





OPEN

DATA DESCRIPTOR

Extracellular circular RNA profiles in plasma and urine of healthy, male college athletes

Elizabeth Hutchins¹ , Rebecca Reiman¹, Joseph Winarta¹, Taylor Beecroft¹, Ryan Richholt¹, Matt De Both¹, Khalouk Shahbender¹, Elizabeth Carlson¹, Alex Janss¹, Ashley Siniard¹, Chris Balak¹, Ryan Bruhns¹, Timothy G. Whitsett¹, Roger McCoy², Matthew Anastasi², April Allen¹, Brian Churas¹, Matthew Huentelman¹ & Kendall Van Keuren-Jensen¹ 

Circular RNA (circRNA) are a recently discovered class of RNA characterized by a covalently-bonded back-splice junction. As circRNAs are inherently more stable than other RNA species, they may be detected extracellularly in peripheral biofluids and provide novel biomarkers. While circRNA have been identified previously in peripheral biofluids, there are few datasets for circRNA junctions from healthy controls. We collected 134 plasma and 114 urine samples from 54 healthy, male college athlete volunteers, and used RNASeq to determine circRNA content. The intersection of six bioinformatic tools identified 965 high-confidence, characteristic circRNA junctions in plasma and 72 in urine. Highly-expressed circRNA junctions were validated by qRT-PCR. Longitudinal samples were collected from a subset, demonstrating circRNA expression was stable over time. Lastly, the ratio of circular to linear transcripts was higher in plasma than urine. This study provides a valuable resource for characterization of circRNA in plasma and urine from healthy volunteers, one that can be developed and reassessed as researchers probe the circRNA contents of biofluids across physiological changes and disease states.

Background and Summary

The advent of next-generation sequencing has spurred the discovery of a growing list of RNA biotypes, many of which are detectable across species, detected in numerous biofluids, and have biological function. While many studies have focused on microRNAs (miRNA), several other small RNA species (e.g. piwi-interacting RNAs (piRNA), tRNA fragments, and Y RNA fragments) have been detected across a range of biofluids and are being developed as clinical biomarkers^{1–4}. In addition to these linear RNAs, the discovery and detection of circular RNAs (circRNA), those with a covalently closed loop structure, have gained attention.

CircRNAs were initially discovered by electron microscopy, in the 1970s, as viroid molecules⁵. Nearly two decades later, circRNA were identified for a handful of mammalian genes^{6–8}. Though initially thought to be rare splicing events, circRNAs have recently been identified as an abundant, endogenous RNA species in a number of organisms from Archaea to yeast, plants, worms, flies, fish, and mammals^{9–11}. Additionally, circRNAs are abundantly expressed in a number of human tissues and cell types, and circRNA expression changes during development, and as a response to extrinsic factors such as stress, immune response, and hormonal stimuli^{12–17}. These endogenous RNAs are characterized by their circular structures, which are formed by a back-splicing event that covalently links the 3' “tail” splice donor with the upstream 5' splice acceptor “head” of the transcript, forming a back-spliced, or “head-to-tail” junction. While circRNA function is still being elucidated, there are examples of circRNA inhibiting microRNA, regulating alternative splicing, and modulating the expression of parental genes^{18–22}.

In comparison to their linear counterparts, circRNA transcripts can be more abundant and have greater stability as they are resistant to linear decay mechanisms and do not contain 5'-3' polarity nor polyadenylated tails^{14,21,23,24}, suggesting feasibility as stable biomarkers. CircRNA stability and detection in biofluids, saliva²⁵, blood^{24,26–30}, and urine^{31–33}, comes, in part, from their being protected in extracellular vesicles^{28,34–36}. Changes in circRNA expression is altered in multiple diseases, including preeclampsia, glioblastoma and colorectal

¹Neurogenomics Division, TGen, 445 N. 5th St., Phoenix, AZ, 85004, USA. ²Arizona State University Sports Medicine, 323 E Veterans Way, Tempe, AZ, 85281, USA. ✉e-mail: kjensen@tgen.org

cancer^{30,37,38}. More recently, circRNAs in tumor tissues, as determined by next-generation sequencing, correlated with disease progression^{39,40}. Urine circRNAs correlated with kidney rejection post-transplant³¹, while differentially expressed circRNAs have been determined in plasma exosomes of lung cancer patients versus controls⁴¹. Most studies of circRNA have small sample sizes or are based on targeted microarray data, rather than discovery-based methods. This dataset includes more than 100 samples from 54 volunteers from two easily accessible biofluids (plasma and urine). In some cases, multiple samples were collected from the same participant longitudinally, allowing us to assess the reliability of circRNA detection in biofluids.

The stability and abundance of circRNAs led us to investigate detection in two easily accessed biofluids: plasma and urine. As the volunteers were part of a larger study elucidating concussion biomarkers in male, college athletes, the samples are derived from young (18–25), healthy, male volunteers as depicted in Table 1. The longitudinal sample collections of plasma and urine are depicted in Online-only Table 1, including the number of circRNAs identified in each biofluid and those circRNAs observed concurrently in the biofluids. We identified circRNA in plasma ($n = 134$) and urine ($n = 114$), using RNAseq data followed by one of six different bioinformatic tools (Fig. 1). The intersection of the 6 bioinformatic tools provides a catalog for circRNA in plasma (Fig. 2a) and urine (Fig. 2c).

As there are few datasets with circular RNAs cataloged in clinically-relevant biofluids, we expect this data to contribute to the characterization of circRNAs in young, healthy males. While this might be a direct comparator for concussions, or other diseases more prevalent in young men, we also expect this dataset to help begin to fill out a broader assessment of circRNAs present in healthy populations.

Methods

Sample collection and participants. Samples were collected from healthy, male volunteers, ages 18–25, with consent and approval from the Western Institutional Review Board (WIRB) study ID #1307009395. All participants provided written consent prior to enrollment. We obtained plasma ($n = 134$) and urine ($n = 114$) samples from 54 healthy male volunteers. In 71.4% of participants, both biofluid types were collected from the same individual. Blood samples were collected in EDTA tubes, and urine was collected in sterile cups. After collection, samples were placed in a cooler with ice packs and transported from Arizona State University to the Translational Genomics Research Institute, within 2–3 hours of collection. Blood samples were spun down at 1320 x G for 10 minutes at 4 °C, and 1 mL aliquots of plasma were collected in RNase/DNase free microcentrifuge tubes (VWR) and stored at –80 °C. Urine samples were spun at 1900 x G for 10 minutes at 4 °C and 15 mL aliquots were collected in 50 mL conical tubes for storage at 80 °C.

