



OPEN ACCESS

EDITED BY

Yong Sun Lee,
National Cancer Center, South Korea

REVIEWED BY

Siheem Cheloufi,
University of California, United States
Aristeidis G. Telonis,
University of Miami, United States

*CORRESPONDENCE

Rune Ougland,
runoug@vestrevik.no
Anindya Dutta,
duttaa@uab.edu

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 01 March 2022

ACCEPTED 27 June 2022

PUBLISHED 18 July 2022

CITATION

Su Z, Monshaugen I, Klungland A, Ougland R and Dutta A (2022), Characterization of novel small non-coding RNAs and their modifications in bladder cancer using an updated small RNA-seq workflow. *Front. Mol. Biosci.* 9:887686. doi: 10.3389/fmolb.2022.887686

COPYRIGHT

© 2022 Su, Monshaugen, Klungland, Ougland and Dutta. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Characterization of novel small non-coding RNAs and their modifications in bladder cancer using an updated small RNA-seq workflow

Zhangli Su^{1,2†}, Ida Monshaugen^{3,4,6†}, Arne Klungland^{3,5}, Rune Ougland^{3,6*} and Anindya Dutta^{1,2*}

¹Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, United States,

²Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA, United States, ³Department of Microbiology, Oslo University Hospital

Rikshospitalet, Oslo, Norway, ⁴Department of Molecular Medicine, Institute of Basic Medical Sciences,

University of Oslo, Oslo, Norway, ⁵Department of Biosciences, Faculty of Mathematics and Natural

Sciences, University of Oslo, Oslo, Norway, ⁶Department of Surgery, Baerum Hospital Vestre Viken

Hospital Trust, Gjettem, Norway

Background: Bladder cancer (BLCA) is one of the most common cancer types worldwide. The disease is responsible for about 200,000 deaths annually, thus improved diagnostics and therapy is needed. A large body of evidence reveal that small RNAs of less than 40 nucleotides may act as tumor suppressors, oncogenes, and disease biomarkers, with a major focus on microRNAs. However, the role of other families of small RNAs is not yet deciphered. Recent results suggest that small RNAs and their modification status, play a role in BLCA development and are promising biomarkers due to their high abundance in the exomes and body fluids (including urine). Moreover, free modified nucleosides have been detected at elevated levels from the urine of BLCA patients. A genome-wide view of small RNAs, and their modifications, will help pinpoint the molecules that could be used as biomarker or has important biology in BLCA development.

Methods: BLCA tumor tissue specimens were obtained from 12 patients undergoing transurethral resection of non-muscle invasive papillary urothelial carcinomas. Genome-wide profiling of small RNAs less than 40 bases long was performed by a modified protocol with TGIRT (thermostable group II reverse transcriptase) to identify novel small RNAs and their modification status.

Results: Comprehensive analysis identified not only microRNAs. Intriguingly, $57 \pm 15\%$ (mean \pm S.D.) of sequencing reads mapped to non-microRNA-small RNAs including tRNA-derived fragments (tRFs), ribosomal RNA-derived fragments (rRFs) and YRNA-derived fragments (YRFs). Misincorporation (mismatch) sites identified potential base modification positions on the small RNAs, especially on tRFs, corresponding to m¹A (N¹-methyladenosine), m¹G (N¹-methylguanosine) and m²₂G (N², N²-dimethylguanosine). We also detected

mismatch sites on rRFs corresponding to known modifications on 28 and 18S rRNA.

Conclusion: We found abundant non-microRNA-small RNAs in BLCA tumor samples. Small RNAs, especially tRFs and rRFs, contain modifications that can be captured as mismatch by TGIRT sequencing. Both the modifications and the non-microRNA-small RNAs should be explored as a biomarker for BLCA detection or follow-up.

KEYWORDS

bladder cancer, non-coding RNA, small RNA, RNA modification, tRNA-derived fragment, rRNA-derived fragment, YRNA-derived fragment

Introduction

Bladder cancer (BLCA) is the sixth most common cancer worldwide with high morbidity and mortality rates. With 550,000 annual new incidents and 200,000 deaths, BLCA poses a significant disease burden globally (Bray et al., 2018). About 75% of incidents present as non-muscle invasive (NMIBC), consisting of a heterogeneous population of tumors (Kirkali et al., 2005; Burger et al., 2013). Currently there is no routine screening for NMIBC or BLCA in general. Patients with NMIBC usually display urinary tract symptoms i.e., hematuria, pain or frequent urination, and is then subject to cystoscopy as the first step in the diagnostic process. If the initial workup reveals a tumor, the affected individual often undergoes surgery. In addition, patients may receive radiation therapy, chemotherapy, immunotherapy and targeted therapy. Despite a 70–80% recurrence rate, NMIBC has a favorable prognosis and a 5-year survival rate greater than 85% (van Rhijn et al., 2009). However, up to 30% of NMIBC cases progress into more advanced stages with less favorable prognosis, and 5-year survival rate drops to about 5% for metastatic disease (Schrier et al., 2004; Sanli et al., 2017; Boegemann and Krabbe, 2020). This lifelong menace necessitates an exhaustive post-operative control scheme burdening both patients and healthcare systems. In fact, BLCA is in the top tier of the most expensive cancer type to treat, both when considering cost per patient and lifetime cost, in addition to the invaluable expense of life quality reduction (James and Gore, 2013). Thus, urgent improvement of diagnostics and follow-up is required. Despite tremendous effort, the development of sensitive biomarkers and non-invasive methods for cost-effective diagnostics and surveillance of patients remains a challenge. However, the family of small non-coding RNAs and their modifications, appear as a promising addition to the future clinical toolbox.

Non-coding RNAs (ncRNAs), including both long non-coding RNAs (lncRNAs) and small RNAs (sRNAs), have gained much attention lately for their key role as mediators of gene expression in cancer (Slack and Chinnaiyan, 2019).

They are considered well-suited as therapeutic targets or agents due to their small size and chemical properties, which allow them to cross tissue barriers and reach tumor cell interior better than macromolecular antibody drugs (Li et al., 2020). In particular, the primary focus of sRNA research in BLCA has been directed towards microRNAs (miRNAs). Extensive RNA sequencing by The Cancer Genome Atlas reported epigenetic regulation of ncRNA, especially miRNAs, in BLCA (Yoshino et al., 2013; Cancer Genome Atlas Research, 2014). In recent years, dysregulated expression of hundreds of miRNAs have been reported in BLCA by large-scale analysis from close to 20 research groups (Lee et al., 2016). Functional studies suggest that miRNAs are involved in different aspects of BLCA development and progression (Li et al., 2011; Morais et al., 2014; Wang et al., 2015). Moreover, miRNAs dysregulated in tumor tissue can also be detected in biological fluids such as serum and urine, suggesting their potential usage as non-invasive diagnostic or prognostic tools (Yun et al., 2012; Armstrong et al., 2015; Fang et al., 2016; Borkowska et al., 2019; Yin et al., 2019).

Besides microRNAs, other emerging small RNAs are detected at high abundance thanks to the technological advances in next-generational sequencing. For example, tRNA-derived fragments (tRFs) have gained attention as their diverse biological functions are being discovered (Magee and Rigoutsos, 2020). These sRNAs have high promise due to their biological functions in different diseases and their high abundance in bodily fluids as recently reviewed (Su et al., 2020b). So far only a few reports focused on tRFs or other non-microRNA-small RNAs in BLCA. tRF expression showed context-dependent association with mRNA expression across 32 cancer types in TCGA, highlighting differences in tRF-mRNA connection by sex in bladder cancer (Telonis et al., 2019). Furthermore, analysis of TCGA data found association between elevated level of a specific tRF (5'-tRF-LysCTT) and early progression and poor outcome in BLCA (Papadimitriou et al., 2020), calling for further investigation of tRF functions in BLCA. In addition to tRFs, other small RNAs such as ribosomal

RNA-derived fragments (rRFs) and Y RNA-derived fragments (YRFs) have also been reported in humans but not investigated as much. Important to note, TCGA small RNA-seq data was collected focusing on microRNAs (~22 nucleotides long) and has a strict size cut-off of 30 nucleotides, losing potential information on RNAs longer than this size range.

Moreover, sRNAs harbor a range of chemical modifications providing a second layer of biological information (Zhang et al., 2016; Li et al., 2021). This modification information is often missing and even leads to under-representation of modification-containing sRNAs during conventional small RNA-seq library preparation (Shi et al., 2021). Enzyme-assisted library preparation improves the cloning of modification-containing sRNAs, and suggests that their abundance was formerly far under-appreciated. Altogether, there is a great need to understand the relative abundance of non-microRNA-small RNAs and modification status on different small RNAs, both of which have not been comprehensively profiled in BLCA samples.

We aimed to establish a workflow that can profile both microRNA (miRs) and non-miR small RNAs in a genome-wide fashion that can be applied to patient samples. Small RNA-seq has been a powerful method for high-throughput profiling and sequence-level information that is important for base-level analysis. However, regular small RNA-seq protocol is known to suffer from the stalling of the reverse transcriptase at sites containing modifications that disrupt Watson-Crick base-pairing, including but not limited to m¹A (N¹-methyladenosine), m¹G (N¹-methylguanosine), and m₂²G (N², N²-dimethylguanosine) (Behrens et al., 2021; Shi et al., 2021). Recently we showed TGIRT (thermostable group II intron reverse transcriptase) can be used in small RNA-seq to overcome under-cloning of m¹A-containing RNAs during regular small RNA-seq protocol, and further be used to identify the modification base position via mismatch (Su et al., 2022). The under-representation of m¹A-containing small RNAs and loss of quantitative mismatch ratio by a commonly used M-MuLV reverse transcriptase (ProtoScriptII) indicates the regular small RNA-seq pipeline is biased against m¹A-modified small RNAs. Intrigued by this result, we wondered whether TGIRT can also overcome and capture the other RNA modifications that disrupt Watson-Crick base-pairing. In addition to A-type mismatch, we noticed TGIRT also produced more G-type mismatch than ProtoScriptII from the same HEK293T RNAs (Supplementary Figure S1A), suggesting TGIRT can potentially capture modifications on guanosine as well. Here we report a comprehensive profiling of small RNAs and their modification status in BLCA patient samples by this modified small RNA-seq pipeline. From 12 tumor samples, we identified non-microRNA-small RNA reads that are comparable in abundance to microRNAs. These non-microRNA-small

TABLE1 NMIBC patient information.

