# A concurrent canonical and modified miRNAome pan-cancer study on TCGA and TARGET cohorts leads to an enhanced resolution in cancer

Rosario Distefano[1,*], Luisa Tomasello[1], Gian Luca Rampioni Vinciguerra[1,2], Pierluigi Gasparini[1,3,4], Yujia Xiang[5], Marina Bagnoli[6], Gioacchino Paolo Marceca[7], Paolo Fadda[8], Alessandro Laganà[9], Mario Acunzo[10], Qin Ma[5], Giovanni Nigita[1,*,†], and Carlo M. Croce[1,*,†]

[1] Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210, USA.
[2] Faculty of Medicine and Psychology, Department of Clinical and Molecular Medicine, University of Rome "Sapienza," Santo Andrea Hospital, Rome, RM, 00189, Italy.
[3] School of Biomedical Sciences and Pharmacy, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, NSW 2308, Australia.
[4] Hunter Medical Research Institute, New Lambton Heights, NSW 2305, Australia.
[5] Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, 43210, USA.
[6] Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, MI, 20133, Italy.
[7] Department of Clinical and Experimental Medicine, University of Catania, Catania, CT, 95100, Italy.
[8] Genomics Shared Resource, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210, USA.
[9] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.
[10] Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, 23298, USA.

* To whom correspondence should be addressed. Tel: +1 614 292 4930; Fax: +1 614 292 3558; Email: Carlo.Croce@osumc.edu
Correspondence may also be addressed to Rosario Distefano. Tel.: +1 614 292 7278; E-mail: Rosario.Distefano@osumc.edu
Correspondence may also be addressed to Giovanni Nigita. Tel.: +1 614 688 1805; E-mail: Giovanni.Nigita @osumc.edu
† These authors contributed equally to this work.

**ABSTRACT**

MiRNA Epitranscriptomics has placed a new layer of complexity in the cancer field. Despite miRNA editing and shifted miRNA isoforms are gaining attention due to recent improvements in next-generation sequencing, a simultaneous study of both modifications in cancer is still missing. Here, we concurrently profiled multiple miRNA modifications, such as A-to-I RNA editing and shifted miRNA isoforms, in >13K adult and pediatric tumor samples across 38 distinct cancer cohorts from The Cancer Genome Atlas and The Therapeutically Applicable Research to Generate Effective Treatments datasets. We investigated the differences among canonical miRNAs and a wider comprehensive miRNAome from the expression, clustering, dysregulation, and prognostic perspective. Interestingly, the wider miRNAome boosted clustering results, uniquely outlining cohorts' clinical-pathological features. The abundance of expressed miRNA isoforms directly related to the activation/deactivation of critical carcinogenesis pathways. We found dysregulated modified miRNAs characterized by an opposite expression trend than their canonical counterparts in cancer, potentially impacting their targetome and function. Our study emphasizes the importance of modified miRNAs as potential cancer biomarkers and gene expression regulators, outlining once more the importance of going beyond the well-established paradigm of one-mature-miRNA per miRNA arm.

## INTRODUCTION

MicroRNAs (miRNAs) are a class of small non-coding RNA (ncRNA) molecules of ~21 nucleotides (nts) in length, expressed in eukaryotes (1, 2), which negatively regulate gene expression at the post-transcriptional level (2). MiRNAs are involved in many developmental and cellular processes (3). Their dysregulation observed in several diseases (4), including cancer (5, 6), prompted the interest in employing such molecules as diagnostic and prognostic cancer biomarkers (7). Until recently, most studies on miRNAs widely relied upon the miRNA biogenesis paradigm "one mature miRNA per miRNA precursor arm" (1, 8). However, the latest advancements in next-generation sequencing (NGS) technologies unveiled a more complex scenario (9), in which some expressed miRNA molecules differ from their reference sequence (miRBase) (10). These miRNA variants, termed *miRNA isoforms* or *isomiRs*, undergo several RNA modification events, such as RNA Editing (11, 12), miRNA sequence alternative cleavage (13–17), and alterations at the DNA level (i.e., Single Nucleotide Polymorphisms – SNPs) (18, 19). The growing interest in such phenomena started to undermine the before mentioned well-established paradigm (10).

The Adenosine-to-Inosine (A-to-I) RNA editing (11, 12) represents the most abundant RNA editing variant in mammals, affecting both splicing and translational machinery. Adenosine deaminases acting on RNA enzymes, ADAR1 and ADAR2, catalyze the A-to-I RNA Editing in the nucleus, binding double-stranded (ds) RNA regions in both coding and non-coding RNAs (12), including miRNAs (20, 21). The RNA editing phenomenon has been associated with several human diseases (22, 23), including cancer (24, 25). Other groups and we assessed how a single RNA editing event occurring within the miRNA seed region (MSR) could compromise the miRNA-mediated gene regulation process, which in turn may drastically alter the miRNA targetome (21, 26, 27). Likewise, *shifted isomiRs*, which may result from imprecise cleavage of the miRNA reference sequence, could induce targetome differentiation (28–31). Shifted isomiRs may likely result from the imprecise cleavage processing by Drosha (13, 14, 32–35), in stark contrast with Dicer, which cleavages at a fixed distance (36). Initially considered as artifacts (37, 38), their function has been recently re-evaluated (14, 16, 17, 35) as they actively interact with mRNAs (13, 15, 16, 33, 39, 40). Like the A-to-I RNA editing occurring within the MSR, the shifting of the 5'-end (addition/trimming of nts at 5'-end) may lead to a diversification of the targetome revealing a more complex role in gene regulation than previously expected (41). The expression profile of shifted isomiRs exhibits high variability that is tissue- and cell-dependent (37). In some cases, the shifted isomiRs expression results higher than their canonical counterparts (42). Shifted isomiRs are classified as "*templated*" and "*non-templated*" forms. The templated form entirely matches its original pre-miRNA sequence (14, 32). In contrast, the non-templated one does not feature a perfect match, possibly due to the addition of nucleotides at the mature level via enzymatic activity, typically at 3'-end (35).

Even though the rising interest in the miRNA Epitranscriptome, most studies have pursued the investigation of a subset of miRNA modifications. Recent efforts on assessing the biological implications of A-to-I RNA Editing affecting canonical miRNAs (43–46), including our work (47), have inflated the interest in employing such molecules as potential biomarkers for cancer prognosis and therapy (44, 46). The existence of studies on the oncogenic role of the edited hsa-miR-200b-3p in

multiple cancers (46), the tumor suppressor role of the edited hsa-miR-589-3p in glioblastoma (43), the supporting role of the edited hsa-miR-455-5p in combination with its canonical form in melanoma (45), together with our work on assaying the consequences of miRNA editing in hypoxia (27), further consolidated the attention on such a phenomenon. Concurrently, we observed a surge of shifted isomiRs-oriented studies, even though mainly focused on detecting templated forms, with some exceptions (no insertion or deletion, up to one mismatch). A study reported the profiling of lymphoblastoid cell lines (LCLs) from 452 patients for the identification of population- and gender-dependent shifted isomiRs (48). In this work, authors considered only shifted isomiRs wholly contained within the mature miRNA sequence extended by six nucleotides at both ends, contributing to the 95th-percentile of reads mapped to each miRNA arm. A similar workflow (49) was applied to profile 316 normal and tumor samples of Breast Cancer Invasive Carcinoma (BRCA) from The Cancer Genome Atlas (TCGA) (50, 51), in which the identified shifted isomiRs easily distinguished between normal and tumor samples, as well as among tumor subtypes. Finally, in recent years, the first pan-cancer study focused on shifted isomiRs saw the light, offering the profiling of 32 tumor cohorts in TCGA (16). In this study, the authors built a "binarized" classifier that labeled shifted isomiRs as "present" or "absent," obtaining overall a better tumor classification in terms of sensitivity. Although the studies mentioned above have proven the importance of investigating such miRNA modifications, there is still a lack of a broader miRNAome analysis in cancer.

In this work, we relied on miRge 2.0 (52), one of the major pipelines for canonical miRNAs/miRNA isoforms profiling (53–56), given its reliability in identifying A-to-I RNA editing sites and accuracy comparable to a well-established miRNA editing detection approach (57, 58). Through miRge 2.0, we concurrently detected canonical miRNAs, templated/non-templated shifted isomiRs, modification events such as A-to-I RNA Editing and SNPs, as well as nucleotide insertions. We processed information at a large scale from the most prominent and reliable public resources, TCGA and The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (59), analyzing >13K adult and pediatric cancer samples spread across 38 distinct cohorts. In our data, the abundance of expressed annotated miRNA isoforms outmatched by *8*-fold the number of expressed canonical miRNAs (miRBase v22). Among the expressed miRNA modifications, we observed a predominance of 3'-end shifted molecules. Both miRNA arms (5p and 3p) displayed higher mobility at 3'-end (large trimmings/additions) over the more conservative 5'-end (small trimmings/additions). Interestingly, most predominant SNP forms exclusively characterized the two arms, while the A-to-I RNA editing sites were mainly located within the 1-10 nts region on both arms. At first, we explored the capability of miRNA isoforms to cluster cancer samples across cohorts. We then examined the differences in the abundance of expressed isomiRs across cohorts' cancer samples, finding several cancer-related pathways enriched significantly. We also investigated miRNA isoforms from a diagnostic and prognostic standpoint. Finally, we experimentally validated gene targeting for two different miRNA isoforms: a shifted (no SNVs) and an edited isomiR (no shifting). By combining canonical miRNAs and miRNA isoforms, we obtained a higher cluster fragmentation that reflected a more profound clinical-pathological stratification. MiRNA isoforms resulted significantly deregulated across cohorts/cancer tissues, showing a distribution of modification types proportional to the one

observed for expressed annotated molecules. Overall Survival (OS) and Relapse Free Survival (RFS) prognostic signatures were significantly enriched with miRNA isoforms over canonical miRNAs, with multiple signatures entirely composed of miRNA isoforms. Finally, we experimentally assessed gene targeting exclusivity by investigating the canonical miR-101-3p and one of its shifted isomiR in lung adenocarcinoma, as well as the canonical miR-381-3p and one of its edited form (A-to-I editing at position 4) in breast cancer.

