Received: 15 January 2023 | Revised: 30 May 2023 | Accepted: 7 June 2023

DOI: 10.1002/appl.202300002

Applied Research

REVIEW

Transcriptome profiling research in urothelial cell carcinoma

Umar Ahmad[1,2] | Buhari Ibrahim[3,4] | Mustapha Mohammed[5,6] | Ahmed Faris Aldoghachi[7] | Mahmood Usman[8,9] | Abdulbasit Haliru Yakubu[10,11] | Abubakar Sadiq Tanko[12] | Khadijat Abubakar Bobbo[13,14] | Usman Adamu Garkuwa[4,15] | Abdullahi Adamu Faggo[16] | Sagir Mustapha[17] | Mahmoud Al-Masaeed[18,19] | Syahril Abdullah[7,13] | Yong Yoke Keong[20] | Abhi Veerakumarasivam[21]

[1]Molecular Genetics Informatics, Department of Anatomy, Faculty of Basic Medical Science, Bauchi State University, Gadau, Bauchi, Nigeria

[2]Institute of Pathogen Genomics, Division of Laboratory Systems and Networks, Africa Centres for Disease Control and Prevention (Africa CDC), African Union Commission, Addis Ababa, Ethiopia

[3]Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[4]Department of Physiology, Faculty of Basic Medical Sciences, Bauchi State University, Gadau, Bauchi, Nigeria

[5]Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmaceutical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria

[6]Vice President for Medical and Health Science Office, QU Health, Qatar University, Doha, Qatar

[7]Medical Genetics Laboratory, Genetics and Regenerative Medicine Research Centre, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[8]Department of Human Anatomy, Faculty of Basic Medical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria

[9]Department of Human Anatomy, Faculty of Basic Medical Sciences, Yusuf Maitama Sule University, Kano, Nigeria

[10]Faculty of Pharmacy, University of Maiduguri, Maiduguri, Borno, Nigeria

[11]Department of Pharmaceutical Service, University of Maiduguri Teaching Hospital, Maiduguri, Borno, Nigeria

[12]Department of Biochemistry, Faculty of Science, Bauchi State University, Gadau, Bauchi, Nigeria

[13]UPM-MAKANA Cancer Research Laboratory, Institute of Bioscience, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[14]Department of Anatomy, Faculty of Basic Medical Sciences, Gombe State University, Gombe, Nigeria

[15]Department of Human Physiology, Faculty of Basic Medical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria

[16]Department of Microbiology, Faculty of Science, Bauchi State University, Gadau, Bauchi, Nigeria

[17]Department of Pharmacology, Universiti Sains Malaysia, Kota Bharu, Kelantan, Malaysia

[18]Department of Nursing, Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia

[19]Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[20]Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[21]Department of Medical Sciences, School of Healthcare and Medical Sciences, Sunway University, Bandar Sunway, Selangor, Malaysia

Correspondence
Umar Ahmad, Molecular Genetics Informatics, Department of Anatomy, Faculty of Basic Medical Science, Bauchi State University, Gadau, Bauchi PMB 65, Nigeria.
Email: umarahmad@basug.edu.ng

Abstract

Urothelial cell carcinoma (UCC) is the ninth most common cancer that accounts for 4.7% of all new cancer cases globally. UCC development and progression are due to complex and stochastic genetic programs. To study the cascades of molecular events underlying the poor prognosis that may lead to limited treatment options for advanced disease and resistance to conventional therapies in UCC, transcriptomics technology (RNA-Seq), a method of analyzing the RNA content of a sample using

modern high-throughput sequencing platforms has been employed. Here, we review the principles of RNA-Seq technology and summarize recent studies on human bladder cancer that employed this technique to unravel the pathogenesis of the disease, identify biomarkers, discover pathways and classify the disease state. We list the commonly used computational platforms and software that are publicly available for RNA-Seq analysis. Moreover, we discussed the future perspectives for RNA-Seq studies on bladder cancer and recommended the application of a new technology called single-cell sequencing to further understand the disease.

KEYWORDS
bioinformatics, bladder cancer, genomics, RNA-sequencing, transcriptome profiling

INTRODUCTION

Cancer represents one of the most life-threatening diseases posing a major health challenge all over the world [1, 2]. Urothelial cell carcinoma (UCC) also known as bladder cancer, is among the principal causes of cancer-related deaths globally [3, 4]. The incidence of bladder cancer is ranked ninth with a total number of 549,393 new cases and 199,922 deaths globally [5, 6]. These numbers are forecasted to double in 2040, particularly in the developed countries, signifying a severe health crisis in the future [5, 7] which can lead to financial burden on countries due to the exorbitant treatments, high recurrence rates and subsequent need for long-term follow-up [8]. UCC is reported to be the second most frequent malignant tumor of the genitourinary tract, with men at risk four times higher than women [4]. The most common bladder cancer is transitional cell carcinoma or UCC, classified into; non-muscle-invasive bladder cancers (NMIBCs) and muscle-invasive bladder cancers (MIBCs). The NMIBCs account for about 70%–80% of all diagnosed patients, with the tendency of future recurrences and may advance into MIBCs [9]. Whereas MIBCs account for about 20%–25% of patients, signifying an active, locally invasive carcinoma with metastatic potential [10].

Despite diagnostic and therapeutic advances for UCC, if it is not detected early, will remain a big challenge. Although there have been great achievements in its management and treatments that include surgical procedures, radiotherapy, pre- and postoperative treatments, and chemotherapy, there has been no significant progress in survival rates for bladder cancer patients [11]. Moreover, these treatments are presented with various side effects that lead to other health problems [4]. As a result, healthcare providers receive a substantial number of cases with relapse and progression, leading to long-term follow-up. All these challenges made bladder cancer an expensive disease to manage [7]. The need to identify novel molecular markers in bladder cancer to predict medical outcomes, especially in patients with relapse has made researchers in recent years to focus more on the molecular aspects of the disease. This development came with the aim of boosting its biological understanding to unravel the molecular factors that can be utilized as targets for therapies and guiding assessment of risk and help in informed decision-making in clinical practice [12].

Transcriptomic technologies employ methods that study and quantify organism's transcriptome [13], a complete set of RNA transcripts in a cell for a specific developmental stage or physiological state. Understanding transcriptomes is essential for interpreting the functional elements of the genome and revealing the molecular constituents of the cells and tissues to better understand the development of diseases [14]. Identifying transcripts and quantifying the level of gene expression have been the major focus in molecular biology since the discovery of RNA as an essential intermediate between the genome and the proteome [15]. The last few decades have witnessed the use of microarray technology and bioinformatic analysis in screening genetic changes at the genome level which has greatly assisted in the identification of differentially expressed genes (DEGs) and functional pathways implicated in cancer [16]. However, with technological advancement and elucidation of noncoding RNAs, transcriptomic approaches have given room for deeper understanding of the intricacies of the regulation of gene expression, alternate splicing events, functions of noncoding RNAs, and ascertaining the importance of such approaches in the correct construction and annotation of complex genomes [1].

Furthermore, high-throughput RNA-sequencing (RNA-Seq) has opened opportunities for transcriptomic studies and has advanced into a standard technique used in biomedical studies. This technique is currently utilized in the estimation of gene expression, identification of noncoding genes, discovery of new genomic characteristics, and drug discoveries. It is well established that RNA-Seq has strong advantages over the previously developed sequencing techniques [17]. This makes RNA-Seq a vital tool in cancer biology that can be exploited for tumor classification, patient stratification, and monitoring of patients' response to therapy [18]. However, despite the high level of correlation that exists between RNA-Seq and other techniques such as microarray, studies have strongly emphasized the merits of RNA-Seq over the other techniques [19]. The discovery of gene fusion and differential expression of RNA

transcripts that are known for causing diseases are some of the prospects of the RNA-Seq [20]. In this review, we summarized the current literature status of transcriptomic studies carried out on UCC and identified the possible areas for future research.

TRANSCRIPTOME SEQUENCING TECHNOLOGIES

Transcriptome sequencing technology is a computational method that determines the identity and abundance of RNA sequences in a biological sample. This technology has been widely used to study dynamic gene expression in diverse human tissues, including human bladder cancer. The general workflow of RNA-Seq begins with the isolation of RNAs from tissue samples, library preparation, sequencing, and computational analysis (Figure 1). Below we summarize and discuss RNA-Seq techniques, narrowing down to its application in human bladder cancer studies.

PRINCIPLE OF RNA-Seq TECHNOLOGY

RNA-Seq is a transcriptome profiling technology that utilizes next-generation sequencing (NGS) platforms [21]. The principle of RNA-Seq involves the reverse transcription of RNA-Seq transcripts into cDNA, ligation of adapters to each end of the cDNA, sequencing and then aligning them to a reference genome or assembling to obtain de novo transcripts, proving a genome-wide expression profile [22]. The quality and quantity of the starting RNA material are the most important aspects to consider when deciding on the methods to generate RNA-Seq libraries [23, 24]. While working with RNA, it is critical to avoid RNase contamination by using sterile, RNase-free solutions, and plastic ware. It is also important to include quality control (QC) in all stages of operation [23]. The method of RNA isolation affects the ability to detect differentially expressed transcripts [25]. Various methods of RNA extraction exist in the literature; the most popular and widely used methods are TRIzol (Life Technologies) and RNAeasy (Qiagen) [25]. Other methods include density gradient centrifugation, magnetic bead technology, lithium chloride, and urea isolation, Oligo(dt)-cellulose column chromatography, and non-column poly (A)+ purification/isolation [26]. TRIzol employs the guanidinium-acid-phenol extraction and regarded as the "gold standard" but the chemicals used are corrosive and toxic, while, the RNAeasy (and variants) employ the glass fiber filter and silica technology that is fast, simple, and safe to use but expensive [24, 25].

