Back-Splicing Transcript Isoforms (Circular RNAs) Affect Biologically Relevant Pathways and Offer an Additional Layer of Information to Stratify NMIBC Patients

Anshita Goel[1], Douglas G. Ward[1], Naheema S. Gordon[1], Ben Abbotts[1], Maurice P. Zeegers[2], K. K. Cheng[3], Nicholas D. James[1], Richard T. Bryan[1*] and Roland Arnold[1*]

[1] Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom, [2] NUTRIM School for Nutrition and Translational Research in Metabolism & CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands, [3] Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom

Circularized transcript isoforms due to back-splicing are increasingly being reported in different tissues types and pathological states including cancer. Since these circular RNAs (circRNAs) are more stable than linear messenger RNA their identification and profiling in tumor tissue could aid in stratifying patients and may serve as biomarkers. In this study, we have investigated the relationship between circRNA expression and tumor grade in a cohort of 58, mostly non-muscle-invasive bladder cancer patients. From 4571 circRNAs detected, we identified 157 that were significantly differentially expressed between tumor grades relative to the linear transcript. We demonstrated that such grade-related differences can be identified in an independent cohort, and that a large fraction of circRNAs can be, in principle, detected in urine. The differentially expressed circRNAs cluster into subgroups according to their co-expression, subgroups which are enriched for DNA repair, cell cycle and intracellular signaling genes. Since one proposed function of circRNAs is to interfere with gene-regulation by acting as microRNA "sponges," candidates which were differentially expressed between tumor grades were investigated for potential miRNA target sites. By investigating the circRNAs from bladder cancer related pathways we demonstrated that the expression of these pathways, the circRNAs, and their parental genes are often decoupled and do not correlate, yet that some circRNAs do not follow this tendency. The present study provides the next step for the comprehensive evaluation of this novel class of RNAs in the context of non-muscle-invasive bladder cancer. Intriguingly, despite their possible function as microRNA sponges, they potentially affect host mRNA levels at the transcriptional stage, as compared to post-transcriptional control by miRNAs. Our analysis indicates differences of their activity between bladder cancer tumor stages, and their relative expression levels may provide an additional layer of information for patient stratification.

Keywords: circular RNA, back-splicing, bladder cancer, tumor grade, NMIBC, biomarkers

INTRODUCTION

Circular RNAs (circRNAs) are a class of single-stranded closed RNA molecules, created by backsplicing events from linear mRNA (1). They were first identified ∼20 years ago, but were initially seen as artifacts of aberrant RNA splicing (2). However, more recently, numerous potential functions of circRNAs have been proposed, such as acting as miRNA sponges, modulating transcription and interacting with RNA-binding proteins (RBPs) (3). Additionally, several genome-wide studies indicate that circRNAs are evolutionary conserved across species (4). Furthermore, an increasing number of studies show that circRNAs are strongly correlated with the proliferation, apoptosis, invasion, and metastasis of human tumors, which indicates the potential of circRNAs to act as novel therapeutic targets and biomarkers (5). The use of circRNAs for the latter is especially intriguing since they are more stable than linear RNA forms, escaping degradation by those processes which are dependent on the recognition of the ends of RNA molecules.

With the help of novel bioinformatics approaches and appropriate sequencing methods, comprehensive studies of circRNA species are possible (6), allowing investigation of the landscape of circRNAs in healthy and cancerous tissues. circRNAs are generally expressed at low levels and often exhibit cell type specific and tissue specific patterns (7). Most circRNAs originate from exons of protein-coding genes and can consist of one or multiple exons; some circRNAs also arise from introns, intergenic regions, non-coding RNA (ncRNA) loci, and other portions of the genome (8). In eukaryotes, the lengths of circRNAs are heterogeneous, ranging from ∼100 to 4000 base pairs (9), and are estimated to account for as many as 0.1–10% of molecules in the transcriptome (10).

Bladder cancer (BC) is the 12th most common cancer worldwide (6th most common in males) with 430,000 new cases reported in 2012 (11). Over three-quarters of new cases are diagnosed as non-muscle-invasive bladder cancer (NMIBC: stages Ta/T1/Tis), with the remainder demonstrating invasion of tumor into the detrusor muscle (muscle-invasive bladder cancer, MIBC: stages T2+) (12). MIBC is a life-threatening disease requiring radical treatment and carrying a poor prognosis (13). Although NMIBC is not immediately life-threatening, recurrence after initial treatment is common and 10–15% of cases will progress to MIBC, necessitating burdensome and expensive long-term surveillance (14, 15). NMIBC is thus a heterogeneous disease with regard to both clinical outcomes (recurrence and progression) and underlying biology, with considerable differences in chromosomal alterations and mutational events between grades (low/high grade or grades 1–3) and stages (Ta/T1/Tis) (16). Furthermore, expression analyses derived from largescale RNA sequencing initiatives have tended to classify NMIBCs into three subtypes: the Lund Group describe urobasal A (UroA), genomically unstable (GU), and infiltrated, overlapping with the UROMOL classification of Class 1 (UroA, -progression signature, -CIS signature), Class 2 (GU, +progression signature, +CIS signature), and Class 3 (pronounced expression of lincRNAs, decreased expression of genes associated with cell cycle and metabolic processes,

and increased expression of genes associate with histone modifications and chromatin remodeling) (17).

More recently, the integration of chromosomal and expression alterations with the mutational landscape has identified six molecular subtypes of BC with different molecular features, prognoses and distributions between NMIBC and MIBC (18): the Neural-like subtype is prevalent in MIBC and characterized by high WNT/β-catenin signaling; HER2-like is distributed evenly across NMIBC and MIBC, with higher *ERBB2* amplification and signaling; Papillary-like is a NMIBC subtype enriched in urothelial differentiation genes with a high frequency of actionable *FGFR3* mutations, amplifications, and *FGFR3-TACC3* fusion; Luminal-like is also predominantly NMIBC, has higher MAPK signaling and more *KRAS* and *KMT2C/D* mutations than other subtypes; Mesenchymal-like and Squamous-cell carcinoma-like are predominant in MIBC. Importantly, about 20% of NMIBCs show MIBC subtype traits and a lower 5-yr OS rate than Papillary-like NMIBC (81 vs. 96%) (18).

Genomic and epigenomic approaches may thus inform the development of more accurate risk stratifiers and non-invasive diagnostics (19), tools that are urgently required by healthcare professionals and patients alike (20). CircRNAs also appear to have a role to play in this already complex setting as a novel class of prognostic biomarkers for NMIBC—in a pioneering study, Okholm et al. (21) evaluated circRNAs in a large cohort (n = 457) of NMIBCs and identified circRNAs that are differently expressed between high and low risk tumors, highlighting two (circHIPK3 and circCDYL) as potential biomarkers (21).