In summary, our findings highlight the importance of considering the broader modified miRNAome, which actively participates in gene regulation and may offer the opportunity to discover novel cancer biomarkers.

## MATERIAL AND METHODS

### Cell lines

HEK-293 (ATCC® CRL-1573™), A549 (ATCC® CCL-185™), and MDA-MB-231 (ATCC® HTB-26™) were seeded and grown in RPMI-1641 medium (Millipore Sigma) supplemented with 10% of fetal bovine serum (Millipore Sigma) and penicillin-streptomycin (100 U/mL penicillin and 0.1 mg/mL streptomycin) (Millipore Sigma), at 5% $CO_2$ and 95% air environment at 37°C, with 100% relative humidity. All cell lines were authenticated through the short-tandem repeat profiling method and tested to be free of mycoplasma contamination.

### Cell transfection

HEK-293, A549, and MDA-MB-231 cell lines were plated in a 6- or 12-wells plate 24 hours before transfection. 100 nM of *miR-101-3p* (full label *miR-101-3p__mir-101-1__0__0__21M*, see *MiRNA Isoform Nomenclature* for more details) mirVana™ miRNA Mimic (Thermo Fisher Scientific, ID# MC11414), *miR-101-3p (-1|-2)* (full label *miR-101-3p__mir-101-1__-1__-2__20M*, see *MiRNA Isoform Nomenclature* for more details) custom mirVana™ miRNA Mimics (Thermo Fisher Scientific, custom), *miR-381-3p* (full label *miR-381-3p__mir-381__0__0__22M*, see *MiRNA Isoform Nomenclature* for more details) mirVana™ miRNA Mimic (Thermo Fisher Scientific, ID# MC10242), and *miR-381-3p_4_A_G* (full label *miR-381-3p__mir-381__0__0__3MG18M*, see *MiRNA Isoform Nomenclature* for more details) custom mirVana™ miRNA Mimics (Thermo Fisher Scientific, custom) were transfected using Lipofectamine™ 2000 Transfection Reagent (Thermo Fisher Scientific) diluted in transfection-medium (RPMI-1641 without FBS or antibiotics). mirVana™ miRNA Mimic, Negative Control #1 (Thermo Fisher Scientific), and Anti-miR™ miRNA Inhibitor Negative Control #1 (Thermo Fisher Scientific) were employed as scrambled controls. After 5 hours, transfection-medium was replaced with RPMI-1641 supplemented with 10% of fetal bovine serum and penicillin-streptomycin. After 24 hours or 48 hours, cells were harvested and subjected to Luciferase assay or RNA isolation and protein lysis.

### Validation of canonical and novel miRNA-target by luciferase assay

PsiCHECK-2 vector (Promega) was employed to generate luciferase-based reporters for miRNA-target validation. 3' UTR of PTGS2 and SYT13 genes, containing miRNAs binding sites and ~200 nt flanking regions, were amplified by PCR from human genomic DNA (Promega), using the primers listed in Supplementary Table 1 and then inserted into the psiCHECK-2 vector, downstream to *Renilla* luciferase open reading frame. All inserted sequences were checked via Sanger Sequencing. 500 ng of psiCHECK™-2 vector holding the specific 3' UTR were transfected together with 100 nM of mirVana™ miRNA Mimics or Negative Controls in HEK-293 cells, as described above. After 24 hours, cells were lysed, and *Firefly* (internal control) and *Renilla* enzymatic activity were measured using Dual-Luciferase® Reporter Assay System (Promega) and detected by GloMax® 96 microplate luminometer (Promega), according to the manufacturer`s protocol. The comparison statistical significance was computed using the two-tailed unpaired Student's t-Test provided by the *t.test* function available in *stats,* an R (v3.4.4) package.

**RNA isolation, reverse transcription, and real-time RT-PCR**

The expression of canonical microRNAs, isomiRs, and target genes was analyzed by real-time RT-PCR after designing custom TaqMan® Small RNA Assays for isomiRs detection (Thermo Fisher Scientific). Total RNA was isolated from cells 48 hours post-transfection using TRIzol™ Reagent (Thermo Fisher Scientific) according to the standard protocol and measured with the Nanodrop 2000c instrument (Thermo Fisher Scientific). For specific microRNAs reverse-transcription, cDNA was synthesized from 5 ng of total RNA using TaqMan® Small RNA Assays RT-primers (Thermo Fisher Scientific) with the High-Capacity cDNA Reverse Transcription Kit, 30 min 16 °C, 30 min 42 °C, 5 min 85 °C. MicroRNA real-Time RT-PCR was performed using TaqMan™ Fast Advanced Master Mix according to the manufacturer`s protocol, with cataloged and custom TaqMan® Small RNA Assays (Thermo Fisher Scientific): miR-101-3p (assay ID 002253), miR-101-3p (-1|-2) (custom assay), miR-381-3p (assay ID 000571), and miR-381-3p_4_A_G (custom assay). The data were normalized using RNU44 (assay ID 001094).

**Protein lysis and western blotting analysis**

Cells were lysed in Lysis buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl, 10% Glycerol, and 0.5% Nonidet P40), supplemented with Protease inhibitors (Millipore Sigma). Then, 25 µg of proteins were loaded onto 4- 12% Mini-PROTEAN Tris-Tricine Precast Gels or Criterion Tris-Tricine Precast Gels (Bio-Rad), and electro-blotted on nitrocellulose membranes (GE Healthcare Life Science). Later, the membranes were blocked in blocking solution (TBS-0.05% Tween®20/fat-free milk 5%) and incubated overnight at 4°C with all the primary antibodies, anti-Cox2 (ABclonal Technologies), anti-SYT13 (Thermo Fisher Scientific), and anti-Vinculin [SPM227] (Abcam), previously diluted in fat-free milk 3-5%. The day after, the membranes were washed three times with TBS-0.05% Tween®20 (TBS-T) and incubated with appropriate HRP-conjugated secondary antibodies (Millipore Sigma), one hour at room temperature. After three additional washes, the membranes were assayed with ECL (Millipore Sigma), and the signal was marked and developed on Blue X-ray film (GeneMate) inside a

dark room. The comparison statistical significance was computed using the two-tailed unpaired Student's t-Test provided by the *t.test* function available in *stats*, an R (v3.4.4) package.

**MiRNA-Seq data**

The TCGA (v20) and TARGET (v20) miRNA-Seq samples (BAM file format) were retrieved via the Genomic Data Commons Data Portal (GDC Data Portal, https://gdc-portal.nci.nih.gov). Following the authorization from the data access committee (DBGap Project IDs: 11332 and 22219 for TCGA and TARGET repositories, respectively), samples were downloaded checking the following options: "*sequencing reads*" (Data Category), "*miRNA-Seq*" (Experimental Strategy), "*bam*" (Data Format), and "*TCGA*" and "*TARGET*" (Program). We adopted the TCGA and TARGET barcode to pair samples and clinical data, processing 33 TCGA (10,977 samples over 10,250 adult patients) and 5 TARGET (2,373 samples over 1,123 pediatric patients) cohorts.

**Clinical data**

The GDC Data Portal provides cohort patients' clinical data in *JavaScript Object Notation* (JSON) file format (e.g., *clinical.cart.XXXX.json*). Clinical data files are available within the GDC *cart* section along with the miRNA-Seq samples. The GDC Legacy Archive (https://portal.gdc.cancer.gov/legacy-archive) offers additional TARGET clinical data in Microsoft Excel file format. Collected clinical data files were downloaded, opportunely parsed, and harmonized across cohorts.

**Single nucleotide DNA/RNA variant data**

After creating the proper login account, COSMIC data (v92) were downloaded by clicking on "*Data*" and "*Downloads*" (main menu), and then on the "*CosmicNCV.tsv.gz*" link located on the right side ("*Non coding variants*" section). The dbSNP (v154) database (NCBI SNP), a VCF file format, was downloaded via the following link: *https://ftp.ncbi.nih.gov/snp/archive/b154/VCF/GCF_000001405.38.gz*.

Finally, A-to-I miRNA editing sites were downloaded from MiREDiBase (60).

**MiRNA-Seq quality check (QC)**

Downloaded BAM files were converted into FASTQ files by leveraging the *bamToFastq* tool (*bedtools* v2.25.0 package) (61) and finally quality-filtered through the ConDeTri tool (v2.3) (62) (parameters: -pb=fq -lq=20 -hq=30 -minlen=15 -sc=33). The workflow is summarized in Figure 1A.

**MiRNA isoforms mapping and quantification**

All quality-filtered sequences were aligned to the Ensembl (www.ensembl.org) human genome (hg38) and annotated via an in-house designed workflow (Figure 1A). The workflow itself leveraged the miRge 2.0 (52), a pipeline for canonical miRNAs/miRNA isoforms annotation based on the latest miRBase (v22) (10) and miRGeneDB 2.0 (63) datasets. The MiRge 2.0 pipeline was executed using the default parameters. Noteworthy, the annotation process covered a wide range of molecules (Figure 1B). Annotated molecules ranged from shifted to non-shifted isomiRs, along with molecules

subjected to single nucleotide variants (SNVs), such as A-to-I RNA Editing sites, Single Nucleotide Polymorphisms (SNPs), and somatic mutations. Shifted isomiRs included molecules characterized by sequence shifting affecting the 5', 3', or both ends (addition or trimming of nucleotides with respect to their canonical miRNA sequence). Once the annotation phase concluded, the workflow enriched annotated molecules with additional information, such as the latest known SNPs (dbSNP), somatic mutations (COSMIC, TCGA, and TARGET), and A-to-I RNA Editing sites collected from 40 studies (Supplementary Table 2), to maximize accuracy. Finally, data were collected into tab-separated text files per cohort (i.e., TCGA/Lung Adenocarcinoma shortened TCGA-LUAD).

