MiREDiBase: a manually curated database of editing events in microRNAs

Gioacchino P. Marceca[1†], Rosario Distefano[2†], Luisa Tomasello[2], Alessandro Lagana'[3], Francesco Russo[4], Federica Calore[2], Giulia Romano[5], Marina Bagnoli[6], Pierluigi Gasparini[2,7], Alfredo Ferro[1], Mario Acunzo[5], Qin Ma[8], Carlo M. Croce[2‡*] and Giovanni Nigita[2‡*]

[1] Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy.

[2] Department of Cancer Biology and Genetics and Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA.

[3] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[4] Translational Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[5] Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, USA

[6] Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy.

[7] The University of Newcastle, School of Medicine and Public Health, Hunter Medical Research Institute, Newcastle, New South Wales, Australia.

[8] Department of Biomedical Informatics, College of Medicine, The Ohio State University, USA.

*Corresponding Authors:

Carlo M. Croce, M.D.
1082 Biomedical Research Tower
460 West 12th Avenue
Columbus, 43210 OH, USA
Phone: (614) 292-4930
Fax: (614) 292-3558
Email: carlo.croce@osumc.edu

Giovanni Nigita, Ph.D.
1096 Biomedical Research Tower
460 West 12th Ave
Columbus, 43210 OH, USA
Phone: (614) 688-5914
Email: giovanni.nigita@osumc.edu

† These authors have contributed equally to this work

‡ These authors have contributed equally to this work

Keywords: database, microRNAs, microRNA editing, A-to-I RNA editing, C-to-U RNA editing, ADAR enzymes, functional characterization.

Abstract

MicroRNAs (miRNAs) are a class of regulatory small non-coding RNAs that modulate gene expression. As such, miRNAs are indirectly involved in most cellular mechanisms, including cell differentiation, proliferation, survival, and homeostasis. Several different mechanisms finely regulate miRNA expression levels and functions. Among these, miRNA editing is a type of epitranscriptomic modification that alters the original nucleotide sequence of selected miRNAs, possibly influencing their conformational state and target-binding ability. To date, adenosine-to-inosine (A-to-I) and cytosine-to-uracil (C-to-U) RNA editing are recognized as the “canonical” editing types, with the A-to-I type being the predominant one. A comprehensive resource explicitly dedicated to miRNA editing data collection is still missing. Here we present MiREDiBase, a manually curated catalog of editing events in miRNAs. Overall, the current version includes 3059 unique A-to-I and C-to-U editing sites occurring in 626 human miRNA transcripts and three different species of non-human primates. Each editing event is provided with essential information and key relevant details, including biological sample, detection method, enzyme-dependent affinity, editing level, and biological function. Editing events in human mature miRNAs were provided with miRNA-target predictions and enrichment analysis. Minimum free energy structures were inferred for editing events falling into pre-miRNA regions, aiming to help users to interpret the biological information. MiREDiBase represents a valuable tool for cell biology and biomedical research, continuously updated, improved, and expanded with new data from the literature. MiREDiBase is available at https://ncrnaome.osumc.edu/miredibase.

Introduction

MicroRNAs (miRNAs) represent the most studied and best-characterized class of small non-coding RNAs, which are involved in gene expression regulation. According to the canonical miRNA biogenesis pathway, miRNAs are initially transcribed into primary transcripts (pri-miRNAs) that present hairpin structures and undergo a double RNase III-mediated cleavage process (Ha and Kim 2014). The first processing step occurs within the nucleus, where the Drosha-DGCR8 enzymatic complex cleaves pri-miRNAs into ~70 nucleotide long transcripts. These typically maintain the stem-loop conformation and represent the direct precursors of miRNAs (pre-miRNAs). Pre-miRNAs are then exported to the cytoplasm, where they are ultimately processed by Dicer into ~22 nucleotides long single-stranded RNAs (mature miRNAs) (Ha and Kim 2014). Mature miRNAs can be found as miRNA-5p or miRNA-3p forms, depending on whether they derive from the pre-miRNA 5' or 3' arm, respectively (Ha and Kim 2014). To date, it is estimated that more than 1900 pre-miRNAs are expressed in humans, giving rise to over 2600 different mature miRNAs (Kozomara et al. 2019).

MiRNAs are important modulators of gene expression as they act as translational repressors (Bartel 2009; Jonas and Izaurralde 2015). The general rule underlying their inhibitory activity consists in a thermodynamically stable base pairing between a specific miRNA region, termed "seed region," and a partially complementary nucleotide sequence of an mRNA transcript, termed "miRNA responsive element" (MRE). Such a mechanism is dependent upon AGO-miRNA complexes and very often causes the enzymatic degradation of the targeted mRNA (Bartel 2009; Jonas and Izaurralde 2015). Conventionally, the seed region consists of nucleotides 2-8 located at the 5' end of miRNAs and is usually assumed to interact with MREs included within the 3' untranslated region (3' UTR) of target mRNAs (Bartel 2009; Jonas and Izaurralde 2015). MiRNAs influence a vast range of physiological processes, including cell cycle control (Liang and He 2011), angiogenesis (Salinas-Vera et al. 2018), brain development (Ziats and Rennert 2014), behavioral changes, and cognitive processes (Wingo et al. 2020). miRNA expression deregulations or mutations in their miRNA seed regions/MREs often leads to the development of several pathologies, including cardiovascular diseases (Ultimo et al. 2018) and cancer (Croce 2009; Liang and He 2011). Thus, changes in miRNAs sequence can have a dramatic impact on cell biology and organ physiology.

RNA editing consists of the co- or post-transcriptional enzymatic modification of a primary RNA sequence through single-nucleotide substitutions, insertions, or deletions (Gott and Emeson 2000). Recent transcriptome-wide analyses have revealed the pervasive presence of RNA editing in the human transcriptome. Currently, adenosine-to-inosine (A-to-I) and cytosine-to-uracil (C-to-U) RNA editing are considered the canonical editing types (Paul et al. 2020) with the A-to-I type being the most prevalent one (Picardi et al. 2015; Zheng et al. 2016; Brümmer et al. 2017; Li et al. 2018).

A-to-I RNA editing is catalyzed by enzymes of the Adenosine Deaminase Acting on RNA (*ADAR*) family, specifically ADAR1 (two isoforms) and ADAR2 (one predominant isoform) (Nishikura 2016). A-to-I RNA editing frequently occurs in non-coding transcripts, including primary miRNA transcripts, and shows tissue-dependent patterns (GTEx Consortium et al. 2017; Picardi et al. 2015; Li et al. 2018). C-to-U RNA editing, instead, is known to be catalyzed by enzymes of the Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide (*APOBEC*) family, specifically APOBEC1 and APOBEC3, at least in the context of mRNAs (Wedekind et al. 2003). However, to date, no proof has been reported concerning the role of APOBECs in miRNA editing, and only a few studies have discussed this type of modification in miRNAs (Joyce et al. 2011; Negi et al. 2015).

