


# Molecular profiles and urinary biomarkers of upper tract urothelial carcinomas associated with aristolochic acid exposure

Sandra Karanović<sup>1</sup> | Maude Ardin<sup>2</sup> | Zuojian Tang<sup>3</sup> | Karla Tomić<sup>4</sup> |  
 Stephanie Villar<sup>2</sup> | Claire Renard<sup>2</sup> | Elisa Venturini<sup>5</sup> | Adam H. Lorch<sup>6</sup> |  
 Daniel S. Lee<sup>5</sup> | Želimir Stipančić<sup>7</sup> | Neda Slade<sup>8</sup> | Ivana Vuković Brinar<sup>1</sup> |  
 Damir Dittrich<sup>9</sup> | Krešimir Karlović<sup>9</sup> | Fran Borovečki<sup>10</sup> |  
 Kathleen G. Dickman<sup>11,12</sup> | Magali Olivier<sup>2</sup> | Arthur P. Grollman<sup>11</sup> |  
 Bojan Jelaković<sup>1</sup> | Jiri Zavadil<sup>2</sup> 

<sup>1</sup>Department of Nephrology, Arterial Hypertension, Dialysis and Transplantation, University Hospital Center Zagreb, School of Medicine, University of Zagreb, Zagreb, Croatia

<sup>2</sup>Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, WHO, Lyon, France

<sup>3</sup>Institute for Systems Genetics, New York University Langone Medical Center, New York, New York, USA

<sup>4</sup>Department of Pathology, General Hospital Dr. Josip Benčević, Slavonski Brod, Croatia

<sup>5</sup>Office for Collaborative Science, New York University Langone Medical Center, New York, New York, USA

<sup>6</sup>Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

<sup>7</sup>Department for Dialysis Odžak, County Hospital Orašje, Odžak, Bosnia and Herzegovina

<sup>8</sup>Division of Molecular Medicine, Institute Ruder Bošković, Zagreb, Croatia

<sup>9</sup>Department of Urology, General Hospital Dr. Josip Benčević, Slavonski Brod, Croatia

<sup>10</sup>Department for Functional Genomics, Center for Translational and Clinical Research, University Hospital Center Zagreb, School of Medicine, University of Zagreb, Zagreb, Croatia

<sup>11</sup>Department of Medicine, Stony Brook University, Stony Brook, New York, USA

<sup>12</sup>Department of Medicine/Nephrology, Stony Brook University, Stony Brook, New York, USA

## Correspondence

Jiri Zavadil, Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, WHO, Lyon, France.  
 Email: zavadilj@iarc.fr

Sandra Karanović and Bojan Jelaković, Department of Nephrology, Arterial Hypertension, Dialysis and Transplantation,

## Abstract

Recurrent upper tract urothelial carcinomas (UTUCs) arise in the context of nephropathy linked to exposure to the herbal carcinogen aristolochic acid (AA). Here we delineated the molecular programs underlying UTUC tumorigenesis in patients from endemic aristolochic acid nephropathy (AAN) regions in Southern Europe.

**Abbreviations:** AA, aristolochic acid; AAN, aristolochic acid nephropathy; AL-dA, aristolactam-I-deoxyadenosine; BAM, binary alignment map; bp, base pair; COSMIC, Catalogue of Somatic Mutations in Cancer; cRNA, complementary RNA; DAPPLE, Disease Association Protein-Protein Link Evaluator; dbSNP, database of single nucleotide polymorphisms; DNA, deoxyribonucleic acid; EN, endemic nephropathy; FDR, false discovery rate; GATK, Genome Analysis Toolkit; GO, Gene Ontology; GSEA, gene set enrichment analysis; HH, hedgehog; IARC, International Agency for Research on Cancer; IPA, Ingenuity Pathway Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAF, minor allele frequency; miRNA, microRNA; mRNA, messenger ribonucleic acid; q-PCR, quantitative polymerase chain reaction; RCF, relative centrifugal force; RNA, ribonucleic acid; RNA-seq, ribonucleic acid sequencing; SAM, sequence alignment map; SBS, single base substitution; TLDA, TaqMan Low Density Array; TMA, tissue microarray; UTUC, upper tract urothelial carcinoma; WES, whole exome sequencing.

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 IARC. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.

University Hospital Center Zagreb, School of Medicine, University of Zagreb, Zagreb, Croatia.

Email: skaranov@kbc-zagreb.hr (S. K.) and bjelakov@kbc-zagreb.hr (B. J.)

Arthur P. Grollman, Department of Medicine, Stony Brook University, Stony Brook, NY, USA.

Email: arthur.grollman@stonybrook.edu

#### Present address

Zuojian Tang, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA.

Karla Tomić, Department of Pathology, Ålesund Hospital, Møre and Romsdal Health Trust, Ålesund, Norway.

Elisa Venturini, Natera, Inc., San Carlos, CA, USA.

#### Funding information

Croatian Science Foundation, Grant/Award Number: 04/38; Ministry of Science and Technology, Croatia, Grant/Award Number: 108-0000000-0329; National Institute of Environmental Health Sciences, Grant/Award Number: P01 ES004068; Fogarty International Center, Grant/Award Number: R03 TW007042; National Cancer Institute, Grant/Award Number: P30 CA016087; Laufer Family Foundation

We applied an integrative multiomics analysis of UTUCs, corresponding unaffected tissues and of patient urines. Quantitative microRNA (miRNA) and messenger ribonucleic acid (mRNA) expression profiling, immunohistochemical analysis by tissue microarrays and exome and transcriptome sequencing were performed in UTUC and nontumor tissues. Urinary miRNAs of cases undergoing surgery were profiled before and after tumor resection. Ribonucleic acid (RNA) and protein levels were analyzed using appropriate statistical tests and trend assessment. Dedicated bioinformatic tools were used for analysis of pathways, mutational signatures and result visualization. The results delineate UTUC-specific miRNA:mRNA networks comprising 89 miRNAs associated with 1,862 target mRNAs, involving deregulation of cell cycle, deoxyribonucleic acid (DNA) damage response, DNA repair, bladder cancer, oncogenes, tumor suppressors, chromatin structure regulators and developmental signaling pathways. Key UTUC-specific transcripts were confirmed at the protein level. Exome and transcriptome sequencing of UTUCs revealed AA-specific mutational signature SBS22, with 68% to 76% AA-specific, deleterious mutations propagated at the transcript level, a possible basis for neoantigen formation and immunotherapy targeting. We next identified a signature of UTUC-specific miRNAs consistently more abundant in the patients' urine prior to tumor resection, thereby defining biomarkers of tumor presence. The complex gene regulation programs of AAN-associated UTUC tumors involve regulatory miRNAs prospectively applicable to noninvasive urine-based screening of AAN patients for cancer presence and recurrence.

#### KEYWORDS

aristolochic acid, massively parallel sequencing, transcriptome profiling, upper tract urothelial carcinoma, urine miRNA

#### What's new?

Ingestion of aristolochic acid (AA) via contaminated wheat-containing food products is a major cause of endemic neuropathy and urologic carcinogenesis in southeastern Europe. Here, using integrated multi-omics analysis, the authors identified molecular programs underlying upper tract urothelial tumors (UTUC) in patients in Southern Europe with past carcinogenic AA exposure. Analyses reveal associations between 89 miRNAs and 1,862 target mRNAs, with confirmation of UTUC-specific transcripts at protein level. AA-specific mutations in UTUC and deleterious mutations were uncovered at both gene and transcript levels. The findings suggest that tumors in the urinary tract can be monitored by urine miRNA signature.

## 1 | INTRODUCTION

Ingestion of *Aristolochia* herbs containing aristolochic acid (AA) leads to aristolochic acid nephropathy (AAN), marked by severe renal damage and cancer formation in the urinary tract, renal cortex, liver, bile duct and possibly other anatomical sites.<sup>1-4</sup> Millions of people worldwide may develop AAN and associated cancers following exposure to AA (International Agency for Research on Cancer [IARC] Group I carcinogen<sup>5</sup>), particularly in Asia where *Aristolochia* is prescribed and used as traditional medicine<sup>2,3,6-8</sup> and where the causal role for AA in urologic carcinogenesis had been supported by

markedly reduced incidence rates of urologic cancers following a ban on the use of select AA-containing herbs.<sup>9</sup> In the endemic nephropathy (EN) regions of Bosnia and Herzegovina, Bulgaria, Croatia, Romania and Serbia, AAN results from environmental contamination of wheat grains used for bread-baking, by the seeds of *Aristolochia clematitis*.<sup>10-12</sup> As patients continue to be admitted for AAN-associated surgery decades after exposure, AAN is a serious public health problem in individuals with past exposure, including the farming families in the Balkan countries or the Belgian female patient cohort identified in the 1990s.<sup>13</sup> AAN-associated upper tract urothelial carcinomas (UTUCs) develop years after exposure

and often recur elsewhere in the urinary tract, including the urinary bladder.<sup>14,15</sup>

Here we report a multiomics analysis of UTUCs from AAN patients from Croatia and Bosnia. We integrate global profiles of the deregulated messenger ribonucleic acid (mRNA) and microRNA (miRNA) transcriptomes, validate the key pathways at the protein level in the source tumors, investigate mutation spectra at both the deoxyribonucleic acid (DNA) and RNA levels and finally investigate urinary miRNAs as markers of urinary tract tumor presence. We describe broad UTUC-specific miRNA:mRNA networks, implicating massive deregulation of mutated oncogenes and tumor suppressor genes, effectors of cell-cell and cell-matrix remodeling, in regulators of chromatin dynamics, and inhibition of developmental morphogenic pathways and mesenchymal programs. We confirm the mutational signature of AA, present in both the DNA and RNA of UTUCs and identify a set of UTUC-specific miRNAs depleted in the patients' urine after surgical tumor resection. By generating a unique catalog of molecular alterations, we provide insights into complex candidate mechanisms of AA-associated UTUC carcinogenesis, and show suitability of urine miRNAs as noninvasive biomarkers of early recurrence of urothelial cancer in AAN patients.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients and clinical data

Fifteen patients from the EN regions in Croatia and Bosnia and Herzegovina were hospitalized with urothelial cancer between 2009 and 2010. All patients had urothelial carcinoma, except for the patient B\_AAN\_27 (Tables S1 and S2) with sarcomatoid urothelial cancer. The patients underwent tumor surgery by radical nephroureterectomy or ureterectomy while two had also undergone cystectomy. The list of patients, their clinical data and information on samples used in molecular analyses is listed in Tables S1 and S2.