RNA isolation, library preparation, and sequencing. For plasma samples, total RNA was isolated from 1 mL plasma using the mirVana PARIS RNA and Native Protein Purification Kit (Thermo Fisher, Cat. No.: AM1556) as in Burgos *et al.*⁴², treated with the DNA-free DNA Removal Kit (Thermo Fisher, Cat. No.: AM1906), and purified and concentrated with RNA Clean & Concentrator – 5 columns (Zymo Research, Cat. No.: R1016) by following Appendix C in the kit's protocol. For urine samples, total RNA was isolated from 15 mL urine using Norgen's Urine Total RNA Purification Maxi Kit (Slurry Format) (Norgen, Cat. No.: 29600), treated with the RNase-Free DNase Set (Qiagen, Cat. No.: 79254), and concentrated with the speed vacuum. The isolated RNA was quantitated with Quant-iT Ribogreen RNA Assay (Thermo Fisher, Cat. No.: R11490). Samples were not ribo-depleted, double-stranded cDNA was synthesized from 10 ng total RNA with the SMARTer Universal Low Input RNA Kit for Sequencing (Clontech, Cat. No.: 634940) using thirteen PCR cycles. The double-stranded cDNA was quantitated with the Qubit dsDNA HS Assay Kit (Thermo Fisher, Cat. No.: Q32854). For each healthy control sample, Illumina-compatible libraries were synthesized from 2 ng double-stranded cDNA with Clontech's Low Input Library Prep Kit (Clontech, Cat. No.: 634947) using four mandatory PCR cycles plus ten additional cycles. Each library was measured for size via Agilent's High Sensitivity D1000 Screen Tape and reagents (Agilent, Cat. No.: 5067–5602 & 5067–5585) and measured for concentration via the KAPA SYBR FAST Universal qPCR Kit (Kapa Biosystems, Cat. No.: KK4824). Libraries were then combined into equimolar pools, and each pool was measured for size and concentration. Pools were clustered onto a paired-end flowcell (Illumina, Cat. No.: PE-401–3001) with a 20% v/v PhiX v3 spike-in (Illumina, Cat. No.: FC-110-3001) and sequenced on Illumina's HiSeq. 2500 with TruSeq v3 chemistry (Illumina, Cat. No.: FC-401-3002). The first and second reads were each 83 bases.

CircRNA prediction. Samples were demultiplexed and raw fastqs generated using CASAVA (v1.8.2, Illumina). Raw fastqs were trimmed using cutadapt (v1.9) with a quality score cutoff of 30 and a minimum length of 30 bp⁴³. For each sample, 6 different algorithms (Table 2) were used to predict circRNA: KNIFE v1.4⁴⁴, find_circ²¹, MapSplice2⁴⁵, CIRCexplorer⁴⁶, CIRI2⁴⁷, and DCC⁴⁸. Indices of the GRCh37/hg19 genome were created using bwa and STAR v2.4.0j using default parameters^{49,50}; bowtie and bowtie2 genome indices were downloaded with the KNIFE package^{51,52}. Reads were mapped to the genome with the recommended aligner and alignment parameters for each program: STAR v2.4.0j for DCC and CIRCexplorer, bowtie2 v2.2.1 for find_circ and KNIFE, bowtie v0.12.9 for MapSplice2, and bwa v0.7.13 for CIRI2. CircRNA prediction was then completed with the suggested parameters for each program, with the exception of incorporating a minimum 18nt overlap on either side of the junction. CircRNAs were kept for downstream analysis if they 1) had 2 or more junction counts and 2) were identified in at least 5 samples for each respective program.

Analysis of predicted circRNA. The version of CIRCexplorer used here does not support paired-end data; therefore, circRNA prediction was performed on each pair separately and then combined for analysis. For each program, BED files containing count expression data were created from the output data. CIRCexplorer, KNIFE, and find_circ output files all produce output files with 0-based coordinates while CIRI2, MapSplice, and DCC output files have 1-based coordinates; therefore, all coordinates were converted to a 0-based system

Participant	# plasma samples	# urine samples	Age	Racial or Ethnic Category	Protocol	dbGaP Participant ID
001	1	1	19	AA	RNAseq	2048063
002	1	1	20	AA	RNAseq	2048064
004	3	3	22	AA	RNAseq	2048065
005	5	1	21	AA	RNAseq	2048066
006	3	3	22	W	RNAseq	2048067
007	3	5	20	AA	RNAseq	2048068
008	2	0	21	H	RNAseq	2048069
010	3	2	21	AA	RNAseq	2048070
011	4	6	20	AA	RNAseq	2048071
013	2	5	20	HAW	RNAseq	2048072
014	3	4	22	W	RNAseq	2048073
015	1	1	20	HAW	RNAseq	2048074
019	3	1	23	AA	RNAseq	2048075
022	2	6	N/A	W	RNAseq	2048076
023	0	1	19	AA	RNAseq	2048077
024	7	6	21	AA	RNAseq	2048078
025	2	2	20	AA	RNAseq	2048079
029	2	2	21	W	RNAseq	2048080
030	3	2	22	AA	RNAseq	2048081
031	5	2	23	W	RNAseq	2048082
036	0	1	21	W	RNAseq	2048083
039	4	3	22	W	RNAseq	2048084
042	1	0	23	AA	RNAseq	2048085
044	1	2	22	W	RNAseq	2048086
045	5	5	20	W	RNAseq	2048087
046	7	4	22	AA	RNAseq	2048088
048	2	1	21	AA and H	RNAseq	2048089
049	5	3	22	AA	RNAseq	2048090
170	1	1	18	Asian	RNAseq	2048091
201	3	5	N/A	AA	RNAseq	2048092
202	0	3	22	AA	RNAseq	2048093
203	9	4	22	AA	RNAseq	2048094
204	1	0	N/A	AA	RNAseq	2048095
205	1	1	N/A	N/A	RNAseq	2048095
206	5	0	20	AA	RNAseq	2048097
207	2	1	20	HAW	RNAseq	2048098
208	1	0	19	AA	RNAseq	2048099
209	1	1	20	W and H	RNAseq	2048100
210	1	1	21	AA	RNAseq	2048101
211	1	0	N/A	N/A	RNAseq	2048102
212	1	0	20	W	RNAseq	2048103
213	1	3	22	AA	RNAseq	2048104
214	0	2	20	W	RNAseq	2048105
215	0	1	20	H	RNAseq	2048106
216	1	3	19	AA	RNAseq	2048107
218	1	1	20	AA	RNAseq	2048109
220	1	0	N/A	N/A	RNAseq	2048111
221	8	3	19	AA	RNAseq	2048112
222	1	0	N/A	N/A	RNAseq	2048113
223	1	3	18	AA	RNAseq	2048114
224	1	2	19	HAW	RNAseq	2048115
226	4	1	21	AA	RNAseq	2048117
227	7	2	21	W	RNAseq	2048118
228	1	3	19	AA	RNAseq	2048119

Table 1. Healthy Participant Characteristics. AA = African American, Asian = Asian or Asian American, H = Hispanic or Latino, HAW = Native Hawaiian, W = White, N/A = not available.