Editing of miRNA precursors exerts significant effects on miRNA biogenesis and function, with profound implications in the human pathophysiology, such as the progression of neurodegenerative diseases and cancers (Han et al. 2015; Nishikura 2016). In certain circumstances, the editing of pri-miRNAs seems to induce a local structural change that avoids the Drosha–DGCR8 complex from recognizing its hairpin conformation. Consequently, Drosha is prevented from cleaving the pri-miRNA and allocate the latter to degradation through endonucleases (Nishikura 2016). Differently, certain editing events that fall within the mature miRNA region do not induce suppression of the primary transcripts but are maintained until miRNA processing is completed. These last events generate miRNAs diversified in their primary sequence, subsequently causing a change in their target repertoire (miRNA re-targeting) (Kawahara et al. 2007; Nigita et al. 2016).

Given the extensive number of high-confidence RNA editing sites identified so far and their established relevance in the biomedical field, several efforts have been made to develop online resources capable of summarizing, contextualizing, and interpreting such data. Databases like DARNED (Kiran and Baranov 2010), RADAR (Ramaswami and Li 2014), REDIportal (Picardi et al. 2017), and EDK (Niu et al. 2019), and more

complex resources such as TCEA (Lin and Chen 2019) are well known. However, no dedicated online resources have been explicitly implemented for the study of miRNA editing until now. Here, we present MiREDiBase, the first comprehensive and integrative catalog of miRNA editing events. MiREDiBase is manually curated and provides users with valuable information for the study of edited miRNAs as potential biomarkers of disease.

Results

Data Collection

The MiREDiBase data processing workflow is depicted in Fig. 1. Firstly, we explored the PubMed literature by searching for specific keywords, such as “microRNA editing” and “miRNA editing,” narrowing the temporal search range between 2000 and 2019. Retrieved articles were then manually filtered, discarding those not containing information on miRNA editing. Among all editing events detected through the exploitation of targeted or wide-transcriptome methods, we retained only those established as “reliable” (based on experimental validation) or “high-confidence” (based on statistical significance) in each study. Statistical significance was taken into consideration when possible, eventually maintaining only significant editing events. For all those putative edited pre-miRNA sequences with no official miRNA name, e.g., “Antisense-hsa-mir-451” in Blow and colleagues (Blow et al. 2006), we employed the BLASTN tool to generate alignments between the putative pre-miRNA sequence and miRBase’s pre-miRNA sequences (v.22) (Griffiths-Jones 2006; Kozomara et al. 2019). Only perfect matches were retained and provided with their respective official name, as indicated by miRBase. Moreover, in case editing positions were presented in the form of coordinates of previous genomic assemblies (i.e., hg19/GRCh37), these were converted to the hg38/GRCh38 assembly using the University of California Santa Cruz (UCSC) *liftOver* tool (Haeussler et al. 2019). Editing sites associated with miRBase’s dead-entries were discarded.

In the second step, we expanded our search by employing the three most prominent online resources for A-to-I events available at present: DARNED (Kiran and Baranov 2010), RADAR (Ramaswami and Li 2014), and

REDIportal (Picardi et al. 2017). Resources were manually screened, removing all those editing sites associated with dead entries and opposite strands. Editing sites falling into misassigned miRNAs in the hg19 genomic assembly (i.e., miRNAs of the mir-548 family and mir-3134 present in DARNED) were excluded from the database. The retained data were then integrated into the initial dataset (Fig. 1).

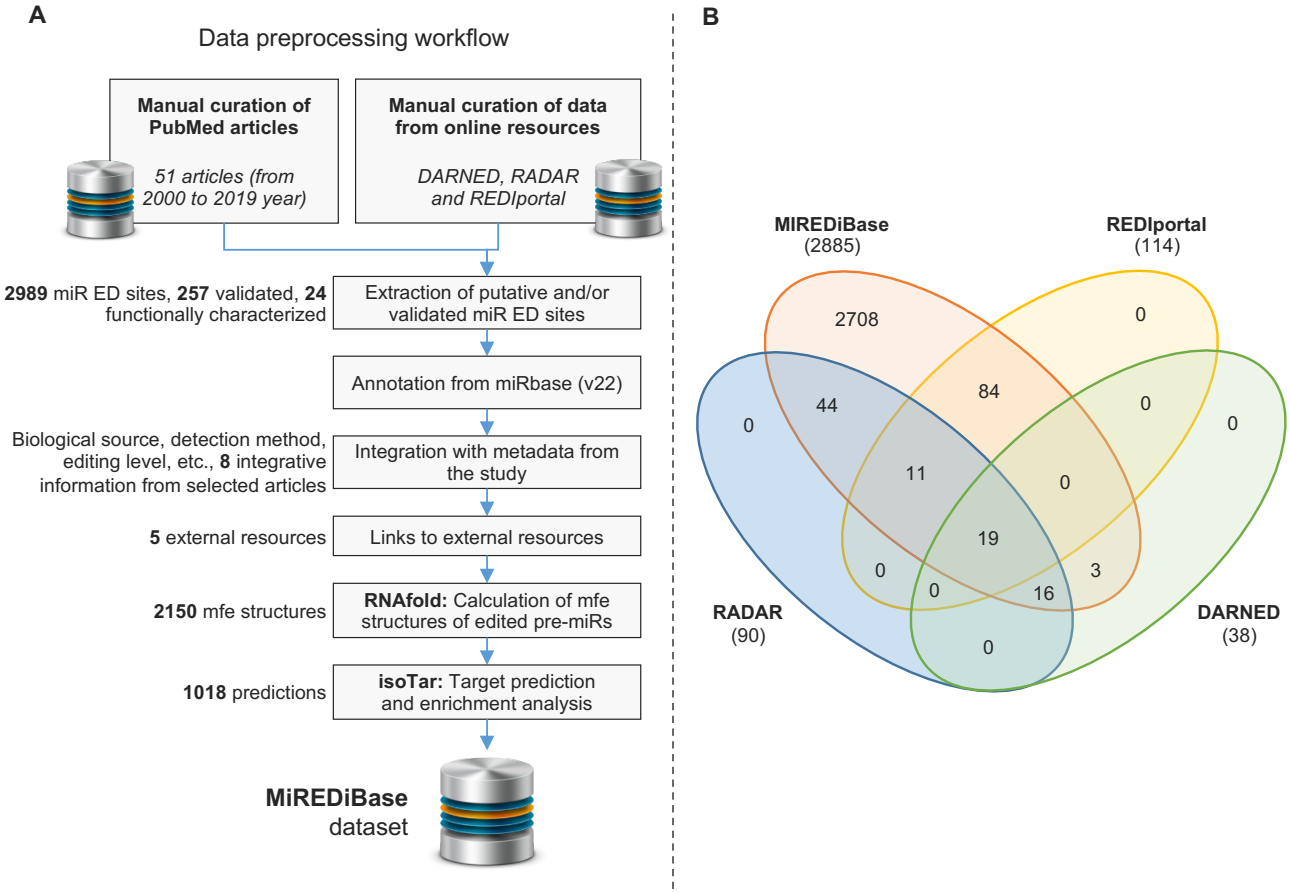

Figure 1. (A) Chart is showing the workflow underlying miRTarBase. (B) Venn diagram showing the intersection of MiREDiBase with the three most prominent online repositories of A-to-I RNA editing events: DARNED, RADAR, and REDIportal. Data used for the intersection were exclusively relative to miRNA editing. The data were filtered for dead entries, opposite strands, and misassigned miRNAs prior to comparison.

