1521-0111/99/6/488–503$35.00
MOLECULAR PHARMACOLOGY

https://doi.org/10.1124/molpharm.120.000225
Mol Pharmacol 99:488–503, June 2021

Circular RNAs of UDP-Glycosyltransferase (*UGT*) Genes Expand the Complexity and Diversity of the UGT Transcriptome[S]

Dong Gui Hu, Peter I. Mackenzie, Julie-Ann Hulin, Ross A. McKinnon, and Robyn Meech

Department of Clinical Pharmacology and Flinders Cancer Centre, Flinders University, College of Medicine and Public Health, Flinders Medical Centre, Bedford Park, South Australia, Australia

Received December 19, 2020; accepted March 26, 2021

ABSTRACT

The human UDP-glycosyltransferase (*UGT*) gene superfamily generates 22 canonical transcripts coding for functional enzymes and also produces nearly 150 variant UGT transcripts through alternative splicing and intergenic splicing. In the present study, our analysis of circRNA databases identified backsplicing events that predicted 85 circRNAs from *UGT* genes, with 33, 11, and 19 circRNAs from *UGT1A*, *UGT2B4*, *UGT8*, respectively. Most of these UGT circRNAs were reported by one database and had low abundance in cell- or tissue-specific contexts. Using reverse-transcriptase polymerase chain reaction with divergent primers and cDNA samples from human tissues and cell lines, we found 13 circRNAs from four *UGT* genes: *UGT1A* (three), *UGT2B7* (one), *UGT2B10* (one), and *UGT8* (eight). Notably, all eight UGT8 circRNAs contain open reading frames that include the canonical start AUG codon and encode variant proteins that all have the common 274–amino acidN-terminal region of wild-type UGT8 protein. We further showed that one UGT8 circRNA (circ_UGT8-1) was broadly expressed in human tissues and cell lines, resistant to RNase R digestion, and predominately present in the cytoplasm. We cloned five UGT8 circRNAs into the Zinc finger with KRAB and SCAN domains 1 vector and transfected them into HEK293T cells. All these vectors produced both circRNAsand linear transcripts with varying circular/linear ratios (0.17–1.14).Western blotting and mass spectrometry assays revealed that only linear transcripts and not circRNAs were translated. In conclusion, our findings of nearly 100 circRNAs greatly expand the complexity and diversity of the UGT transcriptome; however, UGT circRNAs are expressed at a very low level in specific cellular contexts, and their biologic functions remain to be determined.

SIGNIFICANT STATEMENT

The human UGT gene transcriptome comprises 22 canonical transcripts coding for functional enzymes and approximately 150 alternatively spliced and chimeric variant transcripts. The present study identified nearly 100 circRNAs from *UGT* genes, thus greatly expanding the complexity and diversity of the UGT transcriptome. UGT circRNAs were expressed broadly in human tissues and cell lines; however, most showed very low abundance in tissue- and cell-specific contexts, and therefore their biological functions remain to be investigated.

Introduction

The human UDP-glycosyltransferase (UGT) superfamily includes 22 functional enzymes that are divided into four subfamilies (UGT1, 2, 3, 8) (Meech et al., 2019). The UGT1 (1A1, 1A3-1A10) and UGT2 (2A1, 2A2, 2A3, 2B4, 2B7, 2B10, 2B11, 2B15, 2B17, and 2B28) enzymes use UDP-glucuronic acid to conjugate numerous endogenous and exogenous lipophilic compounds (e.g., steroid hormones, carcinogens, and drugs), thus rendering them more water-soluble and facilitating their excretion (Mackenzie et al., 2005). These UGTs are highly expressed in drug metabolism–relevant tissues (liver, kidney, and intestine), reflecting their major role in systemic drug metabolism and clearance (Hu et al., 2014a). We have recently characterized the function of three other UGT enzymes (UGT3A1, UGT3A2, UGT8) and showed that they use alternative UDP-sugars to conjugate a variety of endogenous and exogenous substrates (e.g., ursodeoxycholic acid, bile acids, and ceramide) (MacKenzie et al., 2008, 2011; Meech et al., 2015). Given their overall importance in drug metabolism and signaling molecule homeostasis, *UGT* genes are tightly regulated at transcriptional (Lu et al., 2005; Hu et al., 2014a,b; Hu et al., 2015), splicing (Tourancheau et al., 2016; Hu et al., 2018), post-transcriptional (Dluzen et al., 2014; Margaillan et al., 2016; Papageorgiou et al., 2016; Papageorgiou and Court, 2017; Wijayakumara et al., 2017), and post-translational (Basu et al., 2005, 2008; Hu et al., 2019) levels.

Precursor mRNAs (pre-mRNAs) of protein-coding genes undergo canonical forward splicing (often called *cis*-splicing) that removes introns and joins exons in their genomic order

This study was supported by the National Health and Medical Research Council of Australia [Grant 1143175] (to R.M, R.A.M, P.I.M, D.G.H) and Australia Research Council [Grant DP210103065] (to R.M.).
The authors declare no conflicts of interest.
https://doi.org/10.1124/molpharm.120.000225.

[S] This article has supplemental material available at molpharm.aspetjournals.org.

ABBREVIATIONS: aa, amino acid; BSJ, backsplicing junction; CIRC, circular RNA; CircP, circRNA-encoded protein; circRNA, circular RNA; IRES, internal ribosome entry site; LinearP, linear RNA-encoded protein; MCS, multiple cloning site; ORF, open reading frame; PCR, polymerase chain reaction; pre-mRNA, precursor mRNA; RT-PCR, reverse-transcriptase polymerase chain reaction; RT-qPCR, reverse-transcriptase quantitative real-time polymerase chain reaction; TNBC, triple negative breast cancer; UGT, UDP-glycosyltransferase; ZKSCAN1, Zinc finger with KRAB and SCAN domains 1.

to generate linear RNA transcripts; however, thousands of human protein-coding pre-mRNAs also undergo backsplicing that produces circRNAs through the splicing of a downstream donor splice site to an upstream acceptor splice site (Wilusz, 2018). circRNAs can be classified into three subgroups: exonic, exonic/intronic, and intronic (Memczak et al., 2013). circRNAs have a covalently closed loop structure and therefore lack a 5′-Cap and a 3′-poly(A) tail. Hence, circRNAs are refractory to exonucleases and more stable than linear transcripts. circRNA biogenesis is tightly regulated by *cis*-elements and *trans*-acting factors (Wilusz, 2018). circRNA formation is frequently mediated by base pairing between two inverted *Alu* elements in flanking introns, which brings two splice sites into close proximity (Jeck et al., 2013). Several *trans*-acting factors have been reported to regulate circRNA synthesis, such as Quaking (QKI) (Conn et al., 2015), Muscleblind (Ashwal-Fluss et al., 2014), ADAR1 (adenosine deaminase acting on RNA) (Ivanov et al., 2015),and DExH-Box helicase 9 (Aktas et al., 2017). circRNAs may serve as sponges for miRNAs (e.g., *ciRS-7*, Sry, circZNF91) (Hansen et al., 2013; Kristensen et al., 2018b) and RNA-binding proteins (Abdelmohsen et al., 2017), or serve as templates for protein synthesis, such as circFBXW7 (Yang et al., 2018), circSHPRH (Zhang et al., 2018a), circPINT (Zhang et al., 2018b), circMbl3 (Pamudurti et al., 2017), circβ-catenin (Liang et al., 2019), and circZNF609 (Legnini et al., 2017). However, the biologic function of most circRNAs remains unknown.

In addition to 22 wild-type mRNAs, *UGT* genes generate over 130 variant transcripts via alternative splicing (Girard et al., 2007; Levesque et al., 2007; Tourancheau et al., 2016). Briefly, the nine functional UGT1A enzymes are produced by the splicing of unique first exons (exons 1) to a set of common exons (exons 2–5) (Mackenzie et al., 2005). Alternative splicing of exon 5b generates two sets of nine variant UGT1A proteins (Fig. 1) (Girard et al., 2007; Levesque et al., 2007). Similarly, the UGT2A family has three functional enzymes (2A1, 2A2, 2A3), two of which, UGT2A1 and UGT2A2, are produced by splicing of a unique exon 1 to a set of common exons 2–6 (Mackenzie et al., 2005; Sneitz et al., 2009). The skipping of exon 3 generates two variant proteins: UGT2A1_i2 (Bushey and Lazarus, 2012) and UGT2A2_i2 (Bushey et al., 2013). Extensive alternative splicing of the seven *UGT2B* family genes is also well characterized (Tourancheau et al., 2016; Hu et al., 2019), including unusual intergenic splicing of adjacent *UGT* genes generating chimeric transcripts (Hu et al., 2018). Alternatively spliced transcripts from the *UGT3A* and *UGT8* genes have also been reported (Hu et al., 2019; Meech et al., 2015, 2019). Many of the reported alternatively spliced and chimeric UGT transcripts encode C-terminally truncated UGT proteins that have no glucuronidation activity, but they can inhibit the activity of wild-type UGT enzymes via protein-protein interactions (Bellemare et al., 2010a,b; Bushey and Lazarus, 2012; Bushey et al., 2013; Tourancheau et al., 2016; Hu et al., 2018). However, whether circRNAs are generated from *UGT* genes remains to be investigated.

In the present study, we provide bioinformatic and experimental evidence for circRNAs generated from *UGT* genes. Our analyses of circRNA databases identified backsplicing events that predicted 85 circRNAs from *UGT* genes. Using RT-PCR with divergent primers, we found 13 circRNAs in human tissues and cell lines from four *UGT* genes: *UGT1A*, *UGT2B7*, *UGT2B10*, and *UGT8*. We further assessed the expression profiles and coding and translational potential of UGT8 circRNAs. Our findings of nearly 100 circRNAs from *UGT* genes greatly expand the complexity and diversity of the UGT transcriptome.

Materials and Methods

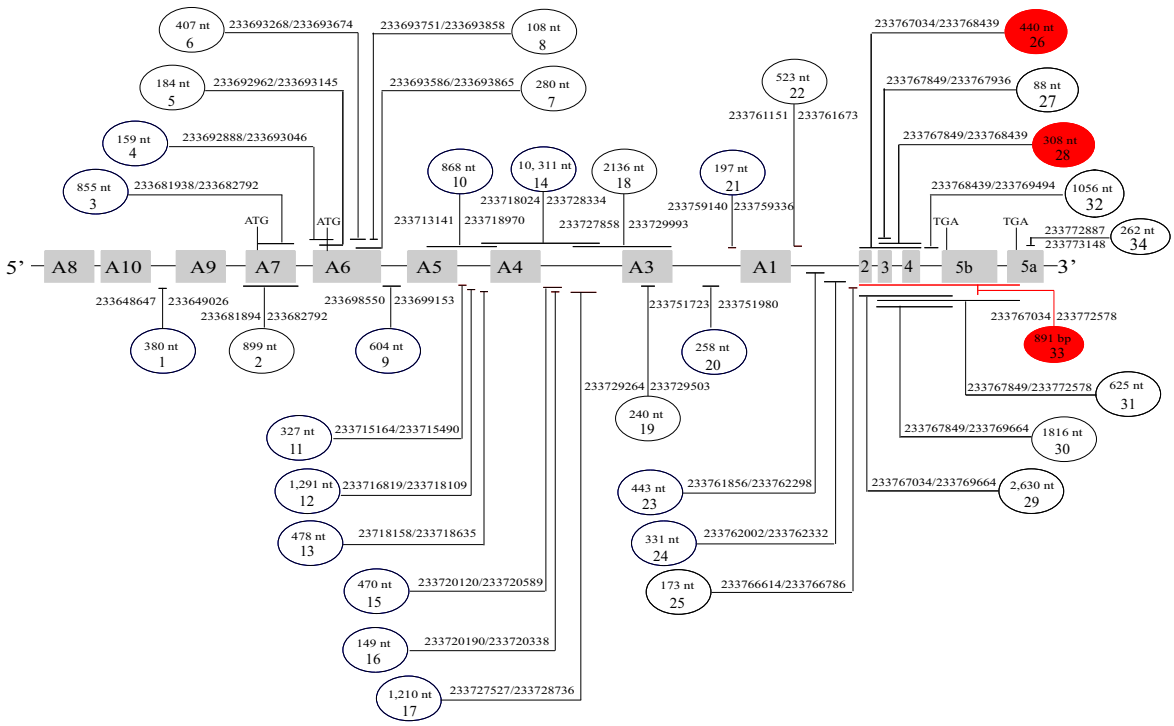
Human Tissues and Cancer Cell Lines. Total RNA samples from a panel of human tissues were purchased from Thermo Fisher Scientific (Ambion brand; Waltham, MA). Human cancer cell lines were purchased from American Type Culture Collection (Manassas, VA) and maintained under media conditions recommended by American Type Culture Collection.

Discovery of UGT circRNAs from circRNA Databases. Previous bioinformatic analysis of large-scale RNA-seq data sets from numerous normal and cancerous human tissues and cell lines has led to the establishment of many publicly accessible circRNA databases (Vromman et al., 2021). These circRNAs are predicted based on backsplicing junctions (BSJ) detected by bioinformatic software, such as CIRCexplorer2 (Zhang et al., 2016), MapSplice (Wang et al., 2010), ACFS (You and Conrad, 2016), find_circ (Memczak et al., 2013), DCC (Cheng et al., 2016), CIRI2 (Gao et al., 2018), and circ RNA_finder (Westholm et al., 2014). Therefore, very few of these predicted circRNAs are experimentally verified, nor are their full-length sequences determined. Hence, several circRNA databases name circRNAs using only their BSJ genomic coordinates, such as CSCD (Xia et al., 2018), TSCD (Xia et al., 2017), and CircRic (Ruan et al., 2019b). Despite this, circRNA databases represent the primary resources for preliminary analysis of circRNAs for genes of interest.