The extracted RNA is treated in solution or on-column with DNase to remove traces of genomic DNA and to prevent contamination of RNA-Seq libraries by DNA. DNA-free RNA could then be assessed for both quality and quantity [24]. In assessing the quality of the RNA, most laboratories utilize the electrophoretic-

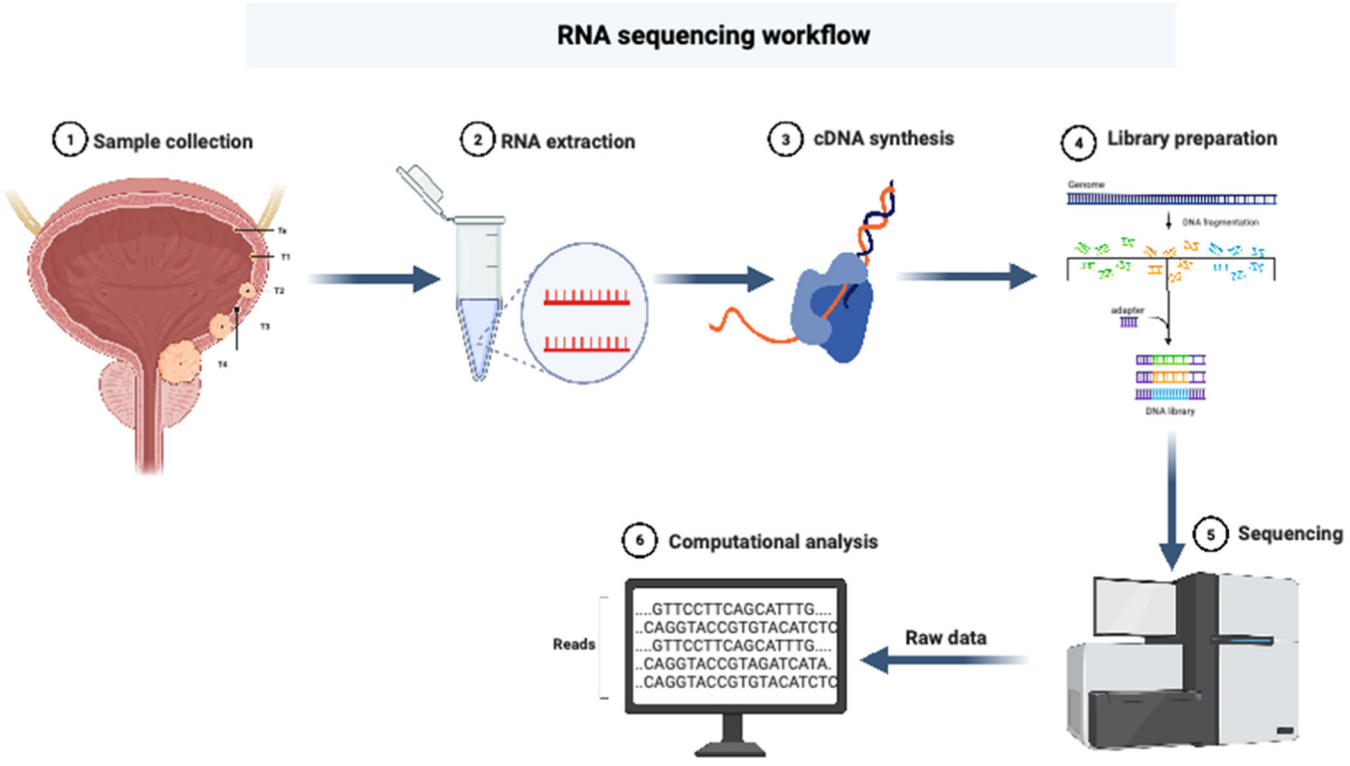

FIGURE 1 Overview of RNA-sequencing on bladder cancer. RNA is first extracted from bladder cancer tissues or cell samples followed by cDNA synthesis through reverse transcription. Then, the cDNAs are amplified for library preparation before subjecting them to next-generation sequencing to generate an output of raw FASTQ data files (reads) that could be used for downstream computational and bioinformatics analyses. The figure is designed and created with BioRender.com.

based system [22, 24]. The quality of the RNA produced is measured through the RNA integrity score, RNA Integrity Numbers (RINs) [23, 24]. The measure reads per kilobase of exon model per million reads (RPKM) and its variance are the most frequently adopted in the measurement of RNA-Seq expression value [15]. The acquisition of RNA-Seq data consists of several steps, and in each of these steps, specific QC checks are applied to monitor the quality of the data output produced [15].

RNA-Seq LIBRARY PREPARATION

In the last decade, there has been a considerable understanding of the dynamic nature of organism's genome [13, 27], this is not limited to the modifications of DNA sequence but also to the quantitative and qualitative measurements of RNA sequences [13]. Before sequencing a sample, a library must be prepared for that sample [13, 27, 28]. A library is a collection of randomly sized DNA fragments representing the sample input [27]. Depending on the type of NGS application, different library preparation steps are available [27–29]. Also, the sample preparation required for RNA sample sequencing varies depending on the type of RNA [27, 29, 30]. The steps that are required for RNA-Seq are; selection of transcript with poly(A) tail, rRNA depletion [27, 29], fragmentation followed by cDNA synthesis, purification, and amplification [29, 31]. mRNA and lncRNA >200 nt contain a poly(A) tail [27, 28], convenient for the enrichment of poly (A) + RNAs from total cellular RNA, which is carried out with oligo-dT nuclease [27, 28]. This is followed by rRNA depletion [31], this can be done based on sequence-specific probes which can hybridize to rRNA then depleted with streptavidin beads [27]. RNA samples are fragmented to a certain size range before reverse transcriptase (RT) [31], this is necessary due to size limitation of most sequencing platforms [27]. Fragmentation of RNA can be done with alkaline solution or divalent cations on an elevated temperature [27]. Enzymes such as RNase III can also be used to fragment RNAs but this can introduce bias due to its preference for double-stranded RNA sequence [27]. cDNA synthesized from the RNA fragments using RT is ligated to DNA adapters before amplification and sequencing [27, 28, 31].

SEQUENCING PLATFORMS

The method that has been in place for the past few decades is Sanger sequencing, this has been characterized by its simplicity but is capital-intensive and the process takes longer time than necessary [32]. Novel sequencing technologies that immediately evolved after Sanger sequencing are collectively termed as NGS and are often described as throughput, accurate, cost-effective, and reliable techniques that can examine the whole genome within a short period of time [32, 33]. Generally, NGS platforms are grouped into second and third generations. The most widely used platforms for second-generation sequencing are Ion Torrent and Illumina, while Pacific Biosystems and Oxford Nanopore Technologies are the most popular platforms for third-generation sequencing [34]. Also, 454 FLX is among the second-generation platforms that were released in 2005 and works based on pyrosequencing (i.e., sequencing-by-synthesis technique) like Illumina [35]. The most established Illumina platforms include MiSeq, HiSeqs [36], and NovaSeq, the newest platform that is used for large-scale whole-genome sequencing analysis [37]. MiSeq works as a personalized sequencer that sequences very small genomes with high speed and completes its operation within 4 h. On the other hand, HiSeqs such as HiSeq. 2500 is designed for "high-throughput" usage and mostly completes its cycle in 6-day period [36]. Unlike Illumina and 454 sequencing platforms that are sequenced by synthesis, sequencing by oligo-nucleotide ligation and detection (SOLiD) developed by Applied Biosystems (which later became Life Technologies) is by ligation through hybridization of short probes with the template DNA strand [35]. Despite its apparent advantages, its reads length and depth are not as sophisticated as that of Illumina, thereby causing major assembly challenges [38, 39]. Throughout all Illumina platforms, only a 1% error rate is recorded and substitution is regarded as their major error [36], providing in-depth sequencing capacity that allows detection of very small quantity of transcripts in the sample [40].

Nevertheless, the third NGS sequencing platforms provide two essential advantages over the second generation, including their ability to increase reads length, avoid partiality of PCR in the amplification process, and the capacity to sequence single molecules at a given time [35]. Despite all the advantages of NGS, the technique has faced numerous challenges including numerous GC content, big size of genome, and the presence of homopolymers. However, various alternatives have been put in place to overcome the current challenges [27]. Moreover, third NGS sequencing platforms such as single-molecule real-time (SMRT) sequencing that was developed by Pacific Biosciences (PacBio), uses long reads that provides solution to the myriads challenges faced by second-generation platforms due to their short reads length that make them unable to accurately detect gene isoform, poor genome assembly and resolution of complicated genomic region. On the contrary, PacBio sequencing (SMRT) is limited by a high percentage of error, high cost per base, and low throughput [41]. These challenges can be overcome by the use of platforms with lower error rates as low as 3% [37], and alternatively, the use of hybrid Sanger sequencing and PacBio sequencing technology has been proposed [41]. In addition, Helicos is one of the single-molecule-based platforms with 5% error rate, reducing the number of usable reads, making the reference genome extremely hard to be matched with the sequence reads, and causing the loss of miRNA reads at the alignment level [40]. Due to their low error rate (>1%), Illumina or SOLiD platforms are regarded as the best alternative for miRNA sequencing because of its small size [40]. Nanopore sequencing platforms such as MinION, PromethION, and GridION [36] belong to third-generation sequencing platforms developed by Oxford Nanopore Technologies, which operate by passing a single-stranded molecule of DNA or RNA via a protein nanopore at the rate of 30 bases per second, through an electrical

current allowing direct sequencing of the molecule, thus offering greater advantages [42]. However, a major setback of these platforms is the high error rate of up to 10%–20%, when compared to other platforms with high throughput [36, 43].