Database Content and Editing Sites Distribution

Considering the recent knowledge about genomic differences and similarities among primates (Rogers and Gibbs 2014), we retained data from *Homo sapiens* and three primate species (*Pan troglodytes*, *Gorilla gorilla*, and *Macaca mulatta*). In particular, the current version of MiREDiBase includes 2,989 unique A-to-I (2,885) and C-to-U (104) editing events occurring in 571 human miRNA transcripts (Fig. 2; Supplemental Data Set 1)

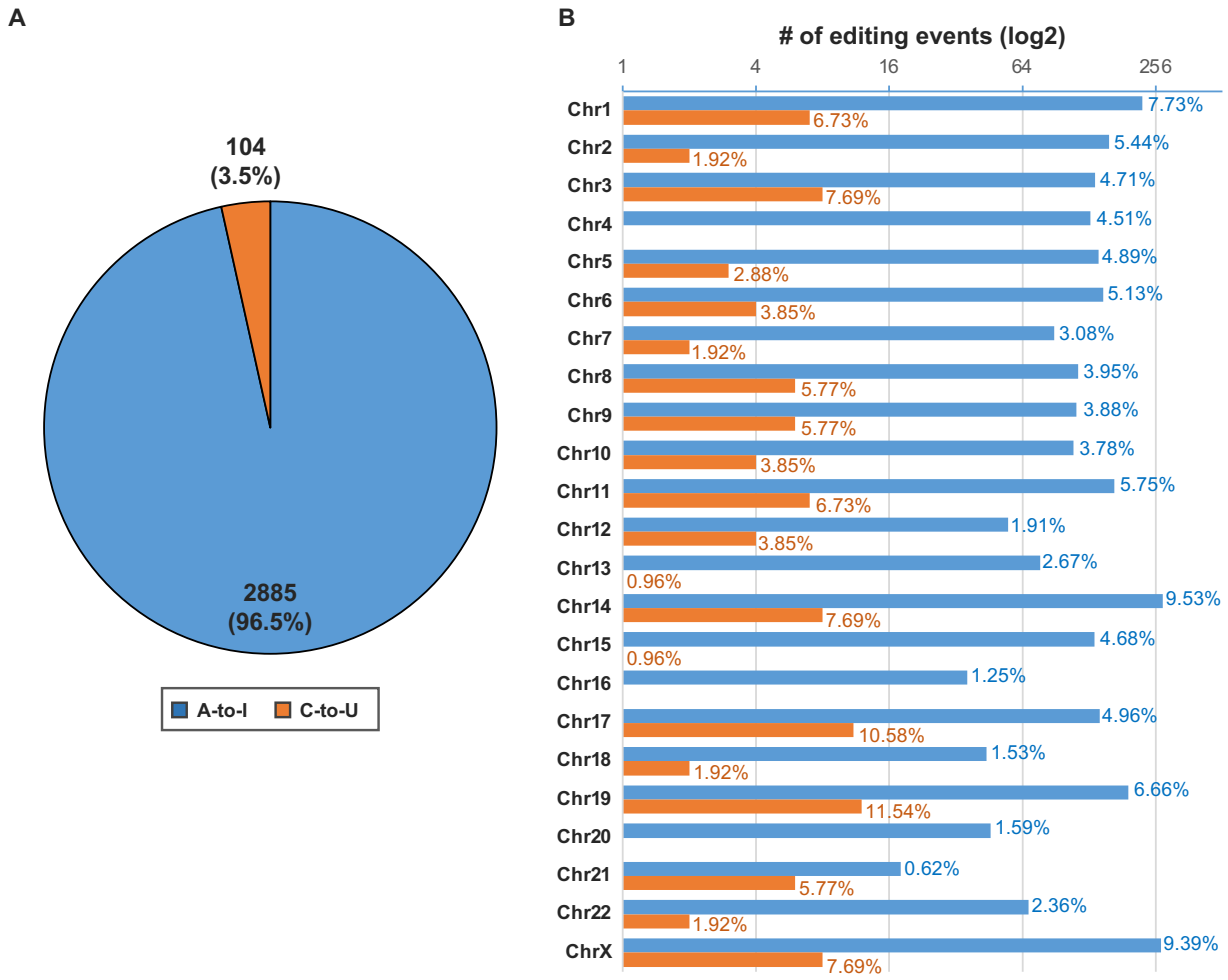

*Figure 2. Descriptive statistics about human data in the current version of MiREDiBase. (**A**) The number of unique A-to-I and C-to-U editing events reported from human tissues. (**B**) Distribution of A-to-I and C-to-U miRNA editing events per chromosome in the human genome. The percentage in each bar represents the percentage of the specific editing type (A-to-I or C-to-U) per chromosome, calculated respect the total number of that type editing event (A-to-I or C-to-U).*

and 70 unique (46 A-to-I and 24 C-to-U) editing events taking place in 55 primate miRNA transcripts (Supplemental Fig. S1; Supplemental Data Set 1). Overall, these data were extracted from 51 original papers (Supplemental Table S1) and refer to 256 biological sources (Supplemental Tables S2-5).

Human editing sites in MiREDiBase are distributed across several genomic positions throughout the human genome, covering all chromosomes (Fig. 2B). However, of the 2,989 unique editing sites, only 257 (8.6%) have been validated by low-throughput methods or ADAR expression perturbation experiments. The majority of human editing events fall into clustered miRNAs located on chromosome (chr) 14 (9.5% of A-to-I and 7.7% of C-to-U editing events), chr X (9.4% of A-to-I and 7.7% of C-to-U editing events), chr 1 (7.7% of A-to-I and 6.7%

of C-to-U editing events), and chr 19 (6.7% of A-to-I and 11.5% of C-to-U editing events), respectively. Such a phenomenon very likely depends on the presence of local structural elements and motifs in these primary transcripts that function as editing inducers (Nigita et al. 2015; Daniel et al. 2017) and would deserve more in-depth investigations.

About primates, the majority of data refer to macaque (*Macaca mulatta*), for which our database reports 40 A-to-I and 24 C-to-U editing sites (Supplemental Fig. S2, S3; Supplemental Data Set 1). Here, 26 (65%) A-to-I editing sites are conserved between human and macaque, whereas only 8 (33%) C-to-U sites are conserved between these two species. This figure might suggest that A-to-I editing of miRNA transcripts is more conserved than the C-to-U type; however, it might also be due to the current low number of C-to-U instances reported for both human and non-human primates. Only three editing sites are reported for both chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*), occurring in one pre-miRNA transcript for each species (Supplemental Data Set 1). All of these six sites are conserved between non-human primates and humans.

Altogether, 909 editing events occur outside of the pre-miRNA sequences, 971 within pre-miRNA sequences, outside of the miRNA mature sequence, and 1,179 within mature miRNA sequences (Fig. 3; Supplemental Fig. S3; Supplemental Data Set 1).

When looking at editing sites falling within the sequence of mature miRNAs, MiREDiBase highlights two distinct patterns for A-to-I and C-to-U editing in humans (Fig. 3 A, B). Concerning the A-to-I type, the majority of edited sites (325 out of 1018, 31.9%) are located at position 2-5 of the seed region. Other hotspots for A-to-I editing seem to be represented by positions 1, 6-9, and 12, which include 325 more edited sites. Differently, in the case of C-to-U miRNA editing, most of the modification sites are located outside of the seed region. In particular, 48 out of 104 edited sites (46.2%) are located at positions 10-12 and 15, whereas only 17 edited sites (16.3%) fall within the seed region. A very similar pattern can be observed for macaque (Supplemental Fig. S3).