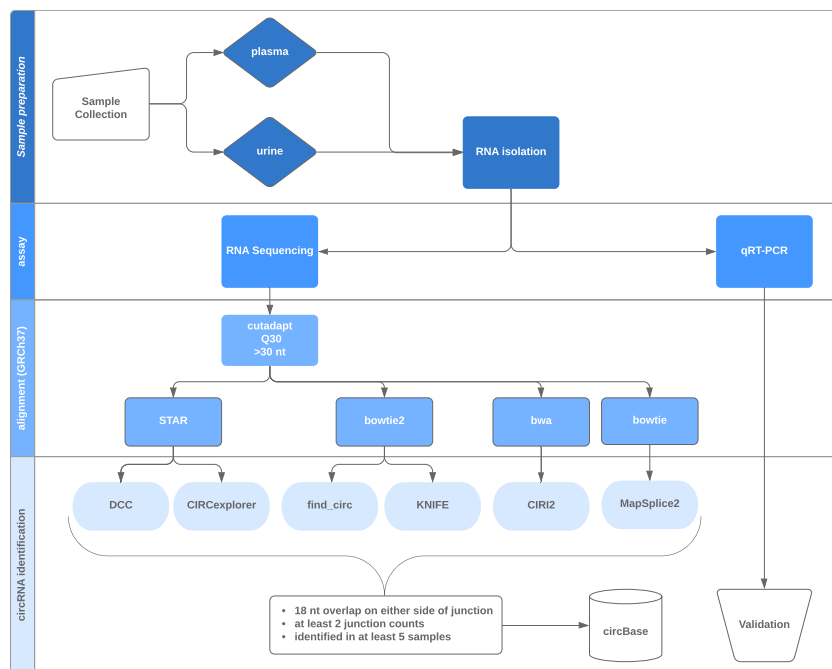


Fig. 1 Study Workflow.

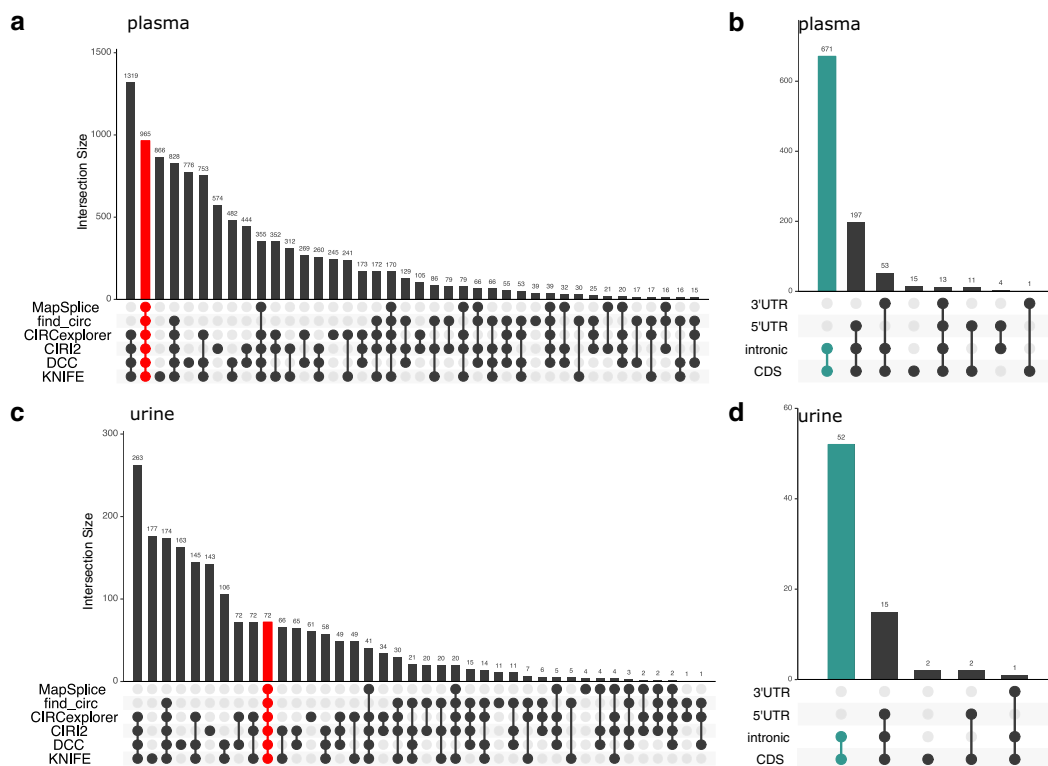


Fig. 2 CircRNAs were predicted from 134 plasma (a,b) and 114 urine (c,d) samples using 6 different bioinformatic tools. 965 circRNA were identified by all 6 tools in plasma (a; red bar), and 72 circRNA were identified by all 6 tools in urine (c; red bar). Genomic features located within predicted back-spliced junctions in plasma (b) and urine (d), respectively.

for comparison. BED12 GRCh37 RefSeq gene annotation files were obtained from UCSC (<http://genome.ucsc.edu/cgi-bin/hgTables>), and bedtools v2.26.0 was used to infer genes from reported backsplice junction genome locations⁵³. Data were analyzed using the R v3.3.2 statistical package (<https://cran.r-project.org>). UpSet plots were generated using the UpSetR v1.3.3 package⁵⁴.

Program	Aligner	Version	Paired-End Read Aware	Annotation Aware	Default Junction Overlap	Adjusted Junction Overlap	Reference
KNIFE	bowtie2	1.4	Yes	Yes	13 nt	18 nt	44
find_circ	bowtie2	1.0	No	No	18 nt	—	21
MapSplice	bowtie	2.1.8	Yes	Yes	10 nt	18 nt	45
CIRCexplorer	STAR	1.1.7	No*	Yes	15 nt**	18 nt	46
CIRI	bwa	2.0.1	Yes	Yes	19 nt***	—	47
DCC	STAR	0.3.2	Yes	Yes	15 nt**	18 nt	48

Table 2. CircRNA program characteristics. *The latest version of CIRCexplorer now supports paired-end reads. **CIRCexplorer and DCC use the STAR chimeric junctions output, so the junction overlap for these tools is set by the splice junction parameters during STAR alignment. ***The default minimum seed length (k) for bwa mem is 19 nucleotides.

