in vivo*. This approach can also be used to define the spatial context of circRNA expression profiles between different cell populations within the same tissue. In conclusion, our approaches represent powerful a toolbox for deep profiling and characterization of circRNA and for unlocking their roles as potential diagnostic biomarkers.

Material and Methods

Samples. Informed consent was obtained from subject or their guardians for the collection of blood and tissue samples. Sample collection, methods and experimental protocols were carried out in accordance with local guidelines and regulations. The use of the samples and experimental protocols were approved by the Uppsala Ethical Review Board (dnr 2010/236 for blood samples and 2012/082 for tissue samples).

Sample preparation for total RNA sequencing. Total RNA from the fetal frontal lobe, fetal liver and fetal heart was purchased from Capital Biosciences. Total RNA from fetal tissues and SHSY-5Y cells was extracted using RiboPure™ RNA purification kit (Ambion) according to the manufacturer instructions.

Sample preparation for sequencing using the circRNA AmpliSeq panel. Adult brain samples (sample 1–5, Supplementary Materials) were obtained from the Harvard Brain Tissue Resource. Total RNA was extracted from 5 adult frontal lobe samples. To obtain total RNA, tissue pieces were first fixed in OCT and two sections were used

for total RNA purification. Total RNA purification for these samples and 6 human blood samples was performed using RiboPure™ RNA purification kit (Ambion) according to the manufacturer instructions. Total RNA from fetal and adult liver, fetal and adult heart, and placenta were purchased from Clonetech. Total RNA from the FFPE tissues was extracted using High Pure FFPET RNA isolation kit (Roche) according to the manufacturer instructions. RNA from the serum was purified using Plasma/Serum Circulating and Exosomal RNA purification Kit (Norgen). RNase R treatment of total RNA was performed after ribosomal RNA depletion. Briefly, $2\,\mu g$ RNA was added to a mixture of 1x RNase R buffer (Epicentre®), 1 mM $MgCl_2$, and $1.0\,\mu l$ (20 U/µl) RNase R enzyme (Epicentre®). The mix was then incubated at 37 C° for 10 minutes. RNase R was then inactivated by adding 1:1 volume of H_2O. All samples are listed in Supplementary Materials.

Total RNA sequencing. RNA was treated with the Ribo-Zero Magnetic Gold Kit for human/mouse/Rat (Epicentre®) to remove ribosomal RNA, and purified using Agencourt RNAClean XP Kit (Beckman Coulter). The rRNA depleted RNA was then treated with RNaseIII according to Ion Total RNA-Seq protocol v2 and purified with Magnetic Bead Cleanup Module (Thermo Fisher). Sequencing libraries were prepared using the Ion Total RNA-Seq Kit for the AB Library Builder™ System and quantified using the Fragment analyzer (Advance Analytical). The quantified libraries were pooled followed by emulsion PCR on the Ion OneTouch 2 system, and sequenced on the Ion Proton System.

Discovery of circular RNAs in total RNA-seq data. To identify circular RNAs in total RNA data, a computational pipeline adapted from[9] was used. In short, reads were aligned using STAR[48]. Reads were aligned using the following parameters to identify chimeric transcripts: ––chimSegmentMin 20 ––chimScoreMin1 ––alignIntronMax 1000000 ––outFilterMismatchNmax 4 ––alignTranscriptsPerReadNmax 10000 ––outFilterMultimapNmax 2. The candidate chimeric junction reads were then filtered to only include cases where the read spanned a junction with the splice acceptor on the same chromosome and strand as the splice donor, at most 100,000 bp upstream. In the subsequent analysis only circular junctions matching GT-AG splice sites we included. All the scripts are available at https://github.com/orzechoj/circRNA_finder.git and further details are provided in[9].

Design of an AmpliSeq panel for circular RNA. Loci to be included in the circRNA Ampliseq panel were selected using the following criteria: First, circRNAs with high expression in total RNA-seq data from brain were chosen, so that each circRNA originated from a different mRNA transcript (i.e. a backs-plice site between annotated exons). We required the length of each circRNA (assuming introns are spliced out) was at least 300 nt. Controls measuring expression of linear RNA isoforms (LinOut) were placed on the same transcript 300 nt upstream from the circRNAs if possible, otherwise 300 nt downstream. As negative controls, possible back splice sites with no support from the RNA-seq data were sampled, from transcripts expressed in the brain. The only exception to this scheme was CDR1as, which was added manually to the panel and doesn't have a control for linear mRNA, as it does not originate from a known mRNA transcript. Design of the circRNA panel was based on the hg19 genome and transcript models from the UCSC genome browser.