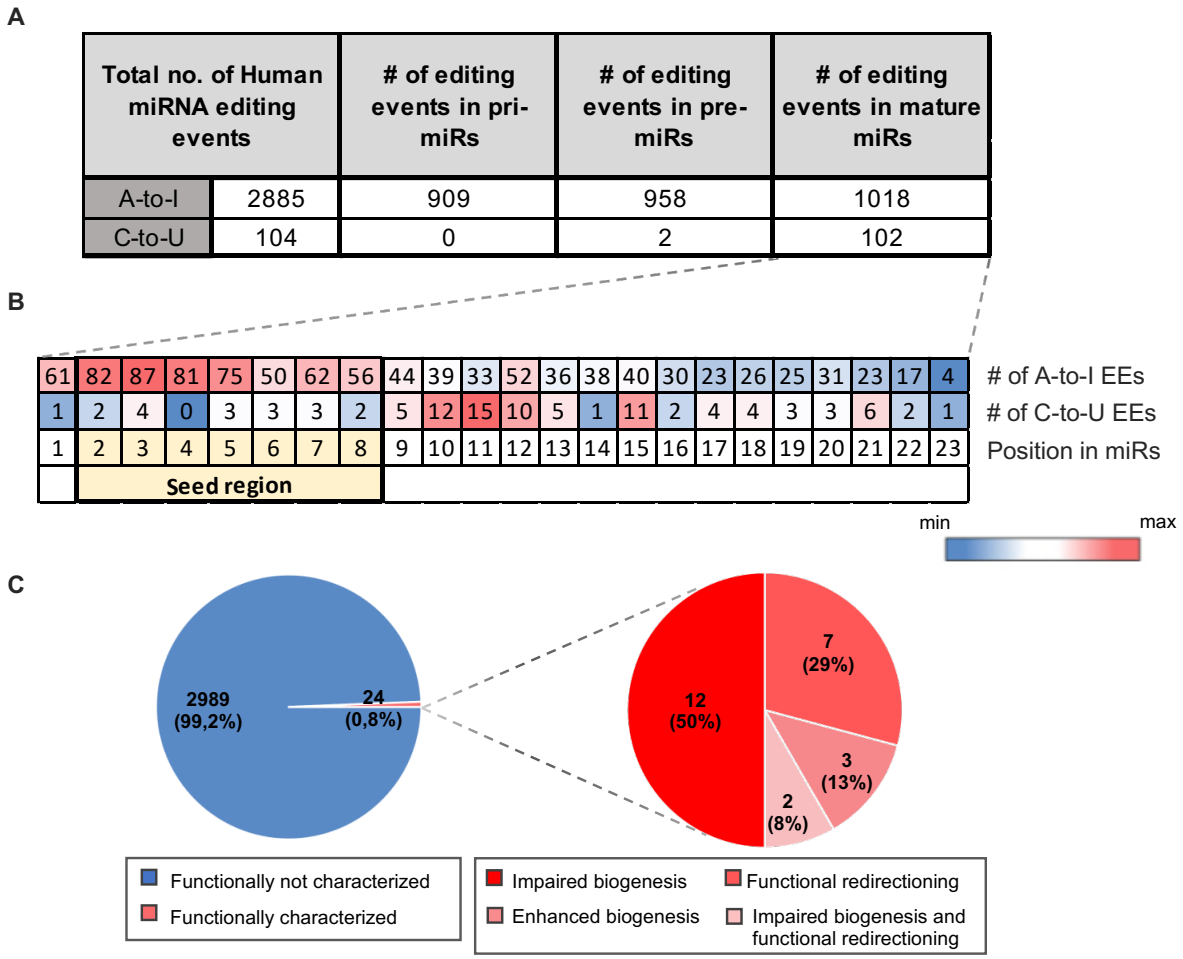

*Figure 3. (**A**) Distribution of A-to-I and C-to-U editing events across the three different regions of primary human miRNA transcripts. (**B**) Heatmap of the distribution of A-to-I and C-to-U falling into human mature miRNAs across nucleotide positions. In each position are presents the number of editing events identified. The canonical seed position is highlighted in yellow. Editing sites falling in positions 24-27 were not shown in the mature sequence. EE=Editing Events. (**C**) Pie charts showing the fraction of functionally characterized editing events in human miRNAs.*

However, for the vast majority of miRNA editing events reported in MiREDiBase, functional consequences have currently remained undetermined. So far, only 24 editing sites (0.8%) were functionally characterized by appropriate techniques (Fig. 3C). Among these, twelve were demonstrated to impair miRNA biogenesis, seven cause functional re-targeting, three simultaneously impair biogenesis and cause functional re-targeting and two lead to an enhancement of biogenesis.

MiREDiBase helps users interpreting and contextualizing data by suppling with *in silico* predictions miRNA editing events occurring within pre-miRNA or mature miRNA sequences. In particular, we computed 2,150 MFE pre-miRNA predictive structures using editing sites internal to pre-miRNA sequences, and 1,018 miRNA-

targeting predictions and enrichment analyses. In both cases, users will have the opportunity to compare the edited miRNAs with their relative wild-type versions.

Biological sources in MiREDiBase can be grouped into three main categories: “normal condition” (both human and primates), “adverse condition” (human only), and “cell line” (human only) (Table 1).

Table 1. Total number and percentage of normal conditions, adverse conditions, and cell lines currently included in MiREDiBase for the four species Homo sapiens, Macaca mulatta, Pan troglodytes, and Gorilla gorilla.

	Species			
Condition	**Human**	**Macaque**	**Chimpanzee**	**Gorilla**
Normal condition	**69**	**23**	**1**	**1**
adult tissues	64 (92,8%)	21 (91.3%)	1 (100%)	1 (100%)
pre-natal tissues	5 (7,2%)	2 (8.7%)	NA	NA
Adverse condition	**68**	**0**	**0**	**0**
oncological diseases	60 (88,2%)	NA	NA	NA
sample subtypes	62	NA	NA	NA
neurological diseases	4 (5,9%)	NA	NA	NA
sample subtypes	6	NA	NA	NA
inflammatory diseases	2 (2,9%)	NA	NA	NA
sample subtypes	3	NA	NA	NA
cardiological diseases	1 (1,5%)	NA	NA	NA
sample subtypes	1	NA	NA	NA
genetic diseases	1 (1,5%)	NA	NA	NA
sample subtypes	1	NA	NA	NA
Cell lines	**88**	**0**	**0**	**0**
non-malignant (commercial)	7 (8,0%)	NA	NA	NA
non-malignant (ex vivo)	9 (10,2%)	NA	NA	NA
malignant (commercial)	71 (80,7%)	NA	NA	NA
malignant (ex vivo)	1 (1,1%)	NA	NA	NA

MiREDiBase allows the user to compare editing levels for an individual miRNA in physiological and pathological conditions, whether simultaneously data are available. The “normal condition” group accounts for 92 different healthy tissues/organs analyzed for miRNA editing. Among them, 85 were obtained from adult individuals and seven from pre-natal developmental stages (Supplemental Table S2, S5). The “adverse condition” group is broadly represented by tumors, with 60 different oncological conditions and 62 different sample subtypes. Users can also find neurological disorders, including four pathological conditions and six sample subtypes. The Inflammatory condition, cardiovascular disease, and genetic disorder are currently the less representative

classes, with two pathological conditions and three sample subtypes for the former and one pathological condition and sample type for the latter two, respectively (Supplemental Table S3). The cell line group accounts for 78 commercially available cell lines and 10 primary human cell samples cultured *in vitro.* Of the 78 commercial cell lines, 71 are malignant, while the remaining represent non-malignant conditions. Among the 10 primary human cell samples, only 1 refers to a malignant condition, while 9 are representative of normal conditions (Supplemental Table S4).