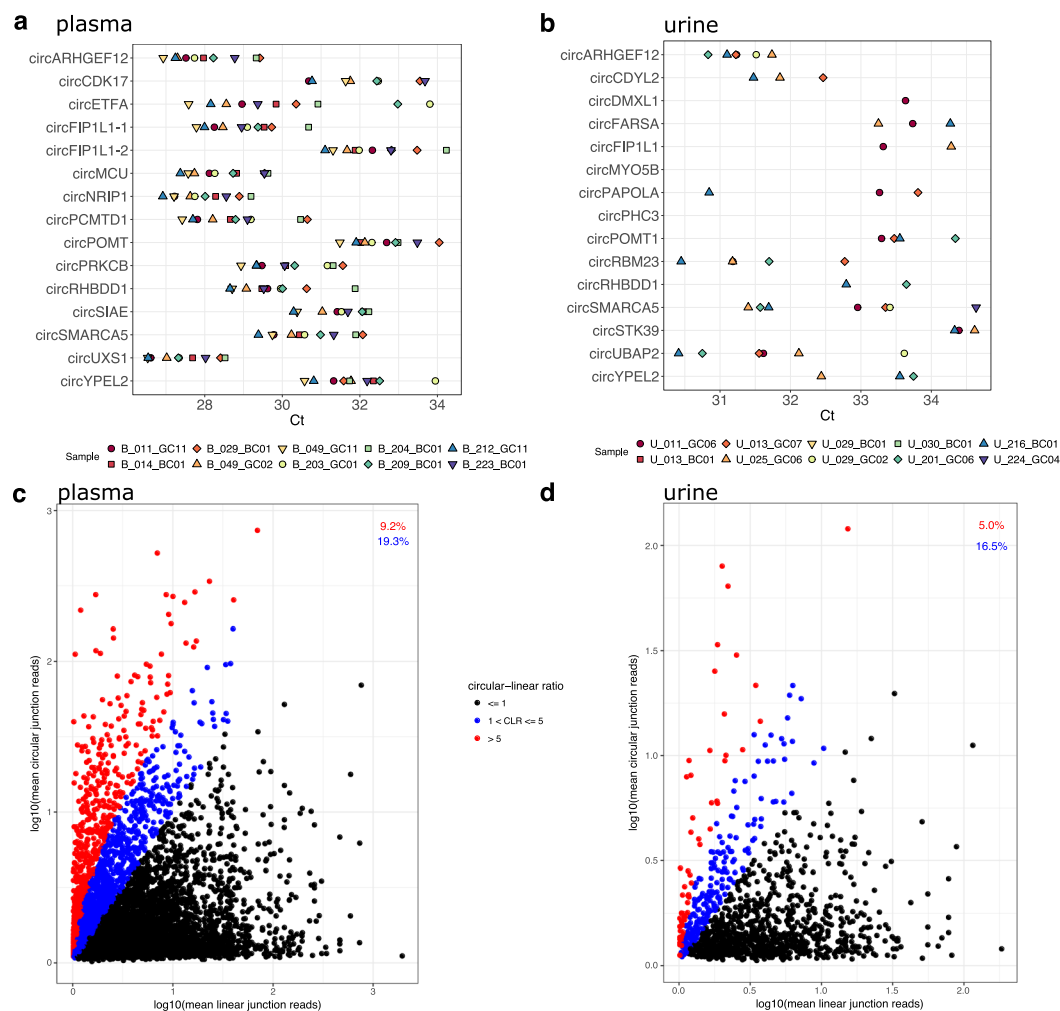


Fig. 3 (a,b) Highly-expressed, predicted back-spliced junctions were validated by qRT-PCR. qRT-PCR validation of the 15 most highly expressed circRNA found in plasma (a) and urine (b), respectively. Each circRNA was examined in 10 cDNA samples from the same source RNA as sequenced samples. (c,d) Circular-linear ratios are higher in plasma than urine. Linear splice junction expression plotted against circular splice junction expression in plasma (c) and urine (d). Points representing circRNA between 1-fold and 5-fold higher than their linear counterparts are blue; 5x or higher are red.

Quantification of circRNA expression. CircRNA count expression data was obtained from each respective bioinformatic program. Junction reads per million (JRPM) were calculated according to the total number of junction reads found in each sample as identified by STAR (both canonical and chimeric); therefore, $JRPM = (\text{circRNA count} / \text{junction reads}) * 1,000,000$. The circular-to-linear ratio (CLR) for each circRNA was calculated as described previously^{13,27}, by counting the linear spliced reads identified by STAR on the 5' and 3' flanks of

	Plasma		Urine	
	total circRNA	mean circRNA/sample	total circRNA	mean circRNA/sample
CIRCexplorer	6,297	909	1,142	119
CIRI2	6,789	1,075	1,205	131
DCC	7,159	1,009	1,287	132
find_circ	2,916	396	438	44
KNIFE	7,462	1,086	1,349	139
MapSplice	1,835	279	163	17

Table 3. CircRNA totals detected across six informatic tools in plasma and urine.

	Plasma (n = 134)	Urine (n = 114)	Both Plasma and Urine
Detected in at least 1 sample	965	72	61
Detected in 10% of samples	964	71	60
Detected in 20% of samples	881	61	51
Detected in 30% of samples	675	41	34
Detected in 40% of samples	538	28	24
Detected in 50% of samples	395	16	16
Detected in 60% of samples	273	14	11
Detected in 70% of samples	177	10	10
Detected in 80% of samples	68	4	2
Detected in 90% of samples	15	2	1
Detected in 100% of samples	0	0	0

Table 4. Number of circRNA detected in plasma and urine by all 6 bioinformatic tools.

Plasma						
	CIRCexplorer	CIRI	DCC	find_circ	KNIFE	MapSplice
CIRCexplorer	1	0.878	0.945	0.838	0.845	0.798
CIRI	0.878	1	0.882	0.836	0.841	0.908
DCC	0.945	0.882	1	0.843	0.887	0.79
find_circ	0.838	0.836	0.843	1	0.82	0.776
KNIFE	0.845	0.841	0.887	0.82	1	0.773
MapSplice	0.798	0.908	0.79	0.776	0.773	1
Urine						
	CIRCexplorer	CIRI	DCC	find_circ	KNIFE	MapSplice
CIRCexplorer	1	0.824	0.916	0.74	0.824	0.767
CIRI	0.824	1	0.869	0.738	0.843	0.817
DCC	0.916	0.869	1	0.793	0.889	0.733
find_circ	0.74	0.738	0.793	1	0.801	0.718
KNIFE	0.824	0.843	0.889	0.801	1	0.709
MapSplice	0.767	0.817	0.733	0.718	0.709	1

Table 5. Pearson's correlation of circRNA expression (JRPm) between informatic tools.

each circRNA junction, and dividing the back-spliced read count by the flank with the highest count; therefore, $CLR = \text{circRNA count} / \max(5' \text{ linear junction count}, 3' \text{ linear junction count})$. In order to avoid division by zero, if no linearly spliced reads were detected, a pseudo count of 1 was added to the denominator. The number of reads assigned to the transcriptome was calculated using featureCounts (subread v1.5.1) with the Ensembl75 gene annotation⁵⁵. Differential expression analysis was performed using DESeq. 2 v1.14.1⁵⁶, after filtering to select samples which had detected at least 300 circRNA/sample as well as exclusion of circRNA that were expressed in less than 50% of samples.