**MiRNA isoform nomenclature**

In this work, we designed a unique human-readable way to label each mapped miRNA isoform. A label is *de facto* a combination of the canonical miRNA, pre-miRNA, 5'- and 3'-end shifting, and a "*Compact Idiosyncratic Gapped Alignment Report*" string, shortened CIGAR (https://samtools.github.io/hts-specs/SAMv1.pdf). Succinctly, the CIGAR string is a standard to indicate base match/mismatch and other operations (e.g., insertion - *I* - and deletion - *D*), used to describe sequence alignments. It is essential to point out that miRge current implementation supports only *insertion* operations, potentially posing some limitations on which miRNA isoforms can or cannot be mapped.

To better understand how a label looks like, let us consider the *miR-21-5p__mir-21__-1__+1__2MG21M* label. It represents a mapped miR-21-5p isoform (pre-miRNA: mir-21), which undergoes the following modifications:

- A single genomic nucleotide shifting (left) at 5'-end, denoted by *-1*;
- A single genomic nucleotide shifting (right) at 3'-end, indicated by *+1*;
- A single nucleotide modification at position 3 (i.e., an A-to-I RNA editing site), represented using the CIGAR string *2MG21M*. The *2M* and *21M* indicate a perfect match between the miRNA isoform and the genomic reference, while at position 3 (after 2M, or two matches), we ended up with a G instead of A (reference).

Above all, we extended the miRge CIGAR string paradigm, replacing each insertion (denoted with *I*), which solely involves the miRNA isoform extremities, with the mapped read's corresponding nucleotide. We adopted such an extension to deal with those miRNA isoforms (same miRNA, pre-miRNA, and shifting) which show identical CIGAR string but undergo different single nucleotide insertions (not reported by miRge). For the sake of clarity, we report below an example of two miRNA isoforms with a single nucleotide insertion at 5'-end and no shifting at 3'-end (represented by *0*), which undergo a single nucleotide insertion at 5'-end:

- **miRge CIGAR**:
    - MiRNA isoform n°1:
        - Canonical miRNA: miR-30a-5p
        - Pre-miRNA: mir-30a
        - Sequence mapped: ATGTAAACATCCTCGACTGGAAG
        - CIGAR: I22M

        - 5'-end shifting: -1
        - 3'-end shifting: 0
        - Tag: miR-30a-5p__ mir-30a__-1__0__ I22M
    - MiRNA isoform n°2:
        - Canonical miRNA: miR-30a-5p
        - Pre-miRNA: mir-30a
        - Sequence mapped: TTGTAAACATCCTCGACTGGAAG
        - CIGAR: I22M
        - 5'-end shifting: -1
        - 3'-end shifting: 0
        - Tag: miR-30a-5p__ mir-30a__-1__0__ I22M
- **miRge CIGAR extension**:
    - MiRNA isoform n°1:
        - Canonical miRNA: miR-30a-5p
        - Pre-miRNA: mir-30a
        - Sequence mapped: ATGTAAACATCCTCGACTGGAAG
        - CIGAR: A22M
        - 5'-end shifting: -1
        - 3'-end shifting: 0
        - Tag: miR-30a-5p__ mir-30a__-1__0__ A22M
    - MiRNA isoform n°2:
        - Canonical miRNA: miR-30a-5p
        - Pre-miRNA: mir-30a
        - Sequence mapped: TTGTAAACATCCTCGACTGGAAG
        - CIGAR: T22M
        - 5'-end shifting: -1
        - 3'-end shifting: 0
        - Tag: miR-30a-5p__ mir-30a__-1__0__ T22M

Given the example above, it is natural and straightforward to link the label *miR-21-5p__mir-21__0__0__22M* to the canonical miRNA form. The way we label mapped miRNA isoforms should give users a clear hint on which modifications occur, along with their precise location within the sequence. Finally, all novel miRNA molecules were labeled as follows: *miR-n##*, where *###* represents a positive incremental integer (e.g., miR-n86-3p, and miR-n136-5p).

**Data aggregation, annotation, and filtering**

For each dataset/cohort, multiple samples per patient and tissue type (i.e., a patient with more than one Primary Solid Tumor sample) were aggregated by calculating the raw read counts expression average. Then, raw read counts tables were enriched with additional information, including known SNVs, such as SNPs from dbSNP, somatic mutations from COSMIC, TCGA, and TARGET, along

with 2,885 A-to-I miRNA Editing sites. A final filter discarded miRNA isoforms with not yet characterized (unknown) SNVs involving the first or last two nucleotides. We removed these molecules because these unknown SNVs could be due to potential sequencing errors or imperfections in the linker ligation during the construction of the cDNA library (64–66). The workflow is summarized in Figure 1A.

**Data normalization**

In this study, every downstream analysis was preceded by a normalization process. In this process, raw read counts were filtered and normalized, applying the criterion described below. Depending on the type of data to be normalized, we used *Reads Per Million miRNA mapped reads (RPM)* for miRNA isoforms and *Fragments Per Kilobase Million (FPKM)* for transcripts. Specifically for miRNA isoforms, we computed the RPM expression out of the raw read counts. An expression filtering was then applied to retain miRNA isoforms/genes having a minimum expression of $\lfloor (RPM \mid FPKM)\ geometric\ mean \rfloor > 1$ . Lastly, the set of expression-filtered molecules was used to extract the corresponding raw read counts from the initial table and finally normalized via the *trimmed mean of M-values* (TMM) method (67) using the *calcNormFactors* function available in *edgeR* (v3.24.3) (68, 69), a *BioConductor* (v3.6) R (v3.4.4) (70) package.

**High dimensionality reduction and clustering**

Aimed to investigate the benefits and drawbacks of using specific miRNA isoforms for clustering purposes, we benchmarked three sets of expressed molecules (minimum expression of $\lfloor RPM\ geometric\ mean \rfloor > 1$) grouped according to their modification type (Supplementary Figure 2A). In the first set, labeled "*CAN*," we considered only canonical miRNAs (miRBase v22). In the second one, marked "*ISO_wo_SNV*," we used both canonical miRNAs and shifted isomiRs with no SNVs. In the last set, labeled "*ISO*," we considered all expressed canonical miRNAs and isomiRs, including the shifted ones. We applied an in-house designed workflow (Supplementary Figure 2B) to each set, aiming to assess molecules' ability to cluster samples across different cohorts/cancer tissues. The workflow starts extracting from each cancer cohort (e.g., cancer samples in TCGA-LUAD) all expressed miRNA isoforms according to a minimum expression of $\lfloor RPM\ geometric\ mean \rfloor > 1$. It then collects the raw reads count expression relative to the extracted miRNA isoforms, collapsing the data into a single massive raw read counts table and normalizing it via TMM. Afterward, a nonlinear dimensionality-reduction technique (Uniform Manifold Approximation and Projection - UMAP) (71) is applied to reduce high-dimensional data (the normalized table) into two-dimensional data, leveraging the Bray-Curtis distance (72). Except for the distance (metric=braycurtis), we performed UMAP (*umap-learn* Python package, v0.4.6) using its default parameters (min_dist=.1, n_neighbors=15, n_components=2), specifying a seed (random_state=99) for reproducibility purposes. The resulting lowered-dimension table is then used for sample visualization and clustering analysis (see Supplementary Figure 3). We applied the DBSCAN algorithm (73) to perform an unsupervised clustering based on UMAP two-dimensional data, leveraging the Euclidean distance. The DBSCAN

algorithm (*scikit-learn* Python package, v0.22.1) mainly relies on two distinct parameters: *eps* and *min_samples*. The *eps* value represents the maximum distance between two points, used to decide whether to group them or not. The *min_samples* value represents the minimum number of nearest neighbors (samples) for a point to be considered as a core point. Due to DBSCAN's high sensibility to the *eps* value, the workflow calculates the optimal value using the *K-Nearest Neighbor Distance*. Having optimal *eps* speeds up the analysis, allowing us to explore only the *min_samples* parameter (from 3 to 25, with incremental steps of 1). All clustering results are evaluated via Adjusted Rand Index (ARI) (74), Adjusted Mutual Information (AMI) (75), and Fowlkes-Mallows Index (FMI) (76) scores. Finally, all results are filtered considering a percentage of mislabeled samples (noise) less than 5%.

**Differentially miRNA isoforms expression analysis**

To detect dysregulated abundant miRNA isoforms across cancer cohorts, we performed a differential expression (DE) analysis with a minimum of 5 samples per cohort/tissue type: i.e., TCGA-LUAD/Solid Normal Tissue, and TCGA-LUAD/Primary Solid Tumor. Input data were normalized according to the normalization criterion described in the "*Data Normalization*" section. For each molecule, *linear fold change* and statistical significance were calculated using the mean and the two-sided unpaired Mann-Whitney U test (77). The resulting p-values were adjusted using Benjamini-Hochberg's correction (78), using the *fdrcorrection* function from *statsmodels* (v0.11.1) Python package. Finally, dysregulated molecules with $adjusted\ p-value < 0.05$ and $|linear\ fold\ change| > 1.5$ were retained.

**Differentially genes expression analysis**

Unlike the previous section, we performed the gene differential expression analysis for the pathways enrichment analysis and each molecule in the two case studies: the canonical miR-101-3p and its shifted isomiR, and the canonical miR-381-3p and its edited form.

For the two case studies, we grouped samples according to the first (Q1) and third (Q3) quartile of each molecule, investigating dysregulated genes to assess potential target variability. Instead, for the pathway enrichment analysis, we grouped, for each cohort/cancer tissue, samples into two groups: low (Q1) and high (Q3) abundant expressed miRNA isoforms.

Input data were normalized according to the normalization criterion described in the "*Data Normalization*" section. The gene differential expression analysis was performed using *edgeR* (v3.32.1), a *Bioconductor* (v3.6) R (v3.4.4) package, keeping all those dysregulated genes with $adjusted\ p-value < 0.05$ and $|linear\ fold\ change| > 1.5$.

**Pathway enrichment analysis**

The analysis relied on differentially expressed genes generated as described above. We then employed the Ingenuity® Pathway Analysis (IPA) software (v01-16) to perform a pathway enrichment analysis for each cohort/cancer tissue. Finally, we retained all those pathways characterized by $|z-score| \geq 2$ and $p-value < 0.01$. We generated a heatmap plotting the z-scores of the most

significant pathways enriched in at least five cohorts/cancer tissues, clustering them via Bray-Curtis distance metric and Complete Linkage method. The heatmap was created using the *clustermap* function from the *seaborn* (v0.10.1) Python package.