User Interface and Data Accessibility

MiREDiBase provides an intuitive and straightforward web interface to access data, requiring no bioinformatics skills to perform accurate searches across the database. Users can explore MiREDiBase by interacting with the Search (Fig. 4) or the Compare module. Each module starts with a modal box by which users can filter out miRNA editing sites.

The Search module provides four filtering fields, including organism (e.g., Human), modification type (e.g., the A-to-I editing), genomic region (e.g., chromosome, pre-miRNA, or miRNA) and, optionally, biological source (e.g., BRCA – breast carcinoma). Based on the selected filtering options, the Search module generates a table listing a set of editing sites supplied with essential information, i.e., organism, modification type, chromosome, strand, genomic position, and position in pre-miRNA and mature miRNA.

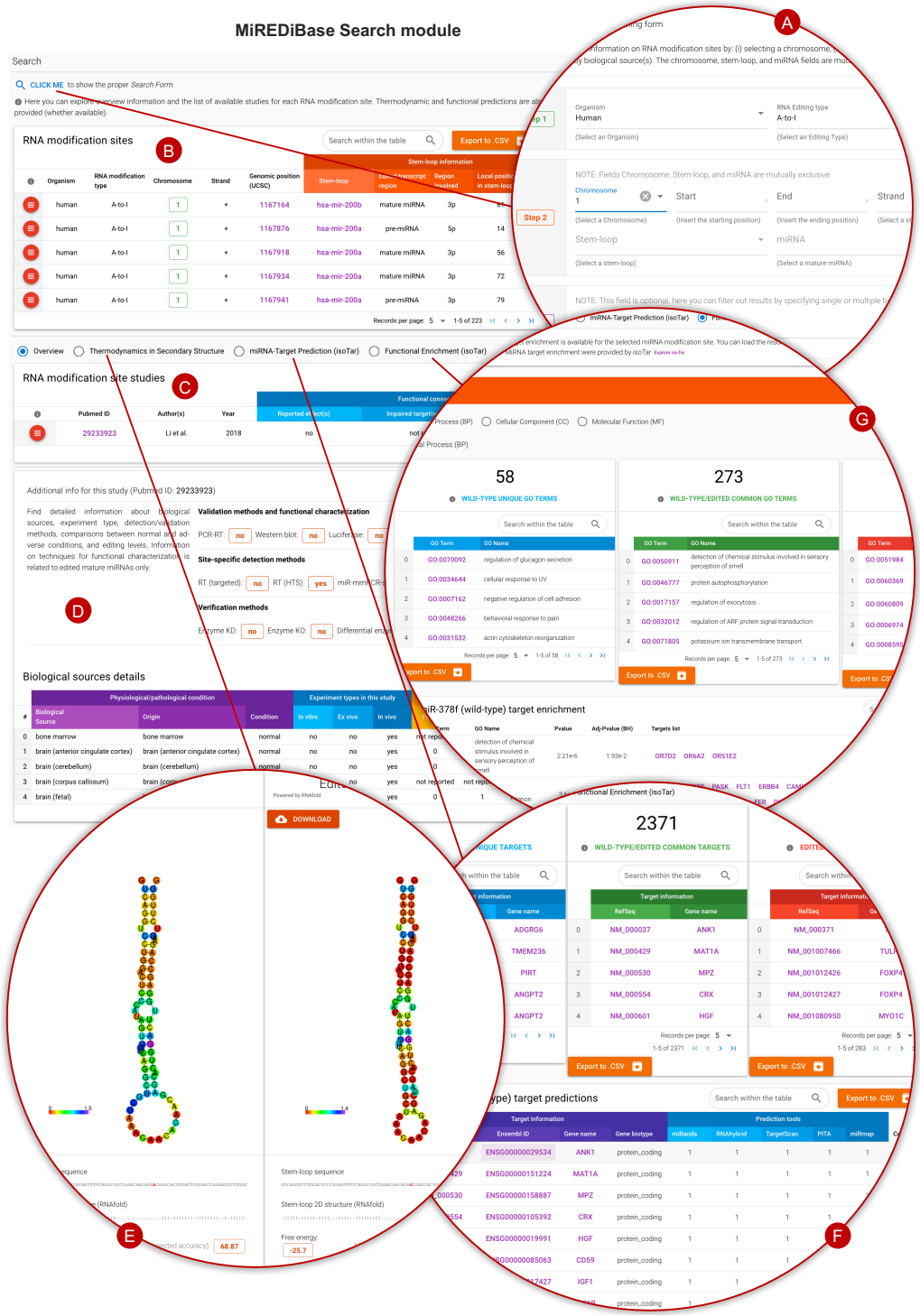

*Figure 4. The MiREDiBase Search module. Users can filter out MiREDiBase data by exploiting the specific modal box (**A**). Then, they can dig into the data by interacting with the filtered editing sites (**B**). The editing site's details (**C**) can be navigated by clicking on the red button placed on its left side. Additional resources include the list of biological sources in which the editing site has been identified (**D**), the thermodynamic comparison of the wild-type and edited pre-miRNA 2D structures (**E**), the miRNA-target predictions (**F**), and functional enrichment (**G**) data. Helpers and downloading buttons are provided throughout the module interface.*

The table also contains links to external resources. By clicking on the dedicated, left-sided buttons, users can dig into each editing site. Additional information covers publications' details, biological sources, 2D structures

of wild-type and edited pre-miRNAs, miRNA-target predictions, and associated functional enrichment data (e.g., GO terms, KEGG pathways) (Fig. 4), which enable ready access to a putative biological interpretation. The results in each module can be easily downloaded through specific buttons.

The Compare module aims to explore differentially edited sites in adverse versus normal conditions. It provides a set of essential information supplied with the editing level for each examined condition. Similarly to the Search module, the Compare module allows users to filter out RNA editing sites by specifying the organism, modification type, disease, and pre-miRNA.

We connected all miRNAs reported in MiREDiBase to their specific miRBase web page. At the same time, all genomic coordinates of edited adenosines were mapped onto the UCSC hg38/GRCh38 genome assembly by external links. Furthermore, editing sites shared with DARNED, RADAR, and REDIportal were annotated with external links, aiming to improve research on miRNA editing.

To encourage users to familiarize themselves with our tool, MiREDiBase offers helpers reporting explanations on how to interpret results, along with complete documentation on how to use each module and statistics. Advanced users can instead exploit the RESTful API (https://ncrnaome.osumc.edu/miredibaseapi/docs), which provides a standalone web interface to explore available methods for extracting data, with the opportunity to embed RESTful API HTTP calls within users' code (Fig. 5).