### 2.2 | Nucleic acid isolation

DNA was extracted from fresh frozen samples of renal cortex, tumor and adjacent nontumor tissues and purified by standard phenol-chloroform extraction techniques. The quantity and purity of DNA were measured by NanoDrop 2000c (NanoDrop Technologies, Wilmington, DE).

Total RNA was isolated from fresh samples of tumor and adjacent normal tissue using miRNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions and stored in RNAlater. All samples were treated with DNase (Qiagen, Valencia, CA). RNA samples were stored at  $-80^{\circ}\text{C}$ . The quantity and purity of RNA was measured using NanoDrop 2000c (NanoDrop Technologies, Wilmington, DE). Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) was used to determine the RNA integrity number which ranged between 8 and 10.

### 2.3 | DNA adduct analysis and TP53 mutation analysis

The level of aristolactam-I-deoxyadenosine (AL-dA) DNA adducts in the renal cortex DNA (10-20 mg) was determined using  $^{32}\text{P}$ -postlabeling polyacrylamide gel electrophoresis as described previously.<sup>10</sup> The UTUC TP53 mutation status was determined as previously described.<sup>16</sup>

### 2.4 | Analysis of mRNA and miRNA abundance in tissues

Gene expression profiling was performed using the Human Genome U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA), measuring over 47,000 transcripts. The RNA was processed using 3' *in vitro* transcription Express kit (Affymetrix, Santa Clara, CA), 500 ng of total purified RNA was transcribed into biotin-labeled complementary RNA (cRNA). Labeled cRNA probe was hybridized overnight to the microarrays. Following hybridization, arrays were washed, fluorescently tagged and scanned, using the Affymetrix GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA). Usual quality measures and normalization for the Affymetrix GeneChip (3'/5' ratios and trimmed mean normalization) were used. Raw data were imported into GeneSpring GX10 software (Agilent Technologies, Santa Clara, CA) and subjected to Robust Multichip Average (RMA) normalization. Differentially expressed genes were identified using the unpaired *t* test ( $P < .05$  with Benjamini-Hochberg false discovery rate (FDR) multiple testing correction). Given the strict statistical parameters, the fold-change cutoff was set to at least 50% change in mRNA abundance.

MicroRNA profiling was performed by subjecting total RNA (500 ng) to high-capacity quantitative polymerase chain reaction (q-PCR) using a megaplex reverse-transcription primer set for  $\sim 754$  human microRNAs (Sanger miRBase database v.14), analyzed by Applied Biosystems TaqMan Low Density Arrays (TLDA, TaqMan Array Human MicroRNA Panel A v2.1 and B v3.0) on the ABI 7900HT Sequence Detection System (Applied Biosystems, Waltham, MA), as described previously.<sup>17</sup> The reactions were incubated in a 384-well plate at  $50^{\circ}\text{C}$  for 2 minutes and  $95^{\circ}\text{C}$  for 10 minutes followed by 40 cycles of  $95^{\circ}\text{C}$  for 15 seconds and  $60^{\circ}\text{C}$  for 1 minute and were then held for 10 minutes at  $72^{\circ}\text{C}$ . The data were collected and processed using GeneSpring GX10 software (Agilent Technologies, Santa Clara, CA). The miRNA abundance was determined by the  $2^{-\text{Ct}}$  value formula which was followed by quantile normalization of all data. A union of unpaired *t* test (corrected *P* value  $< .05$ ) and Pavlidis Template Matching ( $P < .01$ ) yielded the differentially abundant miRNAs. Multiple-testing correction for *t* test was performed using Benjamini-Hochberg FDR. The fold-change cutoff was set at  $1.5\times$ , analogously to the differential mRNA abundance threshold.

### 2.5 | Pathway analysis and bioinformatics

Gene Spring GX10 software (Agilent Technologies, Santa Clara, CA) was used for data management, analysis and visualization. Gene Ontology

(GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used for annotation of specific genes and pathways. Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, Redwood City, CA, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)) based on the TargetScan high-confidence predictions and experimental validation was used to identify the mRNA targets of miRNAs (Table S5). The TargetScan algorithm in IPA searches for the presence of conserved 8-mer and 7-mer sites matching the seed region of a given microRNA, while the experimentally observed miRNA-mRNA relationships are based on the combined TarBase and miRecords database contents as well as the Ingenuity Knowledge Base containing information from published literature manually curated by the Ingenuity expert team. miRconnX was used to identify the top-ranking microRNA-mRNA relationships (<http://mirconnx.csb.pitt.edu>). NIH DAVID (Database for Annotation, Visualization and Integrated Discovery)<sup>18</sup> and gene set enrichment analysis (GSEA), collections C2 and C5<sup>19</sup> were used for pathway analysis. Disease Association Protein-Protein Link Evaluator (DAPPLE) was used to determine and visualize significant physical connectivity among proteins encoded for by miRNA target genes, based on protein-protein interactions reported in public domain.<sup>20</sup> Additional data visualization was done using CIRCOS version 0.63<sup>21</sup> while MutSpec was used for mutation analysis<sup>22</sup> and non-negative matrix factorization method for mutational signatures analysis using an R package.<sup>23,24</sup> Statistical analysis of the tissue microarray (TMA) data were performed using GraphPad Prism version 7.03 (GraphPad Software, La Jolla, CA).

## 2.6 | Whole exome sequencing and RNA sequencing library preparation

Genomic DNAs were prepared for sequencing using the TruSeq LT DNA sample prep kit (Illumina, San Diego, CA). Genomic DNA was sheared using the Covaris S220 (Covaris, Woburn, MA) and the size distribution was verified using a high-sensitivity DNA assay on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Following end repair and 3' adenylation of DNA fragments, a unique index adapter was ligated to each sample along with sequencing adapters. Fragment size selection was performing using a 2% agarose gel, and the excised bands were cleaned up using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). PCR amplification and subsequent Ampure XP (Beckman Coulter, Brea, CA) bead cleanup were performed as described in the TruSeq DNA prep protocol. The resulting library was verified for fragment size using the Bioanalyzer 2100, and quantified via q-PCR on the ViiA7 Real-Time PCR System (Applied Biosystems, Waltham, MA) using the KAPA Library Quantification kit (KAPA Biosystems, Wilmington, MA). The libraries were then enriched using the TruSeq Exome Enrichment kit (Illumina, San Diego, CA) prior to sequencing.

The quality and concentration of RNA were measured by NanoDrop 2000c (NanoDrop Technologies, Wilmington, DE). The starting amount was 3.5  $\mu$ g of total RNA. Ribosomal RNA was removed using Ribo-Zero rRNA removal kits (Epicentre, Illumina, San Diego, CA). RNA-Seq libraries from depleted RNA were prepared using the ScriptSeq v2 RNaseq library prep kit (Epicentre, Illumina, San Diego, CA). In brief, after the addition of fragmentation buffer,

the ribosomal ribonucleic acid-depleted sample is chemically fragmented. The fragmented RNA is reverse-transcribed using random-sequence primers containing a tagging sequence at 5' ends. 3' tagging is accomplished using the Terminal-Tagging Oligo. The di-tagged cDNA is purified by MinElute PCR Purification kit (Qiagen, Valencia, CA) and washed with EB buffer. cDNA is then amplified by limited-cycle PCR using PCR primer pairs that anneal to the tagging sequences of the di-tagged cDNA. Excess nucleotides and PCR primers are removed from the amplified double-stranded, adaptor-tagged cDNA, by Agencourt Ampure XP beads (Beckman Coulter, Brea, CA). The library products were then ready for sequencing analysis via Illumina HiSeq 2000 (Illumina, San Diego, CA).

## 2.7 | RNA sequencing and data processing

RNA sequencing (RNA-seq) libraries were sequenced deeply (for number of generated reads, see Table S6), with either single-end for UTUC\_20 or pair-end for UTUC\_03 and UTUC\_18 (at 51 bp read length) using the Illumina HiSeq 2000 sequencer.