**Risk score-based prognostic signature**

We individually reviewed the clinical data associated with each cohort, considering the patient's survival time independent from each other and treating the censoring time as right-censored data.

We designed a 2-stages workflow (see Supplementary Figure 4) for optimal prognostic signature estimation to assess miRNA isoforms effectiveness as prognostic biomarkers.

The first stage arranges the patients' miRNA isoforms expression data. It starts by checking whether the cohort/cancer tissue provides at least *ten* patients per event type (e.g., *dead* and *alive* for the Overall Survival), representing a minimum number for reliable results. Then, the raw read counts are extracted for the selected patients and finally normalized as per the criterion described in the "Data Normalization" section. The stage ends by calculating a correlation matrix out of the normalized data.

The second stage starts exploring potential prognostic signatures by removing highly correlated miRNA isoforms. The stage iteratively retains those molecules with a correlation coefficient lower than a specific *threshold*, from 0.6 to 0.8, with incremental steps of 0.01. Besides the threshold, each iteration implements the same set of steps. The first step generates a univariate Cox proportional hazard regression model to assess the relationship between miRNA isoforms expression and patients' survival. Observed *Cox P-values* are then adjusted using Benjamini-Hochberg's *correction*. It continues performing a multivariate Cox proportional hazard regression model based on the most significant ($Cox\ FDR < 0.05$) miRNA isoforms extracted from the univariate model. Both univariate and multivariate Cox regression models leverage the *CoxPHFitter* function from the *lifelines* (v0.25.4) Python package. Extracted miRNA isoforms are then narrowed down by keeping those molecules with $Cox\ p-value < 0.05$ (multivariate model) and further reduced by applying a feature selection strategy. The selection strategy relies on Recursive Feature Elimination from the scikit-learn (v0.22.1) Python package, coupled with a Logistic Regression Model used as the estimator.

Next, a *Risk Score* is computed for each patient by linearly combining the expression value and the regression coefficient (univariate model) related to the reduced set of miRNA isoforms (selection strategy), as shown below:

$$Risk\ score = \sum_{i=1}^{n} \exp_i \cdot \beta_i$$

Where *n* represents the number of reduced miRNA isoforms, *exp* the expression, and *β* the regression coefficient (univariate Cox model). Patients are then separated into high- and low-risk groups using the patients' Risk Score median as a cutoff. Finally, the last step ends by generating a Kaplan-Meier Plot (79), a p-value (Log Rank Test), and a prognostic signature accuracy (area under the curve - AUC) based on the two risk groups. For the Kaplan-Meier Plot, we leveraged the *survfit* and *ggsurvplot* functions provided by the *survival* (v3.2-3) R (v3.4.4) package.

The workflow concludes by picking up the prognostic signature with the highest AUC for each cohort/cancer tissue. In this analysis, we investigate both *Overall Survival* (*OS*) (event: *death*, nonevent: *alive*) and *Relapse Free Survival* (RFS) (event: *relapse*, nonevent: *no relapse*).

**MiRNA isoform-target prediction**

All miRNA isoform-target predictions were generated by isoTar v1.2.1 (80). IsoTar leverages state-of-the-art prediction tools, such as miRmap (v1.1) (81), TargetScan (v7.0) (82), PITA (v6) (83), RNAhybrid (v2.1.2) (84), and miRanda (v3.3a) (85). IsoTar focuses on solely seed regions of 7-8 nucleotides in length (7mer-A1, 7mer-m8, 8mer), with no mismatch or G:U base pairs (wobbles). All predictions were performed using isoTar default parameters.

## RESULTS

**MiRNA isoforms profiling**

To investigate miRNA isoforms at a large scale, we used miRNA-Seq sample data from two large cancer datasets, TCGA and TARGET. Both datasets provide insights into a wide range of adult and pediatric cancers and are considered the most prominent and reliable public resources. We processed 33 TCGA (10,977 samples over 10,250 adult cancer patients) and 5 TARGET (2,373 samples over 1,123 pediatric cancer patients) cohorts (Materials and Methods). The cohorts' essential characteristics are summarized in Table 1. An in-house workflow (Figure 1A; Materials and Methods) was employed to perform canonical miRNAs/isomiRs annotation. The annotation process covered both shifted and non-shifted isomiRs, with the last ones characterized by sequence shifting affecting the 5', 3', or both ends (addition/trimming of nucleotides with respect to their canonical miRNA sequence). The process also annotated molecules subjected to Single Nucleotide Variants (i.e., SNPs, A-to-I RNA Editing sites, somatic mutations), hereafter called SNVs. At first, the workflow leveraged the miRge 2.0 pipeline. It then extended the annotation results with additional information, such as the latest known SNPs (dbSNP), somatic mutations (COSMIC, TCGA, and TARGET), and A-to-I RNA Editing sites collected from 40 studies (Supplementary Table 2), to maximize accuracy. Finally, a data filtering phase discarded all those annotated miRNA isoforms having SNVs not yet characterized (unknown) involving the first or last two nucleotides (Figure 1A). On average, we identified 2,569 $\pm$ 303 (*SD*) expressed molecules (minimum expression of $\lfloor RPM\ geometric\ mean \rfloor > 1$, *Reads Per Million*) per cohort. We also identified eight novel expressed miRNA isoforms, labeled as *miR-n###* (Materials and Methods). Notably, TCGA-TGCT (*testicular germ cell tumors*) and TCGA-GBM (*glioblastoma multiforme*) cohorts displayed the highest number of expressed miRNA isoforms (Figure 1C). Overall, the amount of expressed miRNA isoforms was about *8*-fold the number of expressed canonical miRNAs (miRBase v22). Of all identified RNA modifications, the 3'-shifted isomiRs represented the most abundant one, highly present in TCGA-TGCT, TCGA-THYM (*thymoma*), TCGA-UCS (*uterine carcinosarcoma*), TCGA-GBM, TCGA-LAML (*acute myeloid leukemia*), TARGET-RT (*rhabdoid tumors*), TCGA-SKCM (*skin cutaneous melanoma*), TCGA-PCPG (*pheochromocytoma and*

*paraganglioma*), and TCGA-ACC (*adrenocortical carcinoma*). The TCGA-LAML showed the highest number of expressed 5'-shifted isomiRs, along with TARGET-ALL-P3 (*acute lymphoblastic leukemia phase 3*), TCGA-READ (*rectum adenocarcinoma*), TCGA-TGCT, TCGA-COAD (*colon adenocarcinoma*), TCGA-UCEC (*uterine corpus endometrial carcinoma*), TCGA-LUSC (*lung squamous cell carcinoma*), TCGA-OV (*ovarian serous cystadenocarcinoma*), and TCGA-THYM. The TCGA-GBM, TCGA-UVM, TCGA-THYM, TCGA-ACC, TARGET-ALL-P2 (*acute lymphoblastic leukemia phase 2*), TCGA-PCPG, TCGA-LAML, TCGA-TGCT, and TARGET-RT cohorts exhibited the highest number of expressed miRNA isoforms with known SNVs (i.e., SNPs and somatic mutations). Besides, the TCGA-GBM, TCGA-PCPG, TARGET-ALL-P2, TCGA-TGCT, TCGA-UVM (*uveal melanoma*), TARGET-ALL-P3 (*acute lymphoblastic leukemia phase 3*), TCGA-ACC, TCGA-SKCM, and TCGA-THCA (*thyroid carcinoma*) were the most enriched cohorts with A-to-I miRNA editing sites (Supplementary Table 3). Lastly, the TCGA-CHOL (*cholangiocarcinoma*), TCGA-GBM, TCGA-PAAD (*pancreatic adenocarcinoma*), TCGA-MESO (*mesothelioma*), TCGA-ACC, TCGA-PCPG, TCGA-DLBC (*lymphoid neoplasm diffuse large b-cell lymphoma*), TCGA-THYM, and TCGA-ESCA (*esophageal carcinoma*) cohorts displayed the uppermost number of unknown SNVs.
To investigate the potential differences due to Drosha and Dicer cleavage on generating miRNA isoforms across each miRNA arm, we assessed the abundance of modification types (Supplementary Table 3 - "*Expressed molecules across arms*"). Overall, the 3'-shifted isomiR (no SNVs) represented the predominant modification in both arms. We later took a closer look at the 5'- and 3'-end shifting extent for the two arms. On average, the 5p arm 5'-end showed the highest stability (Supplementary Table 3 - "*5p arm - Shifting*"), with ~83% expressed isomiRs characterized by no shifting at all (5'-end untouched, no nucleotide added or trimmed) and ~10% with one nucleotide trimmed (one nucleotide removed at 5'-end). Similarly, the 3p arm 5'-end displayed comparable stability (Supplementary Table 3 - "*3p arm - Shifting*"), although a bit lower (~75%), promoting the addition (~9%) and trimming (~12%) of one nucleotide at 5'-end, respectively. By stark contrast, the 3'-end revealed higher mobility. The percentage of expressed isomiRs with no 3'-end shifting plunged to ~30% (5p arm) and ~33% (3p arm). In much the same way (Supplementary Table 3 - "*Arms shifting comparison*"), the two arms consistently showed trimming of 5 (~4%), 4 (~4%), 3 (~6%), 2 (~10%), and 1 (~21%) nucleotides, along with the addition of 1 (~14%) to 2 (~5%) nucleotides. We examined the SNVs distribution along a hypothetical miRNA isoform sequence of ~26 nts long to include the farthest known SNV observed in our expressed molecules. Interestingly, the A-to-G, C-to-T, G-to-A, and T-to-C represented the most predominant SNP forms in both 5p and 3p arms (Supplementary Table 3 - "*5p and 3p arms – SNPs*"). They were scattered along the whole sequence length, with the last three highly located close to the 3'-end. Switching to the 5p arm, the G-to-T, T-to-A, and T-to-G forms were primarily located near the 3'-end, nearby the 21$^{st}$ nucleotide. Aside, the remaining fewer present forms were somehow scattered along the sequence. Notably, the 5p arm was more susceptible to somatic mutations, even though no particular somatic mutation form emerged above the others (Supplementary Table 3 - "*5p and 3p arms - Somatic mut.*"). A-to-I RNA editing sites were mainly located within the 1-10 nts region (seed region included) in both 5p and 3p arms, with few additional sites involving the 15-24 nts region (Supplementary Table 3 - "*5p and 3p arms - A-to-I RNA Edit.*").