Most recently, several studies have reported remarkable achievements in the diagnosis of different cancers, including bladder cancer using RNA-Seq [44–46]. Transcriptomics (RNA-Seq) refers to the application of any NGS platforms for the examination of RNA [40] and has become the best method for whole-transcriptome profiling over the last decade [47]. The selection criteria for each platform depend on the purpose of the experiments. The techniques operate similarly to DNA sequencing except for the library preparation and their data analysis which comprises assembly of transcript, uncovering novel transcripts, and calculation of transcripts, among others [40]. Additionally, RNA-Seq gives comprehensive, high-throughput, precise, accurate, and impartial view of the transcriptome analysis that overcome the inherent limitations of real-time PCR and microarray techniques [48]. These limitations include: the need for previous knowledge of the sequence in question, inability to calculate low and high-expressed genes with great accuracy, among others [22]. Although RNA-Seq is high- throughput, it is too expensive [49], especially for clinical settings.

RNA-Seq ANALYSIS PIPELINES

Over the years, microarray and gene-chip technologies provide an insight into understanding the genetic changes in biological samples. However, these techniques are known to have certain limitations related to dynamic range, resolution, and accuracy [50] Advances in transcriptome technology have allowed deeper understanding of the intricacies of gene expression regulation, particularly high-throughput RNA sequencing technology that made it possible to observe whole transcriptome variations, discover novel splicing sites and events, functions of noncoding RNAs, as well as proving correct construction and annotation of complex genomes [51]. It also aids to qualitatively ascertain the RNA transcripts present, RNA editing sites, and to quantitatively know how much of the individual transcripts are being expressed [52]. Thus, it is paramount to overview pipelines and workflows applied to bladder cancer RNA-Seq analyses.

A number of computational pipelines and workflows are being used for the pre-processing of RNA-Seq data in cancer studies and other experimental purposes [49, 53, 54]. A typical RNA-Seq workflow consists of seven steps; (1) pre-processing of raw data, (2) alignment of reads to the reference, (3) transcriptome reconstruction, (4) quantifications of transcripts or genes level, (5) differential expression analysis, (6) functional profiling, and (7) advanced analysis (Figure 2) [55]. These stages in the RNA-Seq workflow that includes QC and data analysis can be done using varieties of computational platforms or tools. For example, read counts may be aligned using different tools such as spliced transcript alignment to a reference (STAR) or Tophat [56, 57]. Then, the aligned read counts can be obtained using either HTSeq or Rsubread R/Bioconductor package [58, 59]. The advantage of Rsubread over HTSeq is that the former is faster, requires less memory, and summarizes the read counts that are more closely related to a true value [59, 60].

RNA-Seq raw data often have quality problems that can distort analytical findings significantly and lead to incorrect conclusions [61]. For instance, the quality of raw RNA-Seq data could be altered by residue of ribosomal RNA, degradation of RNA, and variation in read coverage [61]. Hence, to obtain accurate transcripts or gene measurements and proper acquisition of information from the data, raw RNA-Seq data must be reviewed and evaluated by QC measures before subsequent analyses are conducted [15, 61]. Presently, the most widely and commonly used computational tools available for RNA-Seq QC include; FASTQC and MultiQC. FASTQC processes one sample at a time, while MultiQC can generate a single report that visualizes the output of several samples from multiple tools, thereby giving room for easy comparison [62, 63]. Other important and commonly used computational software for QC are RseQC, RNA-seQC, and RNA-QC-Chain [64–66]. Although both RseQC and RNAseQC can offer QC statistics of aligned read counts, RseQC partially relies on the University of California Senta Crus Genome Browser [65]. Moreover, they are slow and unable to provide sequence trimming and filtration of contaminants. However, RNA-QC-Chain can remove low-quality reads and contamination, in addition to providing fast and reliable QC to produce data for downstream analysis [61]. RNA-Seq data analysis steps totally depend on the data quality and specific aims of the study. These analysis steps were reviewed in detail elsewhere [15, 67].

The system of RNA-Seq analysis employs high-computational tool applications for the development of pipelines that orchestrate the entire workflow and optimize the usage of available computational resources [65]. The development of such analytic tools for RNA-Seq data has expanded owing to the complex nature of transcriptome data, and thus, selecting the correct processing pipeline and normalization strategy has a significant impact on downstream analysis [68]. This pipeline consists of multiple independent analytical software packages, tools, and platforms that employ R and Python, Unix/Bash, Javascript, Perl, and C++. As these software are in programmable environment; they provide flexible manipulation of data and methods. However, they required the user to have expertise in programming languages, especially the bash language or Unix Commands Line [69]. With the growing application of RNA-Seq in biomedical research, an integrated user-friendly platform is needed to overcome the barriers encountered when using code-bond platforms. The graphical user interface or web-based platforms provide convenient environment for nonexpert with advantages for quick exploratory analysis, even though not on the scale of large datasets [69]. Supporting Information: Table 1 provides a summary of the various computational tools and their associated platforms used in RNA-Seq analyses.

Variations in the RNA-Seq analysis results might be observed due to usage of different platforms and analytical frameworks. The number of computational tools and bioinformatics methods that are

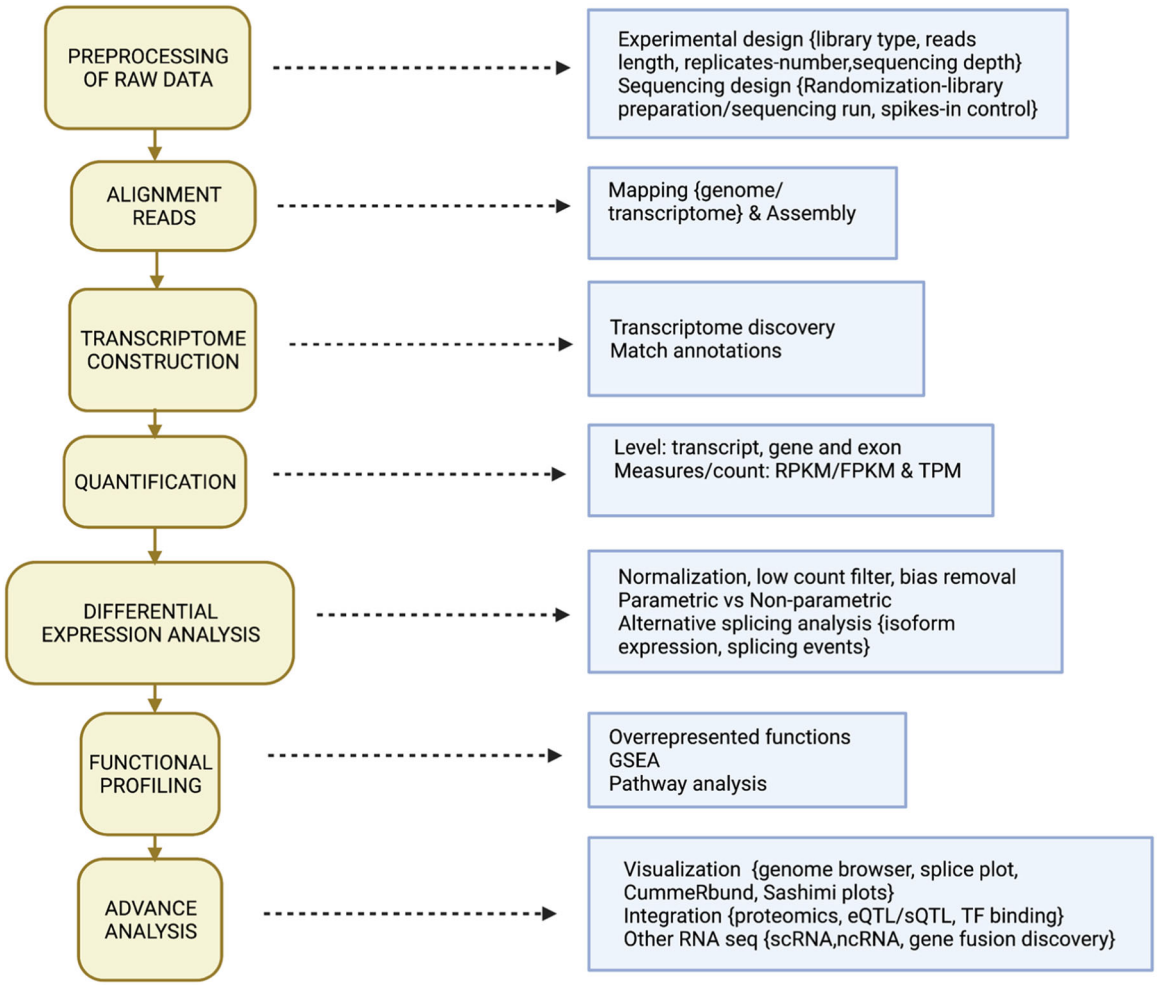

FIGURE 2 General RNA-Seq analysis pipeline. A highlight of the stepwise procedure involved in RNA-Seq data analysis. A typical RNA-Seq analysis workflow consists of these seven steps with each step having some bioinformatics analysis to perform before going downstream to the target objectives of the study. The figure is designed and created with BioRender.com.

currently in use add more challenges to the analysis and interpretation of the RNA-Seq data. To solve these challenges caused by variations in RNA-Seq analysis techniques, standard pipelines need to be enforced and re-designed to integrate analysis of multiple experiments. Workflow construction software packages such as Chipster [70], Anduril [71, 72], and Galaxy [73] could be very much relevant in solving some of these challenges. For example, Anduril was developed for designing complex RNA-Seq pipelines with large-scale datasets, which require automated parallelization. Meanwhile, Chipster and Galaxy are powerful in data integrative visualization, making them very useful for data exploration and interpretation. Other workflows and management frameworks for RNA-Seq analysis are KNIME [74] which aid in visual assembly and interactive execution of data pipeline and Snakemake [75], which is a Python-based workflow management engine that provides a powerful execution environment. Workflow management framework that specifically focuses on RNA-Seq data analysis is (reviewed by [76]). In addition, the large-scale nature of the data analyses associated with RN-Seq brought many challenges that are beyond the scope of this review. Han and colleagues [76] reviewed these challenges comprehensively and proposed solutions. Moreover, results from RNA-Seq study on tumors revealed the presence of molecular subsets of cellular signatures, microenvironment and facilitates choices to circumvent treatment failure [77]. Thus, single-cell sequencing (scRNA-Seq) may prove to be the correct method to understand tumor progression, pathogenesis, and discovery of biomarkers that could lead to a better treatment and management of bladder cancer.