All reads in FASTQ format were generated using Illumina CASAVA version 1.8 by trimming of 2 bp from both the 5' end and 3' end and yielding 47 bp sequence reads. All passed filtered and qualified reads were aligned to human genome version GRCh37 (hg19) using eland software of CASAVA version 1.8 with default settings. Binary alignment map (BAM) files and single base substitution (SBS) calls were generated by variant detection and counting software of CASAVA version 1.8. The bulk sequencing metrics for the RNA-seq are listed in the Table S6.

After removing duplicate reads, postprocessing steps have been applied to the raw SBS calls in order to obtain SBSs with confidence. SBSs with coverage less than 10 $\times$  were removed. SBSs with <10% reads supporting for variant or >90% reads supporting for reference nucleotide per site were also filtered out. Last, we filtered out all SBSs that match the common database of single nucleotide polymorphisms (dbSNP) version 137 with minor allele frequency (MAF)  $\geq$ 1%.

## 2.8 | Whole exome sequencing, data processing and variant calling

Whole exome sequencing (WES) libraries were sequenced (51-bp single-read) using Illumina's HiSeq 2000. FASTQ reads were generated using Illumina CASAVA version 1.8 by trimming of 2 bp from both the 5' end and 3' end and yielding 47 bp sequence reads. All pass-filtered and qualified reads were aligned to human genome version GRCh37 (hg19) in sequence alignment map (SAM) format using BWA with default parameter settings. SAM tools were used to convert SAM files into BAM files. Mapped reads in the raw BAM files were then realigned locally, duplicates marked and base pair score recalibrated using Genome Analysis Toolkit (GATK) version 1.3 and Picard version 1.59/1.79. The UnifiedGenotyper variant caller (within GATK version 1.3) was used to call SBS. The bulk sequencing metrics for the WES analysis are listed in the Table S6.



After removing duplicate reads, several postprocessing steps have been applied to the raw SBS calls in order to obtain SBSs with confidence. SBSs with coverage less than  $10\times$  were filtered out. SBSs with  $<10\%$  reads supporting for variant or  $>1\%$  reads supporting for background control data per site were also removed. Finally, all SBSs that match the common dbSNP version 137 with  $MAF \geq 1\%$  were removed. Annotations of somatic SBS were performed by ANNOVAR and their functional impact was evaluated by the ANNOVAR-embedded SIFT and PolyPhen-2 modules.

## 2.9 | TMA analysis

Immunohistochemistry on TMAs was performed following standard recommendations.<sup>25</sup> TMAs were constructed from formalin-fixed paraffin-embedded blocks using three cores of tumor tissue and three cores of normal urothelium per sample and internal controls (skin, breast, muscle, kidney, thyroid, lymph node), all 1 mm in diameter. Once constructed, the blocks were sectioned using microtome. Slides were deparaffinized, then rehydrated before staining. For the antigen retrieval, slides were carried using Trilogy solution (Cell Marque, Rocklin, CA, ref 920P-04) or Citrate Buffer pH 6 (Vector Laboratories, Burlingame, CA, ref H3300) in a steamer ( $95^{\circ}\text{C}$ ) or by microwave, and maintained at this temperature for 10 minutes followed by a 20 minutes cool-down period at room temperature. Endogenous peroxidases were blocked with  $3\% \text{H}_2\text{O}_2$  in methanol for 5 minutes. Slides were rinsed in distilled water, and immersed in wash buffer before immunostaining. Sections were processed by incubation with primary antibodies overnight at  $4^{\circ}\text{C}$ ; incubation with the secondary antibody for 30 minutes at room temperature, signal amplification with ABS kit and visualization using  $3,3'$ -diaminobenzidine solution for 15 minutes. Slides were subsequently counterstained with hematoxylin, dehydrated, mounted with permanent mounting medium, and coverslipped for visualization.

Antibodies that were used included (listed as the protein name, and in parentheses the antibody clone, type, company and dilution used): Vimentin (N/A, rabbit pAb, Epitomics, Burlingame, CA, 1:2000), Actin (C-4, mouse mAb, MP Biomedicals, Santa Ana, CA, 1:2000), ki67 (MM1, mouse mAb, Novocastra, Leica Biosystems, Wetzlar, Germany, 1:50), p53 (1C12, mouse mAb, Cell Signaling Technology, Danvers, MA, 1:10 000), p21 (Ab-1, mouse mAb, Calbiochem, San Diego, CA, 1:1000), MSH2 (Ab2, mouse mAb, Calbiochem, San Diego, CA, 1:500), IDH1 (H09, mouse mAb, Dianova, Hamburg, Germany, 1:500), PCNA (N/A, mouse mAb, Dako, Glostrup, Denmark, 1:1000), 3meH3K27 (K27, mouse mAb, Abcam, Cambridge, United Kingdom, 1:2000), EZH2 (AC22, mouse mAb, Cell Signaling Technology, Danvers, MA, 1:2000), SMAD3 (N/A, rabbit pAb, Abcam, Cambridge, United Kingdom, 1:2000), pSMAD3 (phospho S208) (N/A, rabbit pAb, Abcam, Cambridge, United Kingdom, 1:2000), S100A4 (N/A, rabbit pAb, Dako, Glostrup, Denmark, 1:2000), TWIST (2C1a, mouse mAb, Abcam, Cambridge, United Kingdom, 1:500), PTEN (A2B1, mouse mAb, Santa Cruz Biotechnology, Dallas, TX, 1:4000), GSK3-beta (N/A, rabbit pAb, Cell Signaling Technology, Danvers, MA, 1:1000), Phospho-GSK3B (Ser9, rabbit pAb, Cell Signaling Technology, Danvers, CA, 1:750), Cyclin D1 (SP4, mouse mAb,

Neomarkers, Portsmouth, NH, prediluted), E-cadherin (NCH-38, mouse mAb, Dako, Glostrup, Denmark, 1:200), Brg1 (N-15, polygoat Ab, Santa Cruz Biotechnology, Dallas, TX, 1:2000), PAI (H135, rabbit pAb, Santa Cruz Biotechnology, Dallas, TX, 1:1000), BRCA1 (D-9, mouse mAb, Santa Cruz Biotechnology, Dallas, TX, 1:1000), CDC25A (F-6, mouse mAb, Santa Cruz Biotechnology, Dallas, TX, 1:1000), Nibrin (Ab-1, rabbit pAb, Oncogene Research Products, San Diego, CA, 1:1000), ATM (233, rabbit pAb, Serotech, Oxford, United Kingdom, 1:500).

Immunoreactivity of the stained slides was visualized by light microscopy and digitally by Leica SCN400 Scanner (Leica Biosystems, Wetzlar, Germany). The quantification of the results was done manually using intensity score ranging 0 to 3 and percentage of the stained cells on high magnification. The final score was calculated by multiplying intensity score and percentage on all three biological replicates. Non-parametric Mann-Whitney test was used to evaluate differences between the pathologic review scores of the tumor and normal tissues.

