isomiRdb: microRNA expression at isoform resolution

Ernesto Aparicio-Puerta [1,*,†], Pascal Hirsch[1,†], Georges P. Schmartz [1],
Tobias Fehlmann [1,2], Verena Keller[1,3], Annika Engel[1], Fabian Kern [1,4],
Michael Hackenberg[5] and Andreas Keller [1,4,*]

[1]Clinical Bioinformatics, Saarland University, 66123 Saarbrücken, Germany, [2]Rejuvenome, Astera Institute, Berkeley, CA 94705, USA, [3]Department for Internal Medicine II, Saarland University Hospital, 66421 Homburg, Germany, [4]Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)–Helmholtz Centre for Infection Research (HZI), Saarland University Campus, 66123 Saarbrücken, Germany and [5]Genetics Department, Faculty of Science, Universidad de Granada, 18071 Granada, Spain

ABSTRACT

A significant fraction of mature miRNA transcripts carries sequence and/or length variations, termed isomiRs. IsomiRs are differentially abundant in cell types, tissues, body fluids or patients' samples. Not surprisingly, multiple studies describe a physiological and pathophysiological role. Despite their importance, systematically collected and annotated isomiR information available in databases remains limited. We thus developed isomiRdb, a comprehensive resource that compiles miRNA expression data at isomiR resolution from various sources. We processed 42 499 human miRNA-seq datasets (5.9 × 10^{11} sequencing reads) and consistently analyzed them using miRMaster and sRNAbench. Our database provides online access to the 90 483 most abundant isomiRs (>1 RPM in at least 1% of the samples) from 52 tissues and 188 cell types. Additionally, the full set of over 3 million detected isomiRs is available for download. Our resource can be queried at the sample, miRNA or isomiR level so users can quickly answer common questions about the presence/absence of a particular miRNA/isomiR in tissues of interest. Further, the database facilitates to identify whether a potentially interesting new isoform has been detected before and its frequency. In addition to expression tables, isomiRdb can generate multiple interactive visualisations including violin plots and heatmaps. isomiRdb is free to use and publicly available at: https://www.ccb.uni-saarland.de/isomirdb.

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNA molecules that mediate gene silencing by promoting translation repression and degradation of target mRNAs (1), mostly by binding to reverse complementary sites in 3′ UTR regions (2). At least 60% of mRNAs are miRNA targets (3) so it's not surprising that their dysregulation is frequently associated with pathologies (4). Early sequencing data revealed that, in their mature form, most miRNAs naturally display sequence and length variations (5) that can arise from imprecise Drosha and Dicer cleavage or non-templated nucleotide additions (6). Said isoforms, commonly termed isomiRs, can affect miRNA targeting (7) and stability and are differentially abundant in developmental stages (7), cell types (8), tissues, bodily fluids (9) or samples from patients, including cancer (10). Additionally, several studies have described important physiological roles of isomiRs based on altered targeting compared to their canonical counterparts (11) as well as biomarker potential for disease detection, diagnosis and prognosis (10).

Despite their apparent importance, systematically collected and annotated isomiR information available in resources remains limited. Of note, a substantial number of miRNA databases exists including those that collect sequence information (12), disease relation (13) and expression across organisms (14), tissues (15) and cell-types (16). Many of these databases compile miRNA-seq datasets from a wide spectrum of origins but only a couple provide isomiR expression information (Table 1). Furthermore, although publicly available miRNA-seq samples are rapidly increasing (17), several current resources compile expression data for a limited number of samples. Here, DIANA-miTED (18) is a remarkable exception hosting over 15 000 datasets. Among the available solutions, only IsomiR Bank (19) and Tumor IsomiR Encyclopedia (TIE) (20) include isomiR

expression data. IsomiR Bank from 2016 excels by covering expression data for eight species and is based on 2 727 samples. TIE is an agile recent resource to explore human isomiRs from over 10 000 samples. It stands out by the clear focus on cancer, relying on TCGA data. Given the importance of isomiRs, a comprehensive resource compiling extensive miRNA profiles at isoform resolution seems to be of high value for the miRNA research community.