Altogether, the distribution of modification types was somehow balanced between 5p and 3p arms, with the 5p arm leading by up to 1% more expressed molecules over the 3p arm. Finally, to gain some insights on the mechanisms that may lead to the isomiRs accumulation in cancer, we explored the differences in the abundance of expressed miRNA isoforms across cohorts' cancer samples. For each cohort/cancer tissue, we compared samples characterized by a low (first quartile) against a high (third quartile) number of expressed miRNA isoforms. Here, we retained significantly dysregulated genes with $adjusted\ p-value < 0.05$ and $|linear\ fold\ change| > 1.5$ (Materials and Methods). We then performed a pathways enrichment analysis via Ingenuity® Pathway Analysis (IPA) software, keeping all those pathways characterized by $|z-score| \geq 2$ and $p-value < 0.01$ (Supplementary Table 4). Finally, in Supplementary Figure 1, we reported the most significant pathways enriched in at least five cohorts/cancer tissues. The difference in the abundance of expressed miRNA isoforms between the two groups, low and high, characterized the activation/deactivation of several critical pathways involved in proliferation, metastasis, tumor immune escape, invasion, and angiogenesis, such as the ILK, HIF1α, and Rac signaling pathways, PD-1/PD-L1 cancer immunotherapy pathway, and regulation of the epithelial-mesenchymal transition (EMT) by growth factors pathway.

**MiRNA isoforms-based clustering reveals unique clinical-pathological stratification**

Purely for clustering purposes, we benchmarked three sets of expressed molecules ($\lfloor RPM\ \ geometric\ mean \rfloor > 1$) grouped according to their modification type (Supplementary Figure 2A; Materials and Methods). We investigated the benefits and drawbacks of using specific sets of molecules by assessing their ability to cluster samples across different cohorts/cancer tissues. We moved from a set to another, increasing the number and type of molecules considered. In the first set, labeled "*CAN*," we considered only canonical miRNAs (miRBase v22). In the second one, marked "*ISO_wo_SNV*," we used both canonical miRNAs and shifted isomiRs without SNVs. In the last set, labeled "*ISO*," we employed all expressed canonical miRNAs and isomiRs, including the shifted ones. We applied an in-house designed workflow to each of the three sets (Supplementary Figure 2B; Materials and Methods), extracting expressed molecules from every cohort and condensing them into a single massive table (expressed molecules as rows, cohorts' cancer samples as columns). To deal with high dimensional data, we applied a nonlinear dimensionality-reduction technique (Uniform Manifold Approximation and Projection - UMAP) (71) (Supplementary Figure 2B; Materials and Methods) to reduce high-dimensional data into a two-dimensional matrix for data visualization and evaluation. Finally, based upon the reduced two-dimensional matrix, we performed an unsupervised clustering by leveraging the DBSCAN algorithm (73) to test the sets' clustering capability. By benchmarking the three sets, we observed a general trend in which the *ISO* set reached a higher clustering ability (Figure 2A), a result assessed by the Adjusted Rand Index (AMI), Adjusted Mutual Information (AMI), and Fowlkes-Mallows Index (FMI) scores. The *ISO*-based clustering outlined a better separation of cohorts' cancer samples than the other two sets (Figure 2B). We then examined

clustering results from a clinical-pathological perspective, focusing solely on available and most significant clinical-pathological features ($Chi-Square\ p-value < 0.01$) (Supplementary Table 5). Overall, the *CAN*-, *ISO_wo_SNV*-, and *ISO*-based findings somehow supported one another in clustering results (Supplementary Figure 3A-C). For instance, in the TCGA-LIHC and TCGA-TGCT cohorts, both *CAN* and *ISO_wo_SNV* sets grouped patients according to their clinical stages (Supplementary Figure 3A-B). At the same time, both *CAN* and *ISO* sets significantly clustered patients in the TCGA-STAD cohort, though reflecting different clinical-pathological features (Supplementary Figure 3A, 3C).

By including miRNA isoforms data (*ISO_wo_SNV* and *ISO* sets), we obtained a more refined classifier, which highlighted, in some cohorts, additional subclusters with clinical-pathological relevance. Unlike the *CAN* set, both *ISO_wo_SNV* and *ISO* sets aggregated the TCGA-ESCA and TARGET-AML cancer samples in the same way. In TCGA-ESCA, patients were split according to their histological type, squamous (*C6* in *ISO_wo_SNV*, *C12* in *ISO*) and adeno (*C11* in *ISO_wo_SNV*, *C13* in *ISO*). In TARGET-AML, cancer samples were grouped in two clusters, *C0* and *C3*, consistently with their cytogenetic complexity, a well-known prognostic marker (Supplementary Figure 3B-C; Supplementary Table 5). Notably, besides having a lower cytogenetic complexity, cluster *C0* included cancer samples that harbored chromosomal translocations commonly associated with good prognosis, t(9;11)(p22;q23) and inv(16) (86, 87). In stark contrast, cluster *C3* was enriched with FLT3-ITD positive samples associated with poor survival (88).

Finally yet importantly, the three sets were able to cluster cohorts' cancer samples exclusively. The *CAN* set uniquely partitioned cancer samples in TCGA-HNSC, distinguishing among patients graded as well (G1) (cluster *C17*) and poorly (G3) (*cluster C5*) differentiated (Supplementary Figure 3A). The *ISO_wo_SNV* set separated the TCGA-KIRP patients according to their clinical stage (Supplementary Figure 3B). Furthermore, the *ISO* set clustered the TCGA-COAD and TCGA-READ cohorts' patients into two groups, in which cluster C23 was characterized by lymphatic invasion and the presence of a history of polyps. In TCGA-LUSC, patients were grouped according to their clinical stages (Supplementary Figure 3C; Supplementary Table 5).

**Differentially abundant miRNA isoforms across cancer tissues**

To determine whether miRNA isoforms were dysregulated across cohorts/cancer tissues, we performed a differential expression (DE) analysis comparing *primary solid*, *recurrent solid*, *metastatic*, and *normal tissues*. Amid expressed molecules ($\lfloor RPM\ geometric\ mean \rfloor > 1$), we estimated both linear fold change and statistical significance for each molecule, using mean and two-sided unpaired Mann-Whitney U test (77), respectively (Materials and Methods), adjusting the resulting p-values via Benjamini-Hochberg's correction (78). The most significantly dysregulated molecules were retained according to an $adjusted\ p-value < 0.05$ and $|linear\ fold\ change| > 1.5$ (Supplementary Table 6). The resulting molecules were characterized by a similar trend outlined in the "*MiRNA Isoforms Profiling*" paragraph (Figure 1B). They were affected mainly by 3'-end shifting in almost all cohorts/comparisons, followed by miRNA isoforms with both 5'- and 3'-end shifting. A summary of the

distribution of dysregulated molecules per cohort/comparison over miRNA isoform modification types is reported in Figure 3A. Notably, we identified four novel dysregulated miRNA isoforms, labeled as *miR-n###* (Materials and Methods). We then examined the modification types distribution across 5p and 3p arms (Supplementary Table 7 - "*Mod. types distribution - Arms*"). The 5p arm showed ~1.12% ±0.46 (*SD*) additional molecules than the other arm in most cohorts/comparisons. Once again, the 3'-end shifting modification (no SNVs) exhibited the highest number of dysregulated molecules in both arms. Considering the subtle difference between the 5p and 3p arms, we decided to aggregate the two arms' contribution, reporting the most noticeable results in Figure 3B. The 5'-end was confirmed to be the most stable of the two ends, with ~70-84% dysregulated molecules characterized by no 5'-end shifting, in sharp contrast to the 3'-end, whose stability sank to ~27-37% (Figure 3B; Supplementary Table 7 - "*5'- and 3'-end shifting*"). The 5'-end was characterized by the addition and trimming of one nucleotide, while the 3'-end displayed a broader shifting, trimming 1 to 5 nts and adding 1 to 2 nts (Figure 3B; Supplementary Table 7 - "*5'- and 3'-end shifting*").

Moving the attention to known DNA SNVs, we identified the G-to-T, C-to-T, and G-to-A forms to be the most frequent variants affecting upregulated molecules, whereas downregulated molecules mainly faced G-to-A, C-to-T, and C-to-A modifications (Figure 3B; Supplementary Table 7 - "*SNPs-Somatic mut. distribution*"). These known DNA SNVs were spread along the whole sequence length in both upregulated and downregulated molecules. Finally, A-to-I RNA editing sites were still primarily located within the 1-10 nts region (seed region included), with few exceptions in the 15-24 nts region (Figure 3B; Supplementary Table 7 - "*A-to-I RNA Edit. distribution*").

**Dysregulated canonical miRNAs and isomiRs with opposite expression trend in cancer reveal different behavior**

Taking a closer look at the dysregulated canonical miRNAs across cohorts, we observed 104 out of 573 canonical miRNAs being characterized by an opposite expression trend compared to their miRNA isoforms (Supplementary Table 8). In this regard, we searched for a candidate among the 104 canonical miRNAs to assess the potential gene targeting shifting between miRNA isoforms and their canonical counterparts. Interestingly, the canonical miR-101-3p appeared to be downregulated in 6 cohorts (TCGA-LUAD, TCGA-LIHC, TCGA-HNSC, TARGET-RT, TARGET-WT, and TCGA-CHOL), with the sole TCGA-LUAD cohort reporting miRNA isoforms lacking SNVs, also confirmed by previous studies (90, 91).