In this study, we investigate circRNAs in a set of 58 tumors and evaluate circRNAs, demonstrating that for a subset, their relative expression is tumor grade dependent and may represent an alternative or additional molecular classification. We also provide an initial analysis of the overlap of the circRNAs detected in our study with the results of Okholm et al. in order to evaluate reproducibility between different tumor sample sets and alternative computational pipelines.

MATERIALS AND METHODS

Samples and Library Preparation for RNA-Sequencing

Fresh-frozen tumors were collected as part of the West Midlands' Bladder Cancer Prognosis Programme (BCPP). This study was carried out in accordance with the recommendations of the International Committee on Harmonization Good Clinical Practice (ICH GCP) guidelines. The protocol was approved by the NHS Health Research Authority East Midlands—Derby Research Ethics Committee (ref: 06/MRE04/65). All participants gave written informed consent for the donation of tissue biospecimens and their subsequent utilization in biomedical research. RNA sequencing was performed on RNA extracted from 58 snap frozen incident bladder cancers (urothelial carcinomas) using RNeasy mini kits (Qiagen). Sequencing libraries were prepared from total RNA using Truseq Stranded Total RNA kits with Ribo-zero Gold ribosomal RNA depletion (Illumina) and paired-end sequenced using the Illumina NextSeq

platform (2 × 75 bp). On average, 48 million reads were obtained per tumor sample.

QC and Alignment

Raw fastq files were first processed using FastQC which makes diagnostic plots; subsequently, for each read, Trimmomatic (version 0.32) (22) was used to trim two bases from the start and clip where average phred score quality fell below 20. Quality-checked fastq reads were then mapped to the human genome (version GRCh37) and Ensembl gene annotation (release 87) using STAR aligner (ver 2.5.2b) (23).

circRNA Prediction

Two different strategies were used to identify potential back-splicing events in the transcriptome which predict circular RNAs (circRNAs) to result in an *in-silico* reconstruction of circRNA coordinates. The CircExplorer2 algorithm (24) was used on the "chimeric.out.junction" file (from STAR aligner) containing information on potential non-linear alignments obtained for each sample. The second method, the DCC algorithm (25), differs from CircExplorer as it applies STAR separately on the R1 and R2 from the paired-end fastq data. To retain a high confidence data set, the output from each algorithm was filtered to select only those cases which had ≥4 reads supporting the back-splicing event in at least one of the samples.

Of the 4,571 circRNA candidates predicted by DCC, ∼66% were also detected by CircExplorer2. Since DCC also provides count data for circRNAs and counterpart linear RNAs (host gene's mRNA) to permit calculation of circular-to-linear ratio, the output of the DCC algorithm was chosen for downstream analyses.

Positional Bias of Exonic circRNAs

To identify any positional bias within the gene body, the genomic coordinates corresponding to RefSeq genes and their untranslated regions (UTRs) and coding sequences (CDS) were retrieved from UCSC Table Browser (hg19). In addition, bedtools (ver. 2.27.1) was used to overlap the exonic circRNA coordinates individually with the CDS, 5′ and 3′ UTR coordinates.

Differential Expression

For the back-splice junctions implicated, read counts were resolved into circRNA and linear RNA (junction count) components using DCC. To control for different library sizes, circular-to-linear ratios were calculated after normalizing for RNA-seq library size resulting in counts per million (CPM); relative expression values were computed by $\log_2$ [(CPM(circularRNA)+1)/(CPM(linearRNA)+1)]. Circular-to-linear ratios were used to perform Analysis of Variance (ANOVA) using R (ver. 3.5.1) to identify tumor grade discriminating circRNA candidates (at adjusted p-value 0.05).

Pathway Analysis

GO enrichment was carried out between a list of differentially expressed genes and the background of all parental genes of circRNAs using Gorilla (26). To compare sets of interests against the complete human background, we used the functional enrichment tool of the STRING database (version 11.0) (27). To