Here, we present isomiRdb, a miRNA expression database at isoform level compiling over 90 483 isomiRs (Figure 1). The resource also provides relevant information for another ∼3 million lower abundance isoforms (all called isomiRs down to a single sample). The database builds on samples from the Sequence Read Archive (SRA) (21), the Cancer Genome Atlas (TCGA) and profiles analyzed by miRMaster (22) reaching a total of 42 499 miRNA-seq datasets. Together, these small RNA sequencing data compile 5.9×10^{11} processed reads, making it a very comprehensive and powerful miRNA expression resource. To present comparable data, we analyzed all data sets from raw fastq files that were uniformly pre-processed with miRMaster followed by downstream processing by sRNAbench (17). The annotation of miRNAs relies on the latest version of miR-base (release 22.1) (12). As described in detail below, we curated isomiRs using different quality control flags depending on the reported activity of enzymes involved in miRNA metabolism to certify the likely biological nature of said isoforms or warn about likely artifacts. Furthermore, we combined metadata from different sources to annotate isomiRdb samples with 52 tissues and 188 cell types. In sum, the database excels by the large number of samples, extensive number of sequencing reads and its detailed annotation. Combining these strengths, we provide a unique resource hosting miRNA expression data and isoform information at isomiR resolution. Access to the data is possible in multiple ways. Users can query expression values by sample, miRNA or isomiR and several interactive visualizations are available as well as tabular files ready for download. In addition to the 90 483 higher abundance isomiRs included in the online expression repository, we decided to release a list of all detected isomiRs. This set of over 3 million potential isomiRs is available on the database for query and download. isomiRdb is free to use and publicly available at: https://www.ccb.uni-saarland.de/isomirdb.

DATA AND METHODS

Generation of miRNA and isomiR profiles

We compiled 42 499 miRNA-seq human datasets from three different sources: SRA, TCGA and samples that were made accessible by miRMaster users who provided consent for aggregated secondary usage of their data, as similarly done before (23). We identified duplicated samples (i.e. SRA samples that were also uploaded to miRMaster) and removed them to avoid bias. To further minimize technical variations, we started all analyses from the raw fastq files. Here, we combined the advantages of different computational tools. First, the raw read files were pre-processed with miRMaster (22) to detect library preparation protocol, trim and collapse reads using default parameters. MiRMaster performs

adapter trimming via fuzzy string matching and read quality control applying a sliding window approach (22). Subsequently, collapsed reads were profiled with sRNAbench (17) in library mode and default parameters using miRBase release 22.1 (12). Essentially, sRNAbench relies on bowtie (24) for read alignment (with parameters *-f -n 1 -l 19 -a –best –strata*) and multiple-mapping reads are evenly split among receiving sites.

MiRNA and isomiR expression data were stored in the database considering different units: read counts (RC), reads per million (RPM, normalized to miRNA reads) and fraction of parent miRNA reads assigned to the isomiR. For miRNAs with several loci, we considered the average value across all loci. In total, isomiRdb offers miRNA expression values for all 2 656 mature miRNAs hosted in miR-Base and for the 90 483 most abundant isoforms (>1 RPM in at least 1% of the samples). A list of over 3 million isoforms with lower abundance but still detected in at least one sample is available for download.

isomiR classification system

isomiRdb relies on sRNAbench's hierarchical isomiR classification system (25). Except for canonical isoforms (i.e. the exact sequence present in the reference database), all other isomiRs are matched to one or several of the following four classes. (A) Nucleotide Variant (NucVar): nucleotide variations in the canonical sequence. (B) 5′ length variants (lv5p): 5′ end is shorter or longer but coincidental with the template sequence. (C) 3′ length variants (lv3p): 3′ end is shorter or longer but coincidental with the template sequence. (D) Non-templated additions (NTA): enzymatic post-transcriptional addition that is not included in the template sequence (e.g. NTA#T denotes an uridylation). Finally, isomiRs belonging to several classes are labelled as multiple variants (mv).

Quality control of putative miRNA isoforms

To provide users with improved confidence on isoforms that are the likely result of miRNA metabolism or to discard potential artifacts we curated all isomiRs using previously available information. As a result of this quality control, 4 tags have been incorporated: New England's Biolabs (NEB) NTA#C (warning), *bona fide* NTA#U, *bona fide* alternative processing by Drosha and unexpected alternative processing by Drosha (warning). Canonical sequences minus 1 nt at the 3′ end followed by a C or NTA#U followed by a C were flagged as likely artifacts from the NEB protocol. 5′ and 3′ length variants were tagged as *bona fide* or likely artifact after checking if their Drosha processing pattern was consistent with previous reports (26). In short, 5′ length variants of *5p* miRNA arms were labelled as *bona fide* if alternative processing for that miRNA had been previously established by Kim *et al.* (26) or as possible artifacts if they had been classified into the single-site processing group (>95% of cuts on the same site). The equivalent assignation was performed on 3′ length variants of *3p* arms. Finally, we reanalyzed data from a set of TUT7 and TUT4 knockdown experiments (27) to determine which NTA#U isoforms could be the result of enzymatic metabolism and label them accordingly.