REVIEW OF RNA-Seq STUDIES ON BLADDER CANCER

RNA sequencing has emerged as a powerful NGS technology tool for unbiased identification of gene expression, discovery and quantification of novel transcripts, and identification of alternatively spliced genes [22, 40]. Consequently, recent progress in RNA-Seq has broadened the understanding of the molecular pathogenesis of

TABLE 1 Summary of the current RNA-Seq studies on bladder cancer.

S/N	Sample	Source	Grade	Library preparation	Library selection	Library layout	Sequencing platform	Instrument model	Average mean reads pair-end length (bp)	References
1.	Tissue	UCC	Ta/T1/T2-4/CIS	ScriptSeq	cDNA	Paired-end	Illumina	Illumina HiSeq. 2000	101 + 7 + 101	Hedegaard et al. [78]
2.	Tissue	UCC	N/A	Ribo-Zero Magnetic Gold Kit (Illumina) and the NEBNext® Ultra™ RNA Library Prep Ki	cDNA	Paired-end	Illumina	HiSeqX	N/A	Li et al. [79]
3.	Tissue	UCC	N/A	TruSeq Stranded mRNA Lib Prep Kit	cDNA	Paired-end	Illumina	NextSeq. 500	72	Maeda et al. [80]
4.	Cell	Urine	N/A	Ovation RNA Seq System V2 kit	cDNA	Paired-end/ single	Illumina	Illumina HiSeq. 2000	N/A	Sin et al. [81]
5.	Cell	N/A	N/A	TruSeq Stranded Total RNA kit	cDNA	N/A	Illumina	Illumina HiSeq. 2500	N/A	Shen et al. [82]
6.	Tissue	N/A	N/A	RiboZero rRNA Removal Kit	cDNA	N/A	Illumina	HiSeq. 2000	N/A	Li et al. [83]
7.	Tissue	UCC	N/A	TruSeq	cDNA	Pair-end	Illumina	Genome Analyzer IIx (GAIIx)	380	Zhang et al. [84]
8.	Tissue/cell	UC	T1, T2-T4a, N0–3, M0–1	DNA/RNA AllPrep kit (QIAGEN)/ mirVana miRNA isolation kit	cDNA, miRNA	N/A	Illumina	Illumina HiSeq. 2500	76	Robertson et al. [85]

cancers, including bladder cancer. RNA-Seq has been successfully applied in bladder cancer research for earlier detection, establishing pathological origin, and defining the aberrant genes and dysregulated molecular pathways across patient groups. Thus far, eight studies have utilized RNA-Seq technology in understanding bladder cancer pathology. These studies are summarized considering the sample utilized, grades of bladder cancer, and the techniques of library preparation employed (Table 1). However, it is important to note that this overview is not exhaustive and does not encompass all available RNA-Seq platforms for sequencing and transcriptomics studies.

PATHOGENESIS OF BLADDER CANCER

The pathogenesis of bladder cancer, despite its noted malignancy and postsurgical relapse rate, remains elusive [86–88]. Evidence suggests the human microbiome, particularly opportunistic bacteria such as the Campylobacter genus, may contribute to bladder cancer development by evading immune response and creating pro-inflammatory conditions favorable for tumorigenesis [89, 90]. This genus has the ability to generate a pro-inflammatory environment that supports tumor progression [89]. Additionally, aberrant expression of noncoding RNAs and protein-coding genes have been linked to bladder cancer [88]. Key findings from Wang et al. indicate that dysregulation of miRNA-93, miRNA-195, and their respective target genes, E2F1, E2F2, and AKT3, could impact cell proliferation and apoptosis in bladder cancer. Moreover, the dysregulation of lncRNA and circRNA is deemed significant in bladder cancer pathogenesis and progression [79].

BIOMARKERS FOR BLADDER CANCER

Cancer biomarkers, including biological molecules found in tissues, body fluids, or blood, play crucial roles in cancer prognosis, diagnosis, treatment response prediction, and disease monitoring [91, 92].

Several studies have attempted to identify bladder cancer biomarkers for improved diagnostic accuracy, but clinical controversy exists due to false-positive and false-negative results [92–95]. However, due to the false-positive and false-negative results, the use of these biomarkers has sparked clinical controversy [96]. Nonetheless, promising biomarkers such as the TERT promoter, genes differentially expressed in bladder cancer (including CDC20, PTTG1, PLK1, SFN, CCNB1, and BUB1B) [97], Sox4, and circular RNA (circRNA), including circHIPK3 and circACVR2A, have been explored. Among these genes, the *TERT* promoter accounted for 73% of alteration in NMBIC [98]. While in MIBC, RNA-Seq analysis on high-grade bladder cancer (urine sample) revealed high expression of 15 genes such as *PLEKHS1*, *CP*, *WNT5A*, *RARRES1*, *MYBPC1*, *AR*, *ROBO1*, *SLC14A1*, *AKR1C2*, *FBLN1*, *IGFBP5*, *STEAP2*, *ENTPD5*, *GPD1L*, and *SYBU* [81]. Sox4, in particular, has shown potential as a biomarker for aggressive bladder cancer phenotypes. Moreover, circHIPK3 and circACVR2A, with their role in gene transcription and translation, could be potential therapeutic targets for bladder cancer [83, 99]. However, none of these potential prognostic and diagnostic biomarkers have yet been clinically validated. Some commercially-available urinary biomarkers such as cytology, UroVysion, BTA, NMP22, uCyt+, and Cxbladder assay hold promising [100].

IDENTIFIED MOLECULAR PATHWAYS IN BLADDER CANCER

Molecular alterations, such as mutations, gene control changes, and epigenetic modifications, drive malignancy and its progression [101]. Notably, perturbations in signaling, metabolic, cytoskeleton, and DNA repair pathways have been associated with tumorigenesis in various cancers [102]. Disruption of pathways controlling cell growth, proliferation, differentiation, and apoptosis may lead to carcinogenesis [102–104]. Specifically in bladder cancer, activation of focal adhesion and MAPK signaling pathways, and ensuing genetic dysregulation, have been identified as contributing factors [105].

RNA sequencing and bioinformatics analyses revealed 17 downregulated and 44 upregulated DEGs in the 5637 cell line compared to T24 [45]. Both cell lines showed downregulation of WNT9A and WNT10A, which impact the Wnt signaling pathway implicated in bladder cancer initiation [1, 9]. The cell adhesion molecules pathway was found to be significantly enriched in bladder cancer (FDR = 2.67E −08) [84]. Genes PTPRF, VEGFA and CLDN7 were identified in the enriched bladder cancer, focal adhesion, and extracellular matrix–receptor interaction pathways [84]. Upregulation of ITGA, F3, ANXA1, and downregulation of GRB7, VEGP were observed in pathways related to cell growth, cell cycle, and apoptosis [106].

Upregulation of IFN-γ, angiogenesis, and inflammatory pathways has been observed in NEURAL, MES, and SCC, while immune activation pathways including NF-κβ, MAPK, and PI3K-Akt showed downregulated DEGs [107]. Essential carcinogenesis stages and tumor progression are influenced by activation of the PIK3/AKT/MTOR pathway by ERBB2 and FGFR3 [108]. Pathways linked to PD-L1 inhibitors, such as IL-10 production, lymphocyte chemotaxis and aberrant IFN-γ, NF-κB and ERK signaling networks, are implicated in HGUC immune evasion [109]. Other pathways, including VEGF/VEGFR and PI3K-Akt-mTOR, are linked with muscle-invasive bladder carcinoma, while JAK/STAT pathway activation by IGFBP4–1 improves UC carcinogenesis [110]. LAT1, controlling the mTOR signaling pathway, is also linked to UC progression [69, 111].

METTL3 knockdown resulted in upregulation of 1793 genes and downregulation of 1759, suggesting METTL3 methylation could influence UC progression [112]. Functional enrichment analysis revealed negative correlations with MYC oncogene and TNF-α/NF-κB target gene pathways, and positive correlation with other pathways. Activation of the RAS pathway was observed more in BC159-T than TCGA-BLCA samples, regulating carcinogenesis via Ras–MEK–ERK and PI3 kinase–AKT–mTOR pathways [112]. In addition, increased activation of RAS pathway was observed in all BC159-T samples than TCGA-urothelial bladder carcinoma (TCGA-BLCA) samples [77].

Efforts to combat bladder cancer development include targeting molecular pathways with anticancer drugs, such as infigratinib and dasatinib for Erk1/2 and Src pathways, and drug combinations like romidepsin, gemcitabine and cisplatin targeting the ERK pathway [113, 114]. Understanding bladder cancer-related metabolomic and transcriptomic pathways could offer alternative treatment avenues [115].