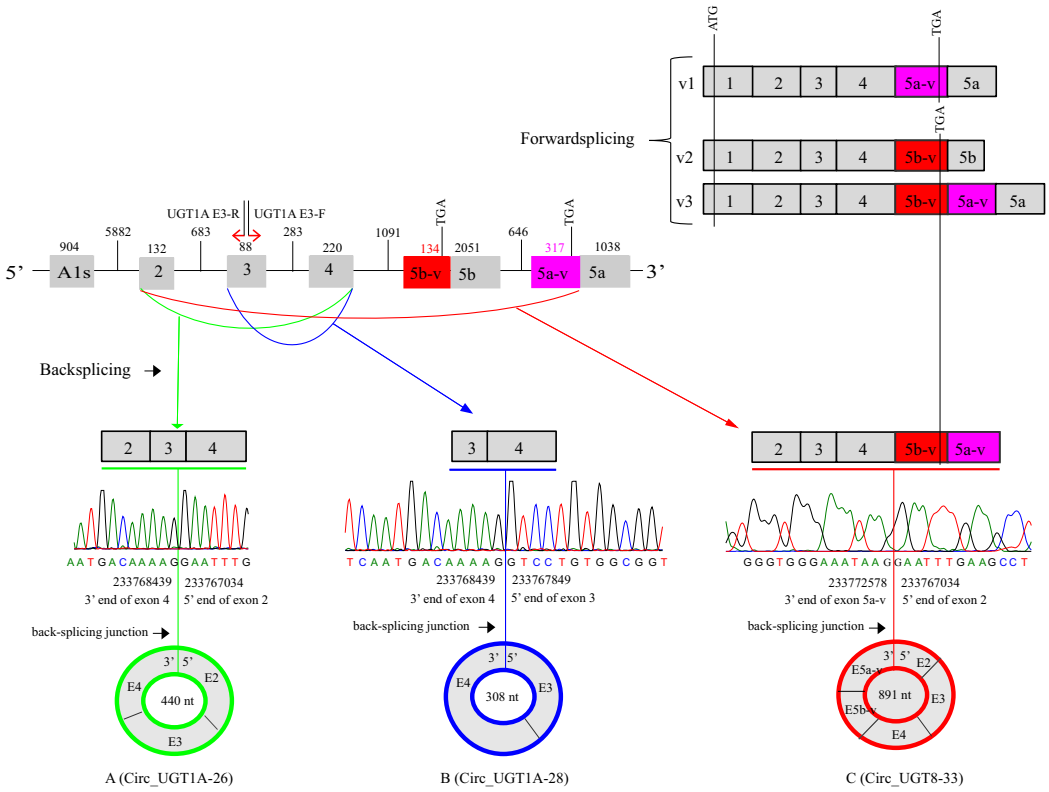
To find circRNAs from *UGT* genes, we searched nine frequently used circRNA databases: MiOncoCirc (Vo et al., 2019), CircRic (Ruan et al., 2019b), CircNet (Liu et al., 2016), exoRBase (Li et al., 2018), CircAtlas (Wu et al., 2020), TSCD (Xia et al., 2017), CSCD (Xia et al., 2018), circBase (Glazar et al., 2014), and CIRCpedia (Dong et al., 2018). As summarized in Supplemental Table 1, these databases predicted circRNAs through bioinformatic analyses of large-scale RNA-seq data sets from major projects (e.g., Encyclopedia of DNA Elements, the Cancer Cell line Encyclopedia, European Bioinformatics Institute), as well as numerous independent NCBI (National Centre for Biotechnology Information) Gene Expression Omnibus data sets. For example, CircRic collected circRNAs from 935 Cancer Cell lines (Ruan et al., 2019b), and MiOncoCirc collected circRNAs from more than 20 different types of cancers (>2000 tumor samples) (Vo et al., 2019).

Most circular RNA databases provided a name and the BSJ genomic coordinates for every circRNA. Some databases also provided the BSJ reads for every circRNA and the tissues and cell lines in which circRNAs were detected. There was no consensus on circRNA nomenclature among these databases. For consistency, we named each UGT circRNA with the prefix (circ), the gene symbol, and a unique number. The number was from one to the total number of circRNAs found for that gene. The order of the numbers was assigned from the 5′- to the 3′-genomic location of the gene according to circRNA BSJ genomic coordinates. All UGT circRNAs are named using this system and listed in Supplemental Table 2.

RNA Preparation. Total RNA was extracted from cell lines using TRIzol reagents according to the manufacturer's protocol (Invitrogen/Thermo Fisher Scientific, Carlsbad, CA). Nuclear and cytoplasmic RNA samples were prepared from cell lines using the PARIS kit (Ambion, Carlsbad, CA) as previously reported (Errichelli et al., 2017; Ruan et al., 2019a). RNase R treatment was conducted as previously reported (Pandey et al., 2019) at 37°C for 30 minutes in a 20-μl reaction containing 1× RNase R reaction buffer, 5 μg total RNA, 10 units of RNase R (Epicentre), 20 units of RiboLock inhibitor (Thermo Scientific), and 1 unit of DNase I (Thermo Scientific). RNase R–treated RNA samples were then purified using RNeasy MinEluteCleanup kit (QIAGEN) according to the manufacturer's protocol.



A



B

Fig. 1. circRNAs of the *UGT1A* gene identified by circRNA databases and RT-PCR with divergent primers. (A) A schematic diagram showing the *UGT1A* gene structure and the locations and genomic coordinates (GRCh38/hg38) of 33 circRNAs identified from circRNA databases. The three circRNAs found by divergent RT-PCR are highlighted in RED. (B) A schematic diagram showing the *UGT1A1* gene, three sets of forward-spliced linear transcripts (v1, v2, v3), and three backspliced circRNAs identified in HT-29 cells using RT-PCR with divergent primers (red arrows) targeting exon 3. The lengths (nt) of exons and introns of the *UGT1A1* gene are given. Sequencing chromatograms show the backsplicing junctions (genomic coordinates, GRCh38/hg38) of the three circRNAs identified by divergent RT-PCR: A (E2/E3/E4), B (E3/E4), and C (E2/E3/E4/E5b-v/E5a-v).

Reverse Transcription. Reverse transcription was conducted using Invitrogen reagents as previously reported (Hu et al., 2018). Briefly, total RNA (1 μg) was treated with DNase I at room temperature for 15 minutes and then reverse-transcribed using random hexamer primers (50 ng) and Superscript III (50 units) at 50°C for 50 minutes in a 20-μl reaction containing 50 mM Tris-HCl (pH 8.0), 75 mM KCl, and 3 mM $MgCl_2$. The resulting cDNAs were diluted in 80 μl RNase-free H_2O prior to polymerase chain reaction (PCR) or real-time quantitative PCR.

Divergent RT-PCR and Cloning of the Resultant Amplicons into the pCR Blunt Vector. We designed four pairs of divergent primers targeting *UGT1A* exon 3 (UGT1A E3-F/UGT1A E3-R), *UGT2B* exon 3 (UGT2B E3-F/UGT2B E3-R), *UGT8* exon 2 (UGT8 E2-F/UGT8 E2-R), or *UGT8* exons 2 and 5 (UGT8 E5-F/UGT8 E2-R) (Supplemental Table 3). PCRs were conducted using cDNA samples from human tissues or cell lines (as indicated) and Phusion High-Fidelity DNA polymerase (New England Biolabs Ltd., Hitchin, UK). PCR products were verified on agarose gels (Supplemental Fig. 1), purified using QIAquick PCR purification kit (QIAGEN, Hilden, Germany), and cloned into the pCR Blunt vector using Zero Blunt PCR cloning kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Inserts were sequenced using primers M13F and M13R (Thermo Fisher Scientific). The Supplemental Figs. 2–14 show the sequencing chromatograms of the clones representing 13 different UGT circRNAs identified in the present study.

Quantitative Real-Time Polymerase Chain Reaction. Reverse-transcriptase quantitative real-time polymerase chain reaction (RT-qPCR) was conducted using cDNAs of human tissues and cell lines and GoTaq qPCR master mix (Promega, Madison, WI) as previously reported (Hu et al., 2018). Briefly, we developed a standard calibration curve for each qPCR analysis using four serial 10-fold dilutions containing known copy numbers (e.g., 6000, 600, 60, and 6) of a plasmid carrying the amplicon. For quantification of circRNAs, we prepared the calibration curves using the pCR Blunt vectors carrying the backsplicing junctions of the respective circRNAs. For quantification of the linear transcripts generated from circRNA expression vectors (below), we prepared the calibration curves using the respective circRNA expression vectors. The standard curves allowed calculation of the absolute copy numbers of each transcript (circRNA or linear RNA) in the experimental samples (Hu et al., 2018). The primers for RT-qPCR are given in Supplemental Table 3.

Preparation of the ZKSCAN1 UGT8 circRNA Expression Vectors. The pcDNA3.1 (+) ZKSCAN1 multiple cloning site (MCS) vector was reported to express circRNAs with high efficiency (Kramer et al., 2015). This vector carries a 53–base pair MCS that is flanked by inverted *Alu* elements from the human *ZKSCAN1* gene for facilitating circRNA synthesis. However, cloning through this MCS site introduces two or more restriction sites between the BSJ of the resultant circRNAs, which alters the circRNA coding frame at the BSJ site. To avoid this, we cloned UGT8 circRNAs into this vector using In-Fusion technology. Briefly, we amplified the ZKSCAN1 vector with primers that excluded the MCS site using the empty ZKSCAN1 MCS vector as template (6701 base pairs) and the UGT8 circRNA cDNA sequence from brain cDNA. The amplified vector and circRNA cDNA were then fused together using In-Fusion HD Cloning Kit (Takara Bio USA, Inc., Mountain View, CA). Using this strategy, we prepared seven ZKSCAN1 UGT8 circRNA vectors for expressing circRNA B (CIRC B, FLAG/CIRC B, FLAG/CIRC B/HA), circRNA C (CIRC C), circRNA D (CIRC D), circRNA E (CIRC E), and circRNA F (CIRC F).

A FLAG epitope sequence (5′-GACTACAAAGACGATGACGACAAG-3′) was inserted in front of *UGT8* exon 1c in both FLAG/CIRC B and FLAG/CIRC B/HA vectors, resulting in the in-frame fusion of a C-terminal FLAG tag to the protein translated from circRNA B (designated FLAG/CircP B, Fig. 6B). An HA epitope sequence (5′-TACCCATACGATGTTCCAGATTACGCT TGA-3′, stop codon underlined) was inserted in the FLAG/CIRC B/HA vector between the 21st and 22nd nucleotides of the ZKSCAN1 donor splice signal sequence, resulting in the in-frame fusion of a C-terminal HA-tag to the protein translated from the linear transcript (designated HA/LinearP B, Fig. 6B).

The identities of all UGT8 circRNA expression vectors were confirmed by DNA sequencing. The sequences of the cloning primers are provided in Supplemental Table 1. However, it is not possible in this table to specify which primers were used for cloning specific circRNA expression vectors. For clarity, we describe below the specific pairs of primers for cloning individual expression vectors.

1. CIRC B: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/UGT8 E1c-R) and 2) CIRC (ZKSCAN1/UGT8 E1c-F; UGT8 E2/ZKSCAN1-R)
2. CIRC C: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/UGT8 E1c-R) and 2) CIRC (ZKSCAN1/UGT8 E1c-F; UGT8 E5/ZKSCAN1-R)
3. CIRC D: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/UGT8 E2-R) and 2) CIRC D (ZKSCAN1/UG E2-F; UGT8 E5/ZKSCAN1-R)
4. CIRC E: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/UGT8 E2-R) and 2) CIRC E (ZKSCAN1/UG E2-F; UGT8 E6/ZKSCAN1-R)
5. CIRC F: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/UGT8 E1c-R) and 2) CIRC F (ZKSCAN1/UG E1c-F; UGT8 E6/ZKSCAN1-R)
6. FLAG/CIRC B: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/FLAG/UGT8 E1c-R) and 2) CIRC (ZKSCAN1/FLAG/UGT8 E1c-F; UGT8 E2/ZKSCAN1-R)
7. FLAG/CIRC B/HA: PCR of 1) vector (UGT8 E2/ZKSCAN1-F; ZKSCAN1/FLAG/UGT8 E1c-R) and 2) CIRC (ZKSCAN1/FLAG/UGT8 E1c-F; UGT8 E2/ZKSCAN1/HA-R)

Quantification of circRNAs and Linear Transcripts Generated from UGT8 circRNA Expression Vectors Using RT-qPCR. HEK293T cells were plated into six-well plates at 50% confluence and incubated overnight. Cells were then transfected with 2 μg of one ZKSCAN1 UGT8 circRNA expression vector using Lipofectamine 2000. At 48 hours post-transfection, total RNA extraction and RT-qPCR were conducted as described above. Supplemental Table 1 lists the primers for qPCR of 1) CIRC B (UGT8 E2-F/UGT8 E1c-R), 2) CIRC C (UGT8 E5 qPCR-F/UGT8 E1c-R), 3) CIRC D (UGT8 E2 qPCR-R/UGT8 E5 qPCR-F), 4) CIRC E (UGT8 E2 qPCR-R/UGT8 E6-146 qPCR-F), and 5) CIRC F (UGT8 E146 qPCR-F/UGT8 E1c-R); Supplemental Table 1 also lists the primers for qPCR of the linear transcripts generated from UGT8 circRNA expression vectors, including 1) CIRC B (UGT8 E2 qPCR-F/ZKSCAN1 qPCR-R), 2) CIRC C/D (UGT8 E5 qPCR-F/ZKSCAN1 qPCR-R), and 3) CIRC E/F (UGT8 E6-146 qPCR F/ZKSCAN1 qPCR-R). The ZKSCAN1 qPCR-R primer targets the vector sequence after the HA-tag.