We elected to investigate the canonical miR-101-3p and one of its isomiRs in lung adenocarcinoma (TCGA-LUAD), comparing *primary solid* against *normal tissue*. Our interest in investigating such a canonical miRNA was corroborated by a recent work in which authors assessed one isomiR of canonical miR-101-3p in the human brain (92). Furthermore, the authors demonstrated the isomiR ability to negatively modulate the expression of five validated miR-101-3p targets, leading them to consider the isomiR as a miR-101-3p functional variant.

In our work, we studied one isomiR, labeled as miR-101-3p (-1|-2), characterized by one nucleotide added at 5'-end, termed "-1," and two nucleotides trimmed at 3'-end, termed "-2" (see Materials and Methods). As shown in Figure 4A, the two molecules were characterized by an opposite expression

trend, with the canonical miRNA resulting downregulated in cancer samples. Dysregulated genes were retained based on their significance (Materials and Methods) and opposite expression trends (i.e., miRNA *up* and genes *down*, or vice versa) (Supplementary Table 9). Finally, the set of dysregulated targets was intersected with the list of predicted gene targets generated by *isoTar* (93) (Supplementary Table 9), requiring a minimum consensus of two prediction tools. Out of the reduced set of genes, we elected *PTGS2* (*COX-2*), a gene studied in several cancers (94–100), which is a validated target for miR-101-3p and overexpressed in lung cancer (Liu et al., 2015). PTGS2 promotes tumor growth, angiogenesis, and tissue invasion. It also induces resistance to therapeutic agents, compromising tumor immunity and apoptosis (101, 102). In line with the literature, we demonstrated the direct binding (Figure 4B) through luciferase assay (Figure 4C) between this oncogene and the downregulated canonical miR-101-3p. After miR-101-3p ectopic overexpression in HEK-293 cells, we observed ~40% reduction of luciferase activity compared to the scramble negative control (SCR) (Figure 4C). On the other hand, overexpression of miR-101-3p (-1|-2) resulted in a minor (20%) reduction of luciferase activity (Figure 4C). After transfecting the canonical miR-101-3p (Figure 4D), western blotting experiments (Figure 4E) highlighted a significant downregulation of endogenous PTGS2 in the A549 lung cancer cell line, while no variation of PTGS2 was observed transfecting *miR-101-3p (-1|-2)* (Figure 4E). Our findings demonstrated that of the two molecules, only the downregulated canonical miR-101-3p exclusively targeted PTGS2, which in turn is upregulated in lung cancer. Even though the two miRNA molecules rise from the same locus, they showed different behavior. In fact, in stark contrast with the shifted isomiR, these results corroborated the tumor suppressor role of the canonical miR-101-3p in lung cancer.

**Dysregulated A-to-I Edited miRNA isoforms in cancer**

In addition to investigating shifted isomiRs, we measured the A-to-I RNA editing abundance across cancer cohorts/comparisons, detecting 169 unique dysregulated A-to-I edited miRNA isoforms (Supplementary Table 6) that originated from 43 distinct miRNA arms. Looking closely, the edited miR-381-3p (A-to-I RNA editing at position 4) resulted in one of the most diffused dysregulated molecules among the before-mentioned ones. Its downregulation interested 11 out of 22 cohorts/comparisons, a trend also confirmed by previous studies (44, 46) and observed in several tumors (103–108), including breast cancer (109–115).

In this work, we examined the canonical miR-381-3p and one of its edited forms in the breast cancer cohort (TCGA-BRCA), labeling both molecules as *miR-381-3p* and *miR-381-3p_4_A_G*, respectively. The expression of the two molecules exhibited a significant downregulation in cancer samples (Figure 4F). In line with the previous section, we applied a similar workflow (Materials and Methods) to assess potential target variability between the two molecules. After retaining significantly dysregulated genes (Supplementary Table 9) characterized by an opposite expression trend (miRNA *down*, genes *up*), we crossed them with the list of gene targets predicted by *isoTar* (Supplementary Table 9). Out of the reduced set of potential direct targets for *miR-381-3p_4_A_G*, we elected to study SYT13, an oncogene involved in different cancers (116, 117). By using luciferase reporter vectors containing the 3' UTR of the gene and the two miRNA molecules (canonical and edited one) in HEK-293 cells, we

demonstrated a direct binding between SYT13 and miR-381-3p_4_A_G (Figures 4G), as highlighted by the reduced luciferase activity of ~80% (Figure 4H). After miR-381-3p_4_A_G overexpression (Figure 4I), western blotting experiments in TNBC MDA-MB-231 cells corroborated our findings (Figure 4J), demonstrating a significant downregulation of SYT13 (~50%), as depicted by densitometry (Figure 4J). Once again, our results pointed out the importance of not limiting studies solely to canonical miRNAs, as outlined by the edited miRNA's ability to target one oncogene exclusively.

**Prognostic miRNA Isoform Signature**

We designed an in-house 2-stages workflow (Supplementary Figure 4; Materials and Methods) to estimate each cohort/cancer tissue's best performing prognostic signature for Overall Survival (OS) and Relapse Free Survival (RFS). The cohorts' clinical characteristics are summarized in Table 2.

Out of the 26 cohorts/cancer tissues examined, only 12 (OS) and 9 (RFS) led to significant (Log Rank Test-based $p-value < 0.01$) prognostic signatures (Figure 5; Supplementary Table 10) with an Area Under the Curve (AUC) $\geq 0.7$. Notably, the TCGA-ACC and TCGA-UVM cohorts provided statistically significant signatures for both OS and RFS. The 21 identified signatures were associated with four distinct cancer tissues (GDC nomenclature): *primary solid tumor* (T), *primary blood-derived cancer – peripheral blood* (PBDP), *recurrent blood-derived cancer – bone marrow* (RBDB), and *primary blood-derived cancer – bone marrow* (PBDB).

Among OS signatures (Supplementary Figure 5A; Supplementary Table 10), TARGET-ALL-P2 was characterized by three signatures of *four* (cancer tissue: PBDP), *eight* (cancer tissue: RBDB), and *eleven* molecules (cancer tissue: PBDB), with the solely PBDB signature including one canonical miRNA (hsa-miR-4772-3p). The *6*- in TARGET-AML (cancer tissue: PBDP), the *8*- in TCGA-ACC (cancer tissue: T), the *17*- in TCGA-LIHC (cancer tissue: T), and the *4*-molecules signature in TCGA-PRAD (*prostate cancer*, cancer tissue: T) they all included the canonical miRNA hsa-miR-1287-3p, hsa-miR-362-5p, hsa-miR-584-5p, and hsa-miR-625-3p, respectively (Supplementary Table 10). Interestingly, TCGA-LGG (*lower-grade glioma*, cancer tissue: T) represented the sole cohort to contain six canonical miRNAs (hsa-let-7f-5p, hsa-miR-99b-3p, hsa-miR-155-5p, hsa-miR-346, hsa-miR-551b-3p, and hsa-miR-6720-3p) out of 17 molecules (Supplementary Table 10). Besides, the TCGA-LAML (cancer tissue: T, nine molecules), the TCGA-MESO (cancer tissue: T, nine molecules), the TCGA-THCA (cancer tissue: T, four molecules), and the TCGA-UVM (cancer tissue: T, two molecules) signatures included no canonical miRNA.

The RFS signatures (Supplementary Figure 5B; Supplementary Table 10) referred to the solely primary solid tumor (cancer tissue: T). Only three out of nine signatures held canonical miRNAs. In particular, the TCGA-ESCA three-molecules signature included hsa-let-7f-5p; the TCGA-HSNC twelve-molecules signature comprised hsa-miR-2355-5p and hsa-miR-4677-3p; the TCGA-KIRP three-molecules signature contained hsa-miR-199b-5p. The remaining signatures in TCGA-ACC (1 molecule), TCGA-KICH (1 molecule), TCGA-PCPG (2 molecules), TCGA-SKCM (1 molecule), TCGA-STAD (5 molecules), and TCGA-UVM (1 molecule) embodied no canonical miRNA. The complete list

of identified signatures, their miRNA isoforms, and additional details are reported in Supplementary Table 10.

## DISCUSSION

The quick brown fox jumps over the lazy dog. Several scientific contributions have magnified our understanding of the so-called miRNA Epitranscriptome, inflating the interest in RNA modifications such as SNPs (19), A-to-I RNA editing (11, 12, 43–45, 47, 118), as well as shifted isomiRs (16, 41, 48, 49, 55), a product of the imprecise miRNA sequence cleavage (13, 14, 32, 35). Although the studies mentioned above have examined miRNA modifications individually, the concurrent occurrence of two miRNA modifications, A-to-I miRNA editing and shifted isomiRs, has yet to be adequately explored.
In this work, we simultaneously estimated the abundance and implications of a broader set of RNA modifications, processing data at a large scale from the most prominent and reliable public resources, TCGA and TARGET, in a pan-cancer setting. Through miRge 2.0 (52), one of the major pipelines for canonical miRNAs/miRNA isoforms profiling (53–56), we profiled >13K adult and pediatric cancer samples spread across 38 distinct cohorts. At first glance, several miRNA isoforms displayed a higher expression than their canonical miRNAs, which are the reference molecules in biological databases such as miRBase (v22). In particular, the abundance of expressed miRNA isoforms exceeded by *8*-fold the number of expressed canonical miRNAs. A closer look at the distribution of modification types among expressed miRNA isoforms outlined a predominance of 3'-end shifts, which equally impacted both 5p and 3p arms due to potentially no differences in Drosha and Dicer cleavage. Affected by the addition/trimming of two or more nucleotides, the 3'-end showed higher mobility than the more conservative 5'-end (15). Interestingly, the lower presence of more extended additions (addition of three or more nts at 3'-end) could be explained via degradation processes carried out by some enzymes that remove the exceeding part spurting out the RISC complex (119). As is well known, both 5'- and 3'-end fulfill different functionalities. The first carries out the miRNA-mRNA partial base-pairing through the MSR, though the 3'-end is proven to be critical for the miRNA-mRNA interaction stabilization (28–30), especially in the presence of mismatches or bubbles (120). In light of this, the main reason for the high number of expressed 3'-end shifted molecules could be explained as the cell's attempt to modulate miRNAs activity, perhaps trying to overcome the weakness of specific miRNA-mRNA bindings under particular conditions (120). Likewise, the 5'-end shifting could be the effort of replacing missing canonical miRNAs or the necessity for a targetome shifting (121). Interestingly, more than 40% of dysregulated edited miRNA isoforms reported at least one A-to-I editing site within the MSR. These observations may indicate the tendency of the A-to-I RNA editing phenomenon to give rise to dysregulated miRNA isoforms in cancer, which could exhibit a different targetome and biological function with respect to their canonical counterpart (12). Besides, the distribution of the most representative known DNA variant forms (top-five) observed across dysregulated miRNA isoforms mostly fall near the 3'-end (between the 21$^{st}$ and 23$^{rd}$ nucleotides), potentially impinging either the miRNA lifespan or the targeting stability (66, 119). Lastly, we explored the underlying differences in the abundance of expressed miRNA isoforms in each cohort/cancer tissue from a functional standpoint. We compared cancer samples characterized by lower (first

quartile) and higher (third quartile) number of expressed molecules. The functional analysis corroborated our hypothesis that miRNA isoforms actively regulate critical genes in cancer. In fact, our results show the activation/deactivation of several critical pathways involved in proliferation, metastasization, tumor immune escape, invasion, and angiogenesis, such as the ILK, HIF1α, and Rac signaling pathways.