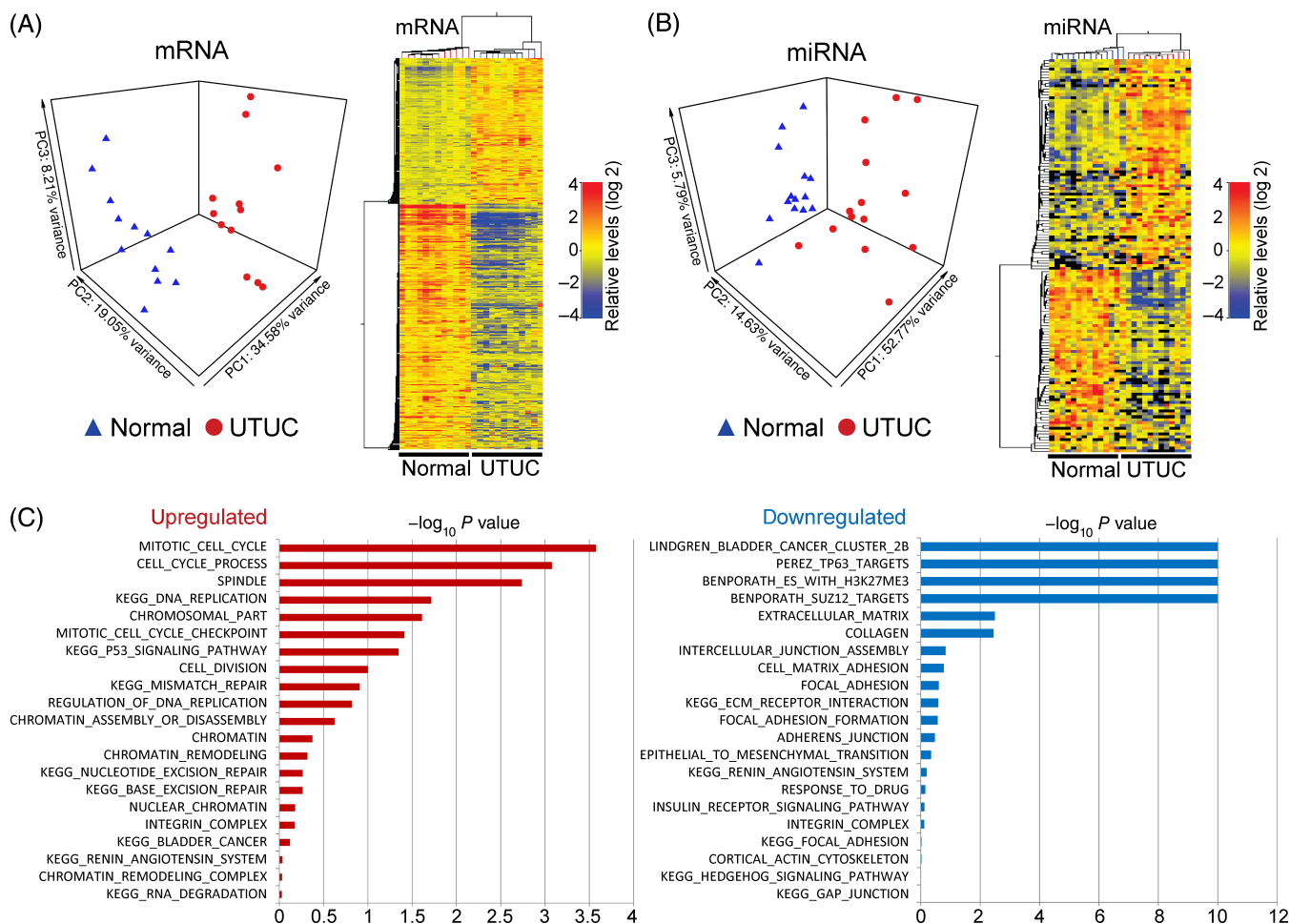
## 2.10 | Profiling of urinary miRNAs

Spot urine samples of additionally involved five UTUC patients were collected prior and at 6 weeks and 3 months after the surgery; 15 mL of each sample was used for further analysis. After centrifugation for 10 minutes at 2,000 RCF, urine supernatant containing cell-free nucleic acids including exosomal miRNAs was kept in long-term preservative (NorgenBiotek, Thorold, ON, Canada). Supernatant was subjected to additional high-speed centrifugation (30 minutes at 10,000 RCF) as to fractionate exosomal miRNAs from cellular debris and was spiked by 10 pM of control miRNA (derived from *Arabidopsis thaliana*). The miRNA content from exosomes was isolated using Urine Exosome RNA Isolation Kit (NorgenBiotek, Thorold, ON, Canada). The miRNA analysis was performed using TaqMan Low Density Array (TLDA, TaqMan Array Human MicroRNA Panel A v2.1 and B v3.0) high-throughput q-PCR as described above for the UTUC and non-tumor control tissues.

## 3 | RESULTS AND DISCUSSION

### 3.1 | miRNA:mRNA networks and target validation by TMAs

Integrated molecular profiling of UTUC tumors and unaffected tissues (Tables S1 and S2) identified consistent gene regulation aberrations. All patients had documented history of exposure to AA, with eight cases testing positive for the aristolactam-I-deoxyadenosine DNA adducts and/or A:T>T:A mutation(s) in the TP53 gene (Table S2). Differential analysis effectively segregated the UTUC and unaffected tissues (Figure 1A,B) involving 5,438 mRNAs (2,021 elevated, 3,417 reduced, see Table S3) and 138 miRNAs (74 elevated, 64 reduced, Table S4) significantly deregulated in UTUC and often comodulated when originating from discrete genomic clusters/families (Figure S1).



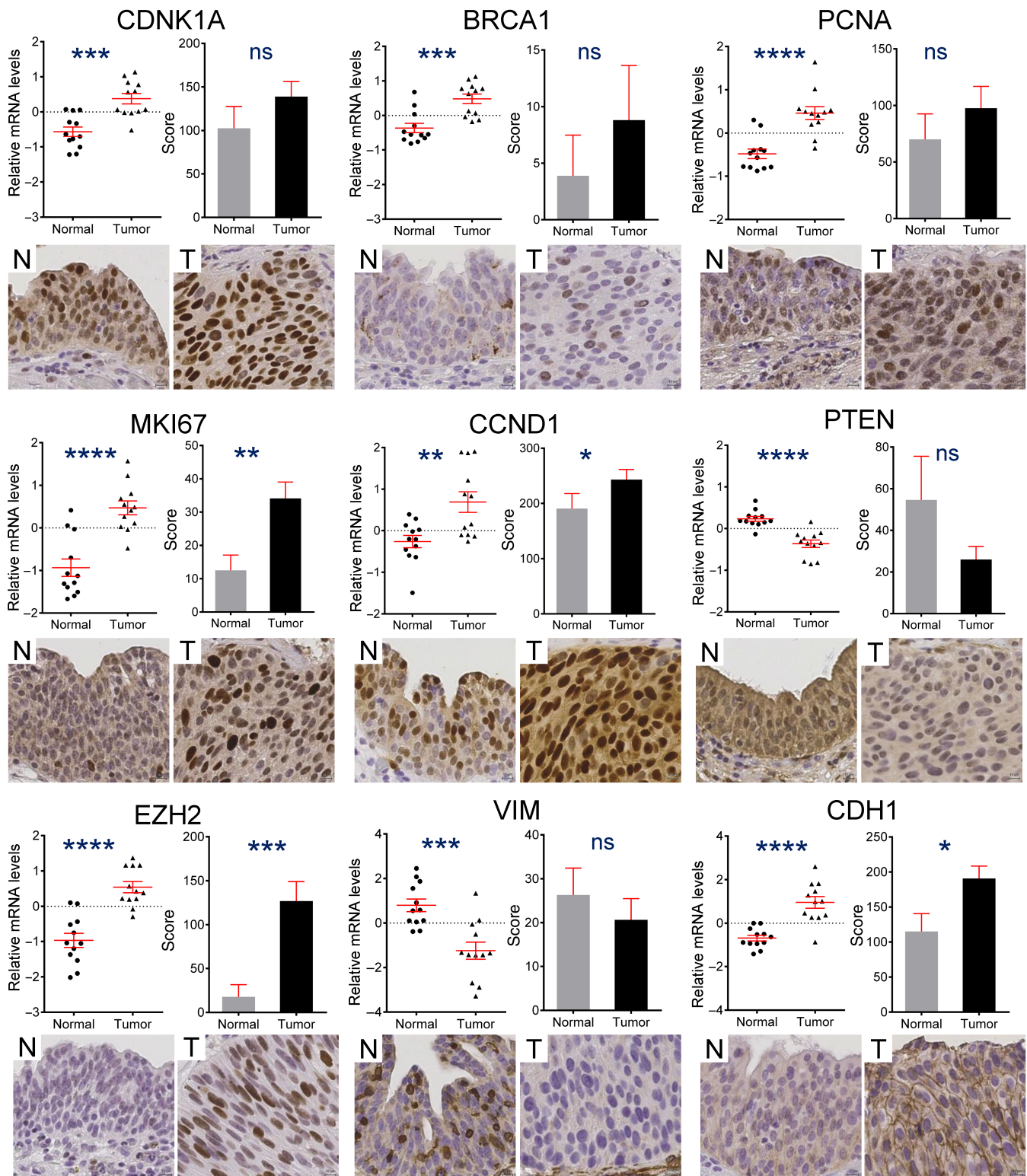
**FIGURE 1** Differential analysis of miRNA and mRNA programs in UTUC. (A) Principal component analysis of global gene expression (500 most variable genes) separates the tumor (red) and normal adjacent samples (blue); the heatmap shows relative abundance levels of significantly modulated genes in the tumor vs normal tissues. (B) Principal component analysis of global miRNA expression (144 most variable miRNAs) separates the tumor (red) and normal adjacent samples (blue); heatmap showing relative abundance levels of significantly modulated miRNAs in normal and tumor tissue. (C) The results of the gene set enrichment analyses performed on significantly deregulated genes showing pathways for genes upregulated (left/red) and downregulated in UTUCs (right/blue). miRNA, microRNA; mRNA, messenger ribonucleic acid; UTUC, upper tract urothelial carcinoma

Increased UTUC mRNAs implicated cell cycle, DNA replication, TP53 signaling, DNA repair, chromatin remodeling and bladder cancer (Figure 1C). Downmodulated UTUC transcripts implicated bladder cancer, silencing by PRC2/3 via histone H3K27 methylation, cell-cell and cell-matrix interactions and epithelial-to-mesenchymal transition (Figure 1C). Collectively, the AAN UTUCs manifest with increased cell proliferation, DNA repair, chromatin remodeling/silencing and reduction of cell motility.

Highest abundance of UTUC-specific miRNAs were observed for the oncogenic miR-17-92 family, the oncogenic and stress response-associated miR-21, and the miR-182 and miR-183 cluster. We also observed prominent upregulation of the antimetastatic miR-200 family. Conversely, tumor suppressor miRNAs, including let-7c, the miR-143/miR-145 cluster and miR-23b, were downmodulated (Table S4). Thus, the UTUC-specific miRNAs indicate roles in tumor growth while limiting its invasive and metastatic capacity.