DNA isolation and qRT-PCR. After centrifugation of blood samples, DNA was isolated from the buffy coat using the DNeasy Kit (Qiagen, Cat. No.: 69504). Previously isolated RNA from samples matching those used for library prep were selected for cDNA synthesis. cDNA was synthesized with random hexamers using the SuperScript III First-Strand Synthesis System for RT-PCR following manufacturer's protocols (Invitrogen,

qRT-PCR	circRNA Detection	RNASeq	circRNA Detection
plasma	X out of 10 samples tested	plasma	X out of 10 samples tested
circARHGEF12	10	circARHGEF12	9
circFIP1L1-1	10	circFIP1L1-1	9
circMCU	10	circMCU	9
circRHBDD1	10	circRHBDD1	9
circSIAE	9	circSIAE	9
circCDK17	8	circCDK17	9
circFIP1L1-2	10	circFIP1L1-2	9
circNRIP1	10	circNRIP1	9
circPOMT1	10	circPOMT1	9
circSMARCA5	10	circSMARCA5	9
circETFA	10	circETFA	7
circPCMTD1	10	circPCMTD1	6
circPRKCB	10	circPRKCB	9
circUXS1	10	circUXS1	9
circYPEL2	10	circYPEL2	9
qRT-PCR	circRNA Detection	qRT-PCR	circRNA Detection
urine	X out of 10 samples tested	urine	X out of 10 samples tested
circPHC3	0	circPHC3	8
circPOMT1	4	circPOMT1	8
circRHBDD1	2	circRHBDD1	6
circSMARCA5	7	circSMARCA5	7
circYPEL2	3	circYPEL2	7
circCDYL2	3	circCDYL2	4
circFARSA	3	circFARSA	7
circPAPOLA	3	circPAPOLA	7
circRBM23	5	circRBM23	4
circUBAP2	6	circUBAP2	6
circARHGEF12	6	circARHGEF12	7
circDMXL1	1	circDMXL1	5
circFIP1L1	2	circFIP1L1	7
circMYO5B	0	circMYO5B	7
circSTK39	3	circSTK39	7

Table 6. circRNA detection in 10 samples by qRT-PCR and RNASeq.

Cat. No.: 18080-051) with three nanograms of total RNA as input, and stored at -20°C . Inward-facing (crossing the back-splice junction) custom primers were designed with Primer3 and LabReady primers (100 μM in IDTE pH 8.0) were ordered from Integrative DNA Technologies with Standard Desalting Purification^{57,58}. Real-time qRT-PCR was performed with SYBR Select Master Mix (Thermo Fisher, Cat. No.: 4472919) on the QuantStudio 7 (Applied Biosystems), with 0.2 μM of primer and 0.2 μL of cDNA template or 2 ng of gDNA template per 10 μL reaction. U6 was used as a positive control and no template controls (NTCs) were used as a negative control. All results are expressed as the mean of three independent reactions, with a standard deviation less than 0.5. The ReadqPCR v1.20.0 and NormqPCR v1.20.0 Bioconductor v3.4 packages were used for qRT-PCR data analysis⁵⁹.

Data Records

Raw FASTQ files for the RNAseq libraries were deposited into dbGap (accession # phs001258.v2.p1) (<https://identifiers.org/dbgap:phs001258.v2.p1>)⁶⁰. Data (circRNAs identified across all informatic tools and raw circRNA expression) are also provided in *figshare*: <https://doi.org/10.6084/m9.figshare.c.5420832>⁶¹.

Technical Validation

CircRNA set size and genomic alignment. The set size (all circRNA in any sample by one tool) ranges from 1,835 to 7,462 and 163 to 1,349 in plasma and urine, respectively (Table 3). 965 and 72 circRNA were detected across all six tools in plasma and urine, respectively (Fig. 2a,c, red bars; Table 4; full list in *figshare* File 1 and 2⁶¹). KNIFE predicted the most circRNA per sample in plasma and urine, while MapSplice predicted the fewest (Table 3). Table 5 displays the correlations between all of the tools, CIRCexplorer and DCC had the highest correlation. 85% (61 of the 72) of the circRNAs found in urine were also detected in plasma (Table 4). Figure 2b (plasma) and 2d (urine) display the number of detected circRNAs and the number that span introns, exons, and UTRs for both plasma and urine. The majority of circRNA identified in plasma and urine contain at least two exons and span an intron; 671 in plasma and 52 in urine; green bars (Fig. 2b, plasma and 2d, urine). A small number of circRNA are transcribed from a single exon (15 in plasma and 2 in urine).

Plasma				
circRNA	mean Ct	qRT-PCR Rank	mean JRPm	RNA-Seq Rank
circUXS1	27.4	1	51.21	15
circNRIP1	27.96	2	89.8	10
circARHGGEF12	28.04	3	110.13	5
circMCU	28.53	4	415.78	1
circPCMTD1	28.8	5	60.09	13
circFIP1L1-1	28.99	6	250.67	2
circRHBDD1	29.75	7	217.81	4
circETFA	30.05	8	66.35	11
circPRKCB	30.16	9	65.81	12
circSMARCA5	30.64	10	99.59	9
circSIAE	31.41	11	234.59	3
circYPEL2	31.88	12	51.44	14
circCDK17	32.12	13	102.04	7
circFIP1L1-2	32.36	14	100.46	8
circPOMT1	32.6	15	106.79	6
Urine				
circRNA	mean Ct	qRT-PCR Rank	mean JRPm	RNA-Seq Rank
circARHGGEF12	31.27	1	5.27	15
circRBM23	31.46	2	8.21	5
circUBAP2	31.68	3	6.49	10
circCDYL2	31.93	4	6.72	8
circPAPOLA	32.64	5	6.61	9
circSMARCA5	32.72	6	7.85	6
circRHBDD1	33.22	7	10.72	3
circYPEL2	33.25	8	9.3	4
circDMXL1	33.63	9	6.29	11
circPOMT1	33.67	10	28.46	2
circFARSA	33.75	11	7.5	7
circFIP1L1.2	33.8	12	5.61	14
circSTK39	34.45	13	6.24	12
circMYO5B	N/A	14	5.7	13
circPHC3	N/A	15	81.19	1

Table 7. qRT-PCR and RNA-Seq expression of the 15 most highly expressed genes in plasma and urine.