Patient #	Sex	Age range	Primary or recidive
Patient #1	Female	40-49	Primary
Patient #2	Male	70-79	Primary
Patient #3	Male	70-79	Primary
Patient #4	Male	60-69	Primary
Patient #5	Male	70-79	Primary
Patient #6	Male	>80	Primary
Patient #7	Male	60-69	Primary
Patient #8	Male	50-59	Primary
Patient #9	Male	40-49	Primary
Patient #10	Male	50-59	Primary
Patient #11	Male	>80	Primary
Patient #12	Female	70-79	Primary

RNAs include tRNA-fragments, rRNA-fragments, Y-RNA-fragments, snoRNA-fragments and more. Their length distribution and cleavage patterns were distinctly different. RNA modification as indicated by TGIRT mismatch pattern was mostly found on tRFs over other small RNA types. Mismatch sites on specific tRFs correspond to known m¹A, m¹G and m₂²G annotations on mature tRNAs, suggest a large proportion of tRFs harbor these modifications. Furthermore, mismatch sites were also identified on rRFs at known modification positions on 28 and 18S rRNAs. This analysis confirms the high potential of using TGIRT to enable modification-friendly profiling of small RNAs in clinical samples.

Materials and methods

Human subject and sample collection

Patients diagnosed and treated for papillary urothelial NMIBC at the Vestre Viken Hospital Trust hospitals were enrolled in the study. Cold cup biopsies were harvested prior to surgical resection of the tumor, and the specimens were kept on -20°C in RNAlater preservation solution (Ambion #AM7020) until preparation and further analyses.

Anonymized collective patient information of the 12 samples used is listed in Table 1.

RNA extraction

Purification of total RNA was done using the RNeasy RT reagent (MRC Inc. #RN190). Subsequently RNA quality was determined using RNA ScreenTape on TapeStation (Agilent Tech. #5067-5576) or Agilent RNA 6000 Pico Kit on Bioanalyzer (Agilent Tech. #5067-1513).

Small RNA library preparation by TGIRT and sequencing

Small RNA-seq library preparation was performed as previously reported (Su et al., 2019; Su et al., 2020a) using NEBNext Small RNA Library Prep Set for Illumina (NEB #7330) with changes to use TGIRT for cDNA synthesis. TGIRT condition is based on m¹A mapping on polyA-enriched RNAs by TGIRT-seq (Li et al., 2017) with the modifications described below. Total RNAs of 0.3–1 µg were ligated with the corresponding 3' and 5' adaptor within the NEBNext kit. Ligated RNAs were converted to cDNA by 1 µL TGIRT-III enzyme (InGex #TGIRT50) per reaction at 60°C for 15 min. TGIRT reaction was carried out in buffer (50 mM Tris, pH 8.3, 75 mM KCl, 3 mM MgCl₂, 1 mM dNTP, 10 mM DTT) with addition of 1 µL RNase Inhibitor. The reaction is stopped by addition of 250 mM (final concentration) NaOH at 95°C for 3 min and 65°C for 15 min. Same amount of HCl was added to neutralize the reaction after the reaction cools down. The cDNA is further purified by QIAquick Nucleotide Removal Kit (Qiagen #28304) or ZYMO oligo clean and concentrator kit (ZYMO #D4060) or Dynabeads MyOne Silane (Thermo Fisher #37002D). cDNA is amplified by 15–16 cycles of PCR with indexed NEBNext primers (NEB #E6609). The individual amplified libraries were purified with ZYMO DNA Clean and Concentrator Kit (ZYMO #D4033) and run on 8% TBE polyacrylamide Novex gel (Thermo Fisher #EC6215). The position corresponding to RNA insert of 15–40 nucleotides long was cut out from the gel and purified *via* crush-and-soak method. Care was taken to cut the region longer than primer dimer and shorter than full-length tRNA. Gel-recovered eluate was purified and concentrated by ethanol precipitation according to NEB kit instruction. Final libraries were quantified by Qubit fluorometer and pooled for sequencing on Illumina sequencer.

HEK293T small RNA-seq data by ProtoScriptII and TGIRT can be accessed from GEO: GSE171040 (GSM5217188 and GSM5217193 for ProtoScriptII; GSM5217184 and GSM5217186 for TGIRT).

General mapping strategy for small RNA TGIRT-seq data analysis

Small RNA-seq data was analyzed similarly as before (Su et al., 2019; Su et al., 2020a). Briefly, cutadapt v1.15 (Martin, 2011) was used to trim 3' adaptor sequence and discard trimmed read length shorter than 15 nt. To avoid mis-annotation of 5' NEBNext adaptor sequence to hsa-miR-3168, reads containing 5' adaptor sequence were discarded with cutadapt. In general, each library has 2–10 million mapped reads. Unitas v1.7.3 (Gebert et al., 2017) with SeqMap v1.0.13 (Jiang and Wong, 2008) was used to map small RNAs. Priority of mapping was given to first map the reads to miRBase Release 22 (Kozomara et al., 2019) human sequence. The remaining reads were mapped to other small RNA sequences including genomic tRNA

database (Chan and Lowe, 2016) and Ensembl Release 97. Additional rRNA and YRNA reference sequences were used for rRF and YRF mapping: 18S (NR_145820.1), 5S (NR_023363.1), 28S (NR_003287.4) and 5.8S (NR_145821.1); RNY1 (NR_004391.1), RNY3 (NR_004392.1), RNY4 (NR_004393.1) and RNY5 (NR_001571.2). miRNA mapping was done allowing 2 non-templated 3' nucleotides addition and 1 internal mismatch. Other nmsRNA mapping was done allowing 1 mismatch and 0 insertion/deletion, unless otherwise specified (for example Supplementary Figure S1). When a given read is mapped to multiple reference RNAs (multi-mapping), fractionated count was calculated assuming even distribution among all possible references. For all the analysis, reads per million (RPM) was calculated to adjust for total mapped reads in each library.

Mismatch calculation for small RNA TGIRT-seq data analysis

After initial mapping as above, tRF/rRF/YRF reads were re-mapped to only the corresponding reference RNAs with unitas v1.7.3 (Gebert et al., 2017) allowing 1 mismatch and 0 insertion/deletion. The mapping start/end position and mismatch position were recorded for each read. The reads were aggregated onto the whole length of reference RNAs into a coverage plot to facilitate visualization, with X axis representing each nucleotide position of the reference RNA and Y axis representing the abundance of all reads that covered that specific position. Mismatch index (on a scale of 0–100%) was calculated for each position by taking the reads with mismatch at that position and divided by total reads that covered that specific position. Mismatch index was visualized by color on the coverage plot (red means higher mismatch). For tRFs (Figure 4), coverage was aggregated by tRNA amino acid groups.

Coverage plot on secondary structure of YRNA-derived fragments

YRNA secondary structures were retrieved from RNAcentral v19 based on Rfam (RF00019). The dot-bracket notation was used to generate secondary structure plot by StructureEditor v6.1. To color each base based on the relative abundance, coverage of each base is normalized to the highest coverage for that YRNA.

Results

An updated small RNA-seq workflow for modification-friendly global analysis

We collected 12 NMIBC tumor samples (patient information summarized in Table 1) to test the updated

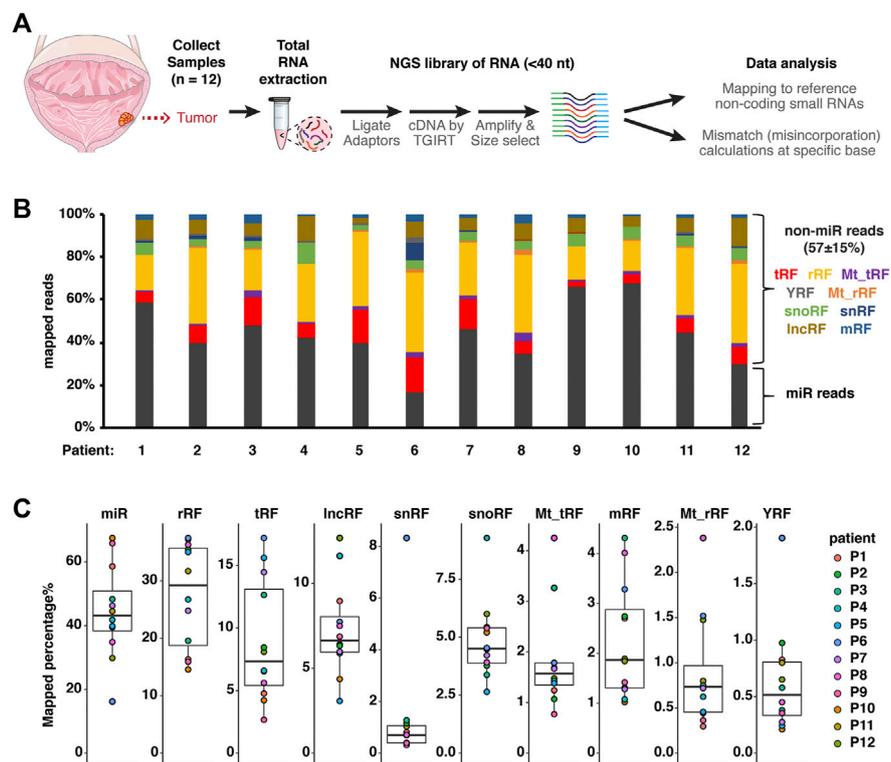


FIGURE 1

An updated small RNA-seq workflow for modification-friendly global analysis. (A) Scheme of collecting tumor samples from 12 NMIBC patients. Small RNA-sequencing libraries by TGIRT were prepared from total RNAs to profile relative abundance and potential RNA modifications (based on mismatch/misincorporation) for small RNAs less than 40 bases long. (B) Overall distribution of total mapped reads between microRNAs (dark grey) and non-microRNA small RNAs (nmsRNAs), including rRFs (yellow), tRFs (red) and more. (C) Distribution of mapped percentage for each subgroup of small RNAs shown as box-whisker plot. Box plot center represents median value, bounds represent upper and lower quartile, whiskers represent 1.5* interquartile range from the bounds.