Table 1. MiRNA and isomiR expression resources. All of them are accessible as online databases except the human cellular microRNAome, which is both an R package and a track at UCSC Genome Browser

Resource	Number of miRNA-seq datasets	Number of organisms	Summary of features	IsomiR data	Last release date
Deepbase	21 235	1 (Human)	Integration of different ncRNA profiles	No	2021
SEAweb	>42 00	10	Also provides pathogen information	No	2020
microRNAome package	2 077	1 (Human)	Manually curated cell types	No	2022
miRNATissueAtlas	188	2 (Human and mouse)	188 tissue samples from 6 individuals	No	2022
DIANA-miTED	15 183	1 (Human)	Manually curated tissues	No	2022
IsomiR Bank	2 727	8	IsomiR target prediction and enrichment analysis	Yes	2016
Tumor IsomiR Encyclopedia	11 667	1 (Human)	IsomiRs from TCGA	Yes	2021
isomiRdb	42 499	1 (Human)	Largest compilation of isomiRs	Yes	2022
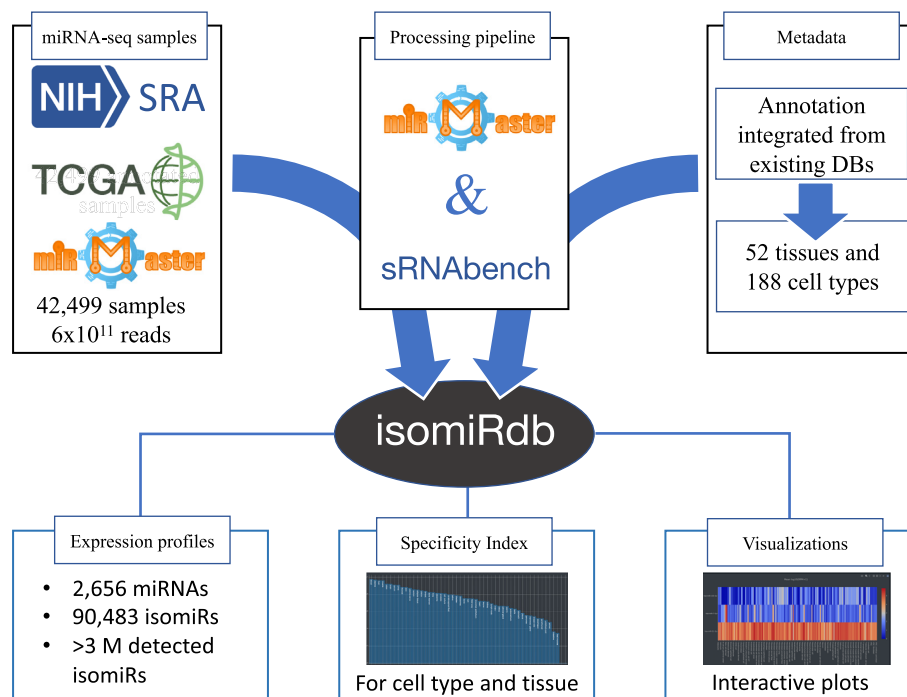
Figure 1. Development process and functionalities of isomiRdb. In short, 42 499 miRNA-seq samples from TCGA, SRA and provided by miRMaster users were uniformly processed to offer tissue and cell type expression levels of miRNAs at isomiR resolution. Tissue/cell specificity index and several types of visualizations are available.

Metadata acquisition

To provide useful results, detailed annotation with metadata is a key step. We thus integrated per sample annotation from several sources to include tissue and cell type information. Tissue annotations are available from previously curated resources including liqDB (28), mirnaQC (29), miR-Switch (23), TCGA and from SRA using OmicIDX. Cell type annotation was obtained from the human cellular microRNAome project (16). In total, 52 different tissues and 188 cell types were annotated as metadata.

Webserver and database implementation

To provide an efficient and appealing web interface for isomiRdb, we used the Django Web Framework (v.2.2), Bootstrap (v5.0.1), javascript and Plotly (v2.12.1) for interactive visualisations. Sparse representation matrix files

(.h5ad) are central for the backend database and loaded with the *anndata* python package (v 0.8.0). To estimate the specificity of tissues, we computed the tissue specificity index (TSI) of each miRNA as previously described (15,30). The same formula was also applied to mean expression of cell types to provide Cell Specificity Index (CSI).