Highly expressed, back-spliced junctions were validated by qRT-PCR. In order to validate predicted back-spliced junctions by qRT-PCR, we designed inward-facing primers for the 15 most highly expressed circRNA in each biofluid and tested each primer pair in samples from 10 different individuals, using the same source RNA for cDNA synthesis that was used for RNAseq (Fig. 3a,b). Figure 3a shows that the 15 circRNAs are detected in most of the 10 plasma samples. The numbers of samples are described in Table 6, and compared with the RNASeq detection for those circRNAs in the same samples. 13 primer pairs were validated in urine. Detection in urine samples was sparse, with fewer samples positive for each circRNA than for plasma (Fig. 3b and Table 6). For the two back-spliced junctions detected in RNASeq data, but not validated by qRT-PCR in urine (circMYO5B and circPHC3), it is possible that the circRNA primers did not work, or there were qPCR inhibitors in the sample, or the circRNA was not present. Two of the samples did not have enough assigned reads via RNASeq to be included, so the total number of samples was 8. In order to rule out chimeric junctions that might be present in DNA or resemble artifacts introduced during library preparation, we also used genomic DNA (gDNA) from each individual as a negative control. All 15 primer pairs used in the plasma and urine samples were not detected in gDNA (data not shown). Table 7 describes the rank from highest to lowest expression for each of the circRNA validated by qRT-PCR, and compares it with the expression detected with sequencing. Their ranks do not correlate well between the two platforms.

Circular-to-linear RNA ratios. While the overall expression of most circRNAs is low compared to their linear counterparts, there are a number of circular RNA transcripts that have been described as more abundant than their linear host, cellularly as well as extracellularly^{23,27,62,63}. We examined the circular-to-linear ratio (CLR) of circRNA transcripts found in plasma and urine as described previously; by taking the ratio of the circular, back-spliced junction counts compared to the linear count of the nearest 5' or 3' splice junction^{13,24,27}. On average, 28.5% of circRNA transcripts in plasma and 21.5% of circRNA transcripts in urine have higher expression than their linear host gene (Fig. 3c, plasma and 3d, urine). Extracellular RNA is often fragmented and may have a 3'

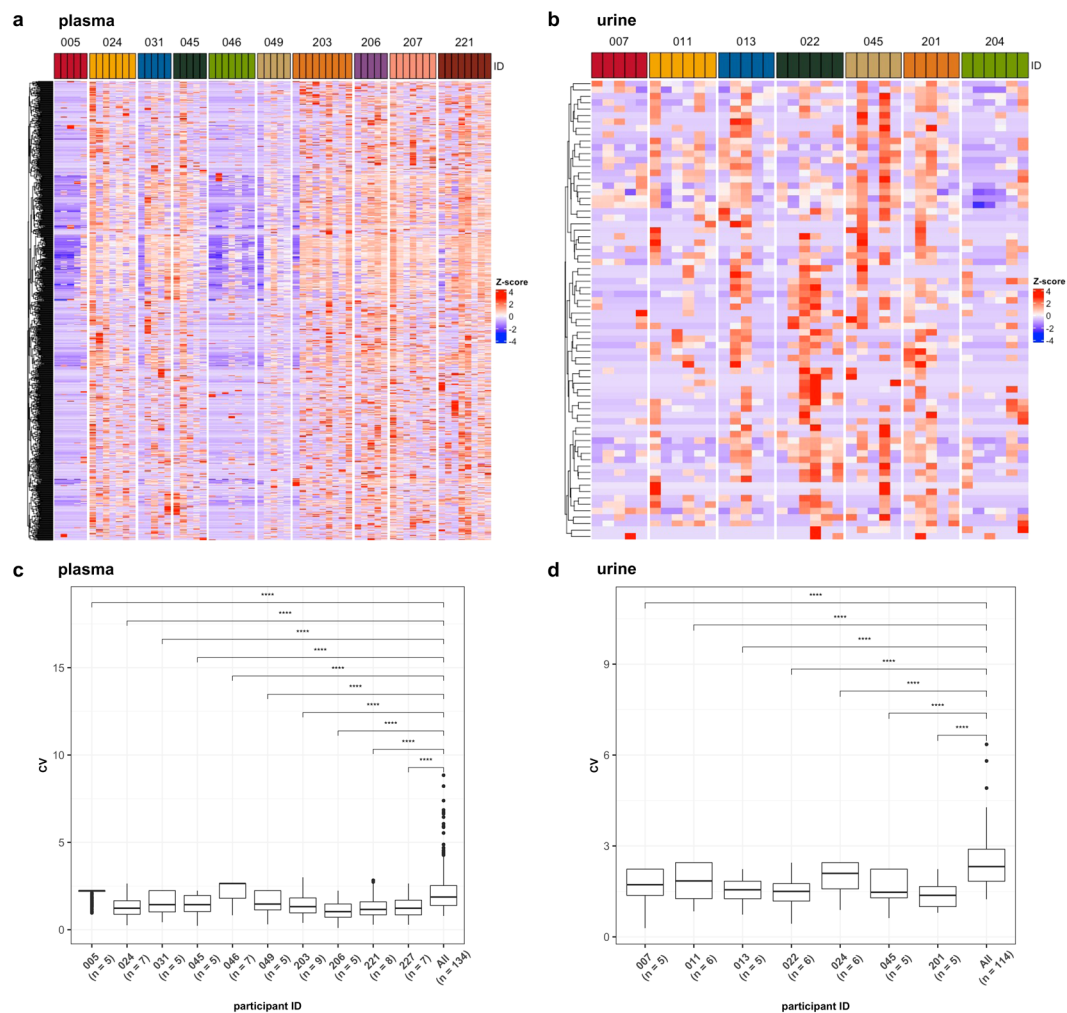


Fig. 4 Participants sequenced five or more times have less inter-sample variation. CircRNA populations identified in plasma (**a,c**) and urine (**b,d**) from participants sampled five or more times. (**a,b**) Heatmaps showing the log-normalized JRPm expression of plasma (**a**) and urine (**b**) samples taken longitudinally from the same participant. The coefficient of variation (CV) of circRNA expression is significantly lower across individual participant samples when compared to the entire dataset (**c**, plasma; and **d**, urine). **** $p < 0.0001$.

bias⁶⁴. Before examining the expression of circular RNA in relation to their host genes, we calculated the overall 5' to 3' coverage of linear transcripts and did not find a bias in our samples.