small RNA-seq workflow (Figure 1A). High-quality total RNAs were used as input and first ligated with 3' adaptor and 5' adaptor. Ligated RNA was converted into cDNA by reverse transcriptase TGIRT, which has been shown to produce more mismatch (misincorporation) than hard-stop products at m¹A modification sites than other reverse transcriptases (Li et al., 2017; Su et al., 2022). When sequencing HEK293T RNAs, we noticed TGIRT protocol is better at capturing non-microRNA-small RNAs than the regular ProtoScriptII protocol (Supplementary Figure S1A). The higher G-type mismatch from the same HEK293T RNAs by TGIRT (Supplementary Figure S1B), suggests TGIRT might detect certain modifications on guanosine as well. To be noted, base modifications that disrupt base pairing such as m¹A, m¹G, m₂²G are more prone to produce RT-induced mismatch, while other RNA modifications including m⁶A and pseudouridine are less affected. Only the fully ligated and converted cDNAs can be further amplified by the next PCR amplification. Lastly, the PCR products are size selected experimentally to enrich for small RNAs of size less than 40 nucleotides long. To avoid ambiguous mapping of very

short sequences, we only mapped reads that are at least 15 nucleotides long. Each clean read was first mapped to human microRNA sequences, and the remaining reads were then mapped to other reference sequences to identify non-microRNA-small RNAs (nmsRNA) (Supplementary Figure S1C). The largest increase in mapping of nmsRNAs (of 25–150%) was seen specifically with tRFs compared to other nmsRNAs (Supplementary Figure S1D). In contrast allowing indels of 1nt did not increase mapping numbers (Supplementary Figure S1E). Overall we found a very high percentage of nmsRNA reads in these libraries, constituting 42–72% of all mapped reads (Figure 1B). The most abundant nmsRNAs include tRNA-derived fragments (tRFs), ribosomal RNA-derived fragments (rRFs), mitochondrial tRFs, mitochondrial rRFs, Y-RNA-derived fragments (YRFs), small nucleolar RNA-derived fragments (snoRFs), small nuclear RNA-derived fragments (snRFs), lncRNA-derived fragments (lncRFs) and protein-coding mRNA-derived fragments (mRFs). Among these, the four most abundant groups are rRFs, tRFs, snoRFs and lncRFs (Figure 1B). It was striking

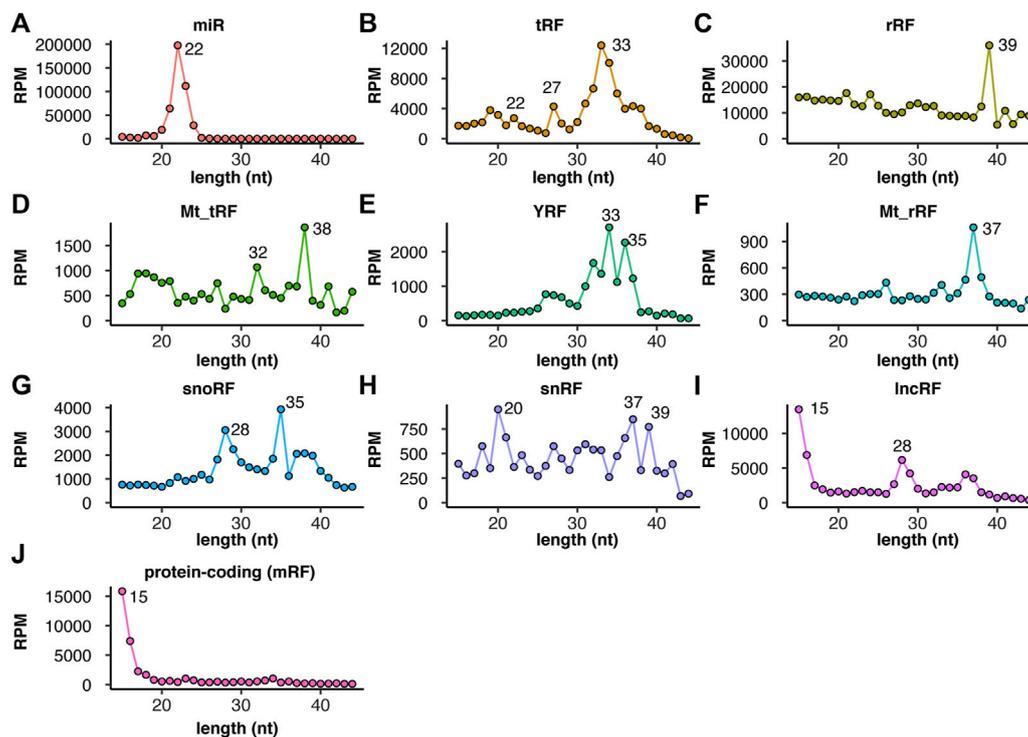


FIGURE 2

microRNAs and non-microRNA small RNAs show distinct size distribution. Size distribution (X-axis in nucleotides, Y-axis in reads per million mapped reads) of each subtype of small RNAs including (A) microRNAs, (B) tRFs, (C) rRFs, (D) mitochondrial tRFs, (E) YRFs, (F) mitochondrial rRFs, (G) snoRFs, (H) snRFs, (I) lncRFs and (J) mRFs. Major peaks in length are labeled for each small RNA subtype. RPM values are averaged from 12 samples (individual samples plotted in [Supplementary Figure S2](#)).

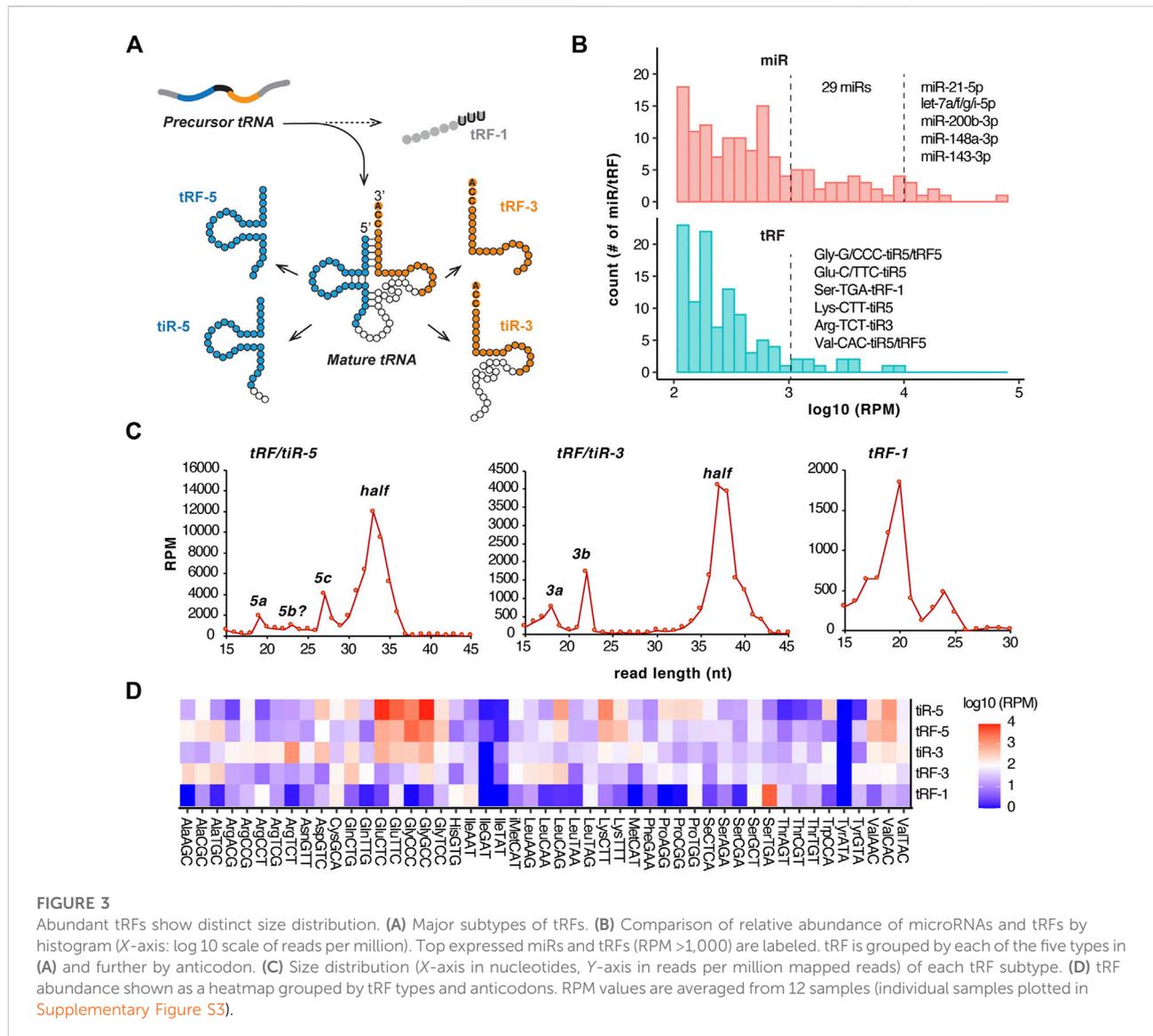
that in some patients (patients 2, 5, 6, 8 and 12), rRF read counts were nearly equal to or more than microRNAs. Interestingly, relative read distribution among different small RNA subgroups was quite variable for different patients ([Figure 1C](#)).

microRNAs and nmsRNAs (non-microRNA-small RNAs) show distinct size distribution

To further understand the characteristics of the nmsRNAs, we checked the length distribution of each subtype. As expected, microRNAs have a specific size of 22 nucleotides (average from 12 samples shown in [Figure 2A](#), individual patient samples shown in [Supplementary Figure S2](#)). Meanwhile, the other small RNAs showed distinct size distributions that were different from microRNAs. For example, tRFs ([Figure 2B](#)), rRFs ([Figure 2C](#)), mitochondrial tRFs ([Figure 2D](#)), YRFs ([Figure 2E](#)), mitochondrial rRFs ([Figure 2F](#)) and snoRFs ([Figure 2G](#)) all have a longer size range than microRNAs. snRFs have a peak at 20 nucleotides and additional peaks at longer size of 37 and 39 nucleotides ([Figure 2H](#)). This also

suggests these longer nmsRNAs were missed or under-represented if a library was size selected around 22 nucleotides.