The MiREDiBase platform adopts a multi-containerized microservice architecture (Fig. 5), which provides user-friendly and efficient ways to access all manually collected data (see Methods section for more details).

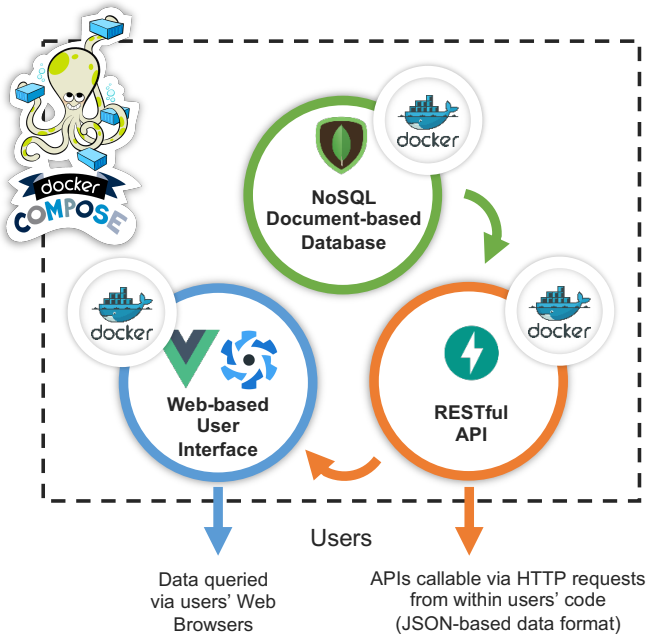

Figure 5. Overview of the MiREDiBase multi-containerized microservice architecture. The MiREDiBase platform provides users different ways to access its data: through a web browser (Web-based User Interface) and RESTful API (HTTP calls).

Discussion

At the beginning of the miRNA editing field, Sanger sequencing represented the standard gold method allowing the accurate sequencing of RNA molecules and reliable identification of editing events, e.g., (Luciano 2004; Yang et al. 2006). Nonetheless, such a low-throughput technique is time-consuming and only allows the sequencing of a restricted set of miRNAs (Kawahara et al. 2008). In later years, the employment of powered high-throughput RNA sequencing (RNA-Seq) technologies and the design of different bioinformatic pipelines have dramatically improved the computational identification of RNA editing events (Diroma et al. 2019), including those occurring in miRNAs.

Given the ever-increasing number of editing sites detected on a genome-wide scale, the need to create a comprehensive catalog of such modification events has become urgent. In this context, Kiran and Baranov published DARNED, the first online repository providing centralized access to published data on RNA editing (Kiran and Baranov 2010). DARNED currently includes ~350,000 predicted RNA A-to-I editing sites from humans, mice (*Mus musculus*), and flies (*Drosophila melanogaster*), together with few C-to-U instances. However, only a small portion of these modification events have been manually annotated, and no information

is provided about editing levels. DARNED last update dates back to 2012 (Kiran et al. 2012). In 2013, Ramaswami and Li presented RADAR, a rigorously annotated database of A-to-I RNA editing containing manually curated annotations for each editing site (Ramaswami and Li 2014).

Similarly to DARNED, RADAR includes data from humans, mice, and flies, and currently accounts for ~1.4 million editing sites, also providing: tissue-specific editing level, conservation in other model organisms, and genomic and transcriptional context. RADAR does not include C-to-U editing data and was last updated in 2014. In 2017, Picardi and colleagues developed REDIportal, which today is the most extensive collection of RNA editing in humans, including more than 4.5 million of A-to-I modification events detected across 55 body sites from thousands of RNA-seq experiments (Picardi et al. 2017). Moreover, with its last update, REDIportal also includes ~90,000 putative A-to-I editing events from the mouse brain transcriptome and incorporates CLAIRE, a searchable catalog of RNA editing levels across cell lines (Schaffer et al. 2020). Although these three mentioned online resources are undoubtedly the most authoritative repositories of RNA editing events, none of them is strictly dedicated to miRNA editing. Instead, the vast majority of the editing events reported in these databases are referred to long RNAs, namely mRNA and long non-coding RNAs (lncRNAs), with only a minority falling into miRNA transcripts.

A few online resources have been later developed that partially focus on the effects of RNA editing on miRNA functionality. For instance, the Editome-Disease Knowledgebase (EDK) (Niu et al. 2019) is a manually curated database that aims to link experimentally validated RNA editing events in non-coding RNAs to various diseases. However, this database currently contains only 16 validated A-to-I instances in miRNAs, and does not provide any information about publications, position of editing sites, or detection/validation methods. The Cancer Editome Atlas (TCEA) is a powerful, user-friendly bioinformatics resource that characterizes more than 192 million editing events at ~4.6 million editing sites from approximately 11,000 samples across 33 cancer types recovered from The Cancer Genome Atlas (Lin and Chen 2019). However, to the same extent as the previously mentioned resources, TCEA is mainly focused on editing events occurring in mRNAs. From the miRNA standpoint, TCEA only allows users to predict the effects of A-to-I editing events occurring in the 3'UTR of mRNAs in terms of miRNA-mRNA interactions.

To fill the gap between the fields of RNA editing and miRNA biology, we have developed MiREDiBase, the first-of-its-kind database specifically dedicated to the field of miRNA modification. In the current version, MiREDiBase includes more than three thousand A-to-I and C-to-U RNA editing events of miRNA editing events manually collected from the literature, occurring in humans and primates. The choice to include primates rather than other model species in this first release was motivated by the fact that primates present the highest genomic and transcriptomic similarity in relation to humans (Rogers and Gibbs 2014). Moreover, primates are recognized as excellent candidates to facilitate investigation on epigenetic control of genome functions and are highly relevant for biomedical studies (Rogers and Gibbs 2014).

MiREDiBase allows users to consult the RNA secondary structure of the wild-type and edited pre-miRNA, as well as to infer the possible function of edited mature miRNA, based on the predicted targetome and subsequent functional enrichment analyses. We implemented a user-friendly interface that allows users to keep track of each search step to improve the user experience. Moreover, MiREDiBase includes a “Compare” section, which enables the comparison of adverse versus normal conditions in a study-specific manner. Finally, the MiREDiBase platform relies on cutting-edge technologies, aiming at providing reliability and continuous operability. The platform represents an orchestration of different containerized services built on top of Docker. Each service fulfills a specific purpose, such as a Web Application Service (Quasar - a Progressive JavaScript Framework), a RESTful API Service (FastAPI - a modern, high-performance, web framework for building secure APIs), and a Database Service (MongoDB - a NoSQL document-based database). The platform is designed to provide the smoothest and user-friendly experience to users.

Besides keeping the database up to date with novel data from the literature, the main future goal is to include A-to-I and C-to-U editing events from other species, primarily those obtained from tissues of model organisms like *Mus musculus* and *Drosophila melanogaster*. Another aim is to add other modification types that may help biological study and interpretation of functional roles of mature and non-mature modified miRNAs in a specific cell context. For example, after analyzing human brain samples for RNA editing events, Paul et al. unexpectedly found that a consistent percentage of editing events are indeed of the non-canonical types, especially C-to-A and G-to-U (Paul et al. 2020). This raised the question of whether these editing events exert any essential function in neurons, and also if there exist specific enzymes capable of catalyzing such

modifications. Similarly, miRNA methylation has recently caught the attention of the scientific community, as it was demonstrated to affect miRNA biogenesis (Alarcón et al. 2015). However, the study of this phenomenon and its potential functional implications have remained widely unexplored. With continuous updating in the coming years, we believe that MiREDiBase will gradually grow as a precious resource for researchers in the field of epitranscriptomics, leading to a better understanding of miRNA modification phenomena and their functional consequences.