further explore the biological processes affected by differentially expressed genes, we carried out pathway analyses using the Enrichr online tool (28). Pathways were adjudged using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database with a p-value threshold of <0.05.
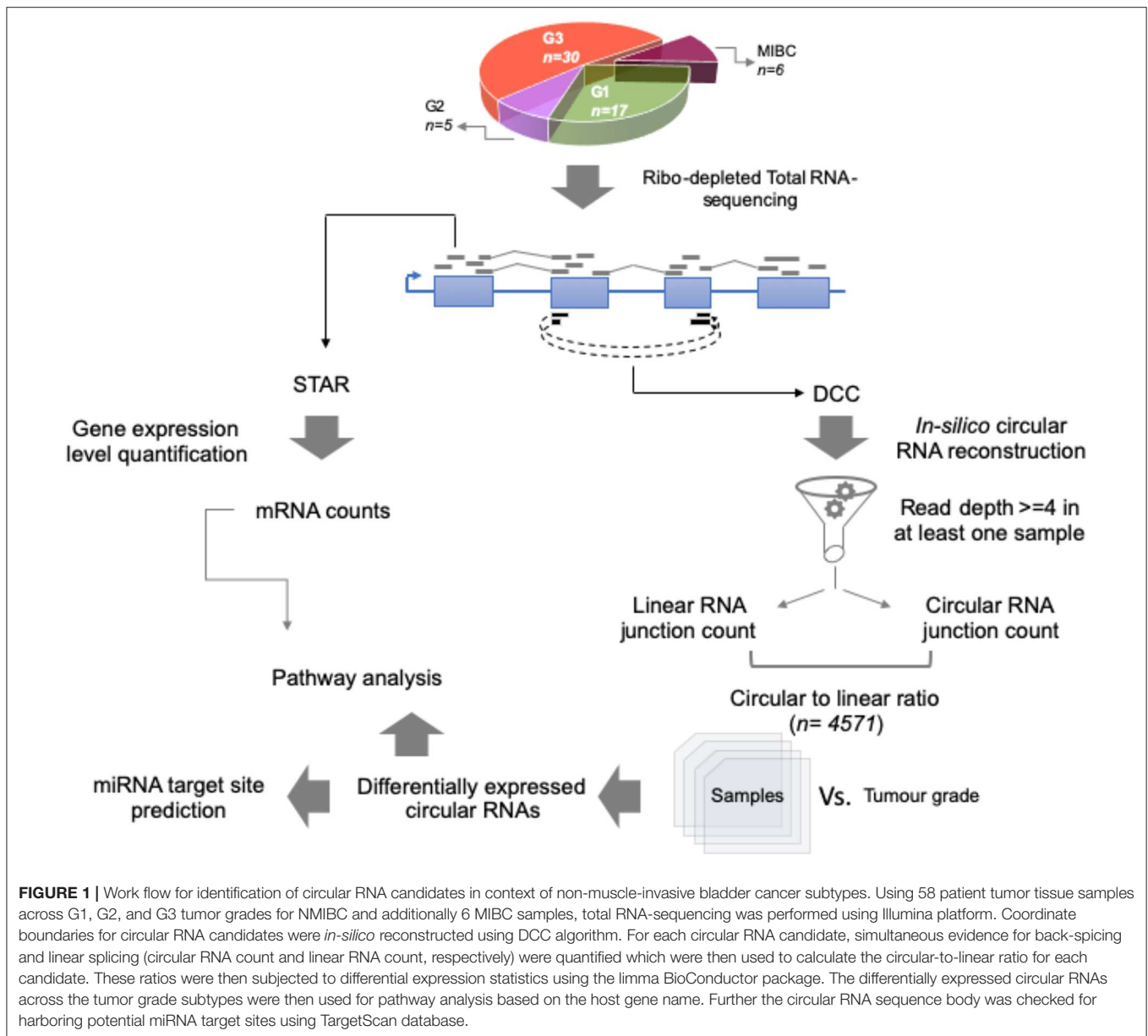
In order to investigate whether the circRNAs showed different behavior compared to host genes in selected pathways, we selected previously-reported important bladder cancer pathways (29, 30). These include the pathways "DNA repair," ERBB signaling, PI3K/AKT signaling, WNT signaling, EGFR signaling, MAPK pathway, and "chromatin remodeling," as defined in the Signatures Database (MSigDB) (ver.7). In order to visualize the relationship of circRNAs to their parental genes in these pathways, we investigated if circRNA events were detected for each pathway member; if not, genes were excluded from further analysis. For the selected genes in the respective pathways, gene expression counts and circRNA counts were tabulated in two separate matrices. These matrices were then individually processed for box plot using ggplot2. To further describe the relationship of gene vs. circRNA expression, we calculated Pearson correlation values between the expression levels of the parental gene to the expression level of the circular RNAs (after normalizing the read counts for the circRNAs by library size). In order to assess potential correlations of circRNA expression to the overall pathway activity, we computed pathway signature scores (using geometric mean over the individual gene expression values, adding a pseudo-count of the lowest expression value in the set). These pathway signature scores were then correlated to the expression of the individual circRNAs in the pathway ($\log_2$ CPM with initial pseudo-count one) and to the relative expression (as used in the differential expression analysis).

microRNA Binding Site Prediction

To identify microRNA (miRNA) binding sites within the predicted circRNAs, we first downloaded the predicted target sites of conserved miRNA families from TargetScan (release 7.2) (31). Of the 122,607 total target sites in the dataset, there were only 192 sites with length >8 bp and, hence, only sites of length 7/8 bp (99.84% of total set) were taken forward for further analysis. This filtered target dataset was then assessed for coordinate level overlap within the circRNA coordinates using bedtools (ver. 2.27.1).

Comparison to Other Datasets

The full set of circRNAs (supported by at least two reads in at least two different samples) from the study of Okholm et al. was kindly provided by the authors. CircRNAs detected in urine by exome capture were downloaded from the supplement from Vo et al. (32) If necessary, circle coordinates were lifted over using a python script based on the python package pyliftover (version 0.4, https://pypi.org/project/pyliftover/) and slightly different coordinate notions (as start coordinate given by either 0, or +1) have been unified and entries mapped using an in-house perl script. In order to compare if the differentially expressed circRNAs identified show the same behavior between grades in the Okholm et al. data, we computed their relative frequency (occurrence in a grade grouping divided by the fraction of that

FIGURE 1 | Work flow for identification of circular RNA candidates in context of non-muscle-invasive bladder cancer subtypes. Using 58 patient tumor tissue samples across G1, G2, and G3 tumor grades for NMIBC and additionally 6 MIBC samples, total RNA-sequencing was performed using Illumina platform. Coordinate boundaries for circular RNA candidates were *in-silico* reconstructed using DCC algorithm. For each circular RNA candidate, simultaneous evidence for back-spicing and linear splicing (circular RNA count and linear RNA count, respectively) were quantified which were then used to calculate the circular-to-linear ratio for each candidate. These ratios were then subjected to differential expression statistics using the limma BioConductor package. The differentially expressed circular RNAs across the tumor grade subtypes were then used for pathway analysis based on the host gene name. Further the circular RNA sequence body was checked for harboring potential miRNA target sites using TargetScan database.

grade in the dataset) in low-grade and high-grade samples from Okholm et al. (removing the 7 papillary urothelial neoplasm samples in the set). We computed the same frequencies in our dataset by combining G1 and G2 into a "low/intermediate grade" set, and keeping G3 as high grade. For both sets, circRNAs with occurrence in at least three samples were selected for analysis. We correlated these relative frequencies using Pearson correlation.