Investigation of the Translational Potential of UGT8 circRNAs Using Western Blotting and Mass Spectrometry Assays. HEK293T cells were transfected with UGT8 circRNA expression vectors as described above. At 48 hours post-transfection, whole-cell lysates were prepared in radioimmunoprecipitation assay buffer [50 mM Tris-HCl (pH 8.0), 1% NP40, 150 mM sodium chloride, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate]. Protein concentrations were measured using the Bradford Protein Assay (Bio-Rad, Gladesville, NSW, Australia). In total, 35 μg of protein of each whole-cell lysate was run on SDS-polyacrylamide gels (10%) and transferred to nitrocellulose membranes. Membranes were incubated with a primary antibody and then with a horseradish peroxidase–conjugated donkey anti-rabbit (or goat anti-mouse) secondary antibody (NeoMarkers; Fremont, CA). Immunosignals were imaged using the SuperSignalWest Pico Chemiluminescent kit (Thermo Fisher Scientific) and an ImageQuant LAS 4000 luminescent image analyzer (GE Healthcare, Chalfont St. Giles, UK). Three primary antibodies were rabbit anti-UGT8 (17982-1-AP; ProteinTech), mouse anti-FLAG (F1804; Sigma), and rabbit anti-HA (H6908; Sigma).

For mass spectrometry assays, the gel piece containing the protein was excised from the gel, and the resultant proteins were digested using trypsin (Sigma) or elastase (Sigma) and then subjected to mass

spectrometry analysis using the Orbitrap Exploris 480 Mass Spectrometer (ThermoFisher Scientific). Data analysis was conducted using Peaks Studio 10.5 (build 202000219) (Bioinformatics Solutions Inc, Waterloo, ON, Canada).

Cloning of UGT8 CIRC D into pGL3 Luciferase Reporter and Luciferase Assays. The UGT8 CIRC D (exons 2–5, 1264 nt) sequence was amplified from HT-29 cDNA using forward (5′TGAGTCTAG ACTATGAAGTCTTACACTC3′) and reverse (5′GCAGTCTAGACTGG GATTATTGATAACCT3′) primers and cloned downstream of the pGL3 reporter (Promega) via the Xba1 site. miRNA mimics corresponding to miR-214-3p, miR-761, miR-3619-5p, and a negative control mimic were purchased from Shanghai GenePharma (Shanghai, China). Luciferase assays were performed in the breast cancer MDA-453 cell line as previously reported (Wijayakumara et al., 2015). Briefly, cells were plated in 96-well plates at 60% confluency in RPMI media containing 5% fetal bovine serum. After overnight culture, cells were transfected in triplicate with 100 ng of a luciferase reporter (either pGL3/CIRC D vector or empty pGL3 vector), 5 ng of the control pRL-null vector, and miRNA mimics at 30 nM or the negative control miRNA. At 48 hours after transfection, cells were assayed for firefly and *Renilla* luciferase activities using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. The firefly luciferase activity was normalized first to the *Renilla* activity and then to that of cells transfected with the empty pGL3 vector, and finally presented relative to that of cells transfected with the negative control miRNA (set as a value of 100%).

Statistical Analysis. Statistical analysis was conducted by one-way analysis of variance with the Tukey's multiple comparison tests or two-way analysis of variance with Bonferroni's multiple comparison tests when appropriate using GraphPad Prism 8.0 software (GraphPad Inc., La Jolla, CA). A P value of less than 0.05 was considered statistically significant. According to recently published guidelines for displaying data and reporting data analysis and statistical methods in experimental biology (Michel et al., 2020), the findings reported in Figs. 4 and 5C and Supplemental Fig. 19B are all considered exploratory.

Results

circRNAs of *UGT* Genes Predicted Based on the Backsplicing Events from circRNA Databases. The human *UGT* gene superfamily has thirteen genes (UGT1A, 2A1/2A2, 2A3, 2B4, 2B7, 2B10, 2B11, 2B15, 2B17, 2B28, 3A1, 3A2, and UGT8) that produce 22 canonical mRNAs encoding functional enzymes (Meech et al., 2019). Our analyses of nine circRNA databases (Supplemental Table 1) identified 85 circRNAs that were derived from all *UGT* genes except for *UGT2B28* (Supplemental Table 2). Three *UGT* genes (*UGT1A*, *UGT2B4*, *UGT8*) generated the highest number (33, 11, 19, respectively) of circRNAs. All UGT circRNAs are schematically presented in Figs. 1 and 2 and Supplemental Figs. 15 and 16. UGT circRNAs were detected in a wide range of human normal and cancerous tissues and cell lines; however, most of them were reported by only a single circ RNA database in tissue- and cell line–specific contexts with very low BSJ reads (Supplemental Table 2). Only six circ RNAs were reported by two or more databases—namely, circ_UGT1A-28 (CSCD, TSCD, CIRCpedia v2, MiOncoCirc, exoRBase), circ_UGT1A-31 (CSCD, CIRCpedia v2), circ_UGT8-1 (circbase, CIRCpedia v2, exoRBase), circ_UGT8-7 (TSCD, CIRCpedia v2, MiOncCirc 2), circ_UGT8-14 (circAtlas v2, MiOncCirc 2), and circUGT8-19 (Circbase, CIRCpedia v2) (Supplemental Table 2). For example, circ_UGT8-7 (CIRC D) was detected in normal tissues (stomach, intestine) and a variety of cancers (BRCA, CHOL, ESCA, HNSC, KDNY, OV, PAAD, PRAD, SECR).

UGT circRNAs can be classified into exonic, exonic/intronic, and intronic circRNAs, with exonic circRNAs being the most common form. Most exonic circRNAs were generated through the backsplicing of canonical donor/acceptor splice sites and contained one or more canonical exons, including a single exon (e.g., A1-27, 2A3-5, 2B10-2, 2B15-2, UGT8-4), two exons (e.g., 1A-28, 2B4-3, 2B4-7, UGT8-1), three exons (e.g., 1A-26, 2B4-4, UGT8-6, UGT8-14), or four exons (e.g., 2B4-5, 2B10-3). However, cryptic donor/acceptor splice sites within exons were also frequently involved in backsplicing, leading to production of exonic circRNAs that contain partial exon sequence (e.g., 1A-29, 2B4-1, 2B7-2, 2B4-8, UGT8-15, UGT8-18). Intronic circ RNAs were all generated via cryptic splice sites within introns. Approximately 99% of exons in human protein-coding genes bear the conserved dinucleotide AG and dinucleotide GT at their acceptor and donor splice sites, respectively (Burset et al., 2000). Most cryptic acceptor/donor splice sites for UGT circRNA formation conform to this AG/GT rule, supporting an authentic backsplicing origin. For example, *UGT1A6* exon 1 was involved in the formation of five circRNAs (UGT1A-4, -5, -6, -7, -8, Fig. 1A) that used cryptic splice sites within exonic and adjacent intronic sequences that all conformed to the AG/GT rule (Supplemental Fig. 17A). Alternative backsplicing was observed for many *UGT* genes in which a donor/acceptor splice site was involved in forming multiple different circ RNAs, such as *UGT1A* (E2, E3), *UGT2B4* (E2, E3, E4), and *UGT8* (E1c, E2, E5). Collectively, circRNA databases reported a large number of UGT circRNAs with diverse exon structures. To independently verify some of these circRNAs and potentially identify new circRNAs, we performed divergent RT-PCR using primers targeting *UGT* genes as described below.

Discovery of circRNAs from the *UGT1A* Gene Using RT-PCR with Divergent Primers Targeting Exon 3. Six *UGT1A* circRNAs (1A-26, -27, -28, -29, -30, -31, -33) reported in databases contained exon 3 (Fig. 1A), two (1A-28, -30) of which were detected in normal kidney and colorectal cancer HT-29 cells (Supplemental Table 2). To validate these circRNAs, we performed RT-PCR with divergent primers (UGT1A E3-F/UGT1A E3-R) targeting exon 3. Multiple similar amplicons were obtained from kidney and HT-29 cells, as well as two other tissues (brain, colon) (Supplemental Fig. 1A). Cloning and sequencing of the amplicons from HT-29 cells revealed BSJs representing three different circular RNAs (Supplemental Figs. 2–4). circRNA A (circ_UGT1A-26) contained three exons (E2/E3/E4) and was previously reported in bladder cancer by MiOncoCirc; circRNA B (circ_UGT1A-28) contained two exons (E3/E4) and was previously reported in various tissues and cell lines by four circRNA databases (Supplemental Table 2). circRNA C (circ_UGT1A-33) contained five exons (E2/E3/E4/E5b-v/E5a-v) and was not previously reported by any circRNA database. Within circRNA C, the first 317 nt (termed E5a-v) of canonical exon E5a (1,044-nt) was backspliced to the exon 2 acceptor splice site via a cryptic donor splice site. The *UGT1A* gene generates three sets of nine transcripts (v1, v2, v3) through forward splicing (*cis*-splicing) (Fig. 1B) (Girard et al., 2007; Levesque et al., 2007). The classic v1 transcripts have E5a (1,038 nt) as the terminal exon; v2 transcripts have E5b (2,185-nt) replacing E5a; v3 transcripts have the first 134-nt of exon E5b (termed E5b-v hereafter) spliced between exons E4 and E5a. Like the forward-spliced v3 transcripts,

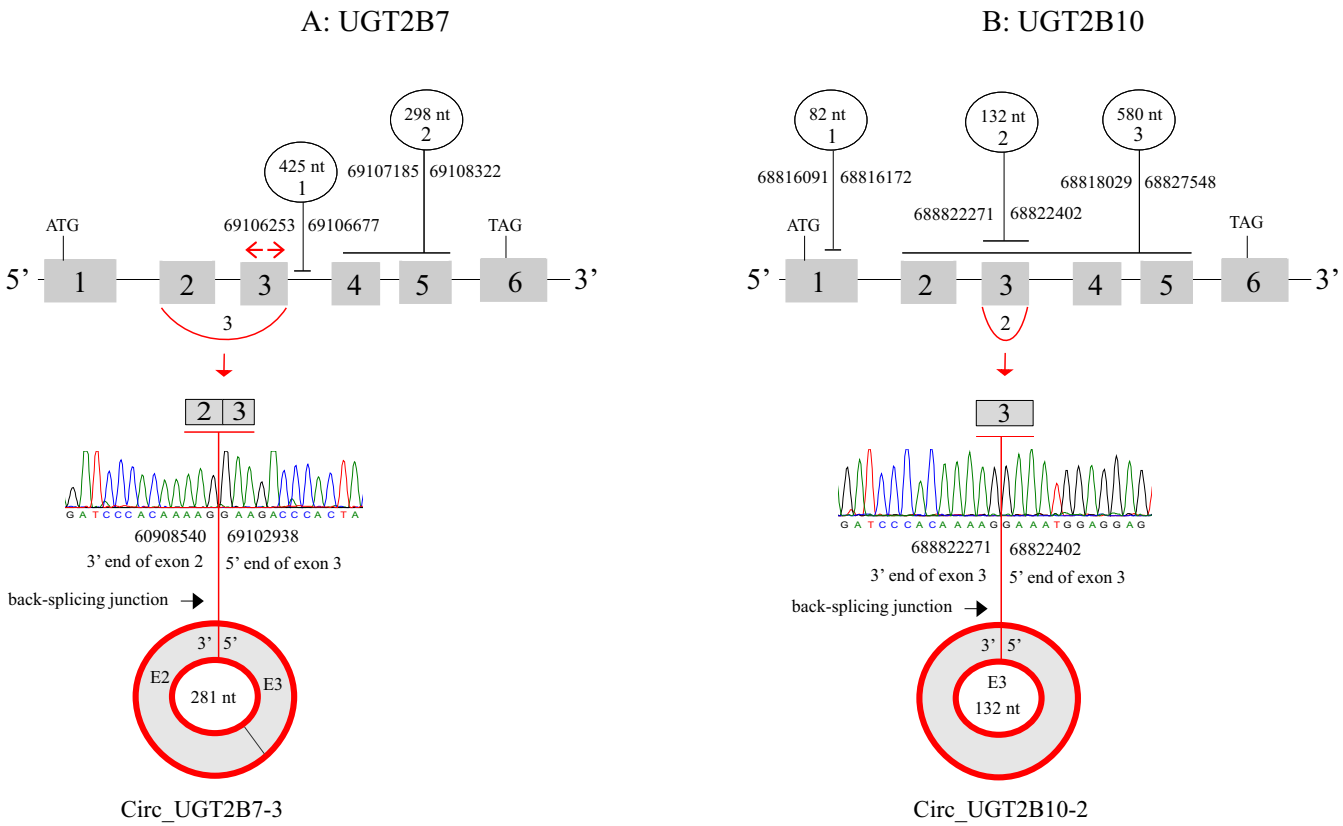


Fig. 2. circRNAs of *UGT2B7* and *UGT2B10* identified by circRNA databases and RT-PCR with divergent primers. (A and B) Two UGT2B7 circRNAs (A) and three UGT2B10 circRNAs (B) were identified in circRNA databases. One UGT2B7 circRNA containing exons 2–3 (A) and one UGT2B10 circRNA containing exon 3 (B) were found in liver cDNA using RT-PCR with divergent primers (red arrows) targeting exon 3. Sequencing chromatograms show the backsplicing junctions (genomic coordinates, GRCh38/hg38) of these two circRNAs.

circRNA C uses the same cryptic donor splice site to fuse E5a-v and E5b-v that all conform to the AG/GT rule (Supplemental Figs. 17, B and C).

Discovery of circRNAs from the *UGT2B* Genes Using RT-PCR with Divergent Primers Targeting Exon 3. As all seven *UGT2B* genes (2B4, 2B7, 2B10, 2B11, 2B15, 2B17, 2B28) have six exons with high sequence similarity, we performed RT-PCR using pan-UGT2B divergent primers targeting exon 2, exon 3, or exon 4. Most UGT2B circRNAs reported by circRNA databases were detected in liver (Supplemental Table 2). PCR of liver cDNA samples with divergent primers (UGT2B E3-R/UGT2B E3-F) targeting *UGT2B* exon 3 produced amplicons that appeared as a faint smear on agarose gels (data not shown). Cloning and sequencing of the PCR products identified BSJs representing two different circular RNAs (Supplemental Figs. 5 and 6). One circRNA (circ_UGT2B10-2) contained *UGT2B10* exon 3, and it was previously detected in liver by CircPdia (Fig. 2B; Supplemental Table 2). The other circRNA (circ_UGT2B7-3) contained *UGT2B7* exons 2–3 (Fig. 2A) and was not previously reported by any circRNA database.