RESULTS

Website sections and types of queries

The main functionality of isomiRdb allows querying of expression data at three different levels: samples, miRNAs and isomiRs. We implemented specific subpages for each of the three to account for different analysis aspects. In the *Samples* view, users can retrieve expression distributions of the top 3, 5, 10, 20 or 50 most expressed miRNAs in any tissue or cell-type contained in isomiRdb. MiRNA expression

values can be retrieved in three different units: RC, RPM or $\log_{10}(\text{RPM} + 1)$. Expression levels are visualised using well-known concepts such as violin plots, boxplots or bar charts. Of course, the displayed data are available for download in tabular format. Conversely, users can also query miRNAs of interest using the *miRNAs* view. In this case, after choosing a miRNA of interest, isomiRdb provides similar graphical information as in the previous view but with samples grouped by tissue or cell-type. Similarly, we implemented the core view, specific for the *isomiR* information. Here, we provide an interactive table with the option to apply several filters. This table also grants quality information about each isoform by means of tags that can indicate possible artifacts (certain NTA cytosines or unexpected Drosha cuts) or likely *bona fide* isoforms (miRNAs with previously reported biological monouridylation or alternative Drosha processing). Adjusting the sequence, parent miRNA or isomiR category allows to identify any isomiR of interest. After selecting an isomiR, the web server generates and displays all corresponding graphical representations.

As a further option, the resource offers the possibility to explore Tissue Specificity Index and Cell Specificity Index, an established estimate of expression specificity (15,30). In summary, this index provides a standardized measure of how tissue/cell-specific a miRNA is. A value of 1 means the miRNA is specific and only expressed in a particular tissue or cell type while a value of 0 means it is present in every tissue/cell type. Additionally, we provide a tool to interactively visualize heatmaps annotated by miRNA/isomiR and tissue/cell-type. Finally, a *Downloads* section allows to batch download expression data and metadata hosted in isomiRdb. Of note, the resource provides expression values for isomiRs that meet the minimum expression threshold of at least 1 RPM in at least 1% of the samples for two reasons. Showing all isoforms would not only lead to performance restrictions but might also introduce noise to the data. More experienced scientists that are interested in detailed and holistic analyses of the isomiR information can download the lower abundance isomiR repertoire from the web page.

Use case: ubiquitously expressed miR-21 and its variants

There are various application scenarios for our database. It allows for very specific considerations. For example, researchers can make use of the resource to find out whether one isomiR is expressed at all and in which tissues. The evaluation of the complete repertoire of isomiRs for one miRNA is feasible too. Finally, complex analyses are also possible, for example users can explore the isomiR repertoire of one cell type or tissue. As use case we provide insights in miR-21-5p. This miRNA is ubiquitously expressed and has been proposed as a biomarker for at least 29 diseases (31). Using the *miRNAs* view, we can easily check that it is indeed present in every tissue and cell type. In fact, its average expression exceeds 1000 RPM in any cell type or tissue (Figure 2A-B). We next checked the miR-21-5p isoform expression pattern by means of the *Heatmaps* view (Figure 2C). On inspection of the isomiR profile, some variation exists on the most abundant isoforms ranging from

canonical sequence to 3′ length variants of ± 1 nt but no general trends are obvious. One isomiR though, *TAGCTT ATCAGACTGATGTTGAT*, an NTA-U, exhibits an interesting expression pattern: while in most tissues its expression is between 0–50 RPM, a few tissues have a considerably higher expression with some exceeding 300 RPM (Figure 2C). These tissues include bone, bone marrow, heart, mammary gland, placenta, liver, bronchial epithelium and the umbilical cord as well as blood and serum. Other tissues have similar miR-21-5p levels without displaying the increased NTA-U, so this pattern cannot be explained by an expression bias of miR-21-5p. Nevertheless, the small fraction of miR-21-5p reads taken by this isoform, <8%, suggests moderate relative abundance only. Recent work on miRNA kinetics has proposed that miRNA transcripts become uridylated after mRNA targeting (32). Unfortunately, this NTA#U isoform is not tagged as validated by previous data because the uridylation decrease detected after TUT4/7 knockdown was not significant for miR-21-5p (27). This kind of observation is prototypic for the usefulness of isomiRdb in generating new hypothesis and in validating the expression background at miRNA and isomiR level.

Use case: a 5′ length variant of miR-181a-5p is almost exclusively detected in urine