Methods

Data Processing

Each editing event was supplied with essential information recovered from miRbase (v22), including the relative position within pre-miRNA and mature miRNA, genomic position, and pre-miRNA region (5'- or 3'-arm, or loop region). For editing events occurring outside the pre-miRNA sequence, we adopted the notation "pri-miRNA." Editing events were then enriched with metadata manually collected from selected publications. Overall, we extracted eight different information types: detection/validation method, experiment type, biological source, correspondent condition (adverse or normal), comparison (pathological versus physiological condition), editing level, enzyme affinity, and functional characterization.

The "detection method" does not specify the method adopted by authors to identify miRNA editing events. Instead, it indicates which kind of methodological approach (targeted, wide-transcriptome, or both) the authors selected for editing detection. Only in a few cases, the method has been specified to highlight particularly sensitive or innovative approaches, e.g., miR-mmPCR-seq (Zhang et al. 2014).

The "validation method" refers to methods confirming sequencing data, especially those obtained by wide-transcription approaches, including enzyme knock-down (only ADAR in the current version), knock-out, differential expression, and modification-specific enzymatic cleavage.

The "experiment type" specifies whether, in a given study, individual editing events were identified *in vitro*, *in vivo*, or *ex vivo*. Editing events obtained by analysis of small RNA-seq data from The Cancer Genome Atlas

(TCGA) (Tomczak et al. 2015) or Genotype-Tissue Expression (GTEx) atlas (The GTEx Consortium et al. 2015) were considered as detected *in vivo*. Editing events obtained by analysis of sequence libraries from the Sequence Read Archive (SRA) database (Kodama et al. 2012) were considered as detected *in vitro*, *in vivo*, or *ex vivo* depending on the library derivation.

The "pathological condition" specifies whether a miRNA editing event was detected in one or multiple diseases. For a given study, physiological and pathological conditions were compared whether editing levels for an individual miRNA were simultaneously available for both conditions.

In studies with multiple editing level values per miRNA editing site, we considered only the minimum and maximum values, rounding them up by multiples of five (e.g., editing levels of 21.1% and 44% were rounded up to 20% and 45%, respectively). Whether a single value was reported for an individual miRNA, this was rounded, creating an interval of 5% (e.g., if a study reported the editing level as 13% for a specific editing site, then the editing level was presented as "from 10% to 15%").

Information concerning enzyme affinity (only ADAR in the current version) was retrieved whether authors carried out enzyme-transfection experiments causing enzyme overexpression. Finally, we annotated all the functionally characterized editing events with information regarding their specific biological function. In the event of functional re-targeting, validation methods were reported along with the set of validated lost and gained targets.

Secondary Structure Prediction Analysis

We generated the minimum free energy (MFE) structures for all those pre-miRNAs subjected to editing and their wild-type (WT) counterparts. The double-stranded RNA structures were created by employing the *RNAfold* tool from the ViennaRNA package (Lorenz et al. 2011) with default settings. Finally, we considered all editing sites occurring within the mature miRNA region to infer possible miRNA target re-direction as well as diversified biological functions.

MiRNA-Target Prediction and Functional Enrichment Analyses

The miRNA-target prediction analysis, for both edited and WT miRNA, was achieved by using our web-based containerized application *isoTar* (Distefano et al. 2019), designed to simplify and perform miRNA consensus target prediction and functional enrichment analyses. For miRNA target predictions, we established a minimum consensus of 3. An adjusted P-value <0.05 was considered as a threshold for the functional enrichment analysis.

Platform Design and Implementation

To achieve reliability and continuous delivery (short-cycle updates), we developed each lightweight, standalone, microservice on top of Docker (v19.03.12) (https://docs.docker.com). The platform itself consists of three microservices, which are orchestrated by Docker Compose (v1.26.0) (https://docs.docker.com/compose), a tool for managing multi-container-based applications. Each microservice provides a specific functionality: a web-based user interface (UI), a RESTful API for data retrieving, and a NoSQL document-based database for data storing. To offer users an engaging and responsive experience, we developed a high-performance platform which relies on cutting-edge open-source technologies, such as Quasar (v1.12.8) (https://quasar.dev) for the UI, FastAPI (v0.55.1) (a Python (v3.8.1) framework, https://fastapi.tiangolo.com) for the RESTful API, and MongoDB (v4.2.8) (https://www.mongodb.com) for the data storage. We have tested MiREDiBase on the following browser: Firefox (80+), Google Chrome (85+), Edge (85+), Safari (13+), and Opera (70+). MiREDiBase is freely accessible to the scientific community through the link: https://ncrnaome.osumc.edu/miredibase, without requiring registration or login.

DATA ACCESS

MiREDiBase is freely available at https://ncrnaome.osumc.edu/miredibase. The whole platform is openly available at https://github.com/ncRNAome-OSU/miredibase.

ACKNOWLEDGMENTS

We thank the Cancer IT Operation Group of The Ohio State University (OSU) and Thomas Moore for his valuable technical assistance. At the same time, we want to special thank Paolo Fadda from The Genomics

Shared Resource (GSR) of OSU, together with Jing Jiang and Cankun Wang from Dr. Ma's Laboratory (Department of Biomedical Informatics at OSU) for their helpful suggestions during the beta testing phase of MiREDiBase.

This work was supported by National Cancer Institute (National Institute of Helth) grant R35CA197706 to C.M.C. F.R. is supported by the BRIDGE - Translational Excellence Programme (bridge.ku.dk), Faculty of Health and Medical Sciences, University of Copenhagen, funded by the Novo Nordisk Foundation with grant agreement no. NNF18SA0034956 and grant agreement no. NNF14CC0001.

AUTHOR CONTRIBUTORS

C.M.C. and G.N. conceived, designed, and supervised the project; G.P.M., R.D., L.T., and G.N. designed the database; G.P.M. collects the data; R.D. designed and developed the platform; L.T., A.L., F.R., F.C., G.R., M.B., P.G., A.F., M.A., and Q.M. contributed to the beta-testing phase of the database. G.P.M., R.D., and G.N. wrote the first draft of the manuscript. L.T., A.L., F.R., F.C., G.R., M.B., P.G., A.F., M.A., and Q.M. C.M.C. participated in manuscript revision. All authors read and approved the final manuscript.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

REFERENCES

Alarcón CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. 2015. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**: 482–485.

Bartel DP. 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**: 215–233.

Blow M, Grocock R, van Dongen S, Enright A, Dicks E, Futreal PA, Wooster R, Stratton M. 2006. RNA editing of human microRNAs. *Genome Biol* **7**: R27.

Brümmer A, Yang Y, Chan TW, Xiao X. 2017. Structure-mediated modulation of mRNA abundance by A-to-I editing. *Nat Commun* **8**: 1255.

Croce CM. 2009. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* **10**: 704–714.