Networks of miRNAs inversely correlated with their predicted, experimentally validated mRNA targets implicated 25 and 14 miRNAs in high-score up- and downmodulated nodes, respectively (Figure S2A). Stringent TargetScan prediction identified a total of 45 upmodulated miRNAs associated with 1,159 downmodulated mRNAs, and 44 downmodulated miRNAs associated with 703 upmodulated mRNAs (Figure S2 and Table S5). The mutual interconnectivity of the miRNA:mRNA network components is summarized in Table S5. Pathway analysis of the UTUC upregulated miRNA targets identified roles in ATM, BRCA2 and CHEK2 networks and cell-cycle deregulation (Figure S2B). Among the prominent miRNA-repressed UTUC, programs were targets of EZH2, SUZ12 and methylated H3K27, consistent with the silencing role of PRC2/3, and also promigratory components of the transforming growth factor- $\beta$  (TGF- $\beta$ ) and integrin signaling and extracellular matrix remodeling (Figure S2B).

TMA analysis confirmed the UTUC deregulation of cell cycle manifesting by elevated Ki67 (MKI67), PCNA and cyclin D1 (CCND1)



**FIGURE 2** Tissue microarray (TMA) immunohistochemistry validation of the mRNA profiling results. The scatter plots show relative mRNA abundance levels in normal and noninvasive component of the tumor tissues, the bar graphs show TMA pathology score results for normal urothelium and tumor tissue for cyclin-dependent kinase inhibitor 1 (CDKN1A), breast cancer type 1 susceptibility protein (BRCA1), proliferating cell nuclear antigen (PCNA), proliferation marker protein Ki-67 (MKI67), G1/S-specific cyclin-D1 (CCND1), histone-lysine N-methyltransferase EZH2 (EZH2), phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN (PTEN), vimentin (VIM) and E-cadherin (CDH1). Error bars represent the standard error of measurement around the mean. Bottom parts of each panel show representative immunohistochemical staining of TMA-derived sections using antibodies against the above proteins: normal urothelium (N), corresponding to patient-matched unaffected upper urinary tract tissue, on the left-hand side, noninvasive component of the tumor tissue (T) on the right-hand side. Original magnification:  $\times 40$ . Nonparametric Mann-Whitney test summaries (one-tailed  $P$  value): \*\*\*\* $P < .0001$ ; \*\*\* $P < .001$ ; \*\* $P < .01$ ; \* $P < .05$ ; ns = not reaching the significance cutoff of  $P < .05$ , trend only. mRNA, messenger ribonucleic acid



proteins (Figure 2). The UTUC-specific DNA damage response and DNA repair manifested by elevated TP53 mRNA/protein (not shown), and by increased mRNA and proteins of the TP53 target p21 (CDKN1A) and BRCA1 (Figure 2). We observed UTUC-specific decrease of the tumor suppressor PTEN, increased levels of EZH2 and augmented UTUC epithelial phenotype (decreased vimentin (VIM), increased E-cadherin (CDH1)). DAPPLE analysis suggested extensive physical connectivity and orchestration among proteins corresponding to the derepressed UTUC genes collectively targeted by 44 reduced miRNAs (Figure S3).

Dysregulation of cancer genes involved 89 oncogenes and 20 tumor suppressor genes (Table S9). The upregulated oncogenes MYC, FGFR3, HRAS and KRAS and downregulated tumor suppressor PTEN (also depleted in UTUC as protein; Figure 2) and PTCH1 were likely coordinated by miRNAs (Figure S4A) with elevated MYC representing a central node.

The upstream miRNA regulators of the PRC2/3 silencing complex included let-7c, miR-1, miR-23b, miR-143, miR-145 and miR-150 (Figure S4B). Additional UTUC-enriched chromatin regulators included lysine demethylases (KDMs) of diverse specificities, IDH1 (implicated in widespread changes in histone and DNA methylation in cancer) and methyl-CpG binding domain protein MBD1. The miRNAs regulating these gene products (including tumor suppressors miR-143, miR-145, miR-23b and let-7c) were collectively depleted in UTUCs. Next, the downregulated epigenome regulators HDAC9 and DNA methyltransferase DNMT3A are targets of elevated miR-21 and the miR-17-92 and miR-200 families (Figure S4B).

The repression of promigratory targets of ZEB1 and ZEB2, inhibitors of E-cadherin (CDH1), correlated with elevated miR-200bc/429 and miR-200a/141 families (Figure S5). ZEB1 and ZEB2 were reduced in UTUC while E-cadherin was increased as mRNA and protein (Figures 2 and S5). TGF- $\beta$  signaling, a positive regulator of both ZEB1 and ZEB2,<sup>26</sup> was downregulated in UTUC, including the TGF- $\beta$  receptors TGFBR1 and TGFBR2, and SMAD3 and phospho-SMAD3 proteins (basal and activated transducer of TGF- $\beta$  signaling, respectively) (Figure S5).

UTUC-specific repression of WNT manifested as reduced components of the KEGG hsa04310:Wnt signaling pathway and GO:0016055:Wnt receptor signaling pathway categories, including the ligands WNT2B, WNT4, the frizzled members of the receptor complexes, signal transducer AXIN and the downstream target LEF1. Additionally, WIF1 (WNT inhibitory factor 1), DKK2 (Dickkopf WNT signaling pathway inhibitor 1) and SFRP (secreted frizzled-related protein 1) were also downregulated, consistent with observed, inversely modulated miRNAs targeting these Wnt components. Furthermore, TMA analysis identified both the glycogen synthase kinase 3 beta (GSK3 $\beta$ ) protein and its phosphorylated form as downregulated in tumors (not shown). Similar to suppression of TGF- $\beta$ , the downmodulation of the WNT, another effector of epithelial-to-mesenchymal transition, may further contribute to the nonmetastatic behavior of UTUCs.

We next observed UTUC-specific downmodulation of PTCH1 and GLI2, the canonical receptor/tumor suppressor protein and a signal transducer in the hedgehog (HH) signaling pathway, and further HH repression was identified by GSEA (Figure 1C). The deregulated miRNAs targeting

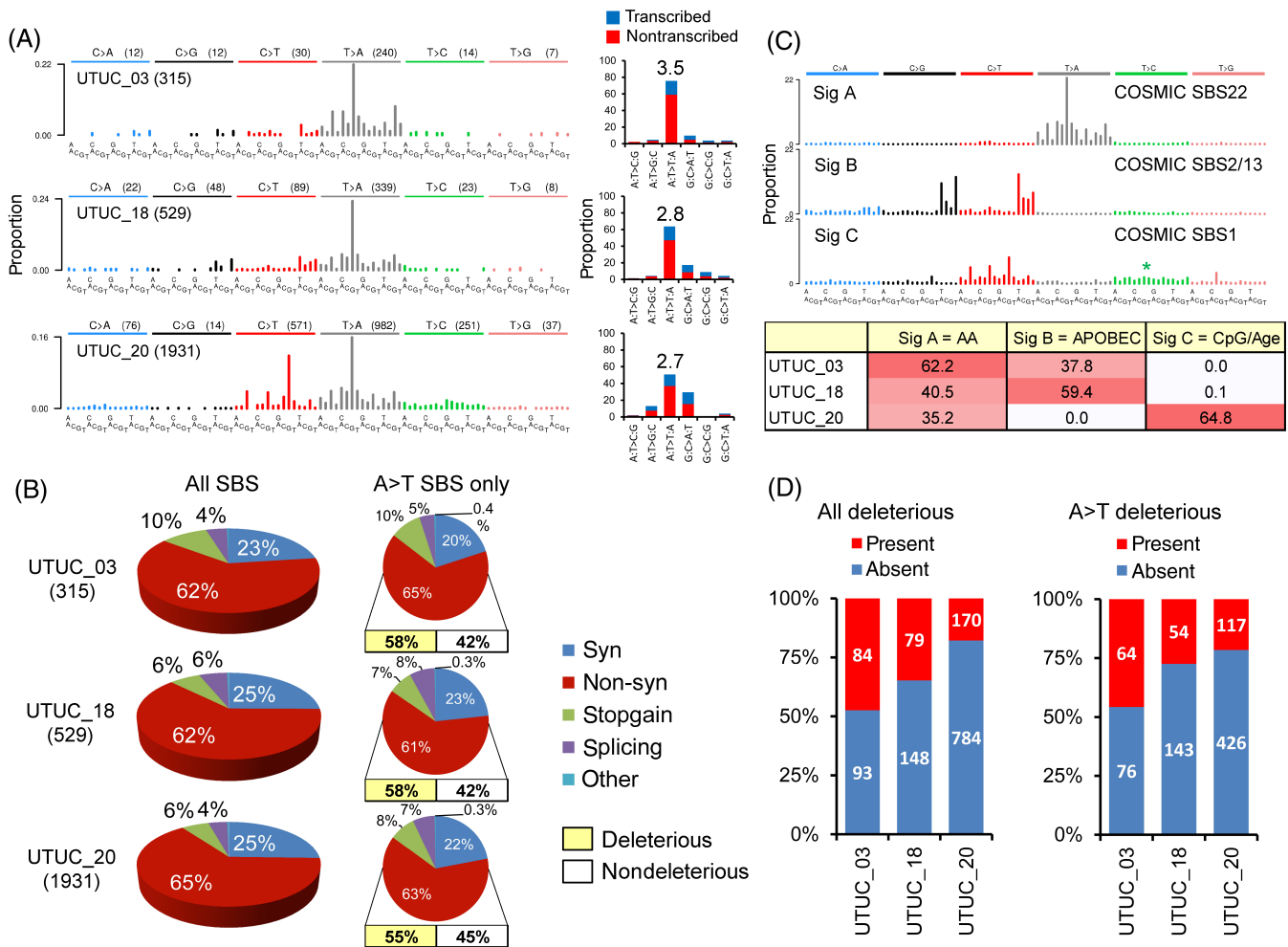
PTCH1 and GLI2 are shown in Table S5. The downregulation of the HH pathway in UTUC suggests a reduction in its tumor suppressor function.