BLADDER CANCER CLASSIFICATION

Bladder cancer has been linked to mutations in 32 genes [116] and APOBEC-cytidine deaminase related genetic metabolism predispositions [10]. Overexpression of APOBEC3B genes contributes to the high APOBEC signature mutations seen in bladder cancer [117, 118], Accurate association of bladder cancer histological variants pathological features and specific molecular mutational pattern might provide promising target for developing therapy [119].

Bladder cancer presents mainly as MIBC and NMIBC. NMIBCs are low-grade tumors [107, 117] accounting for 75% of all diagnosed cases [10, 107] characterized by a stable FGFR genomic mutation [10, 107] and homozygous deletion of *CDKN2A* [120]. Three subtypes of NMIBCs have been identified from gene expression profiling. MIBC, though less common, is genetically unstable and predisposed to gene mutations promoting cell proliferation [121]. Other type of bladder cancer is MIBC, although less frequently diagnosed, it is genetically predisposed as unstable and divergent type [118, 121]. MIBC progresses from a flat lesion and gradually acquires gene mutations that promote cell proliferation and survival [122] and are associated with aneuploidy *TP53* mutations [10, 107]. Key mutations in MIBC include TP53, p53, CK20, HER2/neu genes, PTEN, RB1, and 6P amplification.

Previous anatomical classifications of bladder tumors (NMIBC; Tis, Ta & T1 and MIB; T2, T3 & T4)was insufficient in predicting diverse prognoses had limitations in the understanding of the diverse

prognosis within the same tumor type [123]. Several studies have speculated that tumor heterogeneity might be associated with diagnosis, metastasis and pre-existing treatment resistance especially in the case of MIBCs [123]. Accurate histopathology and specific transcriptomic markers are recommended for optimal therapeutic management [119, 124, 125]

Current classifications include urothelial carcinoma (UC), urothelial carcinoma with variant (UCV), squamous cell carcinoma (SCC), and adenocarcinoma (AD) subtypes [126, 127]. Transcriptomics and other molecular techniques have led to the identification of UC that account for 75% [119] or even up to 90% [128] of the bladder cancer while rare UC cases show chromosomal rearrangements [119]. Invasive UC have a tendency to progress into divergent variants of nonurothelial tumors through progressive histological differentiation [128]. Nonurothelial tumors are also known to consist of different ranges of rare variants based on the WHO 2016 classification that include the pure squamous, glandular, neuroendocrine, sarcomatoid, micropapillary, plasmacytoid, microcystic, clear cell and pure histological variants [119, 129].

Furthermore, UC is more prevalent in developed countries like Japan, North America, and some Western countries [124], while SCC and AD have a higher prevalence in countries like Egypt [127]. SCC of the bladder, though less prevalent, is associated with worse prognosis and is less responsive to chemotherapy [130]. Primary AD represents 0.5%–2% of the all diagnosed bladder cancer, like SCC it has a poor prognosis and has a weak TM immunoexpression score than UC [131]. with small cell carcinoma (SmCC) presenting at late stages and exhibiting certain neuronal marker signatures. Despite their rarity and lethality, some SmCC/NEBC variants respond positively to chemotherapy [128, 132] or radiotherapy in improving survival [132].

CHALLENGES AND LIMITATIONS OF RNA-SEQ METHOD

RNA sequencing (RNA-Seq) has become an essential tool for analyzing differential gene expression utilized in characterizing specific tissues [133]. This development has given room for deeper insight into the complexity of the protein-coding transcriptome than previously understood. Some studies have preferred RNA-Seq over the earlier developed high-throughput RNA analysis by microarrays due to its numerous advantages [17]. One of such promising opportunities is the detection of gene fusions and differential expression of transcripts known to cause disease. Lee et al. [51] reported that dysregulation of long noncoding RNAs had been associated with development of some diseases such as cancer, myocardial infarction and diabetes. Other advantages include increased dynamic range of expression, measurement of focal changes (such as single nucleotide variants, insertions and deletions), detection of rare and novel transcript isoforms, splice variants and chimeric gene fusions (that include genes and transcripts hitherto not identified) [20].

Despite the merits of the RNA-Seq, some limitations especially associated with the library preparation protocols can cause biases and overvaluation of results [15, 134]. Because sequencing is sensitive to the quantity of transcripts, this can make abundant mRNAs to be overly represented in RNA-Seq libraries and can have influence on the majority of the reads. These mRNAs are evaluated with low stochastic variability between samples and can be found significant by differential expression analysis (DEA). Conversely, transcripts low in abundance receive few reads, which can subject them to noise and reduce their chances of being selected [15].

Another factor that influences the transcript detection in RNA-Seq experiments is the length of the transcript itself, because a longer transcript has a higher possibility of being detected in the library, it will be considered significant after DEA which can subsequently affect the functional annotation of significant genes [135, 136]. Excitingly, evidence has proven that microarrays can outperform RNA-Seq in detecting small noncoding RNAs like microRNAs. Another limitation of the RNA-Seq is the errors that arise from the quality of the RNA, as significant percentage of the total RNA come from ribosomal RNA (rRNA) while a very low percentage come from the mRNA, as a result, special considerations must be made in the methods to either enrich the mRNA (polyA selection) or reduce the rRNA levels [17]. It is possible to avoid some of these issues with proper study design. However, bias and inconsistency associated with adapter ligation, cDNA synthesis, and amplification could be primarily dependent on library preparation and pre-processing procedures [137]. Therefore, standardized procedures for addressing the limitations of the RNA-Seq and ascertaining the accuracy, reproducibility and precision in a clinically important settings are pertinent to enable the adoption of RNA-Seq tests.

CONCLUSION AND FUTURE PERSPECTIVE

UCC is among the prime causes of malignancy-associated deaths globally and its etiology is poorly understood. The bladder cancer progression is a multifaceted process that is extremely heterogeneous and complicated. Thus, requiring the use of advanced transcriptome technology like RNA-Seq to elucidate the diverse disease phenotypes more effectively. The RNA-Seq studies of bladder cancer reviewed in this paper reveal genome-wide changes among different classes of bladder cancer and its pathogenicity that includes dysregulated genes and pathways. Moreover, the possibilities to better understand the molecular mechanism underlying UCC progression has been successful through application of RNA-Seq as this helps in understanding pathogenesis, discovering biomarkers and uncovering the mechanism of drug resistance in bladder cancer.

Despite the progressive improvement provided by RNA-Seq technologies in understanding the molecular basis for UCC pathogenesis, there is still a lot to be done to prevent recurrence and further metastasis of the malignant cells, and to reduce cancer-related death due to UCC. A good number of dysregulated genes and biomarkers have been implicated in bladder cancer, and the uses of

these genes clinically have provided controversial results. Hence, there has not been a single prognostic and diagnostic biomarker that is successfully translated into clinical application. This opens up an avenue for a potential research space to fully understand bladder cancer. To address the current limitations, there is a need to have a standard pipeline and integrated multiple experiment analyses as a reference point for all studies due to the high recurrence nature of UCC. The existence of molecular subsets of cellular signatures and microenvironment promotes measures, which could avoid therapeutic failures exposed by RNA-Seq study. Moreover, acceptability of RNA-Seq regarding gene expression in UCC led to the existence of incompatibility due to RNA-Seq techniques and variability in platform application.

In summary, RNA-Seq is regarded as the golden standard for large scale high-throughput genome-wide and gene expression studies. It has provided us an unknown insight into the transcriptional changes in bladder cancer diagnosis and management. As the cost of sequencing is gradually decreasing and more precision is being obtained in most of the computational tools, RNA-Seq technology will continue to be applied in studying bladder cancer transcriptomics and disease state. Consequently, an in-depth exploration of the complex and stochastic genetic nature of bladder cancer could possibly result in a novel biomarker discovery and identifying new therapeutic strategies. Moreover, scRNA-Seq may offer a better option for comprehensive understanding of UCC development and recurrence and could effectively help in improving bladder cancer management.

ACKNOWLEDGMENTS

Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute (NCI), United States. All figures in this manuscript are designed and created with BioRender.com.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Arxiv at https://arxiv.org, reference number https://arxiv.org/abs/2011.10372.

ORCID

Umar Ahmad http://orcid.org/0000-0002-3216-5171
Buhari Ibrahim http://orcid.org/0000-0003-3559-3099
Mustapha Mohammed https://orcid.org/0000-0002-5021-1610
Ahmed Faris Aldoghachi http://orcid.org/0000-0003-3236-1305
Mahmood Usman http://orcid.org/0000-0002-0273-9867
Abdulbasit Haliru Yakubu http://orcid.org/0000-0001-6018-9931
Abubakar Sadiq Tanko http://orcid.org/0000-0001-8415-0740
Khadijat Abubakar Bobbo http://orcid.org/0000-0001-7765-0317
Usman Adamu Garkuwa http://orcid.org/0000-0001-9089-9950
Abdullahi Adamu Faggo http://orcid.org/0000-0001-5413-3093
Sagir Mustapha http://orcid.org/0000-0003-4185-1680
Syahril Abdullah https://orcid.org/0000-0003-2718-6266
Yong Yoke Keong http://orcid.org/0000-0002-9442-7456
Abhi Veerakumarasivam http://orcid.org/0000-0002-2676-1815