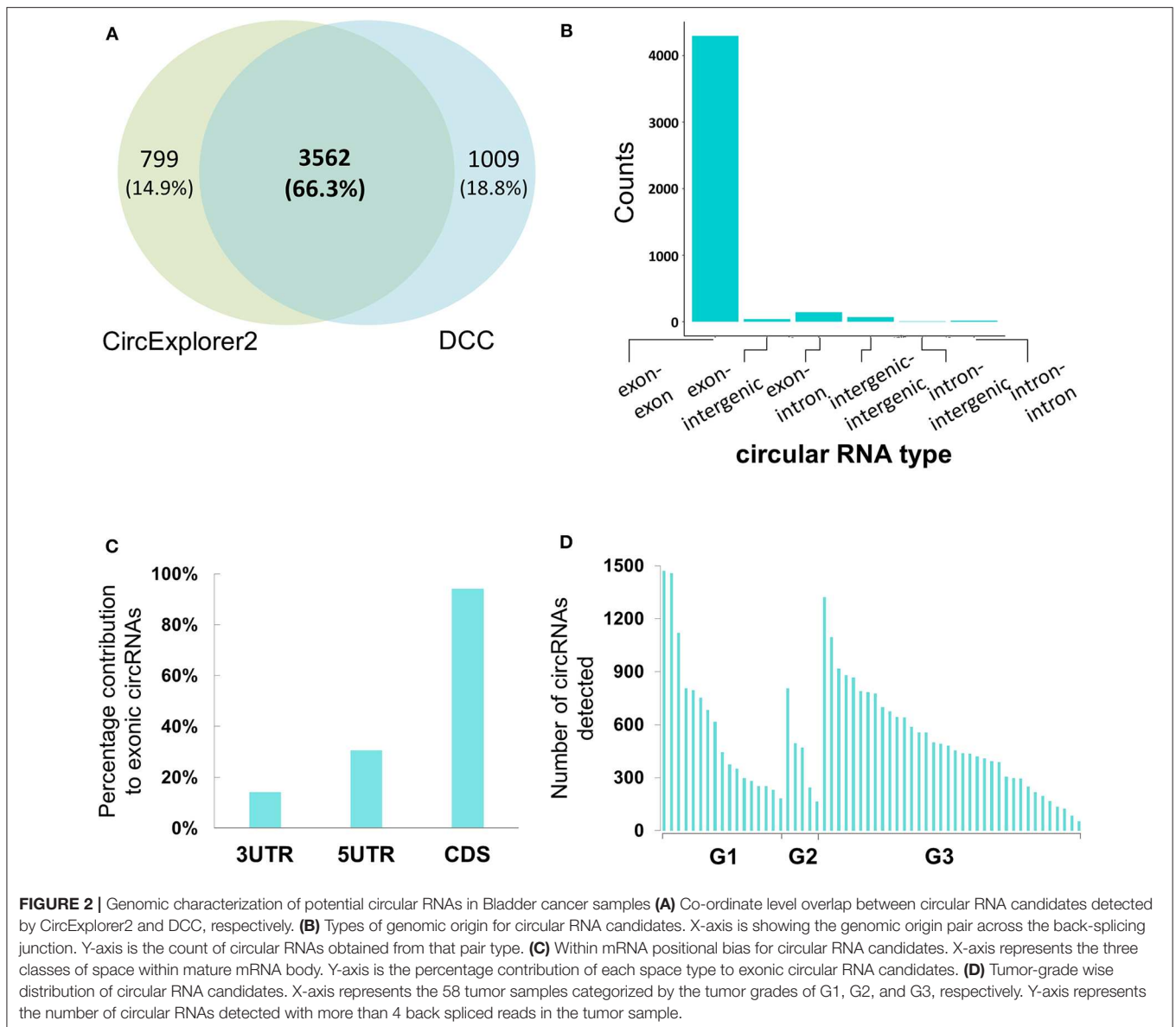## RESULTS

Clinical Phenotype of NMIBC Patient Samples

Our study cohort includes 52 NMIBC samples (grade1/G1 $n = 17$, grade 2/G2 $n = 5$, and grade 3/G3 $n = 30$) and 6 MIBC

samples (all G3). The median age of the patients was 71 years and the male:female ratio was 6:1. There was no statistically significant difference in age distribution between the genders (Mann-Whitney p-value 0.26).

Identification of Circular RNAs

For the total RNA sequencing of all 58 tumor samples, an average of 48 million reads per sample was obtained. To comprehensively identify circular RNA candidates and query the transcriptome status we used two strategies—parallel evaluation of the linear RNA and the circular RNA landscapes (**Figure 1**).

For the circular RNA landscape, two different algorithms, CircExplorer2 and DCC, were used. While individually the two algorithms detected 4,361 and 4,571 candidates, respectively, there was significant overlap in the candidates identified

FIGURE 2 | Genomic characterization of potential circular RNAs in Bladder cancer samples **(A)** Co-ordinate level overlap between circular RNA candidates detected by CircExplorer2 and DCC, respectively. **(B)** Types of genomic origin for circular RNA candidates. X-axis is showing the genomic origin pair across the back-splicing junction. Y-axis is the count of circular RNAs obtained from that pair type. **(C)** Within mRNA positional bias for circular RNA candidates. X-axis represents the three classes of space within mature mRNA body. Y-axis is the percentage contribution of each space type to exonic circular RNA candidates. **(D)** Tumor-grade wise distribution of circular RNA candidates. X-axis represents the 58 tumor samples categorized by the tumor grades of G1, G2, and G3, respectively. Y-axis represents the number of circular RNAs detected with more than 4 back spliced reads in the tumor sample.

(**Figure 2A**). In addition to the junction count for back-splicing events, the output from the DCC algorithm also provided the paired count for the linear events which allowed comparative evaluation; hence, for downstream analyses, results from the DCC were carried forward.

Genomic Characterization of Circular RNAs

When reconstructing the boundaries of the circRNA candidates, the DCC algorithm annotates them on the basis of their genomic context (exonic, intronic, and/or intergenic); amongst all the circRNAs thus annotated, the exonic circRNAs were predominant (∼94%) (**Figure 2B**). We then assessed positional bias (within the mRNA body) of the exons involved in circRNA formation. On further resolution of exonic circRNAs over the mature RNA sections of either the untranslated region (UTR; 5′ or 3′) and coding sequence (CDS), 94.1% of the exonic circRNAs

were found to be within or overlapping the CDS; the values for those within or overlapping 5′ and 3′ UTR were 30.5 and 14.1%, respectively (**Figure 2C**). In addition to these, we looked at the tumor grades for presence of more number of circular RNA candidates in one grade vs. the other. We find that the median number of circular RNAs detected when seen grade-wise are 444, 471, and 468.5 for G1, G2, and G3 respectively (**Figure 2D**). Hence there didn't appear to be any strong bias in the distribution of circRNAs across the tumor grades.