Similar to what we found before ([Kumar et al., 2014](#)), tRFs display specific peaks in length at 33, 27, 22 and 18 nucleotides ([Figure 2B](#)), which will be further discussed in the next section. Intriguingly, mitochondrial tRFs display additional peaks at 38 and 32 nucleotides ([Figure 2D](#)). In addition, both genomic and mitochondrial rRFs are represented by a very specific peak (39 and 37 nucleotides) ([Figure 2C](#) and [Figure 2F](#)). YRFs have peaks of 33, 35 and 31 nucleotides ([Figure 2E](#)), whereas snoRFs have peaks of 35 and 28 nucleotides ([Figure 2G](#)). On the other hand, lncRFs and mRFs have predominantly shorter reads of 15 nucleotides, ([Figures 2I–J](#)) which is the size cut-off for our bioinformatics analysis (we discarded reads shorter than 15 nucleotides to avoid ambiguous mapping). This may suggest more non-specific cleavage on lncRNA and mRNAs than the other RNAs. In general, the pre-dominant size for each small RNA subtype was consistently observed across 12 tumor samples, which shows distinct pattern between different RNA subtypes ([Supplementary Figure S2](#)). Below we describe specific nmsRNA subtypes in more details.



TGIRT-seq detects tRNA halves and tRFs in the microRNA-size range at an abundance comparable to microRNAs

tRFs are grouped by their start and end positions on parental tRNAs, including 5' fragments and 3' fragments from mature tRNAs and tRF-1s from precursor tRNA trailers (Figure 3A). Both 5' and 3' fragments can be further divided into longer fragments (or called tRNA halves, tiRs) and shorter tRFs. Generally, expression levels for microRNAs are higher than that of tRFs (Figure 3B, one-sided Kolmogorov-Smirnov test, $p = 1E-5$). The most abundant microRNAs detected by TGIRT-seq include miR-21-5p, let-7-5p, miR-200b-3p, miR-148a-3p and miR-143-3p (each more than 10,000 reads per million). When compared with these highly abundant microRNAs,

specific tRFs are also expressed at high levels, including 5' halves/fragments from tRNA^{Gly}, tRNA^{Glu}, tRNA^{Lys}, tRNA^{Val}, 3' half from tRNA^{Arg} and tRF-1 from tRNA^{Ser}, all of which exceed 1,000 reads per million averaged from 12 samples (Figure 3B). The abundance (1,000-10,000 RPM) of these five tRFs is comparable to that of 29 unique microRNAs (let-7b-5p, let-7e-5p, miR-100-5p, miR-101-3p, miR-103a-3p, miR-10a-5p, miR-10b-5p, miR-125a-5p, miR-126-3p, miR-148b-3p, miR-151a-3p, miR-191-5p, miR-199a-3p, miR-200a-3p, miR-200c-3p, miR-203a-3p, miR-205-5p, miR-23a-3p, miR-23b-3p, miR-25-3p, miR-26a-5p, miR-26b-5p, miR-27a-3p, miR-27b-3p, miR-30d-5p, miR-378a-3p, miR-92a-3p, miR-98-5p, miR-99b-5p).

Both 5' and 3' fragments have a major peak corresponding to the tRNA halves that are cleaved in the anticodon loop,

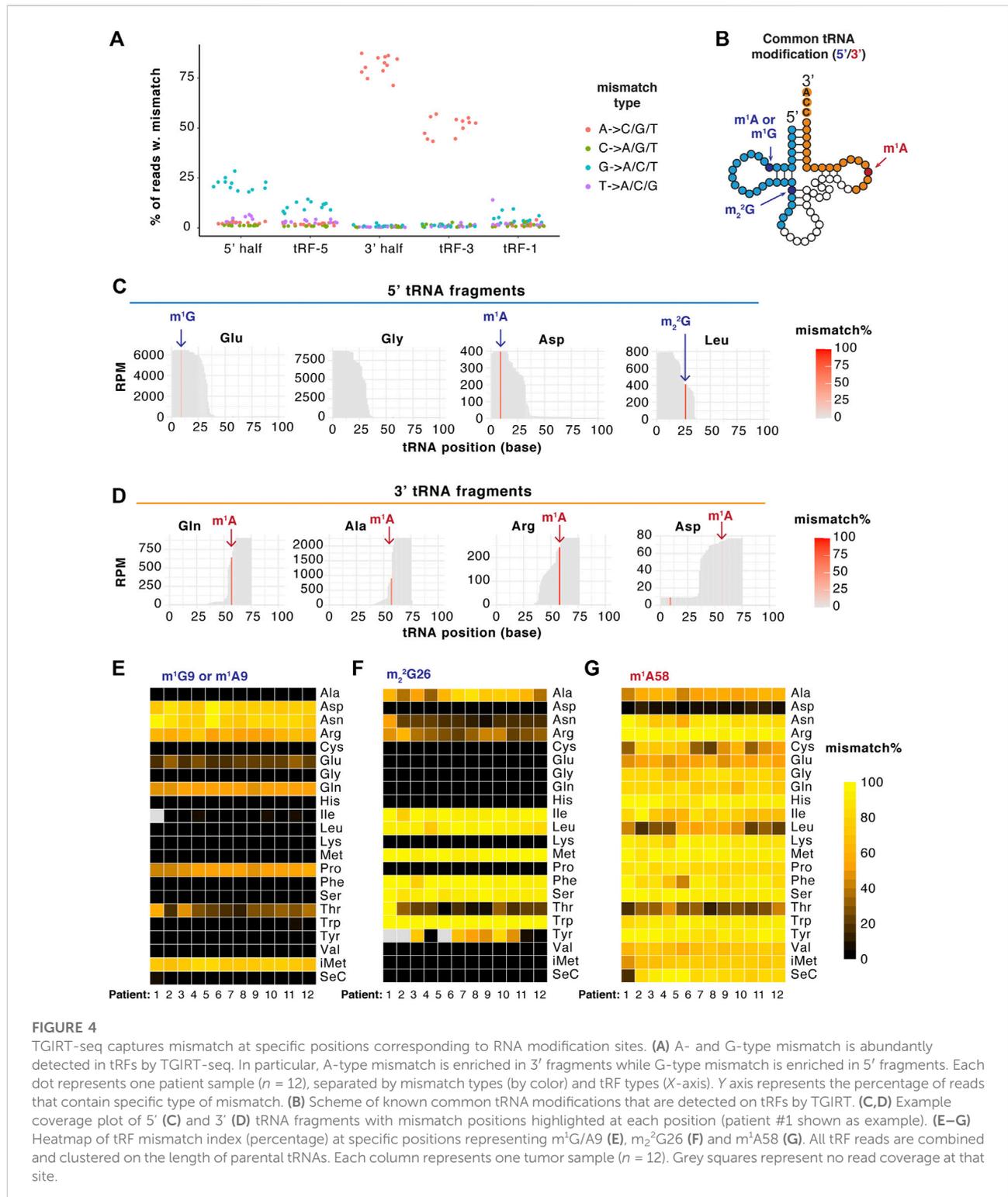


FIGURE 4

TGIRT-seq captures mismatch at specific positions corresponding to RNA modification sites. (A) A- and G-type mismatch is abundantly detected in tRFs by TGIRT-seq. In particular, A-type mismatch is enriched in 3' fragments while G-type mismatch is enriched in 5' fragments. Each dot represents one patient sample ($n = 12$), separated by mismatch types (by color) and tRF types (X-axis). Y axis represents the percentage of reads that contain specific type of mismatch. (B) Scheme of known common tRNA modifications that are detected on tRFs by TGIRT. (C,D) Example coverage plot of 5' (C) and 3' (D) tRNA fragments with mismatch positions highlighted at each position (patient #1 shown as example). (E–G) Heatmap of tRF mismatch index (percentage) at specific positions representing m¹G/A9 (E), m₂G26 (F) and m¹A58 (G). All tRF reads are combined and clustered on the length of parental tRNAs. Each column represents one tumor sample ($n = 12$). Grey squares represent no read coverage at that site.

with other minor peaks representing shorter isoforms (average from 12 samples shown in Figure 3C, individual patient samples shown in Supplementary Figure S3). 5' fragments have dominant size of 32–34 nt (5' halves), 27 nt

(tRF-5c) and 19 nt (tRF-5a). 3' fragments have dominant size of 37–38 nt (3' halves), 22 nt (tRF-3b) and 18 nt (tRF-3a). tRF-1s are generally shorter than 25 nt with a major peak at 20 nt (Figure 3C). Again, the size distribution pattern is

overall consistent among 12 samples (Supplementary Figure 3).

tRF reads are derived from different tRNA genes (Figure 3D). tRF-1 expression shows the lowest correlation with the other tRF types, with the highest tRF-1 expression from tRNA-Ser-TGA (tRF-1001). The most abundant fragments are tiR-5, tRF-5 and tiR-3 from tRNA-Glu-C/TTC and tRNA-Gly-C/GCC. tRF-3s have highest expression from tRNA^{Gln}, tRNA^{Leu} and tRNA^{Ala}.

TGIRT-seq captures mismatch at specific positions corresponding to RNA m¹G, m₂²G and m¹A modification sites

Allowing one nucleotide mismatch increased tRF mapping (Supplementary Figure S1), suggesting tRFs likely bear mismatch-inducing RNA modifications. We checked what type of mismatch was captured in the TGIRT library and found around 15% of tRF reads contain A- > C/G/T mismatch and 15% contain G- > A/C/T mismatch, both of which are much higher than the percentage seen in microRNA reads (Supplementary Figure S4). This is consistent with the fact that tRNAs bear an array of RNA modifications, including modified adenosines and guanosines (Clark et al., 2016; Behrens et al., 2021). Specifically, the G-type mismatch mainly happens on the 5' fragments, whereas the A-type mismatch is strongly enriched in 3' fragments (Figure 4A). Common guanosine or adenosine modification on tRNAs (Figure 4B) include m¹G, m₂²G and m₂²G on the 5' half of tRNA or anticodon loop, and m¹A on the T-loop or on the ninth position of specific tRNAs. The presence and relative abundance of these modifications on tRFs have not been extensively investigated, especially in bladder cancer.