REFERENCES

[1] X. Li, M. Xu, L. Ding, J. Tang, *J. Cancer* **2019**, *10*(12), 2836.
[2] V. Sharma, A. Heer, N. Kour, A. Sharma, S. K. Singh, *Indian J. Tradit. Knowl.* **2019**, *18*(3), 573.
[3] T. Kobayashi, *Asian J. Urol.* **2016**, *3*(4), 211.
[4] M. Garg, *World J. Stem Cells* **2016**, *8*(8), 260.
[5] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, *CA Cancer J. Clin.* **2018**, *68*(6), 394.
[6] K. Saginala, A. Barsouk, J. S. Aluru, P. Rawla, S. A. Padala, A. Barsouk, *Med. Sci.* **2020**, *8*(1), 15.
[7] J. Leal, R. Luengo-Fernandez, R. Sullivan, J. A. Witjes, *Eur. Urol.* **2016**, *69*(3), 438.
[8] F. Soria, M. J. Droller, Y. Lotan, P. Gontero, D. D'Andrea, K. M. Gust, M. Rouprêt, M. Babjuk, J. Palou, S. F. Shariat, *World J. Urol.* **2018**, *36*(12), 1981.
[9] X. Gao, X. Wen, H. He, L. Zheng, Y. Yang, J. Yang, H. Liu, X. Zhou, C. Yang, Y. Chen, M. Chen, S. Zhang, *PeerJ* **2020**, *8*, e9078.
[10] O. Sanli, J. Dobruch, M. A. Knowles, M. Burger, M. Alemozaffar, M. E. Nielsen, Y. Lotan, *Nat. Rev. Dis. Primers* **2017**, *3*(1), 17022.
[11] X. Liu, Y. Chen, W. Tang, L. Zhang, W. Chen, Z. Yan, P. Yuan, M. Yang, S. Kong, L. Yan, J. Qiao, *Sci. China Life Sci.* **2020**, *63*, 1006.
[12] M. R. Hassler, F. Bray, J. W. Catto, A. P. Grollman, A. Hartmann, V. Margulis, S. F. Matin, M. Roupret, J. P. Sfakianos, S. F. Shariat, *Eur. Urol.* **2020**, 209–220.
[13] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, T. Shafee, *PLoS Comput. Biol.* **2017**, *13*(5), e1005457.
[14] Z. Wang, M. Gerstein, M. Snyder, *Nat. Rev. Genet.* **2009**, *10*(1), 57.
[15] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, A. Mortazavi, *Genome Biol.* **2016**, *17*(1), 13.
[16] L. Li, Q. Lei, S. Zhang, L. Kong, B. Qin, *Oncol. Rep.* **2017**, *38*(5), 2607.
[17] P. V. Nazarov, A. Muller, T. Kaoma, N. Nicot, C. Maximo, P. Birembaut, N. L. Tran, G. Dittmar, L. Vallar, *BMC Genomics* **2017**, *18*(1), 443.
[18] M. Bortolomeazzi, M. R. Keddar, F. D. Ciccarelli, L. Benedetti, *Biochim. Biophys. Acta Gene Regul. Mech.* **2020**, *1863*(6), 194445.
[19] M. Suárez-Fariñas, B. Ungar, J. Correa da Rosa, D. A. Ewald, M. Rozenblit, J. Gonzalez, H. Xu, X. Zheng, X. Peng, Y. D. Estrada, S. R. Dillon, J. G. Krueger, E. Guttman-Yassky, *J. Allergy Clin. Immunol.* **2015**, *135*(5), 1218.
[20] S. A. Byron, K. R. Van Keuren-Jensen, D. M. Engelthaler, J. D. Carpten, D. W. Craig, *Nat. Rev. Genet.* **2016**, *17*(5), 257.
[21] R. Stark, M. Grzelak, J. Hadfield, *Nat. Rev. Genet.* **2019**, *20*(11), 631.
[22] K. R. Kukurba, S. B. Montgomery, *Cold Spring Harbor Protocols* **2015**, *2015*(11), pdb.top084970.
[23] N. D. L. Owens, E. De Domenico, M. J. Gilchrist, *Cold Spring Harbor Protocols* **2019**, *2019*(6), pdb.prot098368.
[24] S. Zhao, B. Zhang, Y. Zhang, W. Gordon, S. Du, T. Paradis, M. Vincent, von D. Schack, *Bioinformatics—Updated Features and Applications*, InTech, **2016**, p. 125.
[25] A. N. Scholes, J. A. Lewis, *BMC Genomics* **2020**, *21*(1), 249.
[26] N. Ali, R. C. P. Rampazzo, A. D. T. Costa, M. A. Krieger, *BioMed Res. Int.* **2017**, *2017*, 1.
[27] R. Hrdlickova, M. Toloue, B. Tian, *WIREs RNA* **2017**, *8*(1), e1364.
[28] Y. Song, B. Milon, S. Ott, X. Zhao, L. Sadzewicz, A. Shetty, E. T. Boger, L. J. Tallon, R. J. Morell, A. Mahurkar, R. Hertzano, *BMC Genomics* **2018**, *19*(1), 696.

[29] S. Sekar, P. Geiger, L. Cuyugan, A. Boyle, G. Serrano, T. G. Beach, W. S. Liang, *J. Vis. Exp*. **2019**, *153*, e59981.
[30] M. Vecera, J. Sana, J. Oppelt, B. Tichy, K. Alena, R. Lipina, M. Smrcka, R. Jancalek, M. Hermanova, L. Kren, O. Slaby, *PLoS One* **2019**, *14*(2), e0211978.
[31] Y.-S. Park, S. Kim, D.-G. Park, D. H. Kim, K.-W. Yoon, W. Shin, K. Han, *Genes Genomics* **2019**, *41*(10), 1233.
[32] J. Hodzic, L. Gurbeta, E. Omanovic-Miklicanin, A. Badnjevic, *Med. Arch*. **2017**, *71*(4), 288.
[33] D. Qin, *Cancer Biol. Med*. **2019**, *16*(1), 4.
[34] A. Singh, P. Bhatia, *Int. J. Lab. Hematol*. **2020**, *42*, 833.
[35] E. Lavezzo, L. Barzon, S. Toppo, G. Palù, *Expert Rev. Mol. Diagn*. **2016**, *16*(9), 1011.
[36] J. A. Reuter, D. V. Spacek, M. P. Snyder, *Mol. Cell* **2015**, *58*(4), 586.
[37] K. R. Kumar, M. J. Cowley, R. L. Davis, *Semin. Thromb. Hemost*. **2019**, *45*(7), 661.
[38] J. M. Heather, B. Chain, *Genomics* **2016**, *107*(1), 1.
[39] J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, *Science* **2005**, *309*(5741), 1728.
[40] Y. Chu, D. R. Corey, *Nucleic Acid Ther*. **2012**, *22*(4), 271.
[41] A. Rhoads, K. F. Au, *Genomics Insights* **2015**, *13*(5), 278.
[42] J. Quick, N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouédraogo, B. Afrough, A. Bah, J. H. J. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, Á. Camino-Sánchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallasch, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, T. Racine, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. J. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O. Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Wölfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keïta, A. Rambaut, P. Formenty, S. Günther, M. W. Carroll, *Nature* **2016**, *530*(7589), 228.
[43] J. Quick, A. R. Quinlan, N. J. Loman, *Gigascience* **2014**, *3*(1), 22.
[44] G. Wu, F. Wang, K. Li, S. Li, C. Zhao, C. Fan, J. Wang, *PeerJ* **2019**, *7*, e8261.
[45] X. Xu, X. Wang, B. Fu, L. Meng, B. Lang, *Int. J. Clin. Exp. Pathol*. **2015**, *8*(10), 12678.
[46] T. Kekeeva, A. Tanas, A. Kanygina, D. Alexeev, A. Shikeeva, L. Zavalishina, Y. Andreeva, G. A. Frank, D. Zaletaev, *Cancer Lett*. **2016**, *374*(2), 224.
[47] I. A. Eilertsen, S. H. Moosavi, J. M. Strømme, A. Nesbakken, B. Johannessen, R. A. Lothe, A. Sveen, *Cancer Lett*. **2020**, *469*, 246.
[48] J. Wang, D. C. Dean, F. J. Hornicek, H. Shi, Z. Duan, *Gynecol. Oncol*. **2019**, *152*(1), 194.
[49] D. Castillo, J. M. Gálvez, L. J. Herrera, B. S. Román, F. Rojas, I. Rojas, *BMC Bioinformatics* **2017**, *18*(1), 506.
[50] L.-L. Shi, N. Zhang, X.-M. Xie, Y.-J. Chen, R. Wang, L. Shen, J.-S. Zhou, J.-G. Hu, H.-Z. Lü, *BMC Genomics* **2017**, *18*(1), 173.
[51] H. J. Lee, A. Georgiadou, T. D. Otto, M. Levin, L. J. Coin, D. J. Conway, A. J. Cunnington, *Microbiol. Mol. Biol. Rev*. **2018**, *82*(2), e00071–17.
[52] Y. Hasin, M. Seldin, A. Lusis, *Genome Biol*. **2017**, *18*(1), 83.
[53] M. Callari, A. Guffanti, G. Soldà, G. Merlino, E. Fina, E. Brini, A. Moles, V. Cappelletti, M. G. Daidone, *Oncotarget* **2016**, *7*(1), 976.
[54] N. D. Pennock, S. Jindal, W. Horton, D. Sun, J. Narasimhan, L. Carbone, S. S. Fei, R. Searles, C. A. Harrington, J. Burchard, S. Weinmann, P. Schedin, Z. Xia, *BMC Med. Genomics* **2019**, *12*(1), 195.
[55] I. S. Yang, S. Kim, *Genomics Inform*. **2015**, *13*(4), 119.
[56] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, *Bioinformatics* **2012**, *29*(1), 15.
[57] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, *Nat. Protoc*. **2012**, *7*(3), 562.
[58] S. Anders, P. T. Pyl, W. Huber, *Bioinformatics* **2014**, *31*(2), 166.
[59] Y. Liao, G. K. Smyth, W. Shi, *Nucleic Acids Res*. **2013**, *41*(10), e108.
[60] Y. Liao, G. K. Smyth, W. Shi, *Nucleic Acids Res*. **2019**, *47*(8), e47.
[61] Q. Zhou, X. Su, G. Jing, S. Chen, K. Ning, *BMC Genomics* **2018**, *19*(1), 144.
[62] P. Ewels, M. Magnusson, S. Lundin, M. Käller, *Bioinformatics* **2016**, *32*(19), 3047.
[63] C. M. Ward, T.-H. To, S. M. Pederson, *Bioinformatics* **2019**, *36*(8), 2587.
[64] R. K. Patel, M. Jain, *PLoS One* **2012**, *7*(2), e30619.
[65] L. Wang, S. Wang, W. Li, *Bioinformatics* **2012**, *28*(16), 2184.
[66] C. W. Nazaire, M. Reich, W. Winckler, G. Getz, *Bioinformatics* **2012**, *28*(11), 1530.
[67] J. Costa-Silva, D. Domingues, F. M. Lopes, *PLoS One* **2017**, *12*(12), e0190152.
[68] A. L. P. Reyes, T. C. Silva, S. G. Coetzee, J. T. Plummer, B. D. Davis, S. Chen, D. J. Hazelett, K. Lawrenson, B. P. Berman, S. A. Gayther, M. R. Jones, *BMC Genomics* **2019**, *20*(1), 745.
[69] H. Li, Q. Huang, Y. Liu, L. X. Garmire, *Reproduction* **2020**, *160*, R155.
[70] M. A. Kallio, J. T. Tuimala, T. Hupponen, P. Klemelä, M. Gentile, I. Scheinin, M. Koski, J. Käki, E. I. Korpelainen, *BMC Genomics* **2011**, *12*(1), 507.
[71] A. Cervera, V. Rantanen, K. Ovaska, M. Laakso, J. Nuñez-Fontarnau, A. Alkodsi, J. Casado, C. Facciotto, A. Häkkinen, R. Louhimo, S. Karinen, K. Zhang, K. Lavikka, L. Lyly, M. Pal Singh, S. Hautaniemi, *Bioinformatics* **2019**, *35*(19), 3815.
[72] K. Ovaska, M. Laakso, S. Haapa-Paananen, R. Louhimo, P. Chen, V. Aittomäki, E. Valo, J. Núñez-Fontarnau, V. Rantanen, S. Karinen, K. Nousiainen, A. M. Lahesmaa-Korpinen, M. Miettinen, L. Saarinen, P. Kohonen, J. Wu, J. Westermarck, S. Hautaniemi, *Genome Med*. **2010**, *2*(9), 65.
[73] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, J. Goecks, *Nucleic Acids Res*. **2016**, *44*(W1), W3.
[74] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel, *AcM SIGKDD Explor. Newsl*. **2009**, *11*(1), 26.
[75] J. Köster, S. Rahmann, *Bioinformatics* **2012**, *28*(19), 2520.
[76] Y. Han, S. Gao, K. Muegge, W. Zhang, B. Zhou, *Bioinform. Biol. Insights* **2015**, *9s1*, BBI.S28991.
[77] H. W. Lee, W. Chung, H.-O. Lee, D. E. Jeong, A. Jo, J. E. Lim, J. H. Hong, D.-H. Nam, B. C. Jeong, S. H. Park, K. M. Joo, W. Y. Park, *Genome Med*. **2020**, *12*, 47.
[78] J. Hedegaard, P. Lamy, I. Nordentoft, F. Algaba, S. Høyer, B. Parm Ulhøi, S. Vang, T. Reinert, G. G. Hermann, K. Mogensen, M. Borg, H. Thomsen, M. Muhlig Nielsen, M. Marquez, U. Segersten, M. Aine, M. Höglund, K. Birkenkamp-Demtröder, N. Fristrup, M. Borre, A. Hartmann, R. Stöhr, S. Wach, B. Keck,