Discovery of circRNAs from the *UGT8* Gene Using RT-PCR with Divergent Primers Targeting Exons 2 and 5. Of the 19 *UGT8* circRNAs reported by circRNA databases, eight (UGT8-1, -2, -3, -4, -5, -6, -7, -8) contained exon 2, and seven contained exon 5 (UGT8-5, -7, -8, -14, -15, -16, -18) (Supplemental Fig. 15). Many of these circRNAs were detected in the brain (Supplemental Table 2), consistent with high UGT8 mRNA expression in this tissue (Hu et al., 2019). RT-PCR was performed using divergent primers targeting exon 2 (UGT8 E2-R/UGT8 E2-F) or exons 2 and 5 (UGT8 E2-R/UGT8 E5-F) (Fig. 3A). Multiple amplicons were obtained from brain cDNA; however, similar but less abundant amplicons were also generated from kidney cDNA but not from HT-29 cDNA (Supplemental Fig. 1, B and C). The PCR products from brain cDNA were cloned and sequenced, identifying BSJs representing eight different circular RNAs (Fig. 3; Supplemental Figs. 7–14).

circRNA A and B were detected by the primers targeting exon 2. circRNA B (circ_UGT8-1), comprising E1c and E2, was reported by four circRNA databases (Supplemental Table 2). circRNA A (circ_UGT8-20) contained three exons (1c, 1d, 2), and it was not previously reported by any circRNA database. Both circRNA A and B had the same backsplicing junction (E2/E1c), but circRNA B contained an additional exon (1d), indicating alternative splicing (Fig. 3B). Of the six circRNAs (C–H) detected by the primers targeting exon 2/exon 5, two (D, E) were previously reported by circRNA databases (Supplemental Table 2). circRNAs D (circ_UGT8-7) and E (circ_UGT8-8) contained exons 2–5, but circRNA E had an additional exon (6v) (Fig. 3B). circRNAs C (circ_UGT8-21), F (circ_UGT8-22), and G (circ_UGT8-23) all had the same five exons (1c/2/3/4/5); however, circRNA F also had exon 6v, whereas circRNA G contained three additional exons (6v/6a/6b) (Fig. 3B). circRNA H (circ_UGT8-24) contained seven exons (2/3/4/5/6v/6a/6d). Four *UGT8* exons (1c, 2, 5, 6v) showed alternative backsplicing, with the exon 1c involved in forming five circRNAs (A, B, C, F, G) (Fig. 3A).

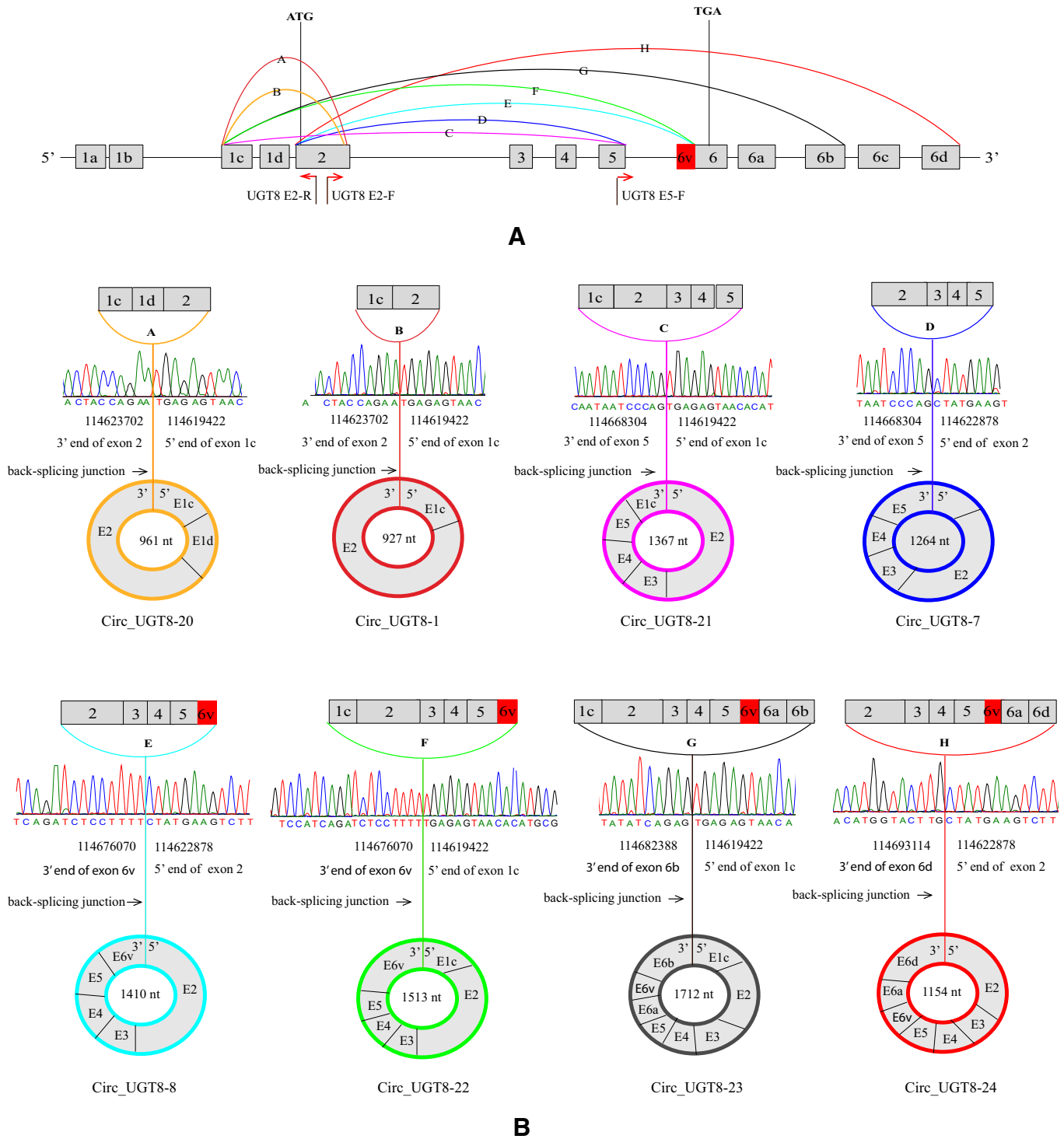


Fig. 3. circRNAs of the *UGT8* gene identified by RT-PCR with divergent primers. (A) A schematic diagram showing the exon/intron structure of the *UGT8* gene and eight backsplicing events (A–H) identified by RT-PCR with divergent primers (red arrows) targeting exon 2 and/or exon 5. The positions of the start ATG codon (exon 2) and the stop TGA codon (exon 6) are also indicated. (B) Two circRNAs (A and B) were found in brain cDNA by RT-PCR with divergent primers targeting exon 2; six circRNAs (C–H) were identified in brain cDNA by RT-PCR with divergent primers targeting exons 2 and 5. Sequencing chromatograms show the backsplicing junctions (genomic coordinates, GRCh38/hg38) of these eight circRNAs.

Expression Profile, Cellular Localization, and RNase R Resistance of UGT8 circRNA B (CIRC B). Most UGT circRNAs from databases were detected by very low BSJ reads; similarly, most circRNAs that we identified by divergent PCR, cloning, and sequencing were rare (e.g., represented by a single clone). However, three UGT8 circRNAs (B, D, F) were observed repeatedly by multiple clones with the same BSJs, thus suggesting relatively high abundance. circRNA B (CIRC B) was also previously reported by four databases (Supplemental Table 2). Therefore, we performed RT-qPCR to

assess the CIRC B expression profiles in 22 human tissues and 22 human cancer cell lines (Fig. 4). CIRC B was detected in all tissues except for the placental, with the highest expression in stomach, kidney, colon, intestine, thyroid, and brain (Fig. 4A). CIRC B was also found in 14 cell lines, with the highest levels in Caco-2, HEK293T, U118, HK-2, HT-29, and VCaP (Fig. 4B). The expression ratio of the CIRC B to the canonical UGT8 mRNA (CIRC B/UGT8 mRNA) was less than 5% in cell lines (Fig. 4D) but was relatively higher in tissues, with the highest ratio (nearly 60%) observed in the thyroid (Fig. 4C).

Linear mRNAs are degraded by exonucleases such as RNase R, whereas circRNAs are resistant because of their covalently closed loop structure (Wilusz, 2018). RNase R treatment significantly reduced linear UGT8 mRNA levels in RNA samples from brain and cell lines (Caco-2, HEK293T, U118, HT-29, VCaP) (Fig. 4F); however, circRNA B levels remained unchanged in RNA samples from four cell lines (Caco-2, HEK293T, U118, HT-29) and was significantly increased in RNA samples from VCaP cells and brain (Fig. 4G). DNA sequencing verified that the RT-qPCR amplicons from RNase R–treated samples had the CIRC B BSJ (Supplemental Fig. 18).

The cellular location of circRNAs is proposed to relate to biologic function. Cytoplasmic circRNAs may serve as translation templates or miRNA sponges, whereas nuclear circRNAs could be involved in transcription regulation (Wilusz, 2018). Using the PARIS kit, we prepared cytoplasmic and nuclear RNA samples from four cell lines (HEK293T, HT-29, U118, VCaP) with the highest CIRC B expression (Fig. 4B) and performed RT-qPCR to quantify CIRC B. CIRC B was detected in the cytoplasm but not the nucleus in U118 and VCaP cells, and CIRC B levels were significantly higher in the cytoplasm than in the nucleus in HEK293T and HT-29 cells (Fig. 4E). The localization of UGT8 circRNAs to the cytoplasm suggests possible roles in miRNA sponging or translation that were investigated as described below.

The Expression Correlation between UGT8 CIRC B and Wild-Type UGT8 mRNA in Human Tissues and Cell lines. Of the assayed 44 human tissues and cell lines (Figs. 4, A and B), 27 expressed both UGT8 CIRC B and wild-type UGT8 mRNA. Correlation analysis showed that the expression of CIRC B and UGT8 mRNA was positively correlated in these samples (Fig. 4H). UGT8 is reported to attenuate the sulfatide-αVβ5 axis and thus suppress tumor progression in basal-like breast cancer, including triple negative breast cancers (TNBC) (negative for estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2) (Cao et al., 2018). Our preliminary screening of 44 breast cancer tumors showed that higher expression of both UGT8 mRNA and CIRC B in TNBC compared with other types of breast cancers (data not shown). This positive correlation between UGT8 CIRC B and UGT8 mRNA was in contrast to previous reports that the expression of circRNAs and their cognate linear mRNAs are either uncorrelated (Barrett and Salzman, 2016; Kristensen et al., 2018a,b) or negatively correlated (Ashwal-Fluss et al., 2014). This positive correlation of CIRC B with UGT8 mRNA suggests that it might serve as a biomarker for UGT8 expression.

The ZKSCAN1 UGT8 Expression Vectors Expressed Both circRNAs and Linear Transcripts. Plasmids designed to express circRNAs generally produce both backspliced circRNAs and unbackspliced linear transcripts (Kramer et al., 2015). We used the ZKSCAN1 plasmid, which is reported to produce a high circular/linear RNA expression ratio (Kramer et al., 2015), to prepare seven vectors expressing five UGT8 circRNAs: B (CIRC B, FLAG/CIRC B, FLAG/CIRC B/HA), C (CIRC C), D (CIRC D), E (CIRC E), and F (CIRC F), as described in *Materials and Methods*. We transfected these vectors into HEK293T cells and performed RT-qPCR to quantify both circRNAs and linear transcripts. All seven vectors generated linear transcripts (Fig. 5B), but only five vectors produced expected circRNAs (CIRC B, FLAG/CIRC B, FLAG/CIRC B/HA, CIRC C, CIRC D) (Fig. 5A). The nonspecific PCR products with larger sizes than the expected circRNAs (Fig. 5A, lanes 1, 2, 3, 5, 9 11) might represent other unexpected cryptic backsplicing events as previously reported (Starke et al., 2015; Pamudurti et al., 2017). The vector CIRC D had the highest circular/linear RNA expression ratio (1.14); the four other vectors (CIRC B, FLAG/CIRC B, FLAG/CIRC B/HA, CIRC C) had significantly reduced ratios (0.17–0.41) (Fig. 5C).

The Coding and Translational Potential of UGT8 circRNAs. The *UGT8* gene codes for a 541-aa protein with a start AUG codon in exon 2 and a stop UGA codon in exon 6 (Fig. 6A). Exon 2 (824 nt) encodes the UGT8 274-aa N-terminal region. All eight UGT8 circRNAs (A–H) contain ORFs that include exon 2, and they are predicted to encode a series of variant UGT8 proteins (CircPs) of varying sizes that all include the 274-aa N-terminal domain. Predicted CircPs A (274 aa), B (274 aa), G (469 aa), H (469 aa) would all share the common 274-aa N-terminal region. Three CircPs [C (432 aa), D (451 aa), and F (535 aa)] would share a common 420-aa N-terminal region (encoded by exons 2–5) with specific C-terminal peptides [C (12 aa), D (31 aa), F (115 aa)]. circRNA E (1410 nt) contains an infinite ORF without a stop codon and thus could potentially generate a concatemer (470 aa per monomer) through rolling cycle translation (Abe et al., 2015; Mo et al., 2019). Translatable circRNAs often have an internal ribosome entry site (IRES) motif for translation initiation (Lei et al., 2020). As shown in Fig. 6A, exon 1c, which is upstream of the start codon in five *UGT8* circRNAs (A, B, C, F, G), has two reported functional IRES motifs [(5′-UUCCUUU-3′) and (5′-UAUCCAG-3′)] (Nicholson et al., 1991; Weingarten-Gabbay et al., 2016). Based on these observations, we assessed the translational potential of UGT8 circRNAs using CIRC B (which includes the exon 1c IRES motifs) and CIRC D (which lacks these IRES motifs) as described below.