High mismatch rate was detected by TGIRT-seq at specific positions on specific tRFs (patient #1 shown as example in Figures 4C,D). For example, G-type mismatch was detected at the ninth position on the highly abundant 5' fragment from tRNA^{Glu} (Figure 4C), consistent with the known m¹G site on the parental tRNAs. Interestingly, another highly abundant 5' fragment, from tRNA^{Gly}, does not have high mismatch rate detected, despite having guanosine at its ninth position. Across 12 tumor samples, the mismatch pattern at ninth position (Figure 4E) corresponds very well with previous measurements of mismatch on mature tRNAs: high mismatch rate on tRF^{Asn}, tRF^{Arg}, tRF^{Gln}, tRF^{Pro} and tRF^{iMet}, moderate mismatch rate on tRF^{Glu} and tRF^{Thr}.

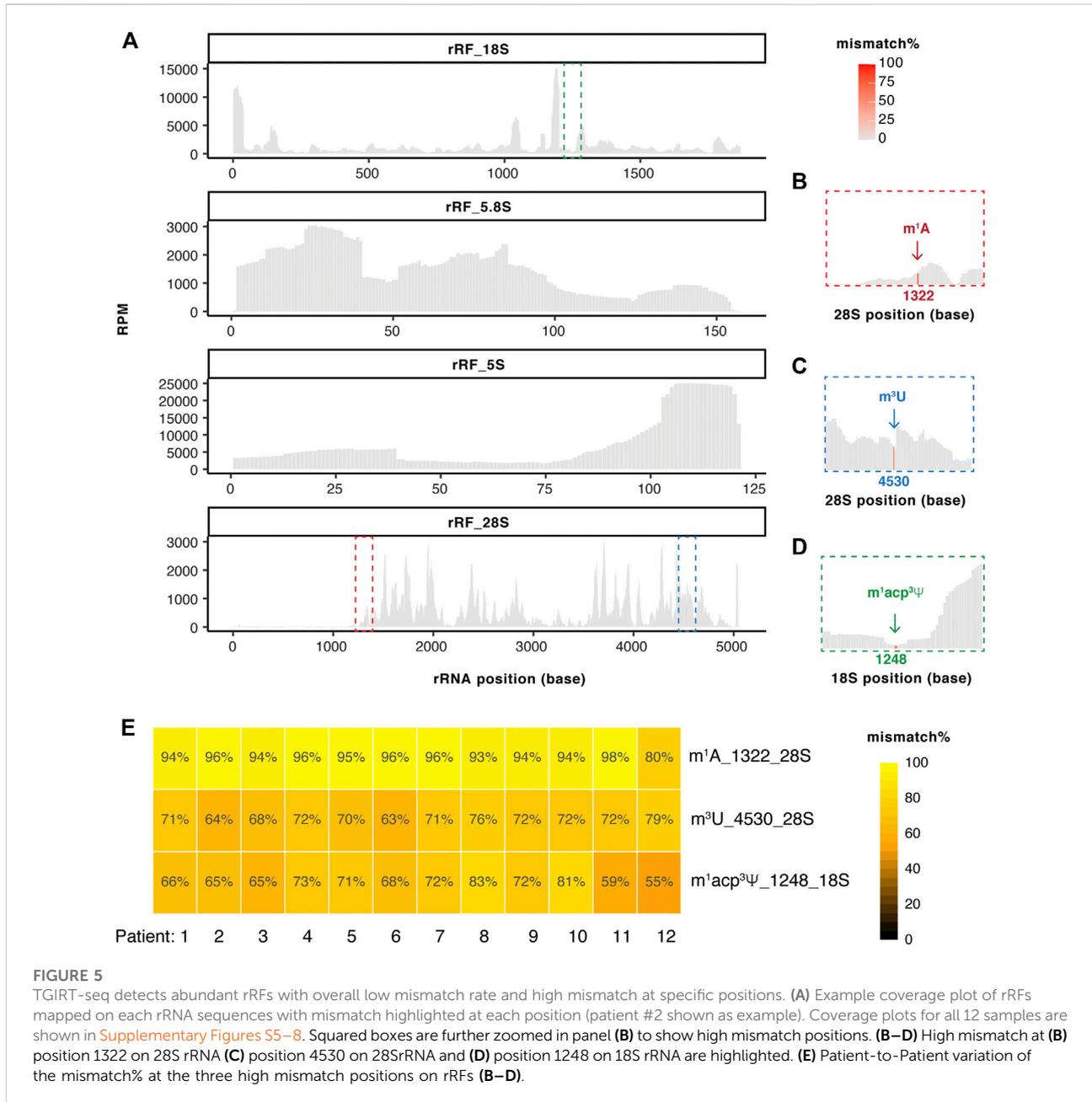
A-type mismatch at the ninth position was also detected on tRF^{Asp} (Figure 4C), corresponding to the reported m¹A modification on tRNA^{Asp}. Similarly, we detected G-type mismatch frequently at the 26th position on specific 5' fragments across 12 samples (Figure 4F): high mismatch rate on tRF^{Ile}, tRF^{Leu}, tRF^{Met}, tRF^{Phe}, tRF^{Ser} and tRF^{Trp},

moderate mismatch rate on tRF^{Ala}, tRF^{Asn}, tRF^{Arg} and tRF^{Tyr}. Lastly, TGIRT detects overall high mismatch rate at m¹A58 position on 3' tRNA fragments across 12 samples with slightly lower rate on tRF^{Ala}, tRF^{Cys}, tRF^{Glu}, tRF^{Leu} and tRF^{Thr} (Figure 4G) and very low mismatch on tRF^{Asp} (Figures 4D, G). Overall TGIRT-seq captures mismatch at specific positions on tRFs, with a mismatch pattern similar to that expected from the mismatch pattern of the corresponding tRNAs. This suggests modifications like m¹G, m₂²G and m¹A are highly prevalent on tRFs.

TGIRT-seq detects abundant rRFs with overall low mismatch rate but high mismatch at specific positions

Another group of abundant nmsRNAs is rRFs (Figure 1). The rRF reads are mapped to all four mature rRNA sequences, 18, 28, 5.8 and 5S. rRFs are highly abundant with comparable RPMs to abundant miRs or tRFs. The rRF coverage along the length of rRNAs is not evenly distributed, as would be expected if they were random degradation products, but interestingly concentrated at discrete regions (Figure 5A, patient #2 shown as an example, all 12 samples shown in Supplementary Figure S5). Consistent with the dominant peak at 39 nt of all rRFs (Figure 2C), these discrete regions show up as peaks of around 39 nt at various sites on the rRNAs (Figure 5A). The rRFs from 18S RNA have the two highest peaks at 0, 1200 bases along the length of the RNA, and this general pattern is seen across 12 patients, with new rRF source sites towards 3' end seen in patient #6 (Supplementary Figure S5). We do not know the explanation for the different pattern in individuals, but there may be interesting biological differences in the tumor accounting for the difference. Similar analysis was done for rRFs from 5.8S (Supplementary Figure S6), 5S (Supplementary Figure S7) and 28S (Supplementary Figure S8), which shows generally conserved patterns across 12 patients.

Unlike tRFs, rRFs are not associated with high mismatch reads on an overall view (Figure 5). However, mismatch is observed at specific position, for example position 1322 on fragments from 28S rRNA (Figure 5B). This position is known to bear m¹A but has not been reported on the rRFs. We were able to detect high mismatch at position 4530 of 28S rRF (Figure 5C) corresponding to known m³U modification and position 1248 of 18S rRF (Figure 5D) corresponding to m¹acp³Ψ. All these modifications are known to disrupt base pairing therefore induce mismatch during reverse transcription. We didn't observe high mismatch (>10%) at positions consistent in multiple patients from 5.8S to 5S rRFs (Supplementary Figures S6, 7). Lastly, these three rRF mismatch sites are consistently detected among 12 patients with overall very high mismatch rate (Figure 5E). This suggests specific rRFs could harbor

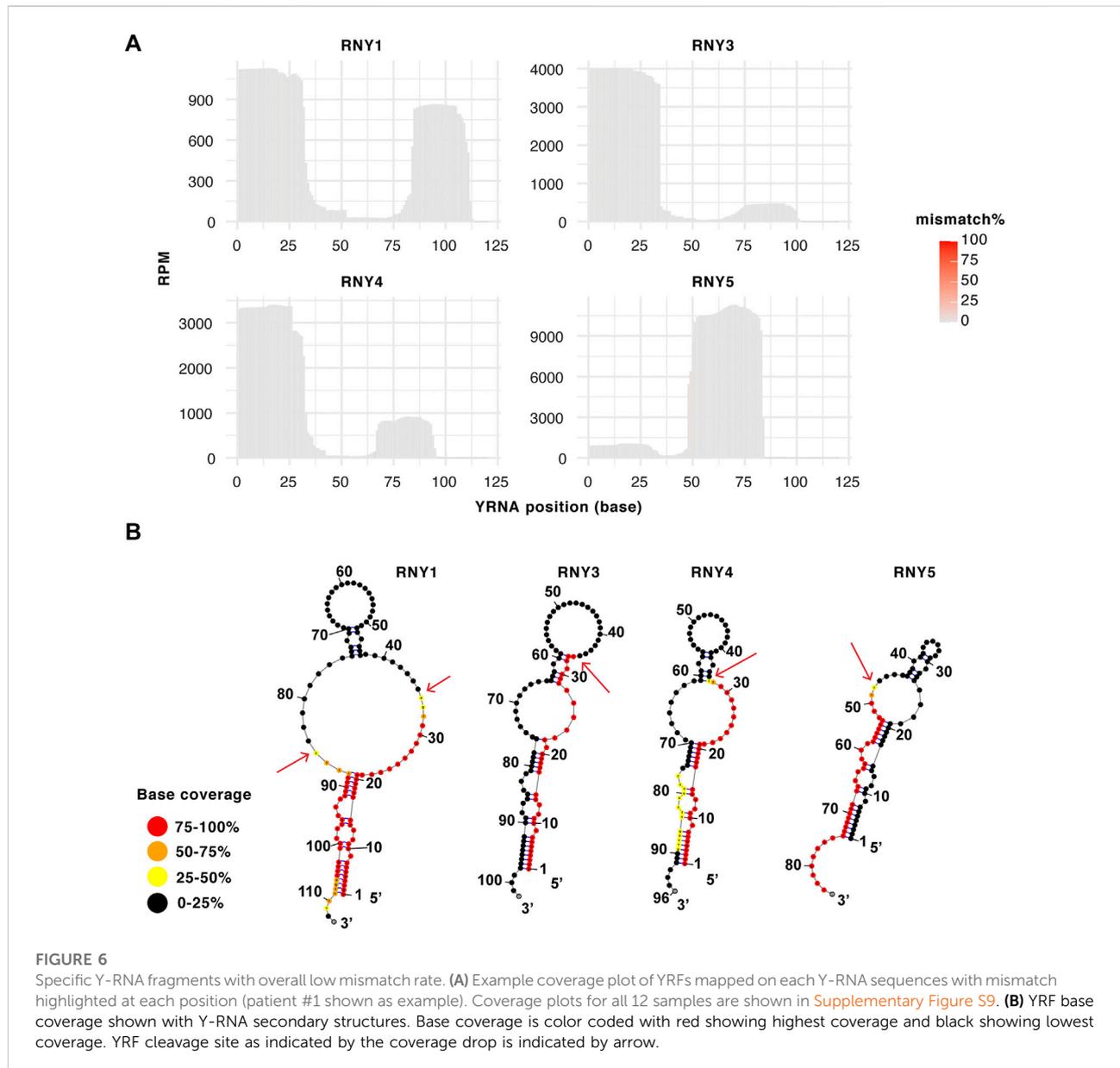


modifications from the parental rRNAs, which will require further future investigation.