A. Katharina Seitz, R. Nawroth, T. Maurer, C. Tulic, T. Simic, K. Junker, M. Horstmann, N. Harving, A. C. Petersen, M. Luz Calle, E. W. Steyerberg, W. Beukers, K. E. M. van Kessel, J. B. Jensen, J. S. Pedersen, P.-U. Malmström, N. Malats, F. X. Real, E. C. Zwarthoff, T. F. Ørntoft, L. Dyrskjøt,*Cancer cell* **2016**, *30*(1), 27.

[79] M. Li, Y. Liu, X. Zhang, J. Liu, P. Wang, *Gene* **2018**, *677*, 189.

[80] S. Maeda, H. Tomiyasu, M. Tsuboi, A. Inoue, G. Ishihara, T. Uchikai, J. K. Chambers, K. Uchida, T. Yonezawa, N. Matsuki, *BMC cancer* **2018**, *18*, 1.

[81] M. L. Y. Sin, K. E. Mach, R. Sinha, F. Wu, D. R. Trivedi, E. Altobelli, K. C. Jensen, D. Sahoo, Y. Lu, J. C. Liao, *Clin. Cancer Res*. **2017**, *23*(14), 3700.

[82] H. Shen, M. Blijlevens, N. Yang, C. Frangou, K. E. Wilson, B. Xu, Y. Zhang, L. Zhang, C. D. Morrison, L. Shepherd, Q. Hu, Q. Zhu, J. Wang, S. Liu, J. Zhang, *Int. J. Biol. Sci*. **2015**, *11*(12), 1363.

[83] Y. Li, F. Zheng, X. Xiao, F. Xie, D. Tao, C. Huang, D. Liu, M. Wang, L. Wang, F. Zeng, G. Jiang, *EMBO Rep*. **2017**, *18*(9), 1646.

[84] S. Zhang, Y. Liu, Z. Liu, C. Zhang, H. Cao, Y. Ye, S. Wang, Y. Zhang, S. Xiao, P. Yang, J. Li, Z. Bai, *PLoS One* **2014**, *9*(3), e91466.

[85] A. G. Robertson, J. Kim, H. Al-Ahmadie, J. Bellmunt, G. Guo, A. D. Cherniack, T. Hinoue, P. W. Laird, K. A. Hoadley, R. Akbani, M. A. A. Castro, E. A. Gibb, R. S. Kanchi, D. A. Gordenin, S. A. Shukla, F. Sanchez-Vega, D. E. Hansel, B. A. Czerniak, V. E. Reuter, X. Su, B. de Sa Carvalho, V. S. Chagas, K. L. Mungall, S. Sadeghi, C. S. Pedamallu, Y. Lu, L. J. Klimczak, J, Zhang, C. Choo, A. I. Ojesina, S. Bullman, K. M. Leraas, T. M. Lichtenberg, C. J. Wu, N. Schultz, G. Getz, M. Meyerson, G. B. Mills, D. J. McConkey, TCGA Research Network, J. N. Weinstein, D. J. Kwiatkowski, S. P. Lerner, *Cell* **2017**, *171*(3), 540.

[86] H. Knollman, J. L. Godwin, R. Jain, Y.-N. Wong, E. R. Plimack, D. M. Geynisman, *Ther. Adv. Urol*. **2015**, *7*(6), 312.

[87] A. Mari, S. Kimura, B. Foerster, M. Abufaraj, D. D'Andrea, M. Hassler, A. Minervini, M. Rouprêt, M. Babjuk, S. F. Shariat, *BJU Int*. **2019**, *123*(1), 11.

[88] X. Wang, Y. Ding, J. Wang, Y. Wu, *Front. Genet*. **2020**, *10*, 1398.

[89] V. B. Popović, M. Šitum, C.-E. T. Chow, L. S. Chan, B. Roje, J. Terzić, *Sci. Rep*. **2018**, *8*(1), 1.

[90] W. S. Garrett, *Science* **2015**, *348*(6230), 80.

[91] N. L. Henry, D. F. Hayes, *Mol. Oncol*. **2012**, *6*(2), 140.

[92] A. R. Santosh, T. Jones, J. Harvey, *J. Cancer Res. Ther*. **2016**, *12*(2), 486.

[93] W. S. Tan, W. P. Tan, M.-Y. Tan, P. Khetrapal, L. Dong, P. deWinter, A. Feber, J. D. Kelly, *Cancer Treat. Rev*. **2018**, *69*, 39.

[94] D. Zhu, S. Singh, X. Chen, Z. Zheng, J. Huang, T. Lin, H. Li, *Int. J. Biochem. Cell. Biol*. **2019**, *110*, 50.

[95] M. L. Cheng, G. Iyer, *Urol. Oncol. Semin. Orig. Investig*. **2018**, *36*(3), 115.

[96] F. Ye, L. Wang, M. Castillo-Martin, R. McBride, M. D. Galsky, J. Zhu, P. Boffetta, D. Y. Zhang, C. Cordon-Cardo, *Am. J. Clin. Exp. Urol*. **2014**, *2*(1), 1.

[97] C. Sucularlı, *Acta Medica. Cordoba*. **2019**, *50*(2), 8.

[98] E. J. Pietzak, A. Bagrodia, E. K. Cha, E. N. Drill, G. Iyer, S. Isharwal, I. Ostrovnaya, P. Baez, Q. Li, M. F. Berger, A. Zehir, N. Schultz, J. E. Rosenberg, D. F. Bajorin, G. Dalbagni, H. Al-Ahmadie, D. B. Solit, B. H. Bochner, *Eur. Urol*. **2017**, *72*(6), 952.

[99] W. Dong, J. Bi, H. Liu, D. Yan, Q. He, Q. Zhou, Q. Wang, R. Xie, Y. Su, M. Yang, T. Lin, J. Huang, *Mol. Cancer* **2019**, *18*(1), 95.

[100] L.-M. Krabbe, S. L. Woldu, S. F. Shariat, Y. Lotan, *Expert Rev. Mol. Diagn*. **2016**, *16*(11), 1189.