Gene Expression Differences Delineating NMIBC Tumor Grades

Gene level counts estimated from the total RNA-sequencing of the 58 tumor samples were used to identify genes with differentially expressed patterns between tumor grades. A total of 1,071 genes were identified at p-value threshold of 0.05 (after multiple-testing correction). Functional annotation of the gene
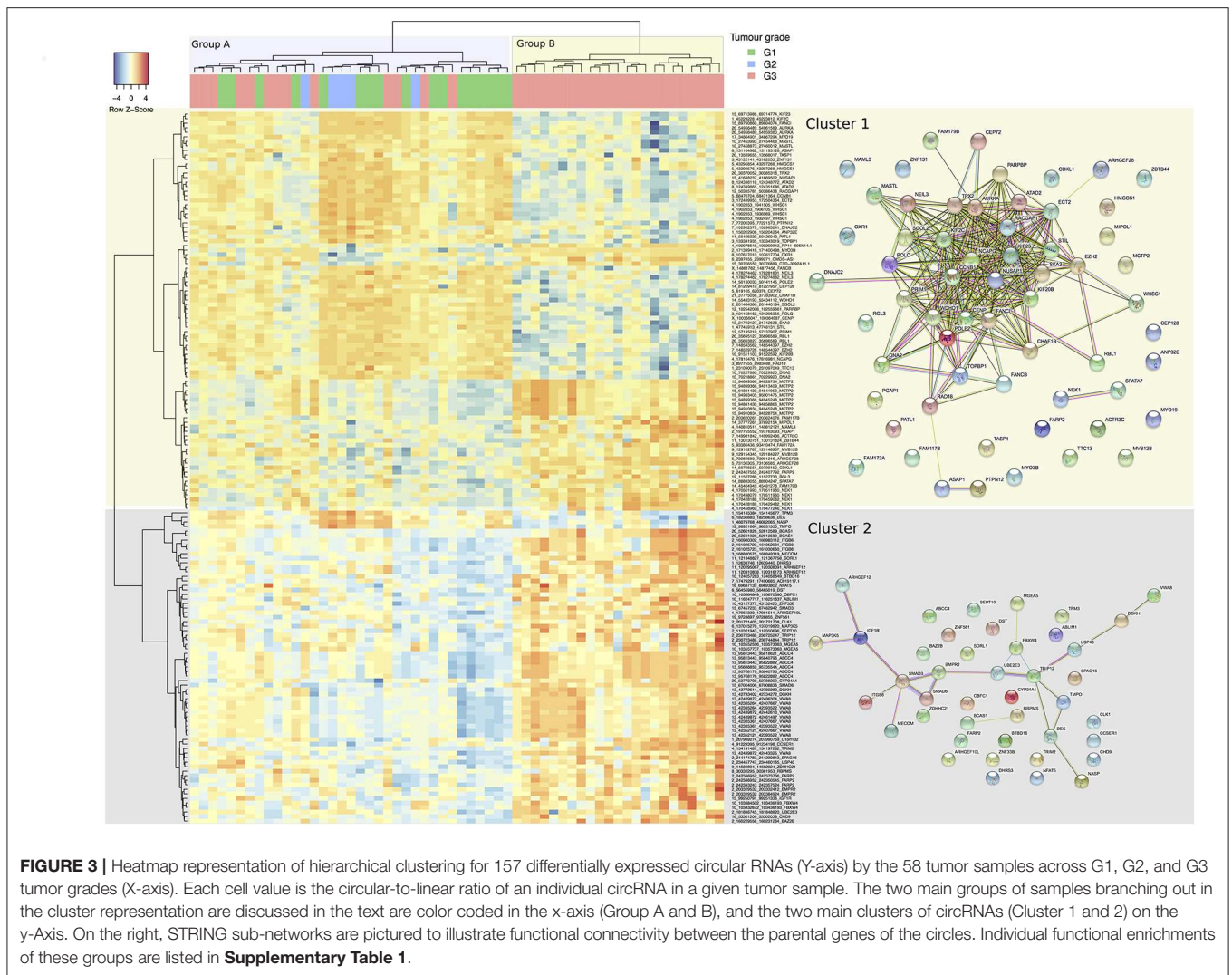
FIGURE 3 | Heatmap representation of hierarchical clustering for 157 differentially expressed circular RNAs (Y-axis) by the 58 tumor samples across G1, G2, and G3 tumor grades (X-axis). Each cell value is the circular-to-linear ratio of an individual circRNA in a given tumor sample. The two main groups of samples branching out in the cluster representation are discussed in the text are color coded in the x-axis (Group A and B), and the two main clusters of circRNAs (Cluster 1 and 2) on the y-Axis. On the right, STRING sub-networks are pictured to illustrate functional connectivity between the parental genes of the circles. Individual functional enrichments of these groups are listed in **Supplementary Table 1**.

set was performed and the pathways found to be significant were predominantly DNA-repair and cell cycle related.

Identification of Dysregulated Circular RNA Between Tumor Grades

Using the DCC algorithm at the confidence threshold of ≥4 supporting reads in at least one sample, 4,571 circRNA candidates were detected. These candidates came from 2,430 unique genes with a ratio of 1.88 circRNAs per gene. The ANOVA analysis, based on relative expression, identified 157 circRNA candidates from 107 unique host genes as having significantly different levels of expression between the tumor grades (set "differentially relative expression of circular RNAs," **Supplementary Table 1**). The parental gene list is functionally enriched in processes such as cell-cycle, DNA-repair, and cytokinesis when compared to the background of all parental genes of circRNAs detected (as reported by a GO enrichment analysis) (**Supplementary Table 1**). A heatmap based on these discriminative circRNAs is shown in (**Figure 3**), indicating their discriminative potential between tumor grades: the samples were

grouped by the clustering into two main subgroups (denoted groups A and B in (**Figure 3**), separated by the initial branch, at the top of the clustering tree, on the X-axis). The first group comprises an inhomogeneous sub-cluster (A) with a mix of different grades, whereas the second main group (B) is homogenous for samples with grade 3. On the Y-axis, the circular RNAs are grouped by the clustering into two main clusters (Cluster 1 and 2). Both clusters exhibit sub-clusters with distinct expression patterns. In Cluster 1, for example, there are two sub-clusters with low expression of circRNAs predominant in the Group B samples, but rather upregulated in most of the Group A ones (mostly in the upper area of the heatmap). The parental genes of these circRNAs are enriched with functional processes relating to DNA replication, cell-cycle, and DNA repair (**Supplementary Table 1**), and reveal a highly connected module in terms of functional interactions (right panel of STRING interactions in **Figure 3**). This pattern is inverted in Cluster 2, for which Group B shows mostly high expression. However, the parental genes of circRNAs in this cluster do not exhibit a large amount of functional connectivity.

Gene Ontology enrichment (albeit not significant after correction for multiple testing) for this cluster suggests involvement of circRNAs in regulative processes including SMAD3 and 6 (indicating a potential connection to TGF-b and BMP signaling) and Insulin-Like Growth Factor 1 Receptor, implying a potential role of these circular RNAs in facilitating progression and evasion of apoptosis (**Supplementary Table 1**), nevertheless, a functional interpretation remains difficult. Compared with a clustering based upon significantly differentially expressed genes (**Supplementary Figure 1**), the clustering appears of similar if not better discrimination between grades: as mentioned, Group B comprises Grade 3 tumors only with a distinct expression pattern of circRNAs. A separate but large fraction of Grade 3 tumors are clustered together with Grade 1 and Grade 2 samples in Group A. Grade 2 exhibits a tendency to cluster together in Group A which shows high expression of a small sub-set of circular RNAs (with parental genes TPM3, DEK, NASP, and TMPO) and which is low for most members of Group B.