Participants sequenced 5 or more times have less inter-sample variation. A notable feature of this dataset is that many participants were sampled longitudinally, allowing for analysis of circRNA stability in individuals versus the entire dataset. Figure 4a,b show longitudinal circRNA expression in the same participants in plasma and urine, respectively. Broadly speaking, the heatmaps demonstrate similar expression patterns in the same participant over time. In order to assess variability within individuals, we calculated the coefficient of variation (CV) of circRNA expression, normalized to junction reads per million (JRPm). Here, we focus on participants sampled on 5 or more occasions over approximately one year. In both plasma and urine, the CV for each individual participant is displayed along with the CV for all participant samples. The data indicate that individuals have a statistically-significant consistency in circRNA expression pattern over time (Fig. 4c, plasma and 4d, urine).

Usage Notes

As the approach to detecting circRNA from RNA-Seq data differ with available tools, we employed 6 different bioinformatic tools: CIRCexplorer, CIRI2, DCC, KNIFE, find_circ, and MapSplice, in two clinically relevant biofluids, plasma and urine, using 134 and 114 samples, respectively. Most of these circRNA pipelines use an external aligner, such as bowtie, STAR, or bwa, to align reads to the genome and/or transcriptome (Table 2). After alignment, reads that contiguously align to the genome and/or transcriptome are filtered out, and the remaining unmapped reads are further filtered to identify back-spliced junctions. Differences in circRNA identification algorithms include: 1) how paired-end reads and gene annotations are used, if at all, 2) the amount of overlap

over the junction that a read must contain, 3) the types of junctions considered, and 4) various filtering steps (Table 1)⁶⁵. We sought to generate a high confidence set of circRNA expressed in plasma and urine with the following requirements for each circRNA: 1) detection in at least 5 samples for each respective biofluid, 2) a minimum 18 nt overlap on either side of the junction, 3) at least two reads spanning the back-spliced junction, and 4) identification by all 6 tested bioinformatic tools as identification can vary widely between tools^{66–68}.

We tested alignment parameters and their influence on the detection rate of circRNA and found that the number of input reads, genome mapped reads, and junction reads did not correlate well with the number of circRNA detected per sample; rather the number of reads assigned to the transcriptome had the greatest correlation with the number of circRNA ($R^2 = 0.805$; data not shown).

Code availability

Code used for circRNA identification is available in the Supplemental data. Software versions used for analysis are as follows:

STAR v2.4.0j for DCC and CIRCexplorer
bowtie2 v2.2.1 for find_circ and KNIFE
bowtie v0.12.9 for MapSplice2
bwa v0.7.13 for CIRI2
bedtools v2.26.0

Data were analyzed using the R v3.3.2 statistical package (<https://cran.r-project.org>). UpSet plots were generated using the UpSetR v1.3.3 package⁵⁴. The ReadqPCR v1.20.0 and NormqPCR v1.20.0 Bioconductor v3.4 packages were used for qRT-PCR data analysis⁵⁹.

Received: 19 October 2020; Accepted: 10 September 2021;

Published online: 28 October 2021

References

- Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D. & Craig, D. W. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **17**, 257–271, <https://doi.org/10.1038/nrg.2016.10> (2016).
- Stępień, E. *et al.* The circulating non-coding RNA landscape for biomarker research: lessons and prospects from cardiovascular diseases. *Acta Pharmacol Sin* **39**, 1085–1099, <https://doi.org/10.1038/aps.2018.35> (2018).
- Murillo, O. D. *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**, 463–477 e415, <https://doi.org/10.1016/j.cell.2019.02.018> (2019).
- Pardini, B., Sabo, A. A., Birolo, G. & Calin, G. A. Noncoding RNAs in Extracellular Fluids as Cancer Biomarkers: The New Frontier of Liquid Biopsies. *Cancers (Basel)* **11**, <https://doi.org/10.3390/cancers11081170> (2019).
- Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. & Kleinschmidt, A. K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci USA* **73**, 3852–3856, <https://doi.org/10.1073/pnas.73.11.3852> (1976).
- Nigro, J. M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
- Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
- Cocquerelle, C., Mascrez, B., Hetuin, D. & Baillet, B. Mis-splicing yields circular RNA molecules. *FASEB J* **7**, 155–160, <https://doi.org/10.1096/fasebj.7.1.7678559> (1993).
- Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* **40**, 3131–3142, <https://doi.org/10.1093/nar/gkr1009> (2012).
- Lu, T. *et al.* Transcriptome-wide investigation of circular RNAs in rice. *RNA* **21**, 2076–2087, <https://doi.org/10.1261/rna.052282.115> (2015).
- 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).
- Dang, Y. *et al.* Tracing the expression of circular RNAs in human pre-implantation embryos. *Genome Biol* **17**, 130, <https://doi.org/10.1186/s13059-016-0991-3> (2016).
- Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* **58**, 870–885, <https://doi.org/10.1016/j.molcel.2015.03.027> (2015).
- 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).
- Tan, W. L. *et al.* A landscape of circular RNA expression in the human heart. *Cardiovasc Res* **113**, 298–309, <https://doi.org/10.1093/cvr/cvw250> (2017).
- Veno, M. T. *et al.* Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol* **16**, 245, <https://doi.org/10.1186/s13059-015-0801-3> (2015).
- You, X. *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* **18**, 603–610, <https://doi.org/10.1038/nn.3975> (2015).
- 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).
- Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388, <https://doi.org/10.1038/nature11993> (2013).
- 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).
- 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).
- 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).
- 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).
- Maass, P. G. *et al.* A map of human circular RNAs in clinically relevant tissues. *J Mol Med (Berl)* **95**, 1179–1189, <https://doi.org/10.1007/s00109-017-1582-9> (2017).
- Bahn, J. H. *et al.* The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin Chem* **61**, 221–230, <https://doi.org/10.1373/clinchem.2014.230433> (2015).
- Alhasan, A. A. *et al.* Circular RNA enrichment in platelets is a signature of transcriptome degradation. *Blood* **127**, e1–e11, <https://doi.org/10.1182/blood-2015-06-649434> (2016).