[101] S. Pineda, F. X. Real, M. Kogevinas, A. Carrato, S. J. Chanock, N. Malats, K. Van Steen, *PLoS Genet*. **2015**, *11*(12), e1005689.

[102] M. A. Zolotovskaia, V. S. Tkachev, A. P. Seryakov, D. V. Kuzmin, D. E. Kamashev, M. I. Sorokin, S. A. Roumiantsev, A. A. Buzdin, *Cancers* **2020**, *12*(2), 271.

[103] I. Petrov, M. Suntsova, E. Ilnitskaya, S. Roumiantsev, M. Sorokin, A. Garazha, P. Spirin, T. Lebedev, N. Gaifullin, S. Larin, O. Kovalchuk, D. Konovalov, V. Prassolov, A. Roumiantsev, A. Buzdin, *Oncotarget* **2017**, *8*(48), 83768.

[104] S. Whittaker, R. Marais, A. X. Zhu, *Oncogene* **2010**, *29*(36), 4989.

[105] D. Liu, B. Zhou, R. Liu, *PeerJ* **2020**, *8*, e9422.

[106] X. S. Lin, L. Hu, K. Sandy, M. Correll, J. Quackenbush, C.-L. Wu, W. S. McDougal, *Urol Oncol* **2014**, *2014*, 327.

[107] T. Z. Tan, M. Rouanne, K. T. Tan, R. Y.-J. Huang, J.-P. Thiery, *Eur. Urol*. **2019**, *75*(3), 423.

[108] X. Chen, F. Jiang, C. Jia, M. Liu, Y. Nan, L. Qu, Q. Kong, F. Hou, W. Luo, W. Na, X. Jin, J. Tan, *Front. Oncol*. **2019**, *9*, 523.

[109] E. Olkhov-Mitsel, A. Hodgson, S. K. Liu, D. Vesprini, J. Bayani, J. Bartlett, B. Xu, M. R. Downes, *J. Clin. Pathol*. **2021**, *74*, 53.

[110] K. Inamura, *Cancers* **2018**, *10*(4), 100.

[111] M. Maimaiti, S. Sakamoto, Y. Yamada, M. Sugiura, J. Rii, N. Takeuchi, Y. Imamura, T. Furihata, K. Ando, K. Higuchi, M. Xu, T. Sazuka, K. Nakamura, A. Kaneda, Y. Kanai, N. Kyprianou, Y. Ikehara, N. Anzai, T. Ichikawa, *Sci. Rep*. **2020**, *10*(1), 1292.

[112] W. Zheng, X. Dong, Y. Zhao, S. Wang, H. Jiang, M. Zhang, X. Zheng, M. Gu, *Front. Oncol*. **2019**, *9*, 1403.

[113] N. C. Lima, E. Atkinson, T. D. Bunney, M. Katan, P. H. Huang, *Int. J. Mol. Sci*. **2020**, *21*(9), 3214.

[114] P. Pattarawat, T. Hong, S. Wallace, Y. Hu, R. Donnell, T.-H. Wang, C.-L. Tsai, J. Wang, H.-C. R. Wang, *Br. J. Cancer* **2020**, *123*, 226.

[115] F.-C. von Rundstedt, K. Rajapakshe, J. Ma, J. M. Arnold, J. Gohlke, V. Putluri, R. Krishnapuram, D. B. Piyarathna, Y. Lotan, D. Gödde, S. Roth, S. Störkel, J. M. Levitt, G. Michailidis, A. Sreekumar, S. P. Lerner, C. Coarfa, N. Putluri, *J. Urol*. **2016**, *195*(6), 1911.

[116] The Cancer Genome Atlas Research Network, *Nature* **2014**, *507*(7492), 315.

[117] M. Rouprêt, M. Babjuk, E. Compérat, R. Zigeuner, R. J. Sylvester, M. Burger, N. C. Cowan, P. Gontero, B. W. G. Van Rhijn, A. H. Mostafid, J. Palou, S. F. Shariat, *Eur. Urol*. **2018**, *73*(1), 111.

[118] W. Choi, A. Ochoa, D. J. McConkey, M. Aine, M. Höglund, W. Y. Kim, F. X. Real, A. E. Kiltie, I. Milsom, L. Dyrskjøt, S. P. Lerner, *Eur. Urol*. **2017**, *72*(3), 354.

[119] M. Moschini, D. D'Andrea, S. Korn, Y. Irmak, F. Soria, E. Compérat, S. F. Shariat, *Nat. Rev. Urol*. **2017**, *14*(11), 651.

[120] M. Aine, P. Eriksson, F. Liedberg, G. Sjödahl, M. Höglund, *Sci. Rep*. **2015**, *5*, 10957.

[121] M. Jung, J. Lee, B. Kim, J. Park, K. Moon, *Int. J. Mol. Sci*. **2019**, *20*(3), 570.

[122] D. J. McConkey, W. Choi, Y. Shen, I.-L. Lee, S. Porten, S. F. Matin, A. M. Kamat, P. Corn, R. E. Millikan, C. Dinney, B. Czerniak, A. O. Siefker-Radtke, *Eur. Urol*. **2016**, *69*(5), 855.

[123] X. Zhang, M. Zhang, Y. Hou, L. Xu, W. Li, Z. Zou, C. Liu, A. Xu, S. Wu, *Oncotarget* **2016**, *7*(40), 66069.

[124] T. Tanaka, K. Miyazawa, T. Tsukamoto, T. Kuno, K. Suzuki, *J. Oncol*. **2011**, *2011*, 1.

[125] T. Inamoto, H. Uehara, Y. Akao, N. Ibuki, K. Komura, K. Takahara, T. Takai, T. Uchimoto, K. Saito, N. Tanda, *Dis. Markers* **2018**, *2018*, 5468672.

[126] S. Bertz, A. Hartmann, R. Knüchel-Clarke, N. T. Gaisa, *Pathologe* **2016**, *37*(1), 40.

[127] J. W. Martin, S. L. Vernez, Y. Lotan, A. Abdelhalim, R. Dutta, A. Shokeir, H. Abol-Enein, A. Mosbah, M. Ghoneim, R. F. Youssef, *World J. Urol*. **2018**, *36*(11), 1835.

[128] G. Wang, L. Xiao, M. Zhang, A. M. Kamat, A. Siefker-Radtke, C. P. Dinney, B. Czerniak, C. C. Guo, *Hum. Pathol*. **2018**, *79*, 57.

[129] V. Genitsch, A. Kollár, G. Vandekerkhove, J. Blarer, M. Furrer, M. Annala, C. Herberts, A. Pycha, de J. J. Jong, Y. Liu, F. Krentel, E. Davicioni, E. A. Gibb, M. Kruithof-de Julio, A. W. Wyatt, R. Seiler, *Urol Oncol* **2019**, *2019*, 826.

[130] M. Owyong, Y. Lotan, P. Kapur, V. Panwar, T. McKenzie, T. K. Lee, X. Zi, J. W. Martin, A. Mosbah, H. Abol-Enein, M. Ghoneim, R. F. Youssef, *Urol Oncol* **2019**, *37*, 478.

[131] J. Watt, D. G. Maguire, C. N. Reid, J. V. Lamont, S. P. Fitzgerald, M. W. Ruddock, *Res. Rep. Urol*. **2020**, *12*, 157.
[132] Z. Chen, Q. Liu, R. Chen, Z. Liu, M. Li, Q. Ling, L. Wu, J. Yang, X. Liu, T. Wang, Z. Hu, X. Guo, S. Wang, W. Yang, J. Liu, *World J. Surg. Oncol*. **2017**, *15*(1), 33.
[133] D. M. F. Francisco, L. Marchetti, S. Rodríguez-Lorenzo, E. Frías-Anaya, R. M. Figueiredo, M. Heymanns, M. Culot, A. R. Santa-Maria, M. A. Deli, R. F. V. Germano, B. Vanhollebeke, N. Kakogiannos, M. Giannotta, E. Dejana, A. Dominguez-Belloso, S. Liebner, M. Schuster, H. A. Klok, M. Wiatr, H. Schroten, T. Tenenbaum, G. Kooij, P. Winter, I. A. Romero, H. E. de Vries, B. Engelhardt, R. Bruggmann, *Fluids Barriers CNS* **2020**, *17*(1), 51.
[134] N. F. Lahens, I. Kavakli, R. Zhang, K. Hayer, M. B. Black, H. Dueck, A. Pizarro, J. Kim, R. Irizarry, R. S. Thomas, G. R. Grant, J. B. Hogenesch, *Genome Biol*. **2014**, *15*(6), R86.
[135] J. A. Timmons, K. J. Szkop, I. J. Gallagher, *Genome Biol*. **2015**, *16*(1), 186.
[136] W. R. Swindell, X. Xing, J. J. Voorhees, J. T. Elder, A. Johnston, J. E. Gudjonsson, *Physiol. Genomics* **2014**, *46*(15), 533.
[137] C. Wright, A. Rajpurohit, E. E. Burke, C. Williams, L. Collado-Torres, M. Kimos, N. J. Brandon, A. J. Cross, A. E. Jaffe, D. R. Weinberger, J. H. Shin, *BMC Genomics* **2019**, *20*(1), 513.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: U. Ahmad, B. Ibrahim, M. Mohammed, A. F. Aldoghachi, M. Usman, A. H. Yakubu, A. S. Tanko, K. A. Bobbo, U. A. Garkuwa, A. A. Faggo, S. Mustapha, M. Al-Masaeed, S. Abdullah, Y. Y. Keong, A. Veerakumarasivam, *Appl. Res*. **2023**, e202300002. https://doi.org/10.1002/appl.202300002