Affected bladder cancer programs involved deregulated cancer genes and upstream miRNA regulators (Figures 1C and S2B). The KEGG bladder cancer pathway hsa05219 was enriched with upregulated HRAS, CDKN1A, E2F3, CCND1, FGFR3, KRAS, VEGFA and CDH1. The networks with key downregulated miRNAs putatively targeting these genes are summarized in Figure S4 and in Table S5. We observed significant enrichment (48 of 141 [34%],  $P$  value 4.12E-32, FDR  $q$ -value 1.42E-30) of downmodulated UTUC genes previously observed as depleted in bladder cancer (GSEA set M11896<sup>27</sup>). We also identified deregulated bladder cancer-specific miRNAs: upregulated miR-103 and downmodulated miR-99a, miR-100, miR-125, miR-145 and miR-143 consistent with observations from papillary stage Ta bladder carcinoma studies.<sup>28</sup> Additionally, miR-21 and miR-221 elevated in invasive bladder cancer (KEGG MicroRNAs in cancer) were upregulated in UTUCs.

In sum, our study identifies broad gene regulation programs in AA-associated UTUCs, together with a signature of 138 deregulated miRNAs including known cancer miRNAs.<sup>29</sup> This data set partly overlapped with a smaller 40-miRNA set identified previously in AAN/EN associated UTUCs,<sup>30,31</sup> including the shared miR-205-5p, miR-1290 (also observed in our study as an urinary tumor marker) and miR-127-3p. We built mRNA:miRNA networks that generally characterized the phenotype of UTUC by revealing deregulation of cell cycle, DNA damage response and repair, chromatin regulation, bladder cancer, developmental signaling and a number of oncogenes and tumor suppressor genes. Importantly, the validation at the protein level using TMAs showed considerable concordance with the mRNA expression of key miRNA targets.

### 3.2 | UTUC-specific DNA mutational signatures and RNA mutation spectra

WES was performed in three representative UTUC and matched nontumor renal cortex DNAs. A total of 2775 UTUC-specific somatic SBSs were identified (Figure 3A and Table S6) with varying SBS counts across samples (Figure 3A). However, the distribution of SBS classes remained consistent: 23% to 25% synonymous, 62% to 65% nonsynonymous, 7% to 10% introducing a stop codon and 5% to 8% affecting mRNA splicing (Figure 3B). Full description of UTUC-specific DNA mutations and their predicted impact on protein function is summarized in Table S7. The UTUC exome data showed a predominance of A:T>T:A transversions (56% of all mutation types) (Figure 3A and Table S7) with a prominent transcription-associated strand bias (2.7- to 3.5-fold enrichment on the nontranscribed DNA strand) (Figure 3A). The distributions of the A:T>T:A mutations into synonymous, nonsynonymous, stop-gain and affecting splicing matched that of all SBS and the impact of mutations was predicted as deleterious for most nonsynonymous A:T>T:As (Figure 3B). We identified distinct mutational signatures corresponding to the Catalogue of Somatic Mutations in Cancer (COSMIC) signature SBS22 (AA), signatures SBS2/13 (APOBEC)



**FIGURE 3** UTUC DNA and RNA somatic mutation spectra analysis. (A) Somatic mutation spectra of three UTUC is shown as frequencies of six SBS types (C>A stands for C:G>A:T, C>G for C:G>G:C, etc) in a specific trinucleotide context (x-axis). Bar graphs show transcriptional stand bias of all mutation types, with numerical ratios of SBS in nontranscribed vs transcribed strand shown for AA-specific A>T transversions. (B) Distribution of functional mutation types for all SBS (left pie charts, total numbers of SBS shown under the sample name) and for the A>T SBS subset only (right pie charts). Further distribution into deleterious and nondeleterious SBS is shown as percentages. (C) Non-negative matrix factorization identifies three essential mutational signatures whose spectra are shown together with matching COSMIC SBS signatures (\* = possible admixture of T>C-based SBS5). The table shows relative contribution of each signature to the mutation load of each sample, as percentage. (D) Percentages and counts of all deleterious mutations found in DNA exomes and their presence in RNA is shown (left graph); same is shown for the subset of deleterious A>T transversions only (right graph). Present/absent refers to presence or absence of DNA-identified mutations in the mRNA. COSMIC, Catalogue of Somatic Mutations in Cancer; Non-syn, nonsynonymous; SBS, single base substitution; Sig, signature; Syn, synonymous; UTUC, upper tract urothelial carcinoma

and signature SBS1 (age, clock-like, with a possible admixture of clock-like SBS5; Figure 3C), in keeping with previous reports.<sup>32,33</sup> The identified signature of AA exposure was nearly identical (cosine similarity of 0.97-0.99, with no remarkable peak differences) to the AA signatures determined previously by whole exome sequencing of UTUC tumors from East Asia and Southern Europe,<sup>32,34,35</sup> including the consistent and significant transcription strand bias of A:T>T:A transversions accumulating prominently on the non-transcribed DNA strand (Figure S6).

To examine the somatic alterations in the tumor RNA, deep whole transcriptome sequencing (RNA-seq) was performed on the UTUC\_03, UTUC\_18 and UTUC\_20 samples used for WES. In the corresponding mRNA samples, we observed the respective counts of

143, 148 and 344 of the WES-identified SBS (Table S6). The SBS type distribution (34% synonymous, 63% nonsynonymous, 2% stop-gain, 1% affecting splicing) was consistent across all UTUC RNA samples (data not shown). The number of deleterious mutations detected in both DNA and mRNA was 79 to 170 (range across samples) for any mutation type, and 54 to 117 for A:T>T:As (Figure 3D). The mutations shared between DNA and RNA are listed in Table S7. Thus, a large proportion of deleterious DNA mutations in UTUCs are propagated at the mRNA level, and the AA-specific A:T>T:As are likely to have a considerable impact on normal protein function. Integrated exome and RNA-seq of a subset of UTUC thus revealed the presence of the AA mutational signature SBS2<sup>32,34,35</sup> and extended the limited available data<sup>35</sup> on the propagation of signature-like