Sequencing using the circRNA panel. RNA was reverse transcribed to cDNA, and the acquired cDNA was amplified using the custom made Ion AmpliSeq™ circRNA panel. The sequencing libraries were prepared using the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher). Sequencing libraries were purified using the Agencourt AMPure XP reagent (Beckman Coulter) and amplified. The amplicons were quantified using the Fragment Analyzer instrument (Advanced Analytical). Samples were then pooled, followed by emulsion PCR using the Ion Chef System, and sequenced on the Ion Proton System.

Quantification of circRNAs using the circRNA panel. Analysis data from the circRNA panel was performed using the Coverage Analysis plugin in the Torrent Suite Software. In each of the samples, this analysis counts the number of reads originating for the different targets in the panel (Circ, LinOut and Neg). To make the circRNA panel data comparable between samples, a normalization was performed by dividing each read count with the total number of reads generated for the sample.

qrtPCR validation. Starting with $1\,\mu g$ of total, cytoplasmic, nuclear RNA, cDNA was synthesized using the RevertAid First strand cDNA synthesis kit (Fermentas) with random hexamers according to the manufacturer's recommendations. $1\,\mu l$ of the resulting cDNA was used for qrtPCR to measure the relative circRNA/linear RNA ratio in each sample. The qrtPCR was performed with Stratagene Mx3000P in 96-well plates. The reactions were carried out with an initial denaturation at 95 °C for 10 min followed by 40 cycles of denaturation at 95 °C for 15 s, primer annealing at 60 °C for 30 s and extension at 72 °C for 30 s. The qrtPCR contained 12.5 ng single stranded cDNA, $0.4\,\mu M$ for each primer and $12.5\,\mu l$ Maxima SYBR Green/ROX qPCR Master Mix (Fermentas) in $25\,\mu l$ reactions. All samples were amplified in triplicate and the mean values were used to calculate the expression level of each target. The circRNA RNA expression level were determined using the $\Delta\Delta CT$ method. CT values were first normalized to the levels of B-actin. Then, expression values for circRNA expression after normalization were calculated as a ratio between the circRNA expression values and the expression values of the negative control. Raw data were analyzed using MxPro Mx3000P software (Stratagene).

Preparation of cells and tissue sections for circRNA visualization. SHSY-5Y cells were cultured in 1:1 mixture of Ham's F12 and DMEM (Gibco) medium (Gibco) with L-glutamine and phenol red, supplemented with 2 mM L-glutamine (Sigma), 10% FBS (Sigma) and $1 \times$ PEST (Sigma). All cell lines were incubated at 37 °C, 5% CO2. To prepare cell slides for the padlock experiment, cells with >80% confluence were treated with 0.25% (w/v) trypsin-EDTA (Sigma) and resuspended in culturing medium. Thereafter, cells were seeded on Superfrost Plus

slides (ThermoFisher) and placed in a 150 mm × 25 mm Petri dish (Corning), and culturing medium was added to a final volume of 25 ml. Cells were then incubated for 12 h at 37 °C, 5% CO2. Cell fixation was then performed in 3.7% (w/v) paraformaldehyde (Sigma) in DEPC-treated PBS for 20 min at RT. Following the fixation, cell slides were washed twice in DEPC-treated PBS and dehydrated in an ethanol series of 70%, 85% and 100% for 5 min each. Then cell slides were kept at −80 °C until use. Tissue sections (4 μm) from human fresh/frozen frontal lobe (Sample 5, Supplementary Materials) were prepared with microtome and mounted on a Superfrost Plus slides (ThermoFisher). Tissue samples were stored at −80 °C until fixation. Fixation was performed in 3.7% (w/v) para-formaldehyde (Sigma) in DEPC-treated PBS with 0.05% Tween-20 (Sigma) (DEPC-PBS-T) for 45 min at RT and two washes in DEPC-PBS-T. The tissue sections were then permeabilized in 2 mg/ml pepsin (Sigma) in 0.1 M HCl at 37 °C for 2 min. After two washings in DEPC-treated PBS and dehydration in an ethanol series, as in the cell line, tissue sections on slides were then covered with Secure-Seal hybridization chambers (Invitrogen).