We then explored the canonical miRNAs/miRNA isoforms' ability to cluster cancer samples across cohorts, benchmarking three different sets of molecules according to specific RNA modifications. Moving on from using only canonical miRNAs (*CAN* set) to employ all expressed molecules (*ISO* set - canonical miRNAs and miRNA isoforms) allowed us to gain a higher cluster fragmentation that reflected an in-depth clinical-pathological stratification. Notably, in the *ISO_wo_SNV- and ISO*-based clustering results, the TARGET-AML cancer samples were significantly separated into two subclusters characterized by patients without (better prognosis) and with FLT3-ITD mutation (increased relapse risk and reduced overall survival). In the TCGA-COAD and TCGA-READ cohorts, cancer samples were clustered according to lymphatic invasion. Nonetheless, the three sets exclusively clustered cancer samples in TCGA-HNSC (*CAN* set), TCGA-KIRP (*ISO_wo_SNV* set), TCGA-COAD, TCGA-READ, and TCGA-LUSC (*ISO* set). Overall, the combination of canonical miRNAs/miRNA isoforms (*ISO* set) boosted the quality of our results, uniquely outlining clinical-pathological features in cohorts where the other sets failed. Altogether, our results depicted a more complex scenario in which canonical miRNAs and miRNA isoforms seemed to work tightly together to uncover the underlying histopathological differences among cancers. The exclusion of one of the two may substantially limit our understanding of tumor heterogeneity.

Both canonical miRNAs and miRNA isoforms resulted significantly dysregulated across all cohorts/cancer tissues. These results may suggest that miRNA isoforms are not the product of Drosha or Dicer's arbitrary cleavage, but they are actively expressed and dysregulated across several human cancers. Of the 573 canonical miRNAs dysregulated across cohorts/cancer tissues, we identified 104 characterized by an opposite expression trend compared to their miRNA isoforms. Supported by previous studies (90, 91), we investigated as the first case study the downregulated canonical miR-101-3p and its upregulated shifted isomiR (one nt longer at 5'-end, and two nts shorted at 3'-end) in lung adenocarcinoma primary tumors (TCGA-LUAD). Aiming at assessing differences in targeting efficiency, we examined dysregulated and predicted gene targets for the two molecules. Our choice fell on PTGS2 (COX-2), an oncogene in lung cancer (Liu et al., 2015), which is a validated canonical miR-101-3p target gene in different cancers (94–100). Although the predicted binding sites for both miRNA molecules and PTGS2's 3' UTR were comparable in terms of binding free energy, we experimentally proved that the sole canonical miR-101-3p was able to target PTGS2, corroborating the difference in terms of gene targeting for the miRNA molecules. In the second case study, we assessed the targetome shifting between two downregulated molecules, the canonical miR-381-3p and its edited form (A-to-I editing site at position 4). In our results, the edited miR-381-3p resulted among the most downregulated molecules across cohorts/cancer tissues. While the role of the canonical miR-381-3p is broadly acknowledged as a tumor suppressor (107, 108), particularly in breast cancer (112, 115), very little is known about the edited form in cancer (44, 46). We investigate

SYT13 in breast cancer (TCGA-BRCA), an oncogenic gene in different cancers (116, 117, 122–127). Unlike the canonical miR-381-3p, which did not show any binding site, our predictions and experiments outlined the ability of the edited miR-381-3p to exclusively regulate the SYT13 expression, suggesting it as a potential tumor suppressor in breast cancer. Finally, the survival analysis results highlighted once again the miRNA isoforms predominance over canonical miRNAs. Of all 21 significant prognostic signatures identified, only nine contained at least one canonical miRNA. Interestingly, almost all signatures contained unique molecules, with a few exceptions.

In conclusion, even though their role is still not well understood, miRNA isoforms may somehow work together with canonical miRNAs to support their function. Our study emphasized the importance of moving forward from the paradigm of one-single-mature-miRNA per miRNA arm to the wider miRNAome. Through these novel potential diagnostic and prognostic cancer biomarkers, we may be able to shine additional lights on those mechanisms related to cancer progression through the study of gene regulation via the wider miRNAome.

## DATA AVAILABILITY

The data and source code developed in this work will be stored on GitHub and Zenodo repositories. The TCGA (v20) and TARGET (v20) miRNA-Seq samples, patients' clinical-pathological data, and somatic mutations used during this study can be downloaded via http://portal.gdc.cancer.gov. Additional single nucleotide DNA variants were downloaded from COSMIC (v92) at https://cancer.sanger.ac.uk/cosmic and dbSNP (v154) at https://ftp.ncbi.nih.gov/snp/archive/b154/VCF, while A-to-I miRNA Editing sites were retrieved from MiREDiBase (v1).

## ACKNOWLEDGEMENT

The results published here are in part based upon data generated by The Cancer Genome Atlas (TCGA) Research Network (https://cancer.gov/tcga, DBGap Project ID: 11332) and by the Therapeutically Applicable Research to Generate Effective Treatments initiative (TARGET) of the NCI (http://ocg.cancer.gov/programs/target, DBGap Project ID: 22219). We thank Dr. Andrea Ventura for his valuable comments and suggestions during the development of this study and the scientific discussions during the preparation of the manuscript. We want to thank the Cancer IT Operation Group of The Ohio State University and Mr. Thomas Moore for his valuable technical assistance. We also want to thank the Ohio Supercomputer Center for its resources and technical support. We thank the Genomics Shared Resource at The Ohio State University Comprehensive Cancer Center (OSU CCC), Columbus, OH, for conducting the qRT-PCR experiments, which were supported in part by the OSU CCC and the National Institutes of Health grant P30 CA16058. M.A. was supported by CTSA award No. UL1TR002649 and NCATS 5KL2TR002648 grant.

## FUNDING

This work was supported by the National Cancer Institute (National Institute of Health) [R35CA197706 to C.M.C]. Funding for open access charge: National Cancer Institute (National Institute of Health).

## CONFLICT OF INTEREST

The authors declare no competing interests.

## REFERENCES


1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

2. Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*, **15**, 509–524.

3. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet*, **11**, 597–610.

4. Mendell,J.T. and Olson,E.N. (2012) MicroRNAs in Stress Signaling and Human Disease. *Cell*, **148**, 1172–1187.

5. Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer*, **6**, 857–866.

6. Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*, **10**, 704–714.

7. Iorio,M.V. and Croce,C.M. (2012) MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med*, **4**, 143–159.

8. Bartel,D.P. (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, **136**, 215–233.

9. Lee,L.W., Zhang,S., Etheridge,A., Ma,L., Martin,D., Galas,D. and Wang,K. (2010) Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA*, **16**, 2170–2180.

10. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucl. Acids Res.*, **42**, D68–D73.

11. Nigita,G., Veneziano,D. and 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**.

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

13. Bofill-De Ros,X., Kasprzak,W.K., Bhandari,Y., Fan,L., Cavanaugh,Q., Jiang,M., Dai,L., Yang,A., Shao,T.-J., Shapiro,B.A., *et al.* (2019) Structural Differences between Pri-miRNA Paralogs Promote Alternative Drosha Cleavage and Expand Target Repertoires. *Cell Reports*, **26**, 447-459.e4.

14. Neilsen,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, **28**, 544–549.

15. 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.

16. 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 Research*, **45**, 2973–2985.

17. Zhou,H., Arcila,M.L., Li,Z., Lee,E.J., Henzler,C., Liu,J., Rana,T.M. and Kosik,K.S. (2012) Deep annotation of mouse iso-miR and iso-moR variation. *Nucleic Acids Research*, **40**, 5864–5875.

18. Saunders,M.A., Liang,H. and Li,W.-H. (2007) Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences*, **104**, 3300–3305.

19. Sun,G., Yan,J., Noltner,K., Feng,J., Li,H., Sarkis,D.A., Sommer,S.S. and Rossi,J.J. (2009) SNPs in human miRNA genes affect biogenesis and function. *RNA*, **15**, 1640–1651.

20. Luciano,D.J. (2004) RNA editing of a miRNA precursor. *RNA*, **10**, 1174–1177.

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

22. Gallo,A. and Locatelli,F. (2012) ADARs: allies or enemies? The importance of A-to-I RNA editing in human disease: from cancer to HIV-1. *Biological Reviews*, **87**, 95–110.

23. Maas,S., Kawahara,Y., Tamburro,K.M. and Nishikura,K. (2006) A-to-I RNA Editing and Human Disease. *RNA Biology*, **3**, 1–9.

24. Dominissini,D., Moshitch-Moshkovitz,S., Amariglio,N. and Rechavi,G. (2011) Adenosine-to-inosine RNA editing meets cancer. *Carcinogenesis*, **32**, 1569–1577.

25. Galeano,F., Tomaselli,S., Locatelli,F. and Gallo,A. (2012) A-to-I RNA editing: The “ADAR” side of human cancer. *Seminars in Cell & Developmental Biology*, **23**, 244–250.