Daniel C, Widmark A, Rigardt D, Öhman M. 2017. Editing inducer elements increases A-to-I editing efficiency in the mammalian transcriptome. *Genome Biol* **18**: 195.

Diroma MA, Ciaccia L, Pesole G, Picardi E. 2019. Elucidating the editome: bioinformatics approaches for RNA editing detection. *Briefings in Bioinformatics* **20**: 436–447.

Distefano R, Nigita G, Veneziano D, Romano G, Croce CM, Acunzo M. 2019. isoTar: Consensus Target Prediction with Enrichment Analysis for MicroRNAs Harboring Editing Sites and Other Variations. In *MicroRNA Target Identification* (ed. A. Laganà), Vol. 1970 of *Methods in Molecular Biology*, pp. 211–235, Springer New York, New York, NY http://link.springer.com/10.1007/978-1-4939-9207-2_12.

Gott JM, Emeson RB. 2000. Functions and Mechanisms of RNA Editing. *Annu Rev Genet* **34**: 499–531.

Griffiths-Jones S. 2006. miRBase: The MicroRNA Sequence Database. In *MicroRNA Protocols*, Vol. 342 of, pp. 129–138, Humana Press, New Jersey http://link.springer.com/10.1385/1-59745-123-1:129.

GTEx Consortium, Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, et al. 2017. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**: 249–254.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**: 509–524.

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research* **47**: D853–D858.

Han L, Diao L, Yu S, Xu X, Li J, Zhang R, Yang Y, Werner HMJ, Eterovic AK, Yuan Y, et al. 2015. The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **28**: 515–528.

Jonas S, Izaurralde E. 2015. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**: 421–433.

Joyce CE, Zhou X, Xia J, Ryan C, Thrash B, Menter A, Zhang W, Bowcock AM. 2011. Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. *Human Molecular Genetics* **20**: 4025–4040.

Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, Nishikura K. 2008. Frequency and fate of microRNA editing in human brain. *Nucleic Acids Research* **36**: 5270–5280.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. *Science* **315**: 1137–1140.

Kiran A, Baranov PV. 2010. DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* **26**: 1772–1776.

Kiran AM, O'Mahony JJ, Sanjeev K, Baranov PV. 2012. Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Research* **41**: D258–D261.

Kodama Y, Shumway M, Leinonen R, on behalf of the International Nucleotide Sequence Database Collaboration. 2012. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Research* **40**: D54–D56.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**: D155–D162.

Li L, Song Y, Shi X, Liu J, Xiong S, Chen W, Fu Q, Huang Z, Gu N, Zhang R. 2018. The landscape of miRNA editing in animals and its impact on miRNA biogenesis and targeting. *Genome Res* **28**: 132–143.

Liang L, He X. 2011. Macro-management of microRNAs in cell cycle progression of tumor cells and its implications in anti-cancer therapy. *Acta Pharmacol Sin* **32**: 1311–1320.

Lin C-H, Chen SC-C. 2019. The Cancer Editome Atlas: A Resource for Exploratory Analysis of the Adenosine-to-Inosine RNA Editome in Cancer. *Cancer Res* **79**: 3001–3006.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.

Luciano DJ. 2004. RNA editing of a miRNA precursor. *RNA* **10**: 1174–1177.

Negi V, Paul D, Das S, Bajpai P, Singh S, Mukhopadhyay A, Agrawal A, Ghosh B. 2015. Altered expression and editing of miRNA-100 regulates iTreg differentiation. *Nucleic Acids Res* **43**: 8057–8065.

Nigita G, Acunzo M, Romano G, Veneziano D, Laganà A, Vitiello M, Wernicke D, Ferro A, Croce CM. 2016. microRNA editing in seed region aligns with cellular changes in hypoxic conditions. *Nucleic Acids Res* **44**: 6298–6308.

Nigita G, Veneziano D, Ferro A. 2015. A-to-I RNA Editing: Current Knowledge Sources and Computational Approaches with Special Emphasis on Non-Coding RNA Molecules. *Front Bioeng Biotechnol* **3**. http://journal.frontiersin.org/Article/10.3389/fbioe.2015.00037/abstract.

Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**: 83–96.

Niu G, Zou D, Li M, Zhang Y, Sang J, Xia L, Li M, Liu L, Cao J, Zhang Y, et al. 2019. Editome Disease Knowledgebase (EDK): a curated knowledgebase of editome-disease associations in human. *Nucleic Acids Research* **47**: D78–D83.

Paul D, Ansari AH, Lal M, Mukhopadhyay A. 2020. Human Brain Shows Recurrent Non-Canonical MicroRNA Editing Events Enriched for Seed Sequence with Possible Functional Consequence. *ncRNA* **6**: 21.

Picardi E, D'Erchia AM, Lo Giudice C, Pesole G. 2017. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* **45**: D750–D757.

Picardi E, Manzari C, Mastropasqua F, Aiello I, D'Erchia AM, Pesole G. 2015. Profiling RNA editing in human tissues: towards the inosinome Atlas. *Sci Rep* **5**: 14941.

Ramaswami G, Li JB. 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucl Acids Res* **42**: D109–D113.

Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet* **15**: 347–359.

Salinas-Vera Y, Marchat L, Gallardo-Rincón D, Ruiz-García E, Astudillo-De La Vega H, Echavarría-Zepeda R, López-Camarillo C. 2018. AngiomiRs: MicroRNAs driving angiogenesis in cancer (Review). *Int J Mol Med*. http://www.spandidos-publications.com/10.3892/ijmm.2018.4003.

Schaffer AA, Kopel E, Hendel A, Picardi E, Levanon EY, Eisenberg E. 2020. The cell line A-to-I RNA editing catalogue. *Nucleic Acids Research* **48**: 5849–5858.

The GTEx Consortium, Ardlie KG, Deluca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, et al. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**: 648–660.

Tomczak K, Czerwińska P, Wiznerowicz M. 2015. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *wo* **1A**: 68–77.

Ultimo S, Zauli G, Martelli AM, Vitale M, McCubrey JA, Capitani S, Neri LM. 2018. Cardiovascular disease-related miRNAs expression: potential role as biomarkers and effects of training exercise. *Oncotarget* **9**: 17238–17254.

Wedekind JE, Dance GSC, Sowden MarkP, Smith HC. 2003. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends in Genetics* **19**: 207–216.

Wingo TS, Yang J, Fan W, Min Canon S, Gerasimov ES, Lori A, Logsdon B, Yao B, Seyfried NT, Lah JJ, et al. 2020. Brain microRNAs associated with late-life depressive symptoms are also associated with cognitive trajectory and dementia. *npj Genom Med* **5**: 6.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* **13**: 13–21.

Zhang R, Li X, Ramaswami G, Smith KS, Turecki G, Montgomery SB, Li JB. 2014. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nat Methods* **11**: 51–54.

Zheng Y, Ji B, Song R, Wang S, Li T, Zhang X, Chen K, Li T, Li J. 2016. Accurate detection for a wide range of mutation and editing sites of microRNAs from small RNA high-throughput sequencing profiles. *Nucleic Acids Res* **44**: e123–e123.

Ziats MN, Rennert OM. 2014. Identification of differentially expressed microRNAs across the developing human brain. *Mol Psychiatry* **19**: 848–852.