In situ reverse transcription, barcode padlock probing and rolling circle amplification (RCA). Cells and tissue samples were first rinsed with DEPC-PBS-T. A reversed transcription mix, containing 5 μM of unmodified random decamers, 0.2 μg/μl BSA (NEB) 500 μM dNTPs (Fermentas), 20 U/μl of TranscriptMe reverse transcriptase (Gdansk), and 1 U/μl RiboLock RNase Inhibitor (Fermentas) in the TranscriptMe reaction buffer, was applied to the slides. The incubation was carried out for 3 h at 37 °C for the cell line and overnight for the tissue sections. Slides were washed twice with DEPC-PBS-T, and followed by a post-fixation step in 3.7% (w/v) paraformalde-hyde in DEPC-PBS for 10 min (or 30 min for tissues) at room temperature. After post-fixation, the samples were washed twice in DEPC-PBS-T. Following the reverse transcription, RNA degradation, hybridization and ligation of padlock probe on synthesized cDNA were performed. A mix contained 1 × Ampligase buffer (20 mM Tris-HCl, pH 8.3, 25 mM KCl, 10 mM MgCl$_2$, 0.5 mM NAD and 0.01% Triton X-100), 100 nM of each padlock probe, 50 μM dNTPs, 0.5 U/μl Ampligase, 0.4 U/μl RNase H (Fermentas), 50 mM KCl and 20% formamide was added into the sample reaction chambers on slides. The slide was incubated at 37 °C for 30 min, and 45 °C for 45 min, and then washed twice with 1 × DEPC-PBS-T. For RCA, the reaction chamber was incubated with a RCA mix (1 U/μl phi29 polymerase (Fermentas), 1 × phi29 polymerase buffer, 0.25 mM dNTPs, 0.2 μg/μl BSA and 5% glycerol in DEPC-H$_2$O) for 2.5 h washed twice in DEPC-PBS-T. For tissue slide, RCA was carried out overnight.

Sequencing-by-ligation. An ethanol wash series were firstly used to remove the mounting medium on tissue sections. As previous study[43] slides were washed with DEPC-PBS-T once and treated with UNG buffer (1 × UNG buffer (Fermentas), 0.2 μg/μl BSA, 0.02 U/μl UNG (Fermentas) for 30 min at 37 °C. Slides were washed twice with DEPC-PBS-T, and then washed three times with 65% formamide for 60 s each. In sequencing-by-ligation chemistry, anchor primers and interrogation probes were hybridized separately before the ligation. Briefly, a hybridization mix containing 500 nM of anchor primers in 2 × SSC and 20% formamide were applied to the sample and incubated at RT for 30 min, and followed by DEPC-PBS-T wash twice. A ligation mix containing each interrogation probe, 100 ng/ml of DAPI, 1 mM ATP (Fermentas), 1 × T4 ligase buffer (Fermentas), and 0.1 U/μl of T4 ligase (Fermentas), was added to the samples and incubated for 30 min at RT. Following 3x DEPC-PBS-T wash, the slides were mounted in SlowFade Gold Antifade Mountant (ThermoFisher).

Image acquisition and analysis. Images were acquired using an AxioplanII epifluorescence microscope (20 × objective). Exposure times for all the experiments are listed in Supplementary Materials. After imaging, the slides were prepared for the next three sequencing cycle by UNG treatment buffer as described above followed by repeating the hybridization, ligation and imaging processes. The image analysis was performed as described in[43], a stack of images at different focal depths were captured and merged to a single image, and followed by a automatically stitching in the Zeiss AxioVision software. The fully automated sequence decoding was performed similar as previous study[43]. In short, cell nuclei were separated based on shape descriptors. Definition of cell cytoplasm uses nucleus as a seed and depends on sufficient cytoplasmic auto-fluorescence. The image of the general stain was enhanced by a top-hat filter, and RCPs were separated by watershed segmentation. The images were aligned, and fluorescence intensity from each of the signals representing A, C, T and G were extracted. The optimal transformation between a merged image of all signals (A+C+T+G) and the general stain was determined based on intensity[49]. Analysis were performed with CellProfiler (2.1.1, 6c2d896)[50] calling ImageJ plugins from Fiji for image registration. All intensity information was decoded using a script written in Matlab. Briefly, RCP was assigned for the base with the highest intensity for each RCP in all hybridization steps. A quality score was extracted from each base, and the quality of a transcript was defined as the lowest quality of all the bases in the transcript. The quality score ranges from 0.25 (poor quality) to 1 (good quality). The frequency of each sequence was extracted after a typical quality threshold of 0.4–0.55 was set.