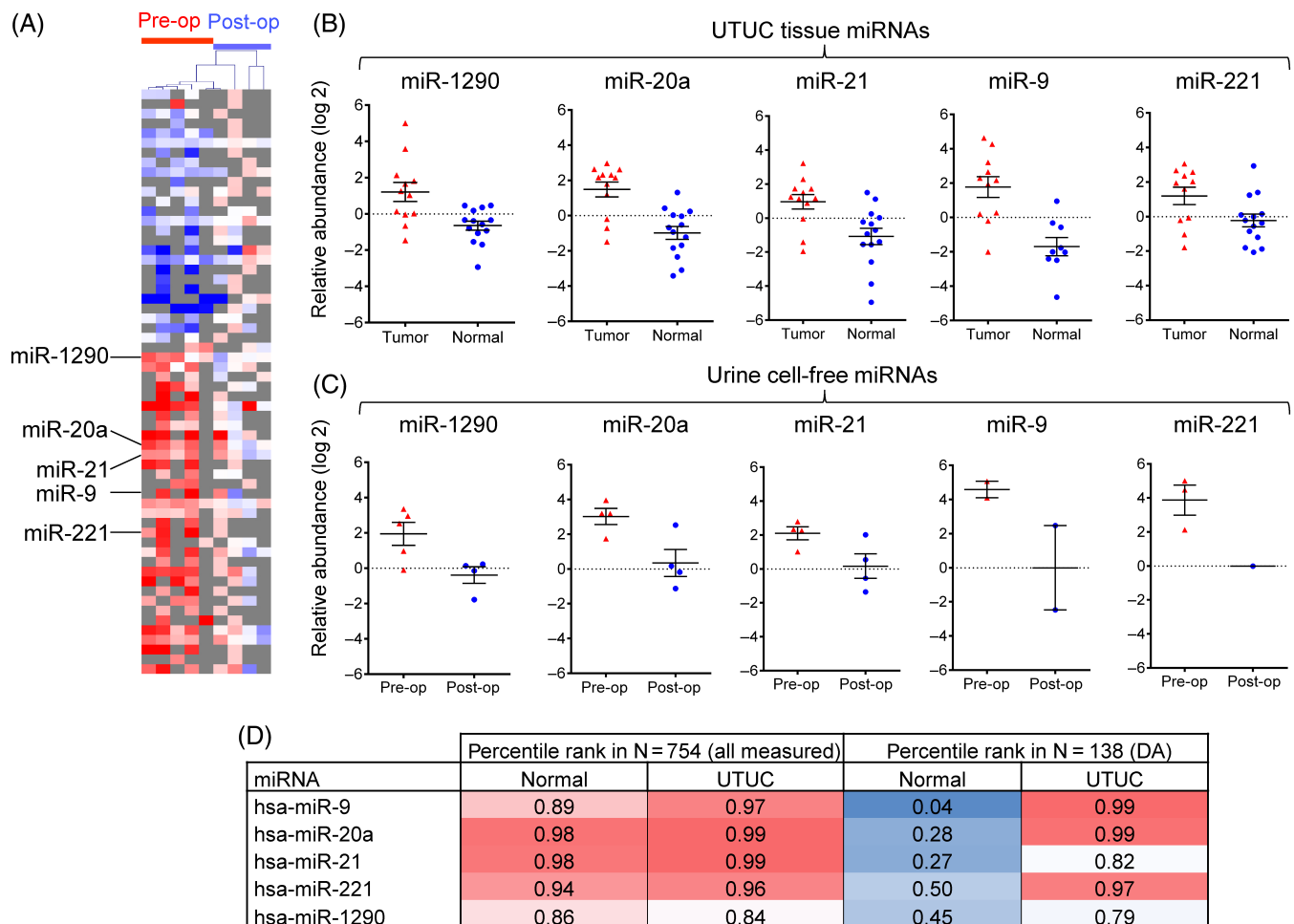


A:T>T:A transversions at the RNA level, by adding hundreds of mutated transcripts, including those deregulated in tumors. Despite the constrained scale, such observation is consistent with a recently published *in silico* prediction of an increased neoantigen load based on genome-scale DNA sequencing of AA signature-positive UTUCs of Chinese patients. The observed increased levels of infiltrating lymphocytes<sup>36</sup> can help select the high mutation-load UTUCs linked to AA exposure for efficient targeting by immunotherapy.<sup>36-38</sup> Additional studies are warranted to further evaluate the impact of AA-directed mutagenesis on the transcriptome and proteome functions, and the potential clinical utility.

### 3.3 | Recurrently mutated cancer genes

Analysis of A:T>T:A counts as a function of chromosome size yielded Pearson product-moment correlation coefficients ranging between

$R = 0.85$  and  $R = 0.89$ , recapitulating previous findings on the stochastic accumulation of A:T>T:A transversions in the genome.<sup>32</sup> However, a set of 63 cancer genes were recurrently mutated (Figure S7 and Table S8). Eight of these (AHNAK, ATM, ARID1A, MKI67, MLL2, PLEC, SYNE1 and TP53) are established cancer driver genes<sup>39</sup> previously reported as recurrently mutated in AAN UTUCs from Taiwan and from EN regions, although these previous studies focused only on mutations in the DNA.<sup>32,34</sup> Recurrently mutated ARID1A, ATM, FAT4, PTPRC, TP53 and UBR5 are listed in the COSMIC Cancer Gene Census.<sup>40</sup> Interestingly, a number of recurrently hit genes (ATM, EPHA3, HMCN1, INPP5B, LYST, MLL2, RELN, TLR6 and USH2A) were mutated by multiple nonsynonymous A:T>T:A transversions in the same tumor. Finally, 35 driver mutations corresponding to 27 recurrently mutated genes (with a subset of 22 genes harboring 27 A:T>T:A transversions) were also detected at the level of mRNA (Figure S7 and in Table S8), a result adding a new dimension to the previous reports of recurrently mutated driver genes in AAN-associated UTUC.<sup>32,34</sup>



**FIGURE 4** Tumor-specific miRNAs in the urine. (A) The heatmap shows relative abundance levels of urinary miRNA in the pre- and postsurgery urine samples (Pre-op and Post-op, respectively). (B) Scatterplots show miRNA relative abundance levels in the tumor and normal adjacent tissues for five distinct UTUC-specific miRNAs. Error bars present mean and SEM. (C) Scatterplots showing miRNA relative abundance levels in the urine collected before and after the tumor removal surgery for five UTUC-specific miRNAs. Error bars present the SE of measurement around the mean. (D) Percentile abundance rank of the 5-miRNA signature components in the normal and UTUC tissues, among all miRNA (N = 754) measurements, and among the N = 138 miRNAs significantly modulated in UTUC vs normal tissues. 1 = the top rank of 100%. DA, differentially abundant tissue miRNAs; miRNA, microRNA; UTUC, upper tract urothelial carcinoma

### 3.4 | Urinary miRNA as biomarkers of tumor presence

We next profiled cell-free miRNAs in the urine of five AAN UTUC patients (Table S10), to explore the utility of this biomaterial for determining the tumor presence in the urinary tract, and by extension for monitoring the tumor recurrence often associated with AAN cancers.<sup>15,41,42</sup> MiRNA profiling of the presurgery and postsurgery urine supernatants revealed a signature of 60 differentially modulated miRNAs (35 high and 25 low in the presurgery urine samples) (Figure 4 and Table S11) of which five cancer-related miRNAs (miR-20a, miR-9, miR-21, miR-221 and miR-1290) were also enriched in primary UTUCs (Figure 1 and Table S4). This subset of urinary miRNAs exhibits high and stable relative abundance rank in both the normal (median percentile of 0.94) and in the tumor tissues (median percentile of 0.97), while it undergoes a major increase in its relative abundance rank in the context of the differentially abundant miRNAs in the UTUC compared to normal tissue (median percentile increase from 0.28 to 0.97; Figure 4D). We propose that due to this shift in abundance in tumors, this 5-miRNA signature in the urine could be UTUC-derived and more easily detectable at elevated levels prior to tumor resection. This is in contrast to other miRNAs that are more abundant in the presurgery urine but either not detected as present or as or differentially modulated in tissues. The signature may therefore serve as a noninvasive biomarker for clinical screening and monitoring of tumor presence and cancer recurrence, upon proper validation for robustness in further studies conducted in an extended set of patients. A number of studies performed on bladder cancer patients similarly implicated urine miR-21 and/or miR-221 as markers of tumor presence,<sup>43-45</sup> and the promise of miRNAs as urinary tumor biomarkers have been extensively discussed.<sup>46-49</sup> To our knowledge, our study is the first to report a specific urine miRNA signature in the context of recurrent upper tract urothelial cancers driven by exposure to AA, a highly potent environmental and iatrogenic carcinogen. As such, it complements recent studies demonstrating the presence of AA mutational signature as a readout of exposure in cell-free DNA in urine,<sup>36</sup> or the sensitive detection of urothelial cancer-specific driver mutations by targeted resequencing.<sup>38,50</sup> Such noninvasive tests have to be further evaluated for the potentially best-suited application in clinical monitoring of AA-exposed individuals without cancer and of AA-associated cancer recurrence in the urinary tract.

## 4 | CONCLUSIONS

While various modes of toxicities of AA have recently been extensively reviewed,<sup>51</sup> the study presented here identifies specific and complex mRNA:miRNA networks of the AAN-associated UTUC, supported by adduct analysis, with components of key affected pathways further validated by high-throughput tissue microarray immunohistochemistry. Sequencing of patient exomes and transcriptomes confirmed a burden of somatic A:T>T:A transversions in UTUC and provided insights into deleteriously mutated expressed genes. Next,

we identified a subset of UTUC-specific urinary miRNAs potentially applicable as noninvasive biomarkers in screening and monitoring the AAN patients for urothelial carcinogenesis. Previous studies focusing on singular-level omic analyses of AA-associated UTUCs have been recently systematically reviewed,<sup>52</sup> yet the integration of various layers of molecular analyses in a single study has not yet been reported. We thus propose our work to be unique in design, scope and comprehensiveness, while offering new insights into the AA UTUC biology and a basis for the development of an important molecular cancer-screening tool.

### ACKNOWLEDGMENTS

In memoriam of Dr Frederick Miller, to whom we are indebted for expert pathology evaluations. We thank Dr Matej Knezevic, Dr Branko Brdar, Dr Elisabetta Kuhn and Christine Carrera for expert assistance with sample collection and processing. Expert technical assistance provided by Andrea Fernandes, Gyongyi Mihalyne and Vincent Cahais is gratefully acknowledged. Funding for this study was provided by the U.S. NIH/National Institute of Environmental Health Sciences (grant P01 ES004068), the U.S. NIH Fogarty International Center (R03 TW007042), the Ministry of Science and Technology, Croatia (108-000000-0329) and Croatian Science Foundation (Grant 04/38). Kathleen G. Dickman and Arthur P. Grollman gratefully acknowledge the financial support provided by Marsha and Henry Laufer (Laufer Family Foundation). The NYU Genome Technology Center received support from the U.S. NIH/National Cancer Institute (P30 CA016087).

### CONFLICT OF INTEREST

The authors declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

The mRNA and miRNA profiling data were deposited to the NCBI Gene Expression Omnibus (GEO) public database ([www.ncbi.nlm.nih.gov/gds](http://www.ncbi.nlm.nih.gov/gds)) under the Series reference ID GSE166912. The WES and RNA-seq data generated in this study were deposited to the European Genome-Phenome Archive (EGA) under the Study ID EGAS00001005363. The list of all somatic SBSs identified by WES and their overlap with the RNA-seq profiling data is available in the Table S7.

### ETHICS STATEMENT

The study protocols included the patients' informed consent and were approved by the IARC Ethics Committee and the Institutional Review Boards of the participating institutions.