We also collected circRNA candidates whose parental genes are affected by NMIBC grade as detected by differential expression analysis on the gene level: the set of all genes with any circRNA detected ($n = 2,430$) was overlapped with the set of differentially expressed genes ($n = 1,071$; **Supplementary Figure 1**). This resulted in an intersection of 143 host genes that are differentially expressed between tumor grades and exhibit circularization (set "gene differentially expressed at mRNA level," **Supplementary Table 1**). Finally, genes where the circRNAs are differentially expressed with little influence of the host gene expression dynamic between the different grades were identified by subtracting the "circles in differentially expressed genes ($n = 107$)" set from the "differentially relative expression of circular RNAs" set ($n = 143$): 24 such circRNAs were identified (**Supplementary Table 1**). These 24 circRNAs did not exhibit functional enrichment.

Pathway Analysis

Visual inspection of the expression levels of linear RNA and circRNA (gene level summarized) for genes with circRNA in bladder cancer relevant pathways (**Supplementary Figure 2**), indicates non-regular patterns of circRNA expression which do not follow obvious trends in relation to the linear form. Whereas, for the linear RNA the spread of expression level per gene tended to be more tightly bound, the circRNA spread was much more variable across genes, samples and tumor grades. In concordance, most circRNA expression levels show only poor correlation to the expression of the parental gene (**Supplementary Table 3**). In general, the expression of individual circRNAs within these pathways does not correlate with the overall pathway activity (**Supplementary Table 2**). This is also the case when correlating the relative circular expression with the pathway activity. However, more significantly correlated instances can be detected in the latter analysis (at a P-value cut-off at 0.05: 66 for the relative vs. 174 for the overall expression, **Supplementary Table 2**), for example, a circRNA originating from FANCI, a gene involved in DNA repair, shows a strong negative correlation with the DNA repair process ($R = -0.85, P < 2.2^{-16}$).

Frequency of MicroRNA (miRNA) Binding Sites Within circRNAs

Using the TargetScan dataset of predicted miRNA target sites, a total of 132 circRNAs (hosted by 99 unique genes) were found to harbor target sites for 141 miRNAs. Amongst the host genes whose circRNAs harbor miRNA target sites, 4 were differentially expressed (both at the mRNA and at circRNA level) between the tumor grade subtypes: *CDKL1*, *HP1BP3*, *MVB12B*, and *TMPO*. The circRNAs from these 4 host genes harbor target sites for 11 different miRNAs: miR-7-5p, miR-30-5p, miR-31-5p, miR-96-5p/1271-5p, miR-139-5p, miR-181-5p, miR-182-5p, miR-433-3p, miR-489-3p, miR-493-5p, and miR-543. The 5 most recurrent miRNAs with target sites within circRNAs are the miR-15/107 family, miR-101-3p, miR-204-5p/211-5p, and miR-203a-3p. Many of these miRNAs have been reported to be implicated in tumor biology and/or be predictive of response to drug treatment (33–38). The results are summarized in (**Supplementary Table 1**).

Comparison to Other Datasets

The comparison between our dataset and that of Okholm et al. identified 3,361 overlapping circRNAs [77% from this study, 22% of the 15,223 instances in the data of Okholm et al. (21)] (all instances listed in **Supplementary Table 4**). For the circRNAs significantly differentially expressed between grades, 108 could be found in both datasets (69%). To each of these 108 circRNAs we assigned their relative frequency to occur in the set of low/medium and high-grade tumors, respectively, to describe their tendency to be grade specific. We compared these between the two datasets and found a strong correlation ($R = 0.47, p = 2.3e-07$; **Supplementary Figure 3**), indicating that their grade-specific behavior can be detected independent of the cohort. Notably, the identified grade-stratifying candidate circRNAs do not include the two proposed progression biomarkers identified by Okholm et al., circHIPK3 and circCDYL.

We identified 552 of our 4,116 circRNAs (12%) in the urine dataset from 13 prostate cancer patients provided by Vo et al. (32). This urine dataset comprises 1,092 circRNAs, resulting in a 50% recall of the urinary circRNAs in at least one of our samples. Within the set of significant differentially expressed genes ($n = 157$), 8% (13 circRNAs) can be found in the urine dataset (**Supplementary Table 4**).

DISCUSSION

In this work we have investigated circRNAs from a cohort of NMIBC patients. The general genomic properties of the circRNAs are in keeping with earlier findings in other studies: circRNAs are mostly connecting exons, and are mostly detected within coding sequences. The total number detected varied between tumors and was lowest in grade 2 tumors (although only 5 grade 2 tumors were analyzed). However, detected number of individual circRNAs was always >100 and up to >1,000, providing a considerable dataset for further investigations. Since our dataset is well-annotated by tumor grade (G1, G2, and G3), we aimed to delineate a subset that stratified the

dataset by grade using a supervised approach. This approach selected 157 circRNAs, and cluster analysis resulted in distinct subgroups of tumors (**Figure 3**). The clustering presented here is a visualization of the results of the differential relative circular expression analysis. The resulting heatmap might slightly vary by choice of parameters and clustering algorithms; however, the result implies the existence of molecular subtypes in terms of circRNA expression, with varying tendencies of certain grades to exhibit the subtype.

Gene expression differences among NMIBCs have been reported previously, and pathways relevant to tumor biology in bladder cancer are ERBB2, PI3K-AKT, cell cycle, MAPK, and DNA-damage repair. One of the primary questions driving investigation into any form of cis-regulatory RNA is how it affects the expression level of the host gene. In this regard, we investigated the comparative expression levels of the linear and circular RNA for selected pathways. We observe that, for a given pathway, the genes (linear RNA) retain a similar trend of expression pattern between tumor grades, whereas the circRNA levels are much more variable. This could be an effect of increased variability in the circRNA count data since the majority are at low expression levels compared to linear RNA. Nonetheless, for a particular grade group, we observe a discernible shift in the expression pattern of some genes. In the ERBB2 pathway, the *ERBB2* gene is consistently the highest expressing linear gene among all the three tumor grades but, for circRNA, *SOS2* has the highest expression across the grades. Similarly, in the PI3K-AKT pathway, *ACTR3*, *SMAD2*, *STAT2,* and *ACACA* are most highly expressed in linear RNA, whereas for circRNA, *MAPK8* appears to have higher expression. These observations indicate that within the same pathway there are likely different regulatory processes acting on linear vs. circular RNA. This is also indicated by the lack of correlation between the expression levels of circRNAs to their parental genes in the investigated pathways.