References

1. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157, https://doi.org/10.1261/rna.035667.112 (2013).
2. Ivanov, A. *et al.* Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* **10**, 170–177, https://doi.org/10.1016/j.celrep.2014.12.019 (2015).
3. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* **15**, 409, https://doi.org/10.1186/s13059-014-0409-z (2014).
4. Starke, S. *et al.* Exon circularization requires canonical splice signals. *Cell Rep* **10**, 103–111, https://doi.org/10.1016/j.celrep.2014.12.002 (2015).
5. Cocquerelle, C., Daubersies, P., Majerus, M. A., Kerckaert, J. P. & Bailleul, B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J* **11**, 1095–1098 (1992).
6. Nigro, J. M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
7. Wang, P. L. *et al.* Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* **9**, e90859, https://doi.org/10.1371/journal.pone.0090859 (2014).

8. Memczak, S. *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338, https://doi.org/10.1038/nature11928 (2013).

9. Westholm, J. O. *et al*. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* **9**, 1966–1980, https://doi.org/10.1016/j.celrep.2014.10.062 (2014).

10. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733, https://doi.org/10.1371/journal.pone.0030733 (2012).

11. Rybak-Wolf, A. *et al*. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Molecular cell* **58**, 870–885, https://doi.org/10.1016/j.molcel.2015.03.027 (2015).

12. You, X. *et al*. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature neuroscience* **18**, 603–610, https://doi.org/10.1038/nn.3975 (2015).

13. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-type specific features of circular RNA expression. *PLoS Genet* **9**, e1003777, https://doi.org/10.1371/journal.pgen.1003777 (2013).

14. Kelly, S., Greenman, C., Cook, P. R. & Papantonis, A. Exon Skipping Is Correlated with Exon Circularization. *J Mol Biol* **427**, 2414–2417, https://doi.org/10.1016/j.jmb.2015.02.018 (2015).

15. Hansen, T. B. *et al*. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388, https://doi.org/10.1038/nature11993 (2013).

16. Yi, Y. *et al*. RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic acids research* **45**, D115–D118, https://doi.org/10.1093/nar/gkw1052 (2017).

17. Chen, L. L. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol* **17**, 205–211, https://doi.org/10.1038/nrm.2015.32 (2016).

18. Zhang, Y. *et al*. Circular intronic long noncoding RNAs. *Mol Cell* **51**, 792–806, https://doi.org/10.1016/j.molcel.2013.08.017 (2013).

19. Li, Z. *et al*. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* **22**, 256–264, https://doi.org/10.1038/nsmb.2959 (2015).

20. Ashwal-Fluss, R. *et al*. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**, 55–66, https://doi.org/10.1016/j.molcel.2014.08.019 (2014).

21. Zhang, X. O. *et al*. Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147, https://doi.org/10.1016/j.cell.2014.09.001 (2014).

22. Pamudurti, N. R. *et al*. Translation of CircRNAs. *Mol Cell*, https://doi.org/10.1016/j.molcel.2017.02.021 (2017).

23. Legnini, I. *et al*. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol Cell*, https://doi.org/10.1016/j.molcel.2017.02.017 (2017).

24. Yang, Y. *et al*. Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res*, https://doi.org/10.1038/cr.2017.31 (2017).

25. van Rossum, D., Verheijen, B. M. & Pasterkamp, R. J. Circular RNAs: Novel Regulators of Neuronal Development. *Front Mol Neurosci* **9**, 74, https://doi.org/10.3389/fnmol.2016.00074 (2016).

26. Gruner, H., Cortes-Lopez, M., Cooper, D. A., Bauer, M. & Miura, P. CircRNA accumulation in the aging mouse brain. *Sci Rep* **6**, 38907, https://doi.org/10.1038/srep38907 (2016).