Besides their function as key regulatory elements of gene expression, miRNAs hold great interest as biomarkers among other properties because of their stability in bodily fluids. Several studies partially compiled elsewhere (28) have attempted to define circulating miRNA disease biomarkers. In this context, urine biomarkers could be particularly non-invasive, an important aspect for continuous monitoring. One possible example of an interesting biomarker candidate is miR-181a-5p, a miRNA that different studies have reported to have predictive value of therapy outcome (33,34). Using the *miRNAs* view of isomiRdb, we can quickly check that miR-181a-5p is highly expressed in a wide range of tissues and particularly in urine. Furthermore, we can also obtain the isomiR profile using the *Heatmaps* page. The expression heatmap reveals that even though miR-181a-5p is already quite urine specific (0.83 TSI) there is a 5′ length variant isoform, *ACATTCAACGCTGTCGGTGAGT*, that is almost exclusively present in this fluid (20 413 RPM versus <50 RPM elsewhere). Furthermore, by inspecting the fraction of miR-181a-5p reads assigned to this isoform, the pattern appears even more striking: the relative abundance in most other tissues comes close to 0% compared to 88% in urine. Using isomiRdb downloadable metadata, we confirmed that urine samples ($n = 63$) came from four different studies, which made us discard the influence of any protocol specific artifact. Moreover, the quality control pipeline tagged this variant as likely *bona fide* because alternative Drosha processing for miR-181a-5p had previously been reported (26). Nonetheless, such prominent differences should be taken with caution. This example illustrates a basic data exploration using isomiRdb and how users can benefit from the hosted quality tags and sample information to increase support of their analysis.
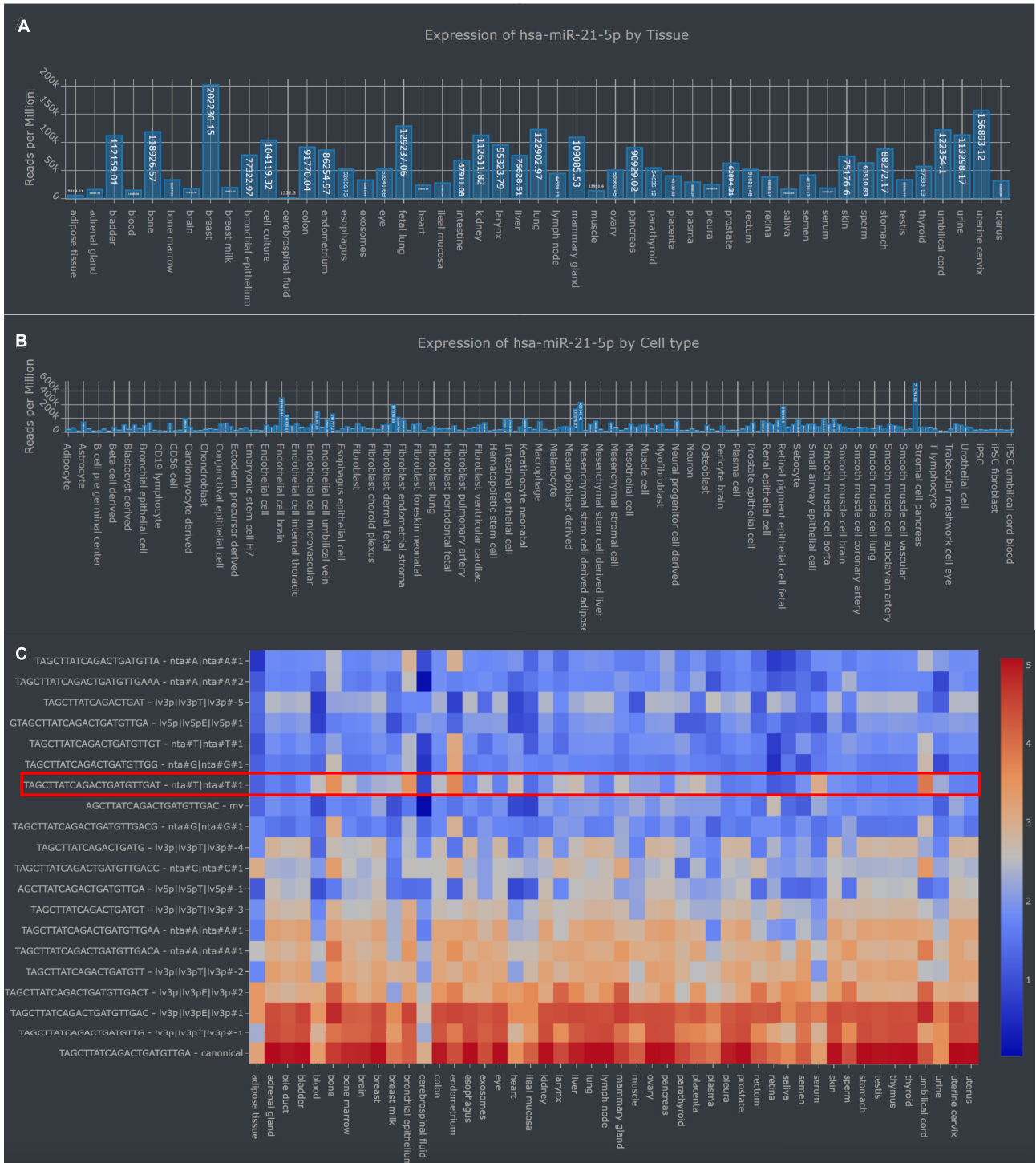
Figure 2. Different visualizations of miR-21–5p expression. (**A**) Mean expression of miR-21-5p per tissue. (**B**) Mean expression of miR-21-5p per cell type. (**C**) Heatmap of miR-21-5p isomiRs. A NTA-U isoform described in the use case is highlighted in red.