We also found that the activation of a pathway (as measured by gene signature score) is mostly independent of the circRNAs within the pathway, but less so when looking at relative expression. These instances, for which the relative expression either positively or negatively correlates with the pathway expression, might be under certain biological constraints. In the case of a negative correlation, the cancer cell seems to avoid upregulation of circRNA and the observation could be explained by a steady level of circRNAs with increasing pathway activity. With a positive correlation there may be a cis-acting role on the pathways, since their relative expression increases with pathway activity but is not bound to the expression of the parental gene. The way by which these circRNAs potentially interact with pathways of interest cannot be directly deduced from this analysis. However, systematic correlation studies may help to identify candidates for functional follow up screens.

Interestingly, four predicted miRNA sponges could be found differentially expressed. We found cancer-related pathway-specific genes being targets of the top recurrent miRNAs (targeting circRNA). The miR-15/107 family is found to have target sites within circRNAs from host genes *CDC42*, *MGEA5*, *CHPT1*, *GPATCH2L*, *TSEN2*, *RPL14*, *PLD1*, *TFRC*, *PDLIM5,* and *PTK2*. The genes *CDC42*, *PLD1,* and *PTK2* are involved in the EGFR signaling

pathway; the miR-15/107 family is reported to have tumor suppressor properties (34, 39), and the circRNAs from the above genes for EGFR signaling pathway can act as "sponges" to exert oncogenic effect in the context of NMIBC. Similarly, miR-204-5p/211-5p has targets within *CDC42*, *KHDRBS1*, *ASH1L*, *MDM2*, *GPATCH2L*, *WSB1*, *ZNF638,* and *AGTPBP1* genes. Along with *CDC42*, *MDM2* is involved in DNA damage responses. Additionally, the duo of miR-204-5p/211-5p has recently been reported to be involved in resistance to BRAF inhibition (40). Hence, circRNAs harboring target sites for miR-204-5p/211-5p can have important implications for tumor treatment and progression (41, 42).

The number of samples in our study is still small, especially for Grade 1 and Grade 2 samples and the reported list of differentially expressed circRNAs provides a list of interesting candidates that can be tested in future studies. However, the existence of another bladder cancer cohort suitable for circRNA detection by Okholm et al. gave us the opportunity to compare two different cohorts and to test our findings in their data. The grade-specific relationships of circRNAs could be identified in the dataset of Okholm et al. (21). This indicates a certain amount of transferability between independent cohorts, despite different computational pipelines, and provides some mutual validation of the two cohorts. The urinary circRNA in the dataset from prostate cancer patients provided by Vo et al. would be expected to comprise both prostate cancer specific circRNAs and circular RNAs from normal bladder tissue, and was therefore suitable to investigate whether circRNAs detected in our study can be detected in urine. Indeed, we found 50% of the circular RNAs in the urine set in at least one of our bladder tumor samples, and 12% of the bladder tumor circRNAs from this study exist in the urine dataset. It is therefore likely that a high proportion of circRNAs expressed by bladder tumors will be detectible in urine, notwithstanding the circRNAs differentiating between grades are slightly under-represented with 8% recall. The latter observation might indicate that the discriminative set comprises instances that are bladder cancer specific; however, given the relatively small size of this dataset, the significance of this finding remains unclear. Nevertheless, these observations indicate a need to further investigate the potential of urinary circRNAs as diagnostic and/or prognostic biomarkers.

CONCLUSION

The present study provides a step in the comprehensive evaluation of circRNAs in the context of bladder cancer. Intriguingly, despite their potential function as microRNA sponges, circRNAs are potentially affecting host mRNA levels at the transcriptional stage, as compared to post-transcriptional control by miRNAs. We have also identified circRNA candidates worthy of further functional investigation, and comprising potential miRNA sponges and circRNAs correlated to pathway expression. Our analysis indicates circRNA differences between bladder cancer grades, and their relative expression levels may provide an additional modality for patient risk stratification. Furthermore, since circRNAs have a longer half-life in the

extracellular milieu than other RNAs and are detectable in urine, they may be useful non-invasive biomarkers for bladder cancer diagnosis and risk stratification.

DATA AVAILABILITY STATEMENT

RNA sequencing data for the bladder cancer patients have been submitted at the European Genome-Phenome Archive (EGA) (https://ega-archive.org/) under accession code: EGAS00001004358.

AUTHOR CONTRIBUTIONS

AG, RA, and RB conceived and designed the study. NG and DW provided technical and material support. BA, MZ, KC, NJ, and RB have been involved in acquisition of funding, samples, and data. AG and RA analyzed the data, AG, RA, DW, and RB wrote and revised the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00812/full#supplementary-material

REFERENCES

1. Qu S, Liu Z, Yang X, Zhou J, Yu H, Zhang R, et al. The emerging functions and roles of circular RNAs in cancer. *Cancer Lett.* (2018) 414:301–9. doi: 10.1016/j.canlet.2017.11.022
2. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol.* (2014) 32:453–61. doi: 10.1038/nbt.2890
3. Liu J, Liu T, Wang X, He A. Circles reshaping the RNA world: from waste to treasure. *Mol Cancer.* (2017) 16:58–58. doi: 10.1186/s12943-017-0630-y
4. Bach DH, Lee SK, Sood AK. Circular RNAs in cancer. *Mol Ther Nucleic Acids.* (2019) 16:118–29. doi: 10.1016/j.omtn.2019.02.005
5. Lei B, Tian Z, Fan W, Ni B. Circular RNA: a novel biomarker and therapeutic target for human cancers. *Int J Med Sci.* (2019) 16:292–301. doi: 10.7150/ijms.28047
6. Zeng X, Lin W, Guo M, Zou Q. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol.* (2017) 13:e1005420. doi: 10.1371/journal.pcbi.1005420
7. Li X, Yang L, Chen LL. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell.* (2018) 71:428–42. doi: 10.1016/j.molcel.2018.06.034
8. Jiang W, Zhang X, Chu Q, Lu S, Zhou L, Lu X, et al. The circular RNA profiles of colorectal tumor metastatic cells. *Front Genet.* (2018) 9:34. doi: 10.3389/fgene.2018.00034
9. Lasda E, Parker R. Circular RNAs: diversity of form and function. *RNA.* (2014) 20:1829–42. doi: 10.1261/rna.047126.114
10. Sekar S, Liang WS. Circular RNA expression and function in the brain. *Noncoding RNA Res.* (2019) 4:23–9. doi: 10.1016/j.ncrna.2019.01.001
11. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
12. Bryan RT, Zeegers MP, van Roekel EH, Bird D, Grant MR, Dunn JA, et al. A comparison of patient and tumour characteristics in two UK bladder cancer cohorts separated by 20 years. *BJU Int.* (2013) 112:169–75. doi: 10.1111/bju.12032
13. Berdik C, Unlocking bladder cancer. *Nature.* (2017) 551:S34. doi: 10.1038/551S34a
14. Wald C. Diagnostics: a flow of information. *Nature.* (2017) 551:S48. doi: 10.1038/551S48a
15. Svatek RS, Hollenbeck BK, Holmang S, Lee R, Kim SP, Stenzl A, et al. The economics of bladder cancer: costs and considerations of caring for this disease. *Eur Urol.* (2014) 66:253–62. doi: 10.1016/j.eururo.2014.01.006
16. Hurst CD, Knowles MA. Mutational landscape of non-muscle-invasive bladder cancer. *Urol Oncol.* (2018). doi: 10.1016/j.urolonc.2018.10.015. [Epub ahead of print].
17. Dyrskjot L, Ingersoll MA. Biology of nonmuscle-invasive bladder cancer: pathology, genomic implications, and immunology. *Curr Opin Urol.* (2018) 28:598–603. doi: 10.1097/MOU.0000000000000543