### ORCID

Jiri Zavadil  <https://orcid.org/0000-0003-0640-5562>

### REFERENCES

- Rosenquist TA, Grollman AP. Mutational signature of aristolochic acid: clue to the recognition of a global disease. *DNA Repair*. 2016;44:205-211.
- Grollman AP. Aristolochic acid nephropathy: harbinger of a global iatrogenic disease. *Environ Mol Mutagen*. 2013;54:1-7.

3. Debelle FD, Vanherweghem JL, Nortier JL. Aristolochic acid nephropathy: a worldwide problem. *Kidney Int.* 2008;74:158-169.
4. Ng AWT, Poon SL, Huang MN, et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci Transl Med.* 2017;9:eaan6446.
5. IARC. *Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 100A: A Review of Human Carcinogens: Pharmaceuticals.* Lyon, France: International Agency for Research on Cancer; 2012.
6. Boot A, Jiang N, Rozen SG. Toward clinical understanding of aristolochic acid upper-tract urothelial carcinoma. *Theranostics.* 2020;10:5578-5580.
7. Chen CH, Dickman KG, Moriya M, et al. Aristolochic acid-associated urothelial cancer in Taiwan. *Proc Natl Acad Sci U S A.* 2012;109:8241-8246.
8. Lai MN, Wang SM, Chen PC, Chen YY, Wang JD. Population-based case-control study of Chinese herbal products containing aristolochic acid and urinary tract cancer risk. *J Natl Cancer Inst.* 2010;102:179-186.
9. Jhuang JR, Chiang CJ, Su SY, Yang YW, Lee WC. Reduction in the incidence of urological cancers after the ban on Chinese herbal products containing aristolochic acid: an interrupted time-series analysis. *Sci Rep.* 2019;9:19860.
10. Jelakovic B, Karanovic S, Vukovic-Lela I, et al. Aristolactam-DNA adducts are a biomarker of environmental exposure to aristolochic acid. *Kidney Int.* 2012;81:559-567.
11. Grollman AP, Shibutani S, Moriya M, et al. Aristolochic acid and the etiology of endemic (Balkan) nephropathy. *Proc Natl Acad Sci U S A.* 2007;104:12129-12134.
12. Stiborova M, Arlt VM, Schmeiser HH. Balkan endemic nephropathy: an update on its aetiology. *Arch Toxicol.* 2016;90:2595-2615.
13. Nortier JL, Martinez MC, Schmeiser HH, et al. Urothelial carcinoma associated with the use of a Chinese herb (*Aristolochia fangchi*). *N Engl J Med.* 2000;342:1686-1692.
14. Zlotta AR, Roumeguere T, Kuk C, et al. Select screening in a specific high-risk population of patients suggests a stage migration toward detection of non-muscle-invasive bladder cancer. *Eur Urol.* 2011;59:1026-1031.
15. Lemy A, Wissing KM, Rorive S, et al. Late onset of bladder urothelial carcinoma after kidney transplantation for end-stage aristolochic acid nephropathy: a case series with 15-year follow-up. *Am J Kidney Dis.* 2008;51:471-477.
16. Moriya M, Slade N, Brdar B, et al. TP53 mutational signature for aristolochic acid: an environmental carcinogen. *Int J Cancer.* 2011;129:1532-1536.
17. Bitzer M, Ju W, Jing X, Zavadil J. Quantitative analysis of miRNA expression in epithelial cells and tissues. *Methods Mol Biol.* 2012;820:55-70.
18. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44-57.
19. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545-15550.
20. Rossin EJ, Lage K, Raychaudhuri S, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 2011;7:e1001273.
21. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639-1645.
22. Ardin M, Cahais V, Castells X, et al. MutSpec: a Galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. *BMC Bioinformatics.* 2016;17:170.
23. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics.* 2010;11:367.
24. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature.* 1999;401:788-791.
25. Ilyas M, Grabsch H, Ellis IO, et al. Guidelines and considerations for conducting experiments using tissue microarrays. *Histopathology.* 2013;62:827-839.
26. Miyazono K. Transforming growth factor-beta signaling and cancer: the 28th Sapporo Cancer Seminar, 25-27 June 2008. *Cancer Sci.* 2009;100:363-365.
27. Lindgren D, Liedberg F, Andersson A, et al. Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. *Oncogene.* 2006;25:2685-2696.
28. Zabolotneva AA, Zhavoronkov A, Garazha AV, Roumiantsev SA, Buzdin AA. Characteristic patterns of microRNA expression in human bladder cancer. *Front Genet.* 2012;3:310.
29. Lujambio A, Lowe SW. The microcosmos of cancer. *Nature.* 2012;482:347-355.
30. Popovska-Jankovic K, Noveski P, Jankovic-Velickovic L, et al. MicroRNA profiling in patients with upper tract urothelial carcinoma associated with Balkan endemic nephropathy. *Biomed Res Int.* 2016;2016:7450461.
31. Tao L, Zeng Y, Wang J, et al. Differential microRNA expression in aristolochic acid-induced upper urothelial tract cancers ex vivo. *Mol Med Rep.* 2015;12:6533-6546.
32. Castells X, Karanovic S, Ardin M, et al. Low-coverage exome sequencing screen in formalin-fixed paraffin-embedded tumors reveals evidence of exposure to carcinogenic aristolochic acid. *Cancer Epidemiol Biomarkers Prev.* 2015;24:1873-1881.
33. Poon SL, McPherson JR, Tan P, Teh BT, Rozen SG. Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. *Genome Med.* 2014;6:24.
34. Hoang ML, Chen CH, Sidorenko VS, et al. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Sci Transl Med.* 2013;5:197ra02.
35. Poon SL, Pang ST, McPherson JR, et al. Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci Transl Med.* 2013;5:197ra01.
36. Lu H, Liang Y, Guan B, et al. Aristolochic acid mutational signature defines the low-risk subtype in upper tract urothelial carcinoma. *Theranostics.* 2020;10:4323-4333.
37. Cheng S, Zhong W, Xia K, et al. Prognostic role of stromal tumor-infiltrating lymphocytes in locally advanced upper tract urothelial carcinoma: a retrospective multicenter study (TSU-02 study). *Onco Targets Ther.* 2021;10:1861737.
38. Rebhan K, Ertl IE, Shariat SF, Grollman AP, Rosenquist T. Aristolochic acid and its effect on different cancers in uro-oncology. *Curr Opin Urol.* 2020;30:689-695.
39. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science.* 2013;339:1546-1558.
40. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer.* 2004;4:177-183.
41. Chen CH, Dickman KG, Huang CY, et al. Aristolochic acid-induced upper tract urothelial carcinoma in Taiwan: clinical characteristics and outcomes. *Int J Cancer.* 2013;133:14-20.
42. Jelakovic B, Nikolic J, Radovanovic Z, et al. Consensus statement on screening, diagnosis, classification and treatment of endemic (Balkan) nephropathy. *Nephrol Dial Transplant.* 2014;29:2020-2027.
43. Matsuzaki K, Fujita K, Jingushi K, et al. MiR-21-5p in urinary extracellular vesicles is a novel biomarker of urothelial carcinoma. *Oncotarget.* 2017;8:24668-24678.
44. Sapre N, Macintyre G, Clarkson M, et al. A urinary microRNA signature can predict the presence of bladder urothelial carcinoma in patients undergoing surveillance. *Br J Cancer.* 2016;114:454-462.
45. Erdmann K, Salomo K, Klimova A, et al. Urinary microRNAs as potential markers for non-invasive diagnosis of bladder cancer. *Int J Mol Sci.* 2020;21:3814-3830.

46. Mlcochova H, Hezova R, Meli AC, Slaby O. Urinary microRNAs as a new class of noninvasive biomarkers in oncology, nephrology, and cardiology. *Methods Mol Biol.* 2015;1218:439-463.
47. Mlcochova H, Hezova R, Stanik M, Slaby O. Urine microRNAs as potential noninvasive biomarkers in urologic cancers. *Urol Oncol.* 2014;32:41.e1-41.e9.
48. Kutwin P, Konecki T, Borkowska EM, Traczyk-Borszynska M, Jablonowski Z. Urine miRNA as a potential biomarker for bladder cancer detection - a meta-analysis. *Cent European J Urol.* 2018;71: 177-185.
49. Amuran GG, Eyuboglu IP, Tinay I, Akkiprik M. New insights in bladder cancer diagnosis: urinary miRNAs and proteins. *Med Sci.* 2018;6: 113-128.
50. Springer SU, Chen CH, Rodriguez Pena MDC, et al. Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy. *Elife.* 2018;7:e32143-e32169.
51. Sidorenko VS. Biotransformation and toxicities of aristolochic acids. *Adv Exp Med Biol.* 2020;1241:139-166.
52. Hassler MR, Bray F, Catto JWF, et al. Molecular characterization of upper tract urothelial carcinoma in the era of next-generation sequencing: a systematic review of the current literature. *Eur Urol.* 2020;78:209-220.

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Karanović S, Ardin M, Tang Z, et al. Molecular profiles and urinary biomarkers of upper tract urothelial carcinomas associated with aristolochic acid exposure. *Int. J. Cancer.* 2022;150(2):374-386. doi: 10.1002/ijc.33827