DISCUSSION

IsomiRs are widespread and expressed on a cell-type dependent manner. Different studies have proposed physiological roles for miRNA variants. For instance, miRNA transcripts seem to get monouridylated because of target regulation, a modification that destabilizes them (32). More recently, 3′ uridylation has been linked to alternative target repression in specific miRNAs (35). Another recent study describes that 1-nt-short isomiRs typically accumulate in several cancer tissues and cancer cell lines (36). Interestingly, the same work explains how RDR1, an exogenous plant immune protein, can add mononucleotides to the 3′ end of these isomiRs to restore the length of their canonical sequence in mammalian cells. Moreover, these modifications resulted in antitumor effects in *in vivo* mouse models. Not surprisingly, several studies consider isomiRs in the quest for disease biomarkers (10,37). Remarkably, some models successfully distinguish between different types of cancer. Although sequencing errors may seemingly inflate the number of isoforms in a sample, there is currently plenty of evidence to support a role of isomiRs beyond mere experimental artifacts. We have taken advantage of previously available evidence on Drosha processing (26) and TUT4/7 monouridylation patterns (27) to curate biologically relevant isomiRs, a feature that radically increases confidence on entries hosted in isomiRdb.

Despite their increasingly apparent importance, until now few resources are available to query detected isomiR sequences as well as their abundance. The two prominent examples are IsomiR Bank (19), with a thoroughly curated dataset, and TIE (20), based on TCGA data. IsomiR Bank that excels by its multi-species support, relies on a relatively low number of samples which complicates tracking of less common miRNA variants. On the contrary, TIE builds on >10 000 samples and has a clear scope to one organism (human) and a disease focus (cancer). Although most isoforms will still be present in cancerous tissue, their expression levels are probably altered by the disease. Besides, TCGA does not cover all human tissues and cell types, making it an insufficient resource to be used as background control for isomiR-based biomarker discovery. IsomiRdb aims to provide a comprehensive view on high- and low-abundance isomiRs. It offers an exceptional range of samples, tissues and cell types with plenty of coverage to detect even the lowliest expressed isoforms. Of course, such a comprehensive compilation of isomiRs unavoidably contains noisy and false positive isoforms (artifacts originated by library construction or sequencing), particularly those that are only present in a handful of experiments. To account for this issue, we performed quality control on all isomiR entries to flag potential artifacts and certify likely biological isoforms. Furthermore, we base the main part of the resource on higher abundance isomiRs by setting a threshold of 1 RPM in 1% of the samples. Although any threshold of this nature can appear arbitrary (we keep approximately the top 3% isoforms), we understand it's necessary to relegate isomiRs that are not widespread enough in the interest of clarity. Nevertheless, users can still explore if any isoform has previously been detected at all and in how many samples, which can be useful in some situations like for isomiRs only expressed in tissues underrepresented in isomiRdb. With 3 million detected isomiRs across 42 499 miRNA-seq samples, isomiRdb comes close to the limit of naturally occurring isoforms in known miRNAs (an average of ∼1200 isomiRs per mature miRNA) and it's unlikely that many true isomiRs have escaped our analysis. The limitations inherent to the annotation or the sequencing process might be an exception in this regard.

Besides the obvious interest of isomiRdb as an isomiR expression database, the very broad collection of miRNA-seq expression data, in terms of the number of samples and sequencing reads, is remarkable. The compiled dataset can be used to answer several research questions related to the expression of miRNAs. One current limitation of isomiRdb is that it does not provide validation information on isomiR processing by Dicer or TENT2, two enzymes that have long been established to participate in miRNA metabolism and that can generate alternative miRNA transcripts. Future releases of isomiRdb should include quality tags supporting or refuting involvement of these enzymes in a similar way to the already present tags for TUT4/7 and Drosha. Another important limitation of isomiRdb compared to similar resources (19) is that it is focused on human samples. Although other species are not nearly as present in SRA or other publicly available sources, it would be of much interest to include comparably annotated datasets from evolutionarily related species to explore conserved patterns in the expression of miRNAs and isomiRs. Such analysis would not be without challenges given the dissimilar quality of miRNA annotations across species, although they could be mitigated by use of an evolution aware database like MirGeneDB (38). In this context, *Mus musculus* is probably the most appropriate candidate model organism to be included in future releases of isomiRdb given its interest to the research community, the quality of its annotation and the abundance of miRNA-seq data available on SRA, only behind *Homo sapiens*. We thus initiated to annotate and process *Mus musculus* miRNA-seq experiments in the same manner that we originally did for human samples with the goal of incorporating them to isomiRdb in the foreseeable future.