We prepared three ZKSCAN1 CIRC B expression vectors to express the predicted CircP B in either an untagged (CIRC B, 274 aa) or FLAG-tagged CircP B (FLAG/CIRC B, FLAG/CIRC B/HA, 282 aa) form. The FLAG/CIRC B/HA vector was also predicted to express HA-tagged LinearP B (290 aa) from the linear transcript (Fig. 6B) as described in *Materials and Methods*. We transfected these three vectors in HEK293T cells and performed Western Blotting assays using an anti-FLAG or anti-HA antibody. As shown in Fig. 6C (left panel), FLAG-tagged CircP B was not observed in HEK293T cells transfected with any of the three vectors: CIRC B (lanes 4–5), FLAG/CIRC B (lanes 6–7), FLAG/CIRC B/HA (lanes 8–11); for positive control, HEK293T cells transfected with the FLAG-UGT2B15 vector produced the expected FLAG-tagged protein (50 kDa) (lane 2). By contrast, the HA-tagged LinearP B (29 kDa) was observed in HEK293T cells transfected with the FLAG/CIRC B/HA vector (lanes 8–10). This HA-tagged protein was not seen in HEK293T cells

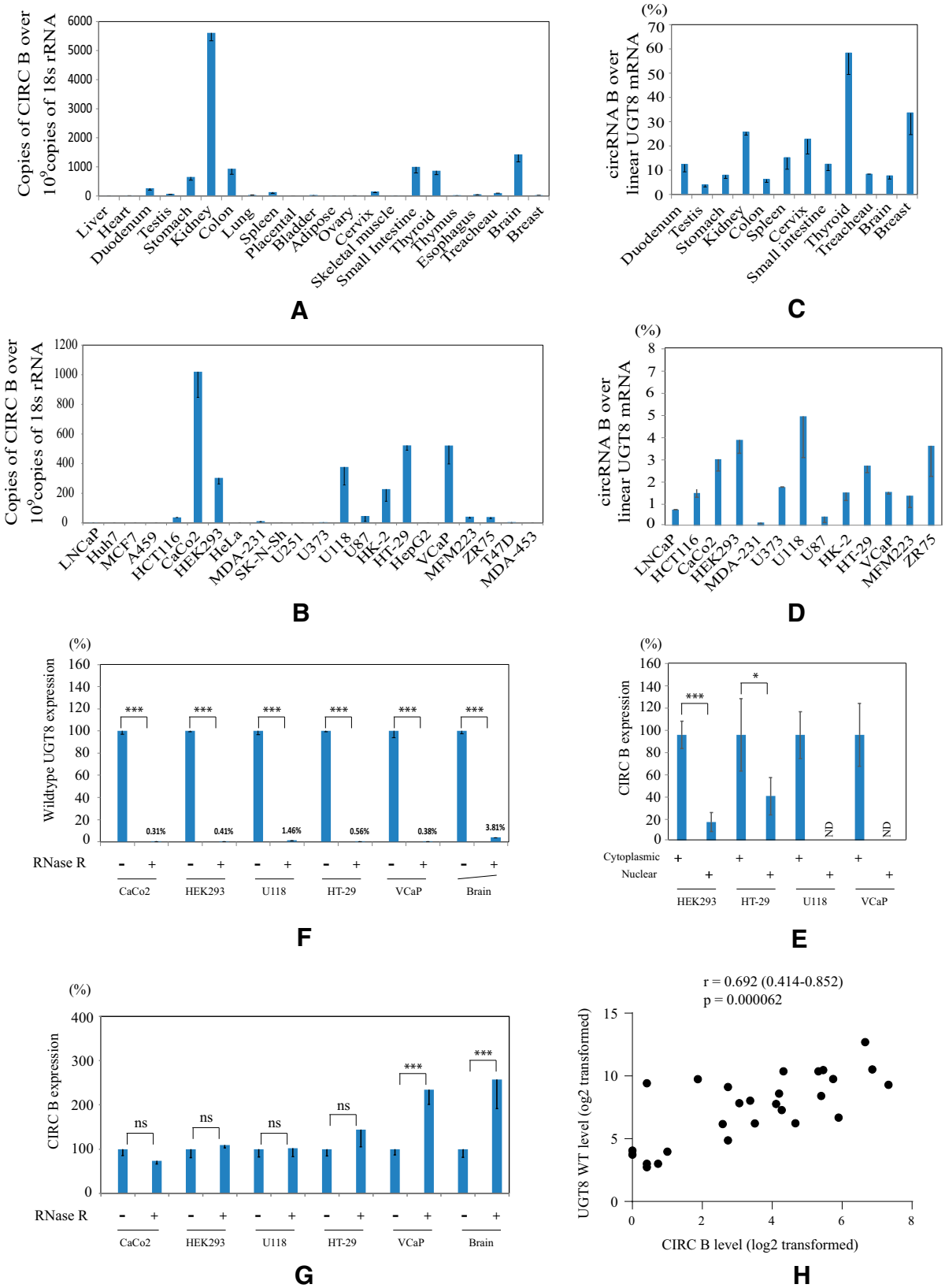


Fig. 4. Expression profile, cellular location, and RNase R resistance of UGT CIRC B. (A–D) RT-qPCR analysis of 22 human normal tissues (A) and 22 cancer cell lines (B) shows broad expression of UGT8 CIRC B. The CIRC B expression levels are presented as the CIRC B copies per 10^9 copies of 18S rRNA (A and B) or are presented relative to those of the canonical UGT8 mRNA (set as a value of 100%) (C and D). Data are the means ± 1 S.D. from three independent experiments. (F and G) RT-qPCR analysis of UGT8 mRNA and CIRC B in RNA samples with or without RNase R treatment. The expression levels of UGT8 mRNA (F) and CIRC B (G) in the RNase R–treated RNA samples are presented relative to those (set as a value of 100%) in RNA samples without RNase R treatment. Data are the means ± 1 S.D. from three independent experiments. (E) RT-qPCR analysis of CIRC B in cytoplasmic RNA and nuclear RNA fractions. Cytoplasmic RNA and nuclear RNA were extracted from cell lines as described in *Materials and Methods*. The

transfected with any other vector (lanes 1–2, 4–7) or in HEK293T cells transfected with an empty ZKSCAN1 vector (lane 3) or untransfected (lane 12). Taken together, these data strongly suggested that proteins can be translated from the linear transcripts that are produced by the CIRC B ZKSCAN1 vector, but not from the circRNAs.

The CIRC D vector generated both circRNA (CIRC D) and linear transcripts, but with a significantly higher circular/linear RNA expression ratio than the CIRC B vectors (Fig. 5, A and B). The CIRC D circRNA and linear transcripts were predicted to encode CircP D (451 aa) and LinearP D (421 aa), respectively (Fig. 7A). Both CircP D and LinearP D contain the common 420-aa N-terminal region (encoded by exons 2–5). HEK293T cells were transfected with the CIRC D vector or the empty ZKSCAN1 vector and analyzed by Western blotting assays with an anti-UGT8 antibody that recognizes the common region of CircP D and LinearP D. Proteins with molecular masses of about 42–45 kDa were detected only in cells transfected with the CIRC D vector (Fig. 7B, lane 2), consistent with the translation of either CircP D (451 aa) or LinearP D (421 aa). The endogenously expressed UGT8 protein (about 51 kDa) was detected in both transfected cells (lanes 1–2). The gel piece containing putative CircP D and/or LinearP D proteins was excised, digested with trypsin or elastase, and analyzed by mass spectrometry as described in *Materials and Methods*. Eight peptides were detected in trypsin-digested samples (GHHTVFLLSEGR, DIAPSNHYSLQR, NTGVYLISR, GMGILLEWK, YLSEDIANK, YNILPEK, MNLLQR, ELYEALVK), and three peptides were detected in elastase-digested samples (SFLVLPK, YNLLPEKS, SPLPE DLQR). All eleven peptides were derived from the common N-terminal region of CircP D and LinearP D. No peptides were identified that were derived from the CircP D–specific 31-aa C-terminal region, suggesting that the protein observed on immunoblots was translated from the linear transcript and not the circRNA.

***Alu* Elements in the *UGT8* Gene and Their Potential for circRNA Biogenesis.** circRNA formation is often mediated by complementary inverted *Alu* pairs flanking the circularized exons (Jeck et al., 2013). Using RepeatMasker (Chen, 2004; Tarailo-Graovac and Chen, 2009), 20 *Alu* elements were identified in *UGT8* introns (Fig. 8). The *AluSg* in intron 1b has 70%–83% similarity to nine downstream inverted *Alu* elements that are present in intron 2 (*AluSz*, *AluY*, *AluSx1*, *AluY*), intron 6a (*AluSx*, *AluSz*), intron 6c (*AluJr*), and intron 6d (*AluYb9*, *AluSp*). Five circRNAs (A, B, C, F, G) (Fig. 5) used the exon 1c splice acceptor site, and three others (D, E, H) used the exon 2 acceptor splice site. The *AluSg* is the only *Alu* element located upstream of exon 1c and exon 2, suggesting that the alternative formation of complementary inverted *Alu* pairs between *AluSg* and nine downstream *Alu* elements might be involved in UGT8 circRNA biogenesis. However, which of the downstream *Alu* elements forms a complementary inverted *Alu* pair with the *AluSg* in the formation of a specific circRNA remains to be determined.

The miRNA-Sponging Potential of UGT8 circRNAs. circRNAs with many binding sites for a single miRNA have been reported to act as miRNA sponges, including*ciRS-7*, Sry, and circZNF91 (Hansen et al., 2013; Kristensen et al., 2018b). Using UGT8 circular RNAs as examples, we assessed whether UGT circRNAs might act as miRNA sponges. Custom analysis using miRDB (http://mirdb.org) (Chen and Wang, 2020) identified potential miRNA binding sites in six (E1c, E2, E3, E5, E6v, E6b) of the 10 exons that are present within UGT8 circRNAs (Supplemental Fig. 19A). Exon 2 (824 nt), which is present within all eight circRNAs, has predicted binding sites for 44 miRNAs. Most of these miRNAs have a single binding site; however, three miRNAs (miR-214-3p, miR-761, miR-3619-5p) that share the same seed sequence (5′cagcagg3′) each have three predicted seed binding sites (5′ccugcug3′) in exon 2.

To test the binding of these miRNAs to UGT8 circRNAs, we generated a luciferase reporter in which the UGT8 CIRC D sequence (exons 2–5, 1264 nt) was inserted downstream of the luciferase gene in the pGL3 vector. Breast cancer MDA-MB-453 cells were cotransfected with the pGL3/CIRC D reporter vector and mimics corresponding to miR-214-3p, miR-761, miR-3619-5p, or a negative control miRNA. Both miR-214-3p and miR-3619-5p mimics significantly reduced the activity of the pGL3/CIRC D reporter, with miR-214-3p having the largest effect (Supplemental Fig. 19B). We also examined previously published PARCLIP data to determine whether Argonaute proteins, core components of (mi)RNA-induced silencing complexes, bind to the exon 2 region of UGT8 transcripts. Argonaute proteins were reported to bind to UGT8 exon 2 within the region spanning the three predicted miR-214-3p binding sites (data not shown) (Kishore et al., 2011; Memczak et al., 2013). Based on these findings and the fact that all eight UGT8 circRNAs contain exon 2, we hypothesize that they may jointly sponge same miRNAs (e.g., miR-214-3p). This hypothesis awaits further investigation.

Discussion

The UGT transcriptome comprises of 22 canonical transcripts coding for functional enzymes and approximately 150 alternatively spliced and chimeric variant transcripts (Girard et al., 2007; Levesque et al., 2007; Bushey and Lazarus, 2012; Bushey et al., 2013; Tourancheau et al., 2016; Hu et al., 2018, 2019). The present study identified 85 UGT circRNAs from circRNA databases and 13 UGT circRNAs using divergent RT-PCR (including seven not reported by circRNA databases), thus greatly expanding the complexity and diversity of the UGT transcriptome. UGT circRNAs were detected in a wide range of human normal and cancerous tissues and cell lines; however, most of them were expressed at very low levels in tissue- and cell-specific contexts. Many alternatively spliced (Bellemare et al., 2010a,b; Bushey and Lazarus, 2012; Bushey et al., 2013; Tourancheau et al., 2016) and chimeric UGT

CIRC B expression levels in nuclear RNA samples are presented relative to those in corresponding cytoplasmic RNA samples (set as a value of 100%). Data are the means ± 1 S.D. from three independent experiments. Two-way analysis of variance and Bonferroni's multiple comparison tests. A P value of less than 0.05 was considered statistically significant. *$P < 0.05$; ***$P < 0.001$; ND, not detected; ns, not significant. (H) Spearman's correlation analysis (GraphPad Prism 8.0) between UGT8 mRNA and UGT8 CIRC B expression from 27 tissues and cell lines that expressed both UGT8 mRNA and CIRC B (A and B). The Spearman coefficient r and 95% confidence interval (CI) are given. A P value of less than 0.05 was considered statistically significant.