Specific and abundant YRNA fragments with overall low mismatch rate

In addition to tRFs and rRFs, we also detected abundant YRNA fragments (YRF) from all four YRNAs, RNY1, RNY3, RNY4 and RNY5 (Figure 6A, patient #1 shown as an example, all 12 patient samples shown in Supplementary Figure S9). The most

abundant YRF is 3' fragment from RNY5 (~10,000 RPM), which is close to the most abundant microRNA level (Figure 3B). Interestingly, the fragmentation pattern is very specific and generates 5' and 3' molecules similar in length to the tRNA halves, although they are not themselves exactly half of a YRNA (Figure 6B). YRNAs share conserved secondary structure with a ~20 bp stem formed by annealing of the 5' and 3' ends, which is adjacent to a loop (preterminal loop) (Figure 6B). The YRF cleavage occurs at single-stranded region of YRNAs, especially within the preterminal loop of RNY1, 4 and 5 (Figure 6B). This cleavage pattern is highly consistent among 12 patient samples



with some patient-to-patient variation in abundance ([Supplementary Figure S9](#)), suggesting specific cleavage. Overall, YRFs are not associated with high mismatch reads ([Figure 6A](#)).

NmsRNAs in the 18–24 base range, that may enter argonaute complexes

Given the emerging reports of tRF involvement in Argonaute-mediated gene silencing activity ([Maute et al., 2013](#); [Kuscu et al., 2018](#); [Ren et al., 2019](#)), it is important to determine how many and which nmsRNAs are likely to enter

Argonaute and potentially affect gene expression. We used the following criteria: 1) 18–24 base long and so expected to enter Argonaute complexes, 2) consistently detected in at least 10 out of the 12 samples, 3) present at an abundance comparable to that of microRNAs. As listed in [Supplementary Table S1](#), 12 unique nmsRNA sequences in this size range were seen at an abundance of 500–15,000 RPM, an abundance at which we see 96 unique microRNA sequences (isomiRs were not combined due to sequence variations). These include tRF-1001, tRF-3001a, tRF-5027b and tRF-5004a (3 nt shorter than annotated sequence). All have been detected in AGO PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) except tRF-1001 ([Kumar et al., 2014](#);

Kumar et al., 2015). The identities of nmsRNAs that have the potential to enter productively into Argonaute complexes by virtue of their length, and are present at an abundance comparable to that of microRNAs, is listed in [Supplementary Table S1](#). Various rRFs have been detected associated with AGO (Guan and Grigoriev, 2021). However, the abundant rRFs that we have identified in this paper have not been reported in association with AGO, but this could either be because these rRFs are not sufficiently abundant in the cell lines where the AGO PAR-CLIP experiments were done, or because there is a mechanism that keeps these fragments away from being loaded into AGO. This analysis suggests that tRF-3001a, tRF-5027b and tRF-5004a should be studied further in BLCA for their microRNA-like activity, and the other nmsRNAs in [Supplementary Table S1](#) may also emerge as being important for BLCA biology through mechanisms waiting to be elucidated.

Discussion

We utilized TGIRT-seq of small RNAs that were size-selected to include RNAs that are usually discarded during microRNA profiling. The results identified a large array of non-microRNA-small RNAs (nmsRNAs) and associated modifications in bladder cancer tumor samples. nmsRNAs are as abundant as the well-studied microRNAs ([Figure 1](#)). nmsRNAs display different size distribution than microRNAs of 22 nucleotides, with a significant portion with a longer length ([Figure 2](#)). General abundance, cleavage patterns and potential modification sites were reported for nmsRNAs, including tRNA-derived fragments (tRFs), rRNA-derived fragments (rRFs) and YRNA-derived fragments (YRFs) ([Figures 3–6](#)). Overall, this highlights the usefulness of TGIRT-seq to profile both abundance and RNA modifications on small RNAs from clinical samples.

Emerging evidence suggests technical biases in small RNA-seq leads to under-representation of certain RNAs. The great abundance of nmsRNAs of length greater than 22 nucleotides ([Figure 2](#)) indicates they are often excluded by the size selection that is used during microRNA profiling. Furthermore, both internal RNA modifications that interfere with reverse transcription and terminal modifications that interfere with ligation could lead to under-cloning (Shi et al., 2021). Here we utilized TGIRT, a thermostable group-II intron reverse transcriptase based on bacterial retrotransposons, that has been developed into a powerful research tool (Belfort and Lambowitz, 2019). TGIRT can mitigate the RT-stalling problem caused by certain internal modifications, but RNAs with other modifications may still be under-cloned. Newer techniques to tackle this gap in true short RNA representation are needed (Alfonzo et al., 2021).

The most abundant nmsRNAs are tRFs and rRFs ([Figures 3–5](#)), both with sequences present at similar abundance as the most abundant microRNAs (500–15,000 RPM). Diverse

biological functions of tRFs have been reported in cancers (Su et al., 2020b). The highly abundant tRFs detected in BLCA samples by this study include 19-nt tRF-1001 from tRNA^{Ser} ([Figure 3](#) and [Supplementary Table S1](#)) that was initially reported in cancer cell lines and associated with cell proliferation (Lee et al., 2009). Another abundant tRF reported here is 18-nt tRF-3001a from tRNA^{Leu}, which has been shown to enter Argonaute complexes (Kumar et al., 2014) and is capable of repressing target gene expression in a seed sequence match manner (Kuscu et al., 2018). In addition, both 5' and 3' tRNA halves from tRNA^{Glu}, tRNA^{Gly}, tRNA^{Lys} and tRNA^{Val} appear to be very abundant in BLCA samples ([Figure 3](#)). Despite their high abundance, the functions of these tRNA halves have not been extensively studied in cancers. tRNA halves can be induced by various stress conditions but can also be detected at endogenous non-stress condition. While 5' tRNA halves have been associated with translational repression (Ivanov et al., 2011), non-coding RNA levels and histone levels (Boskovic et al., 2020) and more recently tRNA transcription (Chen et al., 2021), much less work has been done on 3' tRNA halves. So far, correlational studies based on tRF expression in BLCA patients suggests tRFs very likely play a role in BLCA gene regulation (Telonis et al., 2019; Papadimitriou et al., 2020), however future investigation in a refined experimental system is required to establish a direct association.

In addition to tRFs, we also detected abundant rRFs and YRFs in BLCA samples ([Figures 5, 6](#)). Biological significance and functions are still awaiting investigation for rRFs and YRFs, as recent evidence suggests they are not random degradation products. Profiling of rRFs (<34 nt) in 1000 Genome Project revealed sex- and population-dependent patterns (Cherlin et al., 2020). Furthermore, rRFs ~20 nt are identified associated with Argonaute and paired with cellular transcripts with enriched motifs that are different from microRNA rules (Guan and Grigoriev, 2021). Intriguingly, the abundant rRFs detected in this study were not identified in the Argonaute association studies. Similarly, YRFs ~31 nt could be regulated by stress and were not found associated with Ago, as recently reviewed (Guglas et al., 2020). This could be either because these nmsRNAs are not abundant in the cell lines where the association studies were performed or have been under-cloned due to RT-stalling modifications, or more intriguingly, have other Ago-independent functions. Further investigation is needed to shed light on these abundant nmsRNAs, both the Ago-compatible species and the longer species. Interestingly, relative distribution among different RNA sub-groups is quite variable from sample to sample, with some samples having higher percentage of nmsRNAs than others ([Figure 1](#)). In the future, it will be worthwhile to survey potential causes for such difference in a more systematic analysis. Such causes could be technical (sample handling or contamination) or biological (dysregulation of small RNA homeostasis or correlation with certain clinical features).

We also tested whether TGIRT-mediated mismatches identify known modification sites (m^1G , m^1A and m^2G) on tRFs and rRFs (Figure 4 and Figure 5). In general, the mismatch pattern on tRFs corresponds very well with the sites of modification detected previously on mature tRNAs (Clark et al., 2016; Behrens et al., 2021). The very low m^1A mismatch on tRF^{ASP} (Figures 4D,G) is consistent with the very low m^1A58 on tRNA^{ASP}. Similarly, the A1322 position is known to bear m^1A on large ribosomal subunit RNA across species and is catalyzed by nucleomethylin (also known as RRP8) in humans (Waku et al., 2016; Sharma et al., 2018). The U1248 position of 18S is known to be $m^1acp^3\psi$ modified and is located within the ribosome decoding region (Meyer et al., 2011). The U4530 position of 28S is known to be m^3U modified (Tan et al., 2021). These three rRF modification sites were detected with high mismatch by TGIRT (Figure 5). Interestingly, 18S:1248 ($m^1acp^3\psi$) was suggested to have a lower modification level based on mismatch pattern from long RNA-seq in TCGA tumors, especially READ, UCEC and COAD (Tan et al., 2021). Surprisingly, although rRNA modifications on human ribosomes have very recently been visualized by Cryo-EM (Natchiar et al., 2017), a lot of the rRNA modification enzymatic processes are not well elucidated in humans. The mismatch profile may also be used to identify unannotated modification sites in the parental RNAs in the future but will need to be verified with orthogonal methods. How could these modifications alter in disease conditions and whether they have any impact on ncRNA functions will be an interesting prospective research topic. Recently we reported m^1A impedes tRF-3 gene-silencing activity and is over-expressed in BLCA tumor, coinciding with over-expression of the writer enzyme proteins TRMT6/61A and dysregulation of the tRF-3 targetome (Su et al., 2022). In addition to bladder cancer, TRMT6/61A is also over-expressed in liver cancer and glioma. This is particularly important for cancer since disruption of many RNA modification enzymes has been linked to cancer (Janin et al., 2020; Chujo and Tomizawa, 2021).