REFERENCES

1. Gebert,L.F.R. and MacRae,I.J. (2018) Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.*, **20**, 21–37.
2. Valencia-Sanchez,M.A., Liu,J., Hannon,G.J. and Parker,R. (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.*, **20**, 515–524.
3. Friedman,R.C., Farh,K.K.H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
4. Keller,A., Leidinger,P., Vogel,B., Backes,C., ElSharawy,A., Galata,V., Mueller,S.C., Marquart,S., Schrauder,M.G., Strick,R. *et al.* (2014) miRNAs can be generally associated with human pathologies as exemplified for miR-144. *BMC Med.*, **12**, 224.
5. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
6. Neilsen,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet.*, **28**, 544–549.
7. Llorens,F., Bañez-Coronel,M., Pantano,L., del Río,J.A., Ferrer,I., Estivill,X. and Martí,E. (2013) A highly expressed miR-101 isomiR is a functional silencing small RNA. *BMC Genomics*, **14**, 104.
8. Tan,G.C., Chan,E., Molnar,A., Sarkar,R., Alexieva,D., Isa,I.M., Robinson,S., Zhang,S., Ellis,P., Langford,C.F. *et al.* (2014) 5′ isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.*, **42**, 9424–9435.
9. Rubio,M., Bustamante,M., Hernandez-Ferrer,C., Fernandez-Orth,D., Pantano,L., Sarria,Y., Piqué-Borras,M., Vellve,K., Agramunt,S., Carreras,R. *et al.* (2018) Circulating miRNAs, isomiRs and small RNA clusters in human plasma and breast milk. *PLoS One*, **13**, e0193527.
10. Telonis,A.G., Magee,R., Loher,P., Chervoneva,I., Londin,E. and Rigoutsos,I. (2017) Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res.*, **45**, 2973–2985.
11. Yang,A., Bofill-De Ros,X., Shao,T.J., Jiang,M., Li,K., Villanueva,P., Dai,L. and Gu,S. (2019) 3′ Uridylation confers miRNAs with Non-canonical target repertoires. *Mol. Cell*, **75**, 511.
12. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
13. Huang,Z., Shi,J., Gao,Y., Cui,C., Zhang,S., Li,J., Zhou,Y. and Cui,Q. (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.*, **47**, D1013–D1017.
14. Rahman,R.U., Liebhoff,A.M., Bansal,V., Fiosins,M., Rajput,A., Sattar,A., Magruder,D.S., Madan,S., Sun,T., Gautam,A. *et al.* (2020) SEAweb: the small RNA expression atlas web application. *Nucleic Acids Res.*, **48**, D204–D219.
15. Keller,A., Gröger,L., Tschernig,T., Solomon,J., Laham,O., Schaum,N., Wagner,V., Kern,F., Schmartz,G.P., Li,Y. *et al.* (2022) miRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res.*, **50**, D211–D221.
16. Patil,A.H., Baran,A., Brehm,Z.P., McCall,M.N. and Halushka,M.K. (2022) A curated human cellular microRNAome based on 196 primary cell types. *Gigascience*, **11**, https://doi.org/10.1093/gigascience/giac083.
17. Aparicio-Puerta,E., Gomez-Martín,C., Giannoukakos,S., Medina,J.M., Scheepbouwer,C., García-Moreno,A., Carmona-Saez,P., Fromm,B., Pegtel,M. *et al.* (2022) sRNAbench and sRNAtoolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms. *Nucleic Acids Res.*, **50**, W710–W717.
18. Kavakiotis,I., Alexiou,A., Tastsoglou,S., Vlachos,I.S. and Hatzigeorgiou,A.G. (2022) DIANA-miTED: a microRNA tissue expression database. *Nucleic Acids Res.*, **50**, D1055–D1061.
19. Zhang,Y., Zang,Q., Xu,B., Zheng,W., Ban,R., Zhang,H., Yang,Y., Hao,Q., Iqbal,F., Li,A. *et al.* (2016) IsomiR bank: a research resource for tracking IsomiRs. *Bioinformatics*, **32**, 2069–2071.
20. Bofill-De Ros,X., Luke,B., Guthridge,R., Mudunuri,U., Loss,M. and Gu,S. (2021) Tumor IsomiR encyclopedia (TIE): a pan-cancer database of miRNA isoforms. *Bioinformatics*, **37**, 3023–3025.
21. Kodama,Y., Shumway,M. and Leinonen,R. (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