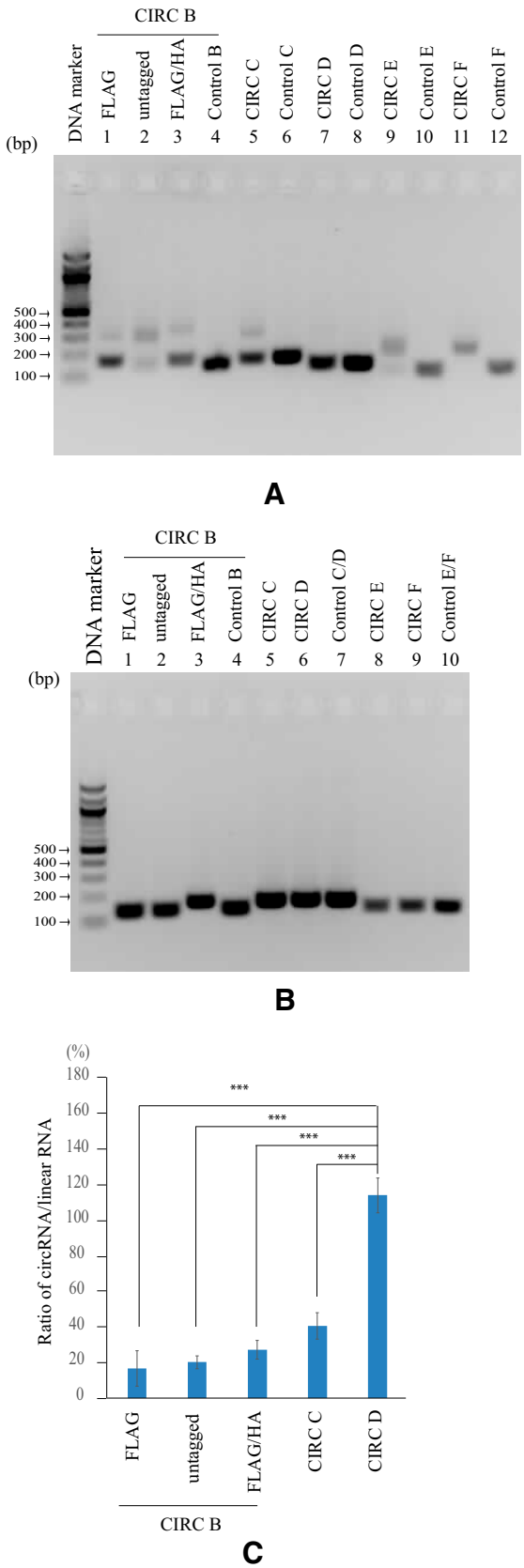


Fig. 5. RT-qPCR analysis of circRNAs and linear transcripts generated from UGT8 circRNA expression vectors. ZKSCAN1 UGT8 circRNA expression vectors were transfected into HEK293T cells, RNA was extracted, and RT-qPCR was performed using primers that specifically quantify either circRNAs or linear transcripts. (A) Agarose gel separation of circRNA-derived PCR amplicons from HEK293T cells transfected with the circRNA vectors FLAG/CIRC B (lane 1), CIRC B (lane 2), FLAG/CIRC B/HA (lane 3), CIRC C (lane 5), CIRC D (lane 7), CIRC E (lane 9), or CIRC F (lane 11). Reference standard amplicons were generated using pCR Blunt vectors containing the BSJs of the respective circRNAs as templates; these standards are designated as Controls B (lane 4), C (lane 6), D (lane 8), E (lane 10), and F (lane 12). (B) Agarose gel separation of linear transcript–derived PCR amplicons from HEK293T cells transfected with the circRNA vectors described in (A). Reference standard amplicons were generated using the respective ZKSCAN1 circRNA expression vectors as templates; these standards are designated as Controls B (lane 4), C/D (lane 7), and E/F (lane 10). (C) circRNA and linear transcript levels were quantified in HEK283T cells transfected with various circRNA vectors using RT-qPCR; data are presented as a ratio of circRNA to linear RNA (the latter set at 100%). Data are the means ± 1 S.D. from three independent experiments. One-way analysis of variance and Tukey's multiple comparison tests. A P value of less than 0.05 was considered statistically significant. ***$P < 0.001$.

transcripts (Hu et al., 2018) code for variant proteins that inhibit the activity of wild-type enzymes through protein-to-protein interactions. As discussed in detail below, we found no evidence for translation of UGT circRNAs into CircPs that might negatively regulate glucuronidation in a similar mechanism. However, we have provided preliminary evidence that UGT8 circRNAs may function as regulatory ncRNAs by sponging specific miRNAs. Notably, one of the miRNAs that appeared to bind to UGT8 circRNAs, miR-214-3p, is regarded as a key hub that controls tumor proliferation, stemness, angiogenesis, invasiveness, extravasation, metastasis, and chemotherapy resistance (Penna et al., 2015). High miRNA-214-3p expression is associated with unfavorable survival in patients with TNBC (Kalniete et al., 2015). These data together with our preliminary observation of high expression of UGT8 CIRC B in TNBC suggest that UGT8 circRNAs may impact survival of patients with TNBC through sponging miRNAs such as miR-214-3p. This hypothesis awaits further investigation.

All *UGT* genes except for *UGT2B28* generate circRNAs (Supplemental Table 2). The diversity of UGT circRNAs is expanded by both alternative backsplicing and forward-splicing events (Zhang et al., 2016). Alternative 5′ backsplicing occurs when multiple downstream donor splice sites are backspliced to the same upstream acceptor splice site. The eight UGT8 circRNAs identified by the present study were all generated by this mechanism, including five circRNAs (A, B, C, F, G) using the same exon 1c acceptor splice site and three circRNAs (D, E, H) using the same exon 2 acceptor splice site (Fig. 3A). Alternative 3′ backsplicing occurs when the same downstream donor splice site is backspliced to multiple upstream acceptor splice sites. The donor splice sites of three *UGT8* exons (E2, E5, E6v) (Fig. 3A), *UGT1A* exon 4 (Fig. 1A), and *UGT2B4* exon 4 (Supplemental Fig. 16A) were all backspliced to two different upstream acceptor splice sites, generating two circRNAs via this mechanism. The alternative backsplicing of the *UGT8* pre-mRNA may be partly explained by the alternative pairing between the upstream *AluSg* element and different downstream inverted *Alu* elements (Fig. 8). Finally, *UGT* genes also produce circRNAs via alternative forward-splicing events, resulting in multiple circRNAs with the same BSJ but a unique set of central exons. For example, UGT8 circRNAs A and B have the same BSJ (E2/E1c), but the former had an additional alternative exon 1d (Fig. 3B).

In almost all exonic circRNAs, only internal exons are subject to backsplicing, with the first and the last exons excluded because of the lack of acceptor and donor splice sites, respectively (Zhang et al., 2014, Wilusz, 2015). We found that the

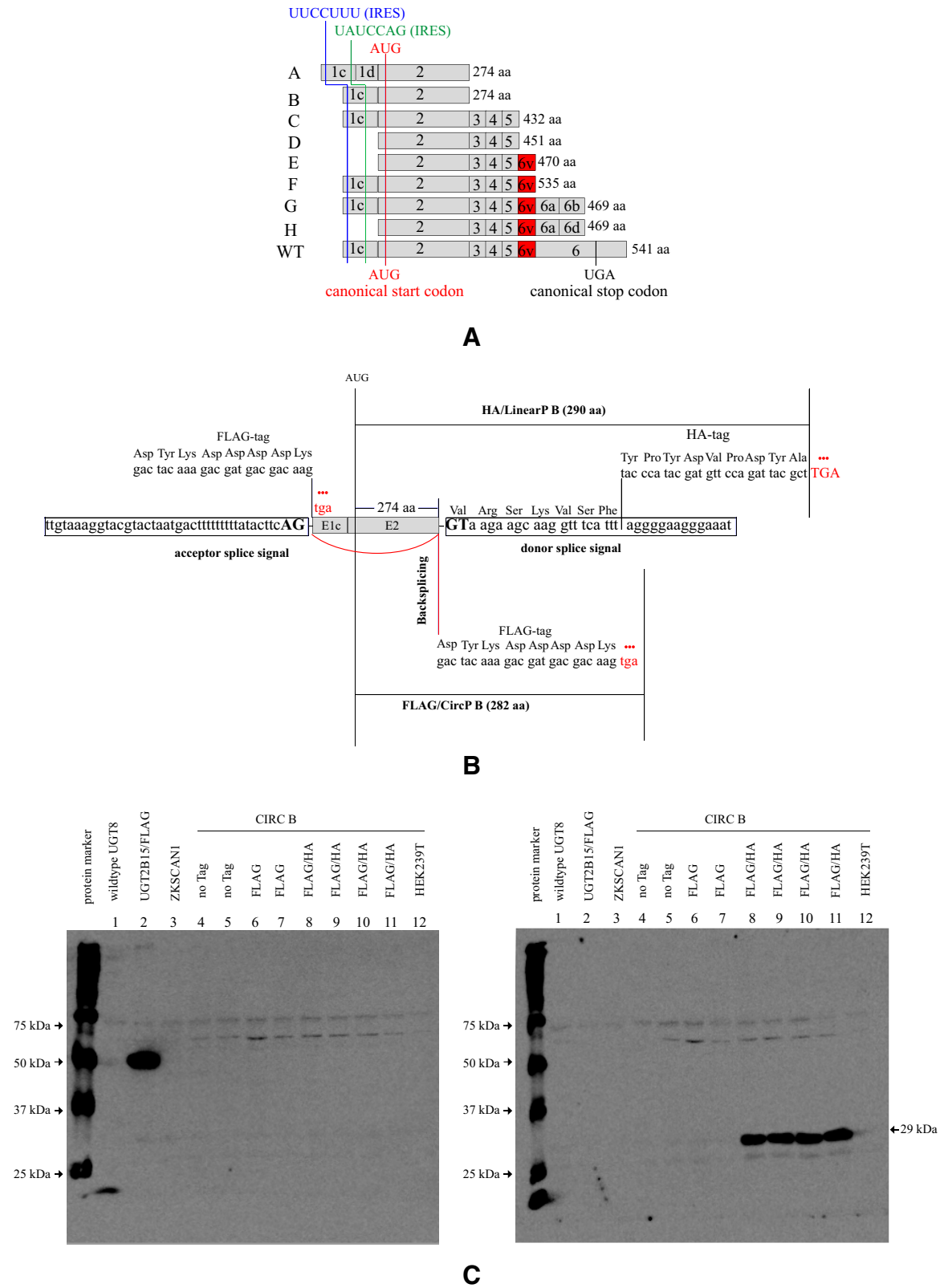
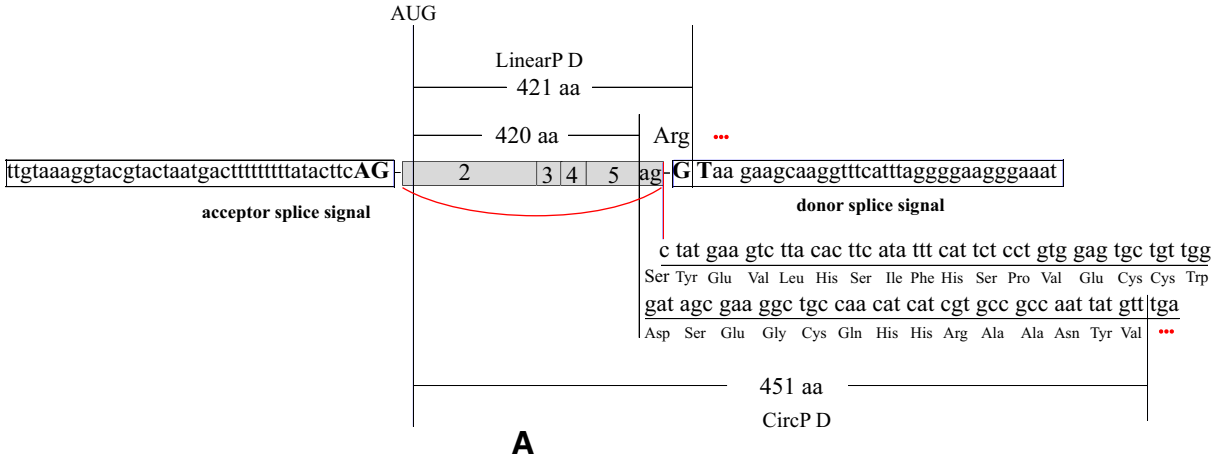
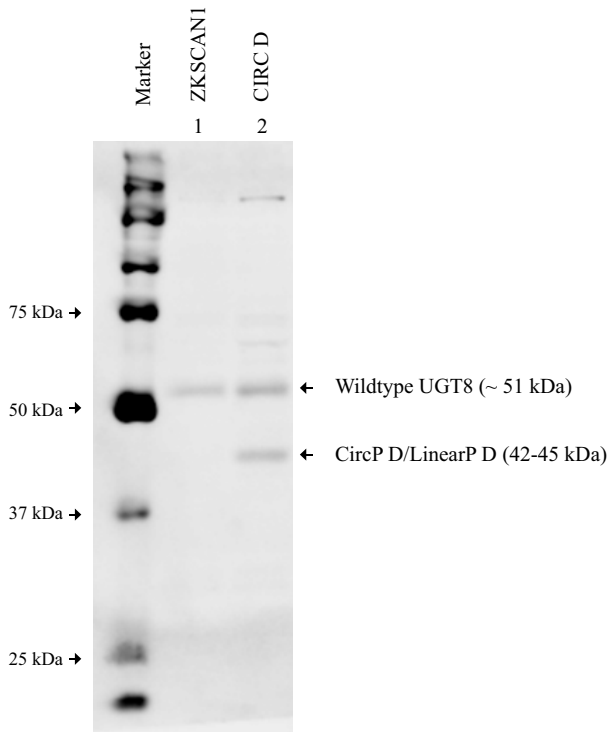


Fig. 6. Prediction of ORFs for UGT8 circRNAs and assessment of circRNA B (CIRC B) translational potential. (A) A schematic diagram shows that all eight UGT8 circRNAs have exon 2 with the canonical start AUG codon and encode for variant UGT8 proteins with the common 274-aa N-terminal region encoded by exon 2. Two IRES motifs (5′-UUCCUUU-3′, 5′-UAUCCAG-3′) are present upstream of the start AUG in five circ RNAs (A, B, C, F, G). (B) A schematic diagram shows the locations of FLAG and HA tags inserted in the ZKSCAN1 circRNA B (CIRC B) vector. In backspliced circRNAs, the CIRC B ORF is placed in frame with the FLAG sequence, encoding the predicted FLAG/CircP B protein (282 aa). In linear transcripts, the CIRC B ORF is contiguous with the first 21 nucleotides of the donor splice signal and in frame with the HA-tag sequence, encoding the predicted HA/LinearP B protein (290 aa). (C) Immunoblotting assays were performed with anti-FLAG (left panel) or anti-HA