Alterations in urinary RNA modification levels hold potential to serve as a non-invasive way to diagnose patients with BLCA and moreover as a monitoring tool to detect disease recurrence. Several studies have reported elevated levels of modified nucleosides detected in urine from BLCA patients, including m^1A (Kvist et al., 1993; Zhang et al., 2014; Sun et al., 2015). The significance of miRNA in BLCA carcinogenesis and as urine cancer biomarkers has been well studied (Yoshino et al., 2013; Hammouz et al., 2021). However, the role of nmsRNAs in BLCA pathogenesis, or as clinically relevant biomarkers, is only beginning to emerge. Interestingly, urine was one of the biofluids with the highest proportion of tRFs detected in healthy donors (Yeri et al., 2017; El-Mogy et al., 2018; Godoy et al., 2018). Both YRNAs and YRFs are recognized as biomarkers in various malignancies as reviewed (Guglas

et al., 2020). They are generally downregulated in BLCA and a low expression of RNY1, 3 and 4 is associated with muscle invasiveness, lymph node metastases, advanced stage, and an unfavorable prognosis (Tolkach et al., 2017). Our observed specific YRNA cleavage pattern taken together with the previous knowledge of YRNA in BLCA suggests a potential regulatory role in the pathogenesis. Yet, whether YRNAs or YRFs are useful urine biomarkers require further investigation.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The raw sequencing data for small RNA TGIRT-seq in BLCA patients are protected by European and national regulations regarding data privacy laws but will be available upon request. Requests should be directed to Rune Ougland (runoug@vestreviken.no) and will be forwarded to the Data Protection Officer and the Ethics Committee for legal- and ethical evaluation. The data will be available for 10 years after publication and if the requesting institution has implemented the European GDPR, or is able to sign the Standard Contractual Clauses for international transfers, the process will take <6 weeks. Otherwise, inter-institutional negotiation is necessary which may prolong the wait time. Meanwhile, we provide the gene counts as Supplementary Data.

Ethics statement

The studies involving human participants were reviewed and approved by the Vestre Viken Hospital Trust and The Regional Ethics Committee South-Eastern Norway Regional Health Authority. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, ZS, RO, and AD; Methodology and Investigation, ZS and IM; Data Curation and Visualization, ZS; Resources, IM, AK, and RO; Writing, ZS, IM, RO, and AD; Funding Acquisition and Supervision, AD and RO.

Funding

The study was supported by the NIH grant AR067712 (to AD), K99 CA259526 (to ZS), and research grants from Vestre Viken Hospital Trust (25C003, to IM) and the Norwegian Cancer Society (216,115, to RO).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alfonzo, J. D., Brown, J. A., Byers, P. H., Cheung, V. G., Maraia, R. J., and Ross, R. L. (2021). A call for direct sequencing of full-length RNAs to identify all modifications. *Nat. Genet.* 53 (8), 1113–1116. doi:10.1038/s41588-021-00903-1
- Armstrong, D. A., Green, B. B., Seigne, J. D., Schned, A. R., and Marsit, C. J. (2015). MicroRNA molecular profiling from matched tumor and bio-fluids in bladder cancer. *Mol. Cancer* 14, 194. doi:10.1186/s12943-015-0466-2
- Behrens, A., Rodschinka, G., and Nedialkova, D. D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell.* 81 (8), 1802–1815. doi:10.1016/j.molcel.2021.01.028
- Belfort, M., and Lambowitz, A. M. (2019). Group II intron RNPs and reverse transcripts: from retroelements to research tools. *Cold Spring Harb. Perspect. Biol.* 11 (4), a032375. doi:10.1101/cshperspect.a032375
- Boegemann, M., and Krabbe, L.-M. (2020). Prognostic implications of immunohistochemical biomarkers in non-muscle-invasive bladder cancer and muscle-invasive bladder cancer. *Mrmc* 20 (12), 1133–1152. doi:10.2174/1389557516666160512151202
- Borkowska, E. M., Konecki, T., Pietrusiński, M., Borowiec, M., and Jablonowski, Z. (2019). MicroRNAs which can prognosticate aggressiveness of bladder cancer. *Cancers* 11 (10), 1551. doi:10.3390/cancers11101551
- Boskovic, A., Bing, X. Y., Kaymak, E., and Rando, O. J. (2020). Control of noncoding RNA production and histone levels by a 5' tRNA fragment. *Genes Dev.* 34 (1-2), 118–131. doi:10.1101/gad.332783.119
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492
- Burger, M., Catto, J. W. F., Dalbagni, G., Grossman, H. B., Herr, H., Karakiewicz, P., et al. (2013). Epidemiology and risk factors of urothelial bladder cancer. *Eur. Urol.* 63 (2), 234–241. doi:10.1016/j.eururo.2012.07.033
- Cancer Genome Atlas Research (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507 (7492), 315–322. doi:10.1038/nature12965
- Chan, P. P., and Lowe, T. M. (2016). tRNADB 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 44 (D1), D184–D189. doi:10.1093/nar/gkv1309
- Chen, L., Xu, W., Liu, K., Jiang, Z., Han, Y., Jin, H., et al. (2021). 5' Half of specific tRNAs feeds back to promote corresponding tRNA gene transcription in vertebrate embryos. *Sci. Adv.* 7 (47), eabh0494. doi:10.1126/sciadv.abh0494
- Cherlin, T., Magee, R., Jing, Y., Pliatsika, V., Loher, P., and Rigoutsos, I. (2020). Ribosomal RNA fragmentation into short RNAs (rRFs) is modulated in a sex- and population of origin-specific manner. *BMC Biol.* 18 (1), 38. doi:10.1186/s12915-020-0763-0
- Chujo, T., and Tomizawa, K. (2021). Human transfer RNA modopathies: diseases caused by aberrations in transfer RNA modifications. *FEBS J.* 288 (24), 7096–7122. doi:10.1111/febs.15736
- Clark, W. C., Evans, M. E., Dominissini, D., Zheng, G., and Pan, T. (2016). tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* 22 (11), 1771–1784. doi:10.1261/rna.056531.116
- El-Mogy, M., Lam, B., Haj-Ahmad, T. A., McGowan, S., Yu, D., Nosal, L., et al. (2018). Diversity and signature of small RNA in different bodily fluids using next generation sequencing. *BMC Genomics* 19 (1), 408. doi:10.1186/s12864-018-4785-8
- Fang, Z., Dai, W., Wang, X., Chen, W., Shen, C., Ye, G., et al. (2016). Circulating miR-205: a promising biomarker for the detection and prognosis evaluation of bladder cancer. *Tumor Biol.* 37 (6), 8075–8082. doi:10.1007/s13277-015-4698-y
- Gebert, D., Hewel, C., and Rosenkranz, D. (2017). Unitas: the universal tool for annotation of small RNAs. *BMC Genomics* 18 (1), 644. doi:10.1186/s12864-017-4031-9
- Godoy, P. M., Bhakta, N. R., Barczak, A. J., Cakmak, H., Fisher, S., MacKenzie, T. C., et al. (2018). Large differences in small RNA composition between human biofluids. *Cell. Rep.* 25 (5), 1346–1358. doi:10.1016/j.celrep.2018.10.014
- Guan, L., and Grigoriev, A. (2021). Computational meta-analysis of ribosomal RNA fragments: Potential targets and interaction mechanisms. *Nucleic Acids Res.* 49 (7), 4085–4103. doi:10.1093/nar/gkab190
- Guglas, K., Kołodziejczak, I., Kolenda, T., Kopczyńska, M., Teresiak, A., Sobocińska, J., et al. (2020). YRNAs and YRNA-derived fragments as new players in cancer research and their potential role in diagnostics. *Ijms* 21 (16), 5682. doi:10.3390/ijms21165682
- Hammouz, R. Y., Kolat, D., Kaluzińska, Ż., Pluciennik, E., and Bednarek, A. K. (2021). MicroRNAs: their role in metastasis, angiogenesis, and the potential for biomarker utility in bladder carcinomas. *Cancers* 13 (4), 891. doi:10.3390/cancers13040891
- Ivanov, P., Villen, J., Gygi, S. P., Anderson, P., and Anderson, P. (2011). Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell.* 43 (4), 613–623. doi:10.1016/j.molcel.2011.06.022
- James, A. C., and Gore, J. L. (2013). The costs of non-muscle invasive bladder cancer. *Urologic Clin. N. Am.* 40 (2), 261–269. doi:10.1016/j.ucl.2013.01.004
- Janin, M., Coll-SanMartin, L., and Esteller, M. (2020). Disruption of the RNA modifications that target the ribosome translation machinery in human cancer. *Mol. Cancer* 19 (1), 70. doi:10.1186/s12943-020-01192-8
- Jiang, H., and Wong, W. H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 24 (20), 2395–2396. doi:10.1093/bioinformatics/btn429
- Kirkali, Z., Chan, T., Manoharan, M., Algaba, F., Busch, C., Cheng, L., et al. (2005). Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology* 66 (6 Suppl. 1), 4–34. doi:10.1016/j.urology.2005.07.062
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47 (D1), D155–D162. doi:10.1093/nar/gky1141
- Kumar, P., Anaya, J., Mudunuri, S. B., and Dutta, A. (2014). Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and

Acknowledgments

We appreciate the invaluable help provided by the clinicians at the Vestre Viken Hospital Trust in enrollment of patients and harvesting surgical specimens. The study was supported by the NIH grant AR067712 (to A.D.), K99 CA259526 (to Z.S.), and research grants from Vestre Viken Hospital Trust (25C003, to I.M.) and the Norwegian Cancer Society (216115, to R.O.).