26. Kawahara,Y., Zinshteyn,B., Sethupathy,P., Iizasa,H., Hatzigeorgiou,A.G. and Nishikura,K. (2007) Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. *Science*, **315**, 1137–1140.

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

28. Bail,S., Swerdel,M., Liu,H., Jiao,X., Goff,L.A., Hart,R.P. and Kiledjian,M. (2010) Differential regulation of microRNA stability. *RNA*, **16**, 1032–1039.

29. Brennecke,J., Stark,A., Russell,R.B. and Cohen,S.M. (2005) Principles of MicroRNA–Target Recognition. *PLoS Biol*, **3**, e85.

30. Brodersen,P. and Voinnet,O. (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol*, **10**, 141–148.

31. Chiang,H.R., Schoenfeld,L.W., Ruby,J.G., Auyeung,V.C., Spies,N., Baek,D., Johnston,W.K., Russ,C., Luo,S., Babiarz,J.E., *et al.* (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & Development*, **24**, 992–1009.

32. Bizuayehu,T.T., Lanes,C.F., Furmanek,T., Karlsen,B.O., Fernandes,J.M., Johansen,S.D. and Babiak,I. (2012) Differential expression patterns of conserved miRNAs and isomiRs during Atlantic halibut development. *BMC Genomics*, **13**, 11.

33. Cloonan,N., Wani,S., Xu,Q., Gu,J., Lea,K., Heater,S., Barbacioru,C., Steptoe,A.L., Martin,H.C., Nourbakhsh,E., *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol*, **12**, R126.

34. Ma,H., Wu,Y., Choi,J.-G. and Wu,H. (2013) Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proceedings of the National Academy of Sciences*, **110**, 20687–20692.

35. Wyman,S.K., Knouf,E.C., Parkin,R.K., Fritz,B.R., Lin,D.W., Dennis,L.M., Krouse,M.A., Webster,P.J. and Tewari,M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Research*, **21**, 1450–1461.

36. Park,J.-E., Heo,I., Tian,Y., Simanshu,D.K., Chang,H., Jee,D., Patel,D.J. and Kim,V.N. (2011) Dicer recognizes the 5′ end of RNA for efficient and accurate processing. *Nature*, **475**, 201–205.

37. 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.

38. Sdassi,N., Silveri,L., Laubier,J., Tilly,G., Costa,J., Layani,S., Vilotte,J.-L. and Le Provost,F. (2009) Identification and characterization of new miRNAs cloned from normal mouse mammary gland. *BMC Genomics*, **10**, 149.

39. Chan,Y.-T., Lin,Y.-C., Lin,R.-J., Kuo,H.-H., Thang,W.-C., Chiu,K.-P. and Yu,A.L. (2013) Concordant and Discordant Regulation of Target Genes by miR-31 and Its Isoforms. *PLoS ONE*, **8**, e58169.

40. Salem,O., Erdem,N., Jung,J., Münstermann,E., Wörner,A., Wilhelm,H., Wiemann,S. and Körner,C. (2016) The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics*, **17**, 566.

41. Guo,L., Li,Y., Cirillo,K.M., Marick,R.A., Su,Z., Yin,X., Hua,X., Mills,G.B., Sahni,N. and Yi,S.S. (2021) mi-IsoNet: systems-scale microRNA landscape reveals rampant isoform-mediated gain of target interaction diversity and signaling specificity. *Briefings in Bioinformatics*, 10.1093/bib/bbab091.

42. Kuo,W.-T., Su,M.-W., Lee,Y., Chen,C.-H., Wu,C.-W., Fang,W.-L., Huang,K.-H. and Lin,W. (2015) Bioinformatic Interrogation of 5p-arm and 3p-arm Specific miRNA Expression Using TCGA Datasets. *JCM*, **4**, 1798–1814.

43. Cesarini,V., Silvestris,D.A., Tassinari,V., Tomaselli,S., Alon,S., Eisenberg,E., Locatelli,F. and Gallo,A. (2018) ADAR2/miR-589-3p axis controls glioblastoma cell migration/invasion. *Nucleic Acids Research*, **46**, 2045–2059.

44. Pinto,Y., Buchumenski,I., Levanon,E.Y. and Eisenberg,E. (2018) Human cancer tissues exhibit reduced A-to-I editing of miRNAs coupled with elevated editing of their targets. *Nucleic Acids Research*, **46**, 71–82.

45. Shoshan,E., Mobley,A.K., Braeuer,R.R., Kamiya,T., Huang,L., Vasquez,M.E., Salameh,A., Lee,H.J., Kim,S.J., Ivan,C., *et al.* (2015) Reduced adenosine-to-inosine miR-455-5p editing promotes melanoma growth and metastasis. *Nat Cell Biol*, **17**, 311–321.

46. Wang,Y., Xu,X., Yu,S., Jeong,K.J., Zhou,Z., Han,L., Tsang,Y.H., Li,J., Chen,H., Mangala,L.S., *et al.* (2017) Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.*, **27**, 1112–1125.

47. Nigita,G., Distefano,R., Veneziano,D., Romano,G., Rahman,M., Wang,K., Pass,H., Croce,C.M., Acunzo,M. and Nana-Sinkam,P. (2018) Tissue and exosomal miRNA editing in Non-Small Cell Lung Cancer. *Sci Rep*, **8**, 10222.

48. Loher,P., Londin,E.R. and Rigoutsos,I. (2014) IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget*, **5**.

49. Telonis,A.G., Loher,P., Jing,Y., Londin,E. and Rigoutsos,I. (2015) Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res*, **43**, 9158–9175.

50. Cancer Genome Atlas Research Network, Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.

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

52. Lu,Y., Baras,A.S. and Halushka,M.K. (2018) miRge 2.0 for comprehensive analysis of microRNA sequencing data. *BMC Bioinformatics*, **19**, 275.

53. Chu,A., Robertson,G., Brooks,D., Mungall,A.J., Birol,I., Coope,R., Ma,Y., Jones,S. and Marra,M.A. (2016) Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res*, **44**, e3–e3.

54. Kanke,M., Baran-Gale,J., Villanueva,J. and Sethupathy,P. (2016) miRquant 2.0: an Expanded Tool for Accurate Annotation and Quantification of MicroRNAs and their isomiRs from Small RNA-Sequencing Data. *Journal of Integrative Bioinformatics*, **13**.

55. Loher,P., Karathanasis,N., Londin,E., Bray,P., Pliatsika,V., Telonis,A.G. and Rigoutsos,I. (2021) IsoMiRmap—fast, deterministic, and exhaustive mining of isomiRs from short RNA-seq datasets. *Bioinformatics*, 10.1093/bioinformatics/btab016.

56. Pantano,L., Estivill,X. and Martí,E. (2010) SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Research*, **38**, e34–e34.

57. Alon,S., Mor,E., Vigneault,F., Church,G.M., Locatelli,F., Galeano,F., Gallo,A., Shomron,N. and Eisenberg,E. (2012) Systematic identification of edited microRNAs in the human brain. *Genome Research*, **22**, 1533–1540.

58. Alon,S., Erew,M. and Eisenberg,E. (2015) DREAM: a webserver for the identification of editing sites in mature miRNAs using deep sequencing data. *Bioinformatics*, **31**, 2568–2570.

59. Gadd,S., Huff,V., Walz,A.L., Ooms,A.H.A.G., Armstrong,A.E., Gerhard,D.S., Smith,M.A., Auvil,J.M.G., Meerzaman,D., Chen,Q.-R., *et al.* (2017) A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat Genet*, **49**, 1487–1494.

60. Marceca,G.P., Distefano,R., Tomasello,L., Lagana',A., Russo,F., Calore,F., Romano,G., Bagnoli,M., Gasparini,P., Ferro,A., *et al.* (2020) MiREDiBase: a manually curated database of editing events in microRNAs Bioinformatics.

61. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

62. Smeds,L. and Künstner,A. (2011) ConDeTri - A Content Dependent Read Trimmer for Illumina Data. *PLoS ONE*, **6**, e26314.

63. Fromm,B., Domanska,D., Høye,E., Ovchinnikov,V., Kang,W., Aparicio-Puerta,E., Johansen,M., Flatmark,K., Mathelier,A., Hovig,E., *et al.* (2020) MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Research*, **48**, D132–D141.

64. de Hoon,M.J.L., Taft,R.J., Hashimoto,T., Kanamori-Katayama,M., Kawaji,H., Kawano,M., Kishima,M., Lassmann,T., Faulkner,G.J., Mattick,J.S., *et al.* (2010) Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. *Genome Research*, **20**, 257–264.

65. Jones,M.R., Quinton,L.J., Blahna,M.T., Neilson,J.R., Fu,S., Ivanov,A.R., Wolf,D.A. and Mizgerd,J.P. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol*, **11**, 1157–1163.

66. Katoh,T., Sakaguchi,Y., Miyauchi,K., Suzuki,T., Kashiwabara,S. -i., Baba,T. and Suzuki,T. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & Development*, **23**, 433–438.

67. Robinson,M.D. and Oshlack,A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**, R25.

68. McCarthy,D.J., Chen,Y. and Smyth,G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, **40**, 4288–4297.

69. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

70. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**, R80.

71. McInnes,L., Healy,J. and Melville,J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*.

72. Bray,J.R. and Curtis,J.T. (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, **27**, 325–349.

73. Ester,M., Kriegel,H.-P., Sander,J. and Xu,X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In.pp. 226–231.

74. Rand,W.M. (1971) Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, **66**, 846–850.

75. Vinh,N.X., Epps,J. and Bailey,J. (2009) Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. ACM Press, Montreal, Quebec, Canada, pp. 1–8.

76. Fowlkes,E.B. and Mallows,C.L. (1983) A Method for Comparing Two Hierarchical Clusterings. *Journal of the American Statistical Association*, **78**, 553–569.

77. Mann,H.B. and Whitney,D.R. (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Statist.*, **18**, 50–60.

78. Hochberg,Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800–802.

79. Kaplan,E.L. and Meier,P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457–481.