A



B

Fig. 7. Prediction of ORF for UGT8 circRNA D and assessment of translational potential. (A) The UGT8 CIRC D vector contains UGT8 exons E2-5 cloned between the acceptor and donor splice sites of the ZKSCAN1 MCS plasmid. This vector generates both circRNA and linear transcripts that are predicted to encode CircP D (451 aa) and LinearP D (421 aa), respectively. CircP D and LinearP D share the 420-aa N-terminal region encoded by exons 2–5. Because of the backsplicing of exon 5 to exon 2, the 3′ end of the ORF within circRNA D extends into the 5′ end of exon 2 in an alternate frame and reaches a stop codon at nucleotides 92–94 (TGA); hence, this predicted protein (CircP D) contains a novel 31-aa C-terminal peptide. The ORF within the linear transcript reaches a stop codon at the first nucleotide (G) of the donor splice signal; hence, this predicted protein (LinearP D) has a single Arg residue appended at the C-terminal end. (B) HEK293T cells were transfected with the empty ZKSCAN1 vector (lane 1) or the CIRC D vector (lane 2). At 48 hours post-transfection, whole-cell lysates were prepared from transfected cells and subjected to immunoblotting assays using an anti-UGT8 antibody targeting the common region of CircP D and LinearP D.

first 317 nt (E5a-v) of the last exon (E5a) of the *UGT1A* gene was backspliced into circ-UGT1A-33 (Fig. 1B) through a cryptic exonic donor splice site (Supplemental Fig. 17). Similarly, the first 146 nt (E6v) of the last exon (E6) of the *UGT8* gene was also backspliced into circ_UGT8-8 and circ_UGT8-22 via a cryptic exonic donor splice site (Fig. 3B). The cryptic donor splice sites for both exon E5a-v and E6v have the conserved splicing signal dinucleotide GT, which conforms to the AG/

(right panel) antibodies using lysates of HEK293T cells transfected with 1) two independent vectors that encode untagged CircP B (lanes 4–5) or FLAG/CircP B (6–7) and 2) four independent vectors (lanes 8–11) that encode both FLAG/CircP B and HA/LinearP B. As positive controls, HEK293T cells were transfected with a vector expressing FLAG/UGT2B15 (lane 2) or the canonical UGT8 cDNA (lane 1); as negative controls, HEK293T cells were transfected with the empty ZKSCAN1 vector (lane 3) or untransfected (lane 12).

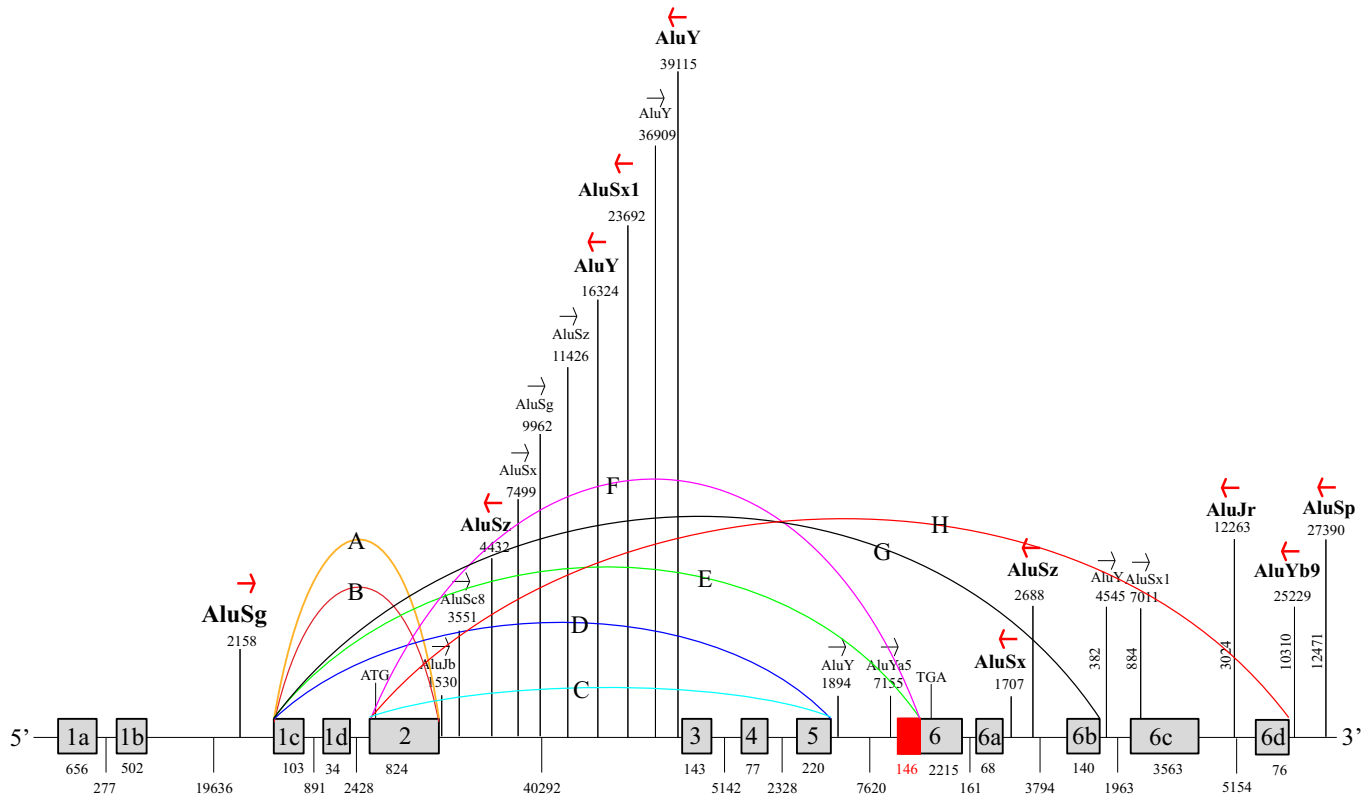


Fig. 8. The Alu elements within the *UGT8* gene may facilitate circRNA biogenesis. A schematic diagram shows the locations of 20 *Alu* elements in the *UGT8* gene and the backsplicing events (A–H) that generate eight circRNAs identified in the present study. The *AluSg* element has high sequence similarity to the nine inverted *Alu* elements (indicated by red arrows) located downstream of exon 2 and may pair with them to facilitate circRNA biogenesis.

GT splicing rule. Taken together, our data indicate that a terminal exon (the first or last exons of known RefSeq genes) can be backspliced into circRNAs through the use of exonic cryptic splice sites.

Exonic circRNAs have thousands of so-called circRNA-predominant cassette exons that are very rarely present in the host mRNAs (Zhang et al., 2016). These exons are generally located in the 5′ and 3′ flanking region of known RefSeq genes. We found five such exons (E1d, E6a, E6b, E6c, E6d) (Fig. 3) within the *UGT8* gene locus that are incorporated into circRNAs but not into any of the five known *UGT8* transcripts (Hu et al., 2019). All five exons are short [E1d (34 nt), E6v (146 nt), E6a (68 nt), E6b (140 nt), E6cv (80 nt), and E6d (76 nt)] and have conserved acceptor and donor splice sites that conform to the splicing AG/GT rule (Supplemental Fig. 20). The *UGT8* gene (RefSeq, NM_001128174) spans ?79 kb of genomic DNA, including three 5′ untranslated exons (1a/1b/1c) and five coding exons (2/3/4/5/6) (Hu et al., 2019). The discovery of five additional exons (one upstream and four downstream of the coding exons) extends this gene approximately 15 kb in the 3′ direction (Fig. 3).

Approximately 16,000 human circRNAs contain ORFs of >100 amino acids that might encode proteins, and nearly half of them have a predicted IRES motif that could drive cap-independent circRNA translation (Chen et al., 2016). However, to date, there is firm evidence for in vivo translation of only several endogenously expressed circRNAs (Wilusz, 2018). AUG circRNAs are a group of circRNAs that contain the canonical translational start AUG codon of the host protein-coding mRNAs. A recent study used ribosome profiling, proteomic analysis, and heterologous circRNA expression to comprehensively assess AUG circRNAs and found no evidence for translation of AUG circRNAs (Stagsted et al., 2019). All eight UGT8 circRNAs (A–H) are AUG circRNAs that encode predicted proteins with the common 274-aa N-terminal region of wild-type UGT8 protein (Fig. 6A). Moreover, five circRNAs (A, B, C, F, G) contain exons 1c, which bears known IRES motifs (5′-UUCCUUU-3′; 5′-UAUCCAG-3′) (Nicholson et al., 1991; Weingarten-Gabbay et al., 2016). We assessed the translational potential of two UGT8 circRNAs (B, D) in a heterologous circRNA expression system (Kramer et al., 2015) using a combination of epitope tagging, immunoblotting, and mass spectrometry analysis. Our results did not provide any evidence for the translation of these two circRNAs, suggesting either that translation does not occur or that its efficiency in this system is below the threshold for detection. Indeed, a previous study suggested that circRNAs may be translated much less efficiently (by ?100-fold) than corresponding linear transcripts in heterologous expression systems (Legnini et al., 2017).

Both the synthesis and translation of circRNAs may be enhanced in tissue-specific contexts by the expression of *trans*-acting factors that promote backsplicing or cap-independent translation events (Wilusz, 2018). Notably, the *UGT8* gene is widely expressed in human tissues (Hu et al., 2019), especially in glial cells in the brain and spinal cord, where it is essential for myelination during development. Overall, although our results are broadly consistent with the report that AUG circRNAs are not translated (Stagsted et al., 2019), the possibility that UGT8 CircPs are generated in specific cellular/developmental contexts warrants further investigation. Production of an antibody with specificity for the novel C-terminal region of the predicted CircP D protein could allow interrogation of endogenous

UGT8 circRNA translation in brain and other tissues during development.

The study of circRNA function continues to be hampered by the limitations of commonly used experimental tools. Loss-of-function studies using siRNA require targeting of the BSJ, with limited opportunity to optimize the siRNA sequence for efficiency or specificity. For example, an siRNA (GACUUCAUAGCUGGGAUUAUU) that we designed to target the BSJ of UGT8 circRNA D did not significantly reduce circRNA levels (data not shown). Gain-of-function studies are limited by the lack of appropriate overexpression vectors that express circRNAs with high efficiency and specificity. As discussed previously (Kramer et al., 2015), almost all reported circRNA expression vectors generated both backspliced circRNAs and unbackspliced linear transcripts, with the circular/linear RNA expression ratio often less than 20% (Hansen et al., 2013; Ashwal-Fluss et al., 2014; Zhang et al., 2014; Starke et al., 2015). Some circRNA expression vectors may also produce *trans*-spliced linear transcripts (Ho-Xuan et al., 2020). Given that linear transcripts produced from circRNA expression vectors generally contain the full sequence as the expected circRNAs, they are likely to have similar biologic functions, including miRNA sponging and translation. As an example, the seven UGT8 circRNA vectors generated in the ZKSCAN1 MCS vector (Kramer et al., 2015) produced both circRNAs and linear transcripts, with only one (CIRC D) producing more than 50% circRNAs (Fig. 5). Moreover, the linear transcripts generated from the UGT8 circRNA expression vectors were robustly translated. Overall, our findings suggest the need for caution when investigating circRNA functions (including translation) using circRNA expression vectors, particularly when there is no evident strategy to distinguish the activities of the linear transcripts and circRNAs. Improving the efficiency and specificity of circRNA synthesis from expression vectors is an important future direction, and a recent report suggests that intron-mediated enhancement may be one strategy to achieve this (Mo et al., 2019).

In conclusion, our discovery of nearly 100 UGT circRNAs greatly expands the complexity and diversity of the UGT transcriptome. With the exception of *UGT8*-derived circRNAs, most UGT circRNAs appear to be expressed at very low levels in tissue- and cell-specific contexts, although the possibility of heterogeneity at the single-cell level, and/or upregulation in response to specific signals or developmental events, remains to be assessed. The biologic functions of UGT circRNAs are yet to be determined, with the present studies suggesting significant limitations of currently applied gain- and loss-of-function approaches and reaffirming a growing consensus that improved circRNA expression systems with a high specificity for circRNA synthesis are necessary for functional characterization.

Acknowledgments

The authors acknowledge Drs. Alex Colella and Timonthy Chataway (Flinders Proteomics Facility, Flinders University, Australia) for conceptual and technical support of mass spectrometry assays.

Authorship Contributions

Participated in research design: Hu, Mackenzie, Hulin, McKinnon, Meech.

Conducted experiments: Hu.

Performed data analysis: Hu, Meech.

Wrote or contributed to the writing of the manuscript: Hu, Mackenzie, Hulin, McKinnon, Meech.

References

Abdelmohsen K, Panda AC, Munk R, Grammatikakis I, Dudekula DB, De S, Kim J, Noh JH, Kim KM, Martindale JL et al. (2017) Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol* **14**:361–369.

Abe N, Matsumoto K, Nishihara M, Nakano Y, Shibata A, Maruyama H, Shuto S, Matsuda A, Yoshida M, Ito Y et al. (2015) Rolling circle translation of circular RNA in living human cells. *Sci Rep* **5**:16435.

Aktaş T, Avşar Ilōk İ, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, and Akhtar A (2017) DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**:115–119.

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, and Kadener S (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**:55–66.

Barrett SP and Salzman J (2016) Circular RNAs: analysis, expression and potential functions. *Development* **143**:1838–1847.

Basu NK, Kole L, Basu M, Chakraborty K, Mitra PS, and Owens IS (2008) The major chemical-detoxifying system of UDP-glucuronosyltransferases requires regulated phosphorylation supported by protein kinase C. *J Biol Chem* **283**:23048–23061.