22. Fehlmann,T., Kern,F., Laham,O., Backes,C., Solomon,J., Hirsch,P., Volz,C., Müller,R. and Keller,A. (2021) miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale. *Nucleic Acids Res.*, **49**, W397–W408.
23. Kern,F., Amand,J., Senatorov,I., Isakova,A., Backes,C., Meese,E., Keller,A. and Fehlmann,T. (2020) miRSwitch: detecting microRNA arm shift and switch events. *Nucleic Acids Res.*, **48**, W268–W274.
24. Langmead,B. (2010) Aligning short sequencing reads with bowtie. *Curr Protoc. Bioinformatics*, **Chapter 11**, Unit 11.7.
25. Aparicio-Puerta,E., Lebrón,R., Rueda,A., Gómez-Martín,C., Giannoukakos,S., Jaspez,D., Medina,J.M., Zubkovic,A., Jurak,I., Fromm,B. *et al.* (2019) sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Res.*, **47**, W530–W535.
26. Kim,K., Baek,S.C., Lee,Y.Y., Bastiaanssen,C., Kim,J., Kim,H. and Kim,V.N. (2021) A quantitative map of human primary microRNA processing sites. *Mol. Cell*, **81**, 3422–3439.
27. Kim,H., Kim,J., Kim,K., Chang,H., You,K.v. and Kim,N. (2019) Bias-minimized quantification of microRNA reveals widespread alternative processing and 3′ end modification. *Nucleic Acids Res.*, **47**, 2630–2640.
28. Aparicio-Puerta,E., Jáspez,D., Lebrón,R., Koppers-Lalic,D., Marchal,J.A. and Hackenberg,M. (2019) liqDB: a small-RNAseq knowledge discovery database for liquid biopsy studies. *Nucleic Acids Res.*, **47**, D113–D120.
29. Aparicio-Puerta,E., Gomez-Martin,C., Giannoukakos,S., Maria Medina,J., Marchal,J.A. and Hackenberg,M. (2020) mirnaQC: a webserver for comparative quality control of miRNA-seq data. *Nucleic Acids Res.*, **48**, W262–W267.
30. Yanai,I., Benjamin,H., Shmoish,M., Chalifa-Caspi,V., Shklar,M., Ophir,R., Bar-Even,A., Horn-Saban,S., Safran,M., Domany,E. *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–659.
31. Jenike,A.E. and Halushka,M.K. (2021) miR-21: a non-specific biomarker of all maladies. *Biomark. Res.*, **9**, 18.
32. Baccarini,A., Chauhan,H., Gardner,T.J., Jayaprakash,A.D., Sachidanandam,R. and Brown,B.D. (2011) Kinetic analysis reveals the fate of a MicroRNA following target regulation in mammalian cells. *Curr. Biol.*, **21**, 369–376.
33. Meijer,L.L., Garajová,I., Caparello,C., le Large,T.Y.S., Frampton,A.E., Vasile,E., Funel,N., Kazemier,G. and Giovannetti,E. (2020) Plasma miR-181a-5p downregulation predicts response and improved survival after FOLFIRINOX in pancreatic ductal adenocarcinoma. *Ann. Surg.*, **271**, 1137–1147.
34. Millán,O., Ruiz,P., Orts,L., Ferré,P., Crespo,G., Santana,M., Fortuna,V., Quintairos,L., Navasa,M. and Brunet,M. (2019) Monitoring of miR-181a-5p and miR-155-5p plasmatic expression as prognostic biomarkers for acute and subclinical rejection in de novo adult liver transplant recipients. *Front. Immunol.*, **10**, 873.
35. Yang,A., Bofill-De Ros,X., Shao,T.J., Jiang,M., Li,K., Villanueva,P., Dai,L. and Gu,S. (2019) 3′ Uridylation confers miRNAs with non-canonical target repertoires. *Mol. Cell*, **75**, 511–522.
36. Qi,Y., Ding,L., Zhang,S., Yao,S., Ong,J., Li,Y., Wu,H. and Du,P. (2022) A plant immune protein enables broad antitumor response by rescuing microRNA deficiency. *Cell*, **185**, 1888–1904.
37. Starlinger,P., Hackl,H., Pereyra,D., Skalicky,S., Geiger,E., Finsterbusch,M., Tamandl,D., Brostjan,C., Grünberger,T., Hackl,M. *et al.* (2019) Predicting postoperative liver dysfunction based on blood-derived microRNA signatures. *Hepatology*, **69**, 2636.
38. Fromm,B., Høye,E., Domanska,D., Zhong,X., Aparicio-Puerta,E., Ovchinnikov,V., Umu,S.U., Chabot,P.J., Kang,W., Aslanzadeh,M. *et al.* (2022) MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res.*, **50**, D204–D210.