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.887686/full#supplementary-material>

- associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* 12, 78. doi:10.1186/s12915-014-0078-0
- Kumar, P., Mudunuri, S. B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.* 43, D141–D145. doi:10.1093/nar/gku1138
- Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., and Dutta, A. (2018). tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* 24 (8), 1093–1105. doi:10.1261/rna.066126.118
- Kvist, E., Sjølin, K.-E., Iversen, J., and Nyholm, K. (1993). Urinary excretion patterns of pseudouridine and β -aminoisobutyric acid in patients with tumours of the urinary bladder. *Scand. J. Urology Nephrol.* 27 (1), 45–53. doi:10.3109/00365599309180413
- Lee, J.-Y., Ryu, D.-S., Kim, W.-J., and Kim, S.-J. (2016). Aberrantly expressed microRNAs in the context of bladder tumorigenesis. *Investig. Clin. Urol.* 57 (Suppl. 1), S52–S59. doi:10.4111/icu.2016.57.S1.S52
- Lee, Y. S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 23 (22), 2639–2649. doi:10.1101/gad.1837609
- Li, X., Chen, J., Hu, X., Huang, Y., Li, Z., Zhou, L., et al. (2011). Comparative mRNA and microRNA expression profiling of three genitourinary cancers reveals common hallmarks and cancer-specific molecular events. *PLoS One* 6 (7), e22570. doi:10.1371/journal.pone.0022570
- Li, X., Peng, J., and Yi, C. (2021). The epitranscriptome of small non-coding RNAs. *Non-coding RNA Res.* 6 (4), 167–173. doi:10.1016/j.ncrna.2021.10.002
- Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., et al. (2017). Base-Resolution mapping reveals distinct m¹ methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell.* 68 (5), 993–1005. doi:10.1016/j.molcel.2017.10.019
- Li, Y., Li, G., Guo, X., Yao, H., Wang, G., and Li, C. (2020). Non-coding RNA in bladder cancer. *Cancer Lett.* 485, 38–44. doi:10.1016/j.canlet.2020.04.023
- Magee, R., and Rigoutsos, I. (2020). On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Res.* 48 (17), 9433–9448. doi:10.1093/nar/gkaa657
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi:10.14806/ej.17.1.200
- Maute, R. L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K., et al. (2013). tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* 110 (4), 1404–1409. doi:10.1073/pnas.1206761110
- Meyer, B., Wurm, J. P., Kötter, P., Leisegang, M. S., Schilling, V., Buchhaupt, M., et al. (2011). The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ 1191 in yeast 18S rRNA. *Nucleic Acids Res.* 39 (4), 1526–1537. doi:10.1093/nar/gkq931
- Morais, D. R., Reis, S. T., Viana, N., Piantino, C. B., Massoco, C., Moura, C., et al. (2014). The involvement of miR-100 in bladder urothelial carcinogenesis changing the expression levels of mRNA and proteins of genes related to cell proliferation, survival, apoptosis and chromosomal stability. *Cancer Cell. Int.* 14 (1), 119. doi:10.1186/s12935-014-0119-3
- Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I., and Klaholz, B. P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. *Nature* 551 (7681), 472–477. doi:10.1038/nature24482
- Papadimitriou, M.-A., Avgeris, M., Levis, P., Papisotiriou, E. C., Kotronopoulos, G., Stravodimos, K., et al. (2020). tRNA-derived fragments (tRFs) in bladder cancer: increased 5'-tRF-LysCTT results in disease early progression and patients' poor treatment outcome. *Cancers* 12 (12), 3661. doi:10.3390/cancers12123661
- Ren, B., Wang, X., Duan, J., and Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science* 365 (6456), 919–922. doi:10.1126/science.aav8907
- Sanli, O., Dobruch, J., Knowles, M. A., Burger, M., Alemozaffar, M., Nielsen, M. E., et al. (2017). Bladder cancer. *Nat. Rev. Dis. Prim.* 3, 17022. doi:10.1038/nrdp.2017.22
- Schrier, B. P., Hollander, M. P., van Rhijn, B. W. G., Kiemeny, L. A. L. M., and Alfred Witjes, J. (2004). Prognosis of muscle-invasive bladder cancer: difference between primary and progressive tumours and implications for therapy. *Eur. Urol.* 45 (3), 292–296. doi:10.1016/j.eururo.2003.10.006
- Sharma, S., Hartmann, J. D., Watzinger, P., Klepper, A., Peifer, C., Kötter, P., et al. (2018). A single N¹-methyladenosine on the large ribosomal subunit rRNA impacts locally its structure and the translation of key metabolic enzymes. *Sci. Rep.* 8 (1), 11904. doi:10.1038/s41598-018-30383-z
- Shi, J., Zhang, Y., Tan, D., Zhang, X., Yan, M., Zhang, Y., et al. (2021). PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nat. Cell. Biol.* 23 (4), 424–436. doi:10.1038/s41556-021-00652-7
- Slack, F. J., and Chinnaiyan, A. M. (2019). The role of non-coding RNAs in oncology. *Cell.* 179 (5), 1033–1055. doi:10.1016/j.cell.2019.10.017
- Su, Z., Frost, E. L., Lammert, C. R., Przanowska, R. K., Lukens, J. R., and Dutta, A. (2020a). tRNA-derived fragments and microRNAs in the maternal-fetal interface of a mouse maternal-immune-activation autism model. *RNA Biol.* 17 (8), 1183–1195. doi:10.1080/15476286.2020.1721047
- Su, Z., Kuscu, C., Malik, A., Shibata, E., and Dutta, A. (2019). Angiogenin generates specific stress-induced tRNA halves and is not involved in tRF-3-mediated gene silencing. *J. Biol. Chem.* 294 (45), 16930–16941. doi:10.1074/jbc.RA119.009272
- Su, Z., Monshaugen, I., Wilson, B., Wang, F., Klungland, A., Ougland, R., et al. (2022). TRMT6/61A-dependent base methylation of tRNA-derived fragments regulates gene-silencing activity and the unfolded protein response in bladder cancer. *Nat. Commun.* 13 (1), 2165. doi:10.1038/s41467-022-29790-8
- Su, Z., Wilson, B., Kumar, P., and Dutta, A. (2020b). Noncanonical roles of tRNAs: tRNA fragments and beyond. *Annu. Rev. Genet.* 54, 47–69. doi:10.1146/annurev-genet-022620-101840
- Sun, X., Wei, P., Shen, C., Yang, Y., Wang, Y., Li, Y., et al. (2015). Prognostic value of the IASLC/ATS/ERS classification and IMP3 expression in lung adenocarcinoma of Chinese cases. *Am. J. Cancer Res.* 5 (7), 2266
- Tan, K.-T., Ding, L.-W., Wu, C.-S., Tenen, D. G., and Yang, H. (2021). Repurposing RNA sequencing for discovery of RNA modifications in clinical cohorts. *Sci. Adv.* 7 (32), eabd2605. doi:10.1126/sciadv.abd2605
- Telonis, A. G., Loher, P., Magee, R., Pliatsika, V., Londin, E., Kirino, Y., et al. (2019). tRNA fragments show intertwining with mRNAs of specific repeat content and have links to disparities. *Cancer Res.* 79 (12), 3034–3049. doi:10.1158/0008-5472.CAN-19-0789
- Tolkach, Y., Stahl, A. F., Niehoff, E.-M., Zhao, C., Kristiansen, G., Müller, S. C., et al. (2017). YRNA expression predicts survival in bladder cancer patients. *BMC Cancer* 17 (1), 749. doi:10.1186/s12885-017-3746-y
- van Rhijn, B. W. G., Burger, M., Lotan, Y., Solsona, E., Stief, C. G., Sylvester, R. J., et al. (2009). Recurrence and progression of disease in non-muscle-invasive bladder cancer: from epidemiology to treatment strategy. *Eur. Urol.* 56 (3), 430–442. doi:10.1016/j.eururo.2009.06.028
- Waku, T., Nakajima, Y., Yokoyama, W., Nomura, N., Kako, K., Kobayashi, A., et al. (2016). NML-mediated rRNA base methylation links ribosomal subunit formation to cell proliferation in a p53-dependent manner. *J. Cell. Sci.* 129 (12), 2382–2393. doi:10.1242/jcs.183723
- Wang, H., Zhang, W., Zuo, Y., Ding, M., Ke, C., Yan, R., et al. (2015). miR-9 promotes cell proliferation and inhibits apoptosis by targeting LASS2 in bladder cancer. *Tumor Biol.* 36 (12), 9631–9640. doi:10.1007/s13277-015-3713-7
- Yeri, A., Courtright, A., Reiman, R., Carlson, E., Beecroft, T., Janss, A., et al. (2017). Total extracellular small RNA profiles from plasma, saliva, and urine of healthy subjects. *Sci. Rep.* 7, 44061. doi:10.1038/srep44061
- Yin, X.-H., Jin, Y.-H., Cao, Y., Wong, Y., Weng, H., Sun, C., et al. (2019). Development of a 21-miRNA signature associated with the prognosis of patients with bladder cancer. *Front. Oncol.* 9, 729. doi:10.3389/fonc.2019.00729
- Yoshino, H., Seki, N., Itesako, T., Chiyomaru, T., Nakagawa, M., and Enokida, H. (2013). Aberrant expression of microRNAs in bladder cancer. *Nat. Rev. Urol.* 10 (7), 396–404. doi:10.1038/nrurol.2013.113
- Yun, S. J., Jeong, P., Kim, W.-T., Kim, T. H., Lee, Y.-S., Song, P. H., et al. (2012). Cell-free microRNAs in urine as diagnostic and prognostic biomarkers of bladder cancer. *Int. J. Oncol.* 41 (5), 1871–1878. doi:10.3892/ijo.2012.1622
- Zhang, X., Cozen, A. E., Liu, Y., Chen, Q., and Lowe, T. M. (2016). Small RNA modifications: integral to function and disease. *Trends Mol. Med.* 22 (12), 1025–1034. doi:10.1016/j.molmed.2016.10.009
- Zhang, Y. R., Shi, L., Wu, H., Tang, D. D., Wang, S. M., Liu, H. M., et al. (2014). Urinary modified nucleosides as novel biomarkers for diagnosis and prognostic monitoring of urothelial bladder cancer. *Tumori* 100 (6), 660–666. doi:10.1700/1778.19274