Basu NK, Kovarova M, Garza A, Kubota S, Saha T, Mitra PS, Banerjee R, Rivera J, and Owens IS (2005) Phosphorylation of a UDP-glucuronosyltransferase regulates substrate specificity. *Proc Natl Acad Sci USA* **102**:6285–6290.

Bellemare J, Rouleau M, Girard H, Harvey M, and Guillemette C (2010a) Alternatively spliced products of the UGT1A gene interact with the enzymatically active proteins to inhibit glucuronosyltransferase activity in vitro. *Drug Metab Dispos* **38**:1785–1789.

Bellemare J, Rouleau M, Harvey M, and Guillemette C (2010b) Modulation of the human glucuronosyltransferase UGT1A pathway by splice isoform polypeptides is mediated through protein-protein interactions. *J Biol Chem* **285**:3600–3607.

Burset M, Seledtsov IA, and Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* **28**:4364–4375.

Bushey RT, Dluzen DF, and Lazarus P (2013) Importance of UDP-glucuronosyltransferases 2A2 and 2A3 in tobacco carcinogen metabolism. *Drug Metab Dispos* **41**:170–179.

Bushey RT and Lazarus P (2012) Identification and functional characterization of a novel UDP-glucuronosyltransferase 2A1 splice variant: potential importance in tobacco-related cancer susceptibility. *J Pharmacol Exp Ther* **343**:712–724.

Cao Q, Chen X, Wu X, Liao R, Huang P, Tan Y, Wang L, Ren G, Huang J, and Dong C (2018) Inhibition of UGT8 suppresses basal-like breast cancer progression by attenuating sulfatide-αVβ5 axis. *J Exp Med* **215**:1679–1692.

Chen N (2004) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*

Chen X, Han P, Zhou T, Guo X, Song X, and Li Y (2016) circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* **6**:34985.

Chen Y and Wang X (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* **48** (D1):D127–D131.

Cheng J, Metge F, and Dieterich C (2016) Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* **32**:1094–1096.

Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, and Goodall GJ (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**:1125–1134.

Dluzen DF, Sun D, Salzberg AC, Jones N, Bushey RT, Robertson GP, and Lazarus P (2014) Regulation of UDP-glucuronosyltransferase 1A1 expression and activity by microRNA 491-3p. *J Pharmacol Exp Ther* **348**:465–477.

Dong R, Ma XK, Li GW, and Yang L (2018) CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinformatics* **16**:226–233.

Errichelli L, Dini Modigliani S, Laneve P, Colantoni A, Legnini I, Capauto D, Rosa A, De Santis R, Scarfò R, Peruzzi G et al. (2017) FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat Commun* **8**:14741.

Gao Y, Zhang J, and Zhao F (2018) Circular RNA identification based on multiple seed matching. *Brief Bioinform* **19**:803–810.

Girard H, Lévesque E, Bellemare J, Journault K, Caillier B, and Guillemette C (2007) Genetic diversity at the UGT1 locus is amplified by a novel 3′ alternative splicing mechanism leading to nine additional UGT1A proteins that act as regulators of glucuronidation activity. *Pharmacogenet Genomics* **17**:1077–1089.

Glažar P, Papavasileiou P, and Rajewsky N (2014) circBase: a database for circular RNAs. *RNA* **20**:1666–1670.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, and Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* **495**:384–388.

Ho-Xuan H, Glažar P, Latini C, Heizler K, Haase J, Hett R, Anders M, Weichmann F, Bruckmann A, Van den Berg D et al. (2020) Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Res* **48**:10368–10382.

Hu DG, Hulin JA, Wijayakumara DD, McKinnon RA, Mackenzie PI, and Meech R (2018) Intergenic splicing between four adjacent *UGT* genes (*2B15, 2B29P2, 2B17, 2B29P1*) gives rise to variant UGT proteins that inhibit glucuronidation via protein-protein interactions. *Mol Pharmacol* **94**:938–952.

Hu DG, Hulin JU, Nair PC, Haines AZ, McKinnon RA, Mackenzie PI, and Meech R (2019) The UGTome: the expanding diversity of UDP glycosyltransferases and its impact on small molecule metabolism. *Pharmacol Ther* **204**:107414.

Hu DG, Mackenzie PI, Lu L, Meech R, and McKinnon RA (2015) Induction of human UDP-glucuronosyltransferase 2B7 gene expression by cytotoxic anticancer drugs in liver cancer HepG2 cells. *Drug Metab Dispos* **43**:660–668.

Hu DG, Meech R, McKinnon RA, and Mackenzie PI (2014a) Transcriptional regulation of human UDP-glucuronosyltransferase genes. *Drug Metab Rev* **46**:421–458.

Hu DG, Rogers A, and Mackenzie PI (2014b) Epirubicin upregulates UDP glucuronosyltransferase 2B7 expression in liver cancer cells via the p53 pathway. *Mol Pharmacol* **85**:887–897.

Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C et al. (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* **10**:170–177.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, and Sharpless NE (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**:141–157.

Kalniete D, Nakazawa-Miklaševiča M, Štrumfa I, Āboliņš A, Irmejs A, Gardovskis J, and Miklaševičs E (2015) High expression of miR-214 is associated with a worse disease-specific survival of the triple-negative breast cancer patients. *Hered Cancer Clin Pract* **13**:7.

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, and Zavolan M (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8**:559–564.

Kramer MC, Liang D, Tatomer DC, Gold B, March ZM, Cherry S, and Wilusz JE (2015) Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev* **29**:2168–2182.

Kristensen LS, Hansen TB, Venø MT, and Kjems J (2018a) Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* **37**:555–565.

Kristensen LS, Okholm TLH, Venø MT, and Kjems J (2018b) Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. *RNA Biol* **15**:280–291.

Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatica A, Santini T, Andronache A, Wade M et al. (2017) Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* **66**:22–37.e9.

Lei M, Zheng G, Ning Q, Zheng J, and Dong D (2020) Translation and functional roles of circular RNAs in human cancer. *Mol Cancer* **19**:30.

Lévesque E, Girard H, Journault K, Lépine J, and Guillemette C (2007) Regulation of the UGT1A1 bilirubin-conjugating pathway: role of a new splicing event at the UGT1A locus. *Hepatology* **45**:128–138.

Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, Zheng Q, Li Y, Wang P, He X et al. (2018) exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res* **46** (D1):D106–D112.

Liang WC, Wong CW, Liang PP, Shi M, Cao Y, Rao ST, Tsui SK, Waye MM, Zhang Q, Fu WM et al. (2019) Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol* **20**:84.

Liu YC, Li JR, Sun CH, Andrews E, Chao RF, Lin FM, Weng SL, Hsu SD, Huang CC, Cheng C et al. (2016) CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* **44** (D1):D209–D215.

Lu Y, Heydel JM, Li X, Bratton S, Lindblom T, and Radominska-Pandya A (2005) Lithocholic acid decreases expression of UGT2B7 in Caco-2 cells: a potential role for a negative farnesoid X receptor response element. *Drug Metab Dispos* **33**:937–946.

Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, and Nebert DW (2005) Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics* **15**:677–685.

MacKenzie PI, Rogers A, Elliot DJ, Chau N, Hulin JA, Miners JO, and Meech R (2011) The novel UDP glycosyltransferase 3A2: cloning, catalytic properties, and tissue distribution. *Mol Pharmacol* **79**:472–478.

Mackenzie PI, Rogers A, Treloar J, Jorgensen BR, Miners JO, and Meech R (2008) Identification of UDP glycosyltransferase 3A1 as a UDP N-acetylglucosaminyltransferase. *J Biol Chem* **283**:36205–36210.

Margaillan G, Levesque E and Guillemette C (2016) Epigenetic regulation of steroid inactivating UDP-glucuronosyltransferases by microRNAs in prostate cancer. *J Steroid Biochem Mol Biol* **155**: 85-93.

Meech R, Hu DG, McKinnon RA, Mubarokah SN, Haines AZ, Nair PC, Rowland A, and Mackenzie PI (2019) The UDP-glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. *Physiol Rev* **99**:1153–1222.

Meech R, Mubarokah N, Shivasami A, Rogers A, Nair PC, Hu DG, McKinnon RA, and Mackenzie PI (2015) A novel function for UDP glycosyltransferase 8: galactosidation of bile acids. *Mol Pharmacol* **87**:442–450.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**:333–338.

Michel MC, Murphy TJ, and Motulsky HJ (2020) New author guidelines for displaying data and reporting data analysis and statistical methods in experimental biology. *Mol Pharmacol* **97**:49–60.

Mo D, Li X, Raabe CA, Cui D, Vollmar JF, Rozhdestvensky TS, Skryabin BV, and Brosius J (2019) A universal approach to investigate circRNA protein coding function. *Sci Rep* **9**:11684.

Nicholson R, Pelletier J, Le SY, and Sonenberg N (1991) Structural and functional analysis of the ribosome landing pad of poliovirus type 2: in vivo translation studies. *J Virol* **65**:5886–5894.

Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E et al. (2017) Translation of CircRNAs. *Mol Cell* **66**:9–21.e7.

Pandey PR, Rout PK, Das A, Gorospe M, and Panda AC (2019) RPAD (RNase R treatment, polyadenylation, and poly(A)+ RNA depletion) method to isolate highly pure circular RNA. *Methods* **155**:41–48.

Papageorgiou I and Court MH (2017) Identification and validation of microRNAs directly regulating the UDP-glucuronosyltransferase 1A subfamily enzymes by a functional genomics approach. *Biochem Pharmacol* **137**:93–106.

Papageorgiou I, Freytsis M, and Court MH (2016) Transcriptome association analysis identifies miR-375 as a major determinant of variable acetaminophen glucuronidation by human liver. *Biochem Pharmacol* **117**:78–87.

Penna E, Orso F, and Taverna D (2015) miR-214 as a key hub that controls cancer networks: small player, multiple functions. *J Invest Dermatol* **135**:960–969.

Ruan H, Deng X, Dong L, Yang D, Xu Y, Peng H, and Guan M (2019a) Circular RNA circ_0002138 is down-regulated and suppresses cell proliferation in colorectal cancer. *Biomed Pharmacother* **111**:1022–1028.

Ruan H, Xiang Y, Ko J, Li S, Jing Y, Zhu X, Ye Y, Zhang Z, Mills T, Feng J et al. (2019b) Comprehensive characterization of circular RNAs in ~1000 human cancer cell lines. *Genome Med* **11**:55.

Sneitz N, Court MH, Zhang X, Laajanen K, Yee KK, Dalton P, Ding X, and Finel M (2009) Human UDP-glucuronosyltransferase UGT2A2: cDNA construction, expression, and functional characterization in comparison with UGT2A1 and UGT2A3. *Pharmacogenet Genomics* **19**:923–934.

Stagsted LV, Nielsen KM, Daugaard I, and Hansen TB (2019) Noncoding AUG circ RNAs constitute an abundant and conserved subclass of circles. *Life Sci Alliance* **2**:e201900398.

Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, and Bindereif A (2015) Exon circularization requires canonical splice signals. *Cell Rep* **10**:103–111.

Tarailo-Graovac M and Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*

Tourancheau A, Margaillan G, Rouleau M, Gilbert I, Villeneuve L, Lévesque E, Droit A, and Guillemette C (2016) Unravelling the transcriptomic landscape of the major phase II UDP-glucuronosyltransferase drug metabolizing pathway using targeted RNA sequencing. *Pharmacogenomics J* **16**:60–70.

Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, Wu YM, Dhanasekaran SM, Engelke CG, Cao X et al. (2019) The landscape of circular RNA in cancer. *Cell* **176**:869–881.e13.

Vromman M, Vandesompele J, and Volders PJ (2021) Closing the circle: current state and perspectives of circular RNA databases. *Brief Bioinform* **22**:288–297.

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM et al. (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* **38**:e178.

Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N, Yakhini Z, Weinberger A, and Segal E (2016) Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**:aad4939.

Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, and Lai EC (2014) Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* **9**:1966–1980.

Wijayakumara DD, Hu DG, Meech R, McKinnon RA, and Mackenzie PI (2015) Regulation of human UGT2B15 and UGT2B17 by miR-376c in prostate cancer cell lines. *J Pharmacol Exp Ther* **354**:417–425.

Wijayakumara DD, Mackenzie PI, McKinnon RA, Hu DG, and Meech R (2017) Regulation of UDP-glucuronosyltransferases UGT2B4 and UGT2B7 by microRNAs in liver cancer cells. *J Pharmacol Exp Ther* **361**:386–397.

Wilusz JE (2015) Repetitive elements regulate circular RNA biogenesis. *Mob Genet Elements* **5**:1–7.

Wilusz JE (2018) A 360° view of circular RNAs: from biogenesis to functions. *Wiley Interdiscip Rev RNA* **9**:e1478.

Wu W, Ji P, and Zhao F (2020) CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* **21**:101.

Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, Jin Y, Gao Y, Xia L, Chang H et al. (2018) CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* **46** (D1):D925–D929.

Xia S, Feng J, Lei L, Hu J, Xia L, Wang J, Xiang Y, Liu L, Zhong S, Han L et al. (2017) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform* **18**:984–992.

Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, Huang N, Yang X, Zhao K, Zhou H et al. (2018) Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. *J Natl Cancer Inst* **110**:304–315.

You X and Conrad TO (2016) Acfs: accurate circRNA identification and quantification from RNA-Seq data. *Sci Rep* **6**:38820.

Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, Chen W, Gao X, Zhao K, Zhou H et al. (2018a) A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* **37**:1805–1814.

Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, Liu H, Xu J, Xiao F, Zhou H et al. (2018b) A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun* **9**:4475.

Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, and Yang L (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* **26**:1277–1287.

Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, and Yang L (2014) Complementary sequence-mediated exon circularization. *Cell* **159**:134–147.

Address correspondence to: Dr. Dong Gui Hu, Department of Clinical Pharmacology and Flinders Cancer Centre, Flinders University College of Medicine and Public Health, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia. E-mail: donggui.hu@flinders.edu.au