27. Lukiw, W. J. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front Genet* **4**, 307, https://doi.org/10.3389/fgene.2013.00307 (2013).

28. Bachmayr-Heyda, A. *et al*. Correlation of circular RNA abundance with proliferation–exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep* **5**, 8057, https://doi.org/10.1038/srep08057 (2015).

29. Hansen, T. B., Kjems, J. & Damgaard, C. K. Circular RNA and miR-7 in cancer. *Cancer Res* **73**, 5609–5612, https://doi.org/10.1158/0008-5472.CAN-13-1568 (2013).

30. Zhang, Y. G., Yang, H. L., Long, Y. & Li, W. L. Circular RNA in blood corpuscles combined with plasma protein factor for early prediction of pre-eclampsia. *BJOG: an international journal of obstetrics and gynaecology* **123**, 2113–2118, https://doi.org/10.1111/1471-0528.13897 (2016).

31. Li, Y. *et al*. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell research* **25**, 981–984, https://doi.org/10.1038/cr.2015.82 (2015).

32. Lin, X., Lo, H. C., Wong, D. T. & Xiao, X. Noncoding RNAs in human saliva as potential disease biomarkers. *Frontiers in genetics* **6**, 175, https://doi.org/10.3389/fgene.2015.00175 (2015).

33. Bahn, J. H. *et al*. The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clinical chemistry* **61**, 221–230, https://doi.org/10.1373/clinchem.2014.230433 (2015).

34. 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* **10**, e0141214, https://doi.org/10.1371/journal.pone.0141214 (2015).

35. Burd, C. E. *et al*. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* **6**, e1001233, https://doi.org/10.1371/journal.pgen.1001233 (2010).

36. Gao, Y., Wang, J. & Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol* **16**, 4, https://doi.org/10.1186/s13059-014-0571-3 (2015).

37. Veno, M. T. *et al*. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome biology* **16**, 245, https://doi.org/10.1186/s13059-015-0801-3 (2015).

38. Edvardson, S. *et al*. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PLoS One* **7**, e36458, https://doi.org/10.1371/journal.pone.0036458 (2012).

39. Stessman, H. A. *et al*. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet* **49**, 515–526, https://doi.org/10.1038/ng.3792 (2017).

40. Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660, https://doi.org/10.1126/science.1262110 (2015).

41. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585, https://doi.org/10.1038/ng.2653 (2013).

42. Larsson, C., Grundberg, I., Soderberg, O. & Nilsson, M. *In situ* detection and genotyping of individual mRNA molecules. *Nat Methods* **7**, 395–397, https://doi.org/10.1038/nmeth.1448 (2010).

43. Ke, R. *et al*. *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10**, 857–860, https://doi.org/10.1038/nmeth.2563 (2013).

44. Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res* **44**, 863–877, https://doi.org/10.1093/nar/gkv1206 (2016).

45. Zhang, T. *et al*. RNALocate: a resource for RNA subcellular localizations. *Nucleic acids research* **45**, D135–D138, https://doi.org/10.1093/nar/gkw728 (2017).

46. Zheng, Q. *et al*. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun* **7**, 11215, https://doi.org/10.1038/ncomms11215 (2016).

47. Chen, L., Huang, C., Wang, X. & Shan, G. Circular RNAs in Eukaryotic Cells. *Curr Genomics* **16**, 312–318, https://doi.org/10.2174/1389202916666150707161554 (2015).

48. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, https://doi.org/10.1093/bioinformatics/bts635 (2013).

49. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* **7**, 27–41, https://doi.org/10.1109/83.650848 (1998).
50. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, R100, https://doi.org/10.1186/gb-2006-7-10-r100 (2006).

Acknowledgements

Author Contributions

A.Z., A.A. and L.F. conceived and designed the study. A.Z. performed sample preparation, extraction and PCR validation. A.Z., C.W. and M.N. designed and performed *in situ* experiments. A.A., J.O.W. and A.N. performed bioinformatic analyses. M.M. and K.B. created the AmpliSeq panel. A.Z., A.A., C.W. and L.F. wrote the paper with help form all authors. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-35001-6.

Competing Interests: The authors Manimozhi Manivannan and Kelli Bramlett are employed at Clinical Sequencing Division, Life Science Solutions Group, Thermo Fisher Scientific, San Francisco, CA, USA. The remaining authors declare no potential conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.