27. 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).
28. Preusser, C. *et al.* Selective release of circRNAs in platelet-derived extracellular vesicles. *J Extracell Vesicles* **7**, 1424473, <https://doi.org/10.1080/20013078.2018.1424473> (2018).
29. Savelyeva, A. V. *et al.* Variety of RNAs in Peripheral Blood Cells, Plasma, and Plasma Fractions. *Biomed Res Int* **2017**, 7404912, <https://doi.org/10.1155/2017/7404912> (2017).
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* **123**, 2113–2118, <https://doi.org/10.1111/1471-0528.13897> (2016).
31. Kolling, M. *et al.* Circular RNAs in Urine of Kidney Transplant Patients with Acute T Cell-Mediated Allograft Rejection. *Clin Chem* **65**, 1287–1294, <https://doi.org/10.1373/clinchem.2019.305854> (2019).
32. Liu, B. *et al.* Characterization of tissue-specific biomarkers with the expression of circRNAs in forensically relevant body fluids. *Int J Legal Med* **133**, 1321–1331, <https://doi.org/10.1007/s00414-019-02027-y> (2019).
33. Ma, H. *et al.* Differential expression study of circular RNAs in exosomes from serum and urine in patients with idiopathic membranous nephropathy. *Arch Med Sci* **15**, 738–753, <https://doi.org/10.5114/aoms.2019.84690> (2019).
34. Dou, Y. *et al.* Circular RNAs are down-regulated in KRAS mutant colon cancer cells and can be transferred to exosomes. *Sci Rep* **6**, 37982, <https://doi.org/10.1038/srep37982> (2016).
35. Lasda, E. & Parker, R. Circular RNAs Co-Precipitate with Extracellular Vesicles: A Possible Mechanism for circRNA Clearance. *PLoS One* **11**, e0148407, <https://doi.org/10.1371/journal.pone.0148407> (2016).
36. Li, Y. *et al.* Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res* **25**, 981–984, <https://doi.org/10.1038/cr.2015.82> (2015).
37. 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).
38. Song, X. *et al.* Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res* **44**, e87, <https://doi.org/10.1093/nar/gkw075> (2016).
39. Chen, S. *et al.* Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell* **176**, 831–843 e822, <https://doi.org/10.1016/j.cell.2019.01.025> (2019).
40. Ding, X. *et al.* Profiling expression of coding genes, long noncoding RNA, and circular RNA in lung adenocarcinoma by ribosomal RNA-depleted RNA sequencing. *FEBS Open Bio* **8**, 544–555, <https://doi.org/10.1002/2211-5463.12397> (2018).
41. Chen, F. *et al.* Circular RNAs expression profiles in plasma exosomes from early-stage lung adenocarcinoma and the potential biomarkers. *J Cell Biochem* **121**, 2525–2533, <https://doi.org/10.1002/jcb.29475> (2020).
42. Burgos, K. L. *et al.* Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. *RNA* **19**, 712–722, <https://doi.org/10.1261/rna.036863.112> (2013).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12, <https://doi.org/10.14806/ej.17.1.200> (2011).
44. Szabo, L. *et al.* Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol* **16**, 126, <https://doi.org/10.1186/s13059-015-0690-5> (2015).
45. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* **38**, e178, <https://doi.org/10.1093/nar/gkq622> (2010).
46. 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).
47. 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).
48. Cheng, J., Metge, F. & Dieterich, C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* **32**, 1094–1096, <https://doi.org/10.1093/bioinformatics/btv656> (2016).
49. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, <https://doi.org/10.1038/nmeth.1923> (2012).
52. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, <https://doi.org/10.1186/gb-2009-10-3-r25> (2009).
53. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
54. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940, <https://doi.org/10.1093/bioinformatics/btx364> (2017).
55. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930, <https://doi.org/10.1093/bioinformatics/btt656> (2014).
56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol* **15**, 550, <https://doi.org/10.1186/s13059-014-0550-8> (2014).
57. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291, <https://doi.org/10.1093/bioinformatics/btm091> (2007).
58. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res* **40**, e115, <https://doi.org/10.1093/nar/gks596> (2012).
59. Perkins, J. R. *et al.* ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics* **13**, 296, <https://doi.org/10.1186/1471-2164-13-296> (2012).
60. Van Keuren-Jensen, K. & Huentelman, M. dbGaP <https://identifiers.org/dbgap:phs001258.v2.p1> (2020).
61. Van Keuren-Jensen, K. *et al.* figshare <https://doi.org/10.6084/m9.figshare.c.5420832> (2021).
62. 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).
63. 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).
64. Batagov, A. O. & Kurochkin, I. V. Exosomes secreted by human cells transport largely mRNA fragments that are enriched in the 3′-untranslated regions. *Biol Direct* **8**, 12, <https://doi.org/10.1186/1745-6150-8-12> (2013).
65. Szabo, L. & Salzman, J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat Rev Genet* **17**, 679–692, <https://doi.org/10.1038/nrg.2016.114> (2016).
66. Chen, I., Chen, C. Y. & Chuang, T. J. Biogenesis, identification, and function of exonic circular RNAs. *Wiley Interdiscip Rev RNA* **6**, 563–579, <https://doi.org/10.1002/wrna.1294> (2015).
67. Hansen, T. B., Veno, M. T., Damgaard, C. K. & Kjems, J. Comparison of circular RNA prediction tools. *Nucleic Acids Res* **44**, e58, <https://doi.org/10.1093/nar/gkv1458> (2016).
68. Zeng, X., Lin, W., Guo, M. & Zou, Q. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol* **13**, e1005420, <https://doi.org/10.1371/journal.pcbi.1005420> (2017).

Acknowledgements

This work was funded by support from the Flinn Foundation (Grant Award #1994 and #2037) and by NIH grant UH2TR000891. We would like to thank Terry Lee, Dan Arment, Thad Ide, Dan Vooletich, Erin Griffin, Taylor Hanohano and Brian Roche from Riddell for their significant input, time, effort, and financial support. There were a large number of individuals that made the collection of these samples possible. We would like to thank the staff at Arizona State University: Todd Graham, Ray Anderson, Jean Boyd, Tim Cassidy, Anikar Chhabra, and Jerry Neilly. We would also like to thank Ann Marie Bothwell (Desert Testing), April Allen, Yana Gadev, Stephanie Buchholtz, Cassandra Lucas, Therese de la Torre, Brian Anderson, Stephanie Althoff, and Brian Churas, Sean Allen, Ryan Bruhns, Ashley Suiter, Brandon Chaves, Mari Turk, Khalouk Shahbander, Michael Schmale, Kirk Ryden, and Alex Starr for assistance with sample collection.

Author contributions

Conceptualization, E.H. and K.V.K.J.; Methodology, E.H. and K.V.K.J.; Validation, E.H. and J.W.; Formal Analysis, E.H.; Investigation, E.H., R.R., J.W., R.R., T.B., E.C., A.J., A.S., C.B., J.A.; Resources, M.A., R.M. Y.K.; Writing - Original Draft, E.H. and K.V.K.J.; Writing, E.H., T.G.W. and K.V.K.J.; Review & Editing, E.H., T.G.W., Y.K., M.J.H., K.V.K.J. Funding Acquisition, M.H., Y.K. and K.V.K.J.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-01056-w>.

Correspondence and requests for materials should be addressed to K.V.K.-J.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, 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 metadata files associated with this article.

© The Author(s) 2021