18. Tan TZ, Rouanne M, Tan KT, Huang RY, Thiery JP. Molecular subtypes of urothelial bladder cancer: results from a meta-cohort analysis of 2411 tumors. *Eur Urol.* (2019) 75:423–32. doi: 10.1016/j.eururo.2018.08.027
19. Ward DG, Gordon NS, Boucher RH, Pirrie SJ, Baxter L, Ott S, et al. Targeted deep sequencing of urothelial bladder cancers and associated urinary DNA: a 23-gene panel with utility for non-invasive diagnosis and risk stratification. *BJU Int.* (2019) 124:532–44. doi: 10.1111/bju.14808
20. Bessa A, Maclennan S, Enting D, Bryan R, Josephs D, Hughes S, et al. Consensus in bladder cancer research priorities between patients and healthcare professionals using a four-stage modified Delphi method. *Eur Urol.* (2019) 76:258–59. doi: 10.1016/j.eururo.2019.01.031
21. Okholm TLH, Nielsen MM, Hamilton MP, Christensen LL, Vang S, Hedegaard J, et al. Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer. *npj Genomic Med.* (2017) 2:36. doi: 10.1038/s41525-017-0038-z
22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
24. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* (2016) 26:1277–87. doi: 10.1101/gr.202895.115
25. Cheng J, Metge F, Dieterich C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics.* (2016) 32:1094–6. doi: 10.1093/bioinformatics/btv656
26. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* (2009) 10:48. doi: 10.1186/1471-2105-10-48
27. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* (2019) 47:D607–13. doi: 10.1093/nar/gky1131
28. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) 44:W90–7. doi: 10.1093/nar/gkw377
29. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Hoyer S, Ulhoi BP, et al. Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer Cell.* (2016) 30:27–42. doi: 10.1016/j.ccell.2016.05.004
30. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell.* (2017) 171:540–56.e25. doi: 10.1016/j.cell.2017.09.007
31. Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* (2015) 4:e05005. doi: 10.7554/eLife.05005

32. Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell.* (2019) 176:869–81. doi: 10.1016/j.cell.2018.12.021

33. Wang WX, Kyprianou N, Wang X, Nelson PT. Dysregulation of the mitogen granulin in human cancer through the miR-15/107 microRNA gene group. *Cancer Res.* (2010) 70:9137–42. doi: 10.1158/0008-5472.CAN-10-1684

34. Wang S, Zhu W, Xu J, Guo Y, Yan J, Meng L, et al. Interpreting the MicroRNA-15/107 family: interaction identification by combining network based and experiment supported approach. *BMC Med Genet.* (2019) 20:96. doi: 10.1186/s12881-019-0824-9

35. Li B, Xie D, Zhang H. MicroRNA-101-3p advances cisplatin sensitivity in bladder urothelial carcinoma through targeted silencing EZH2. *J Cancer.* (2019) 10:2628–34. doi: 10.7150/jca.33117

36. Liu D, Li Y, Luo G, Xiao X, Tao D, Wu X, et al. LncRNA SPRY4-IT1 sponges miR-101-3p to promote proliferation and metastasis of bladder cancer cells through up-regulating EZH2. *Cancer Lett.* (2017) 388:281–91. doi: 10.1016/j.canlet.2016.12.005

37. Jiang N, Jiang X, Chen Z, Song X, Wu L, Zong D, et al. MiR-203a-3p suppresses cell proliferation and metastasis through inhibiting LASP1 in nasopharyngeal carcinoma. *J Exp Clin Cancer Res.* (2017) 36:138. doi: 10.1186/s13046-017-0604-3

38. Aakko S, Straume AH, Birkeland EE, Chen P, Qiao X, Lonning PE, et al. MYC-Induced miR-203b-3p and miR-203a-3p control Bcl-xL expression and paclitaxel sensitivity in tumor cells. *Transl Oncol.* (2019) 12:170–9. doi: 10.1016/j.tranon.2018.10.001

39. Chai L, Kang XJ, Sun ZZ, Zeng MF, Yu SR, Ding Y, et al. MiR-497-5p, miR-195-5p and miR-455-3p function as tumor suppressors by targeting hTERT in melanoma A375 cells. *Cancer Manage Res.* (2018) 10:989–1003. doi: 10.2147/CMAR.S163335

40. Diaz-Martinez M, Benito-Jardon L, Alonso L, Koetz-Ploch L, Hernando E, Teixido J. miR-204-5p and miR-211-5p contribute to BRAF inhibitor resistance in melanoma. *Cancer Res.* (2018) 78:1017–30. doi: 10.1158/0008-5472.CAN-17-1318

41. Kurahashi R, Kadomatsu T, Baba M, Hara C, Itoh H, Miyata K, et al. MicroRNA-204-5p: A novel candidate urinary biomarker of Xp11.2 translocation renal cell carcinoma. *Cancer Sci.* (2019) 110:1897–908. doi: 10.1111/cas.14026

42. Yu Y, Wang Y, Xiao X, Cheng W, Hu L, Yao W, et al. MiR-204 inhibits hepatocellular cancer drug resistance and metastasis through targeting NUAK1. *Biochem Cell Biol.* (2019) 97:563–70. doi: 10.1139/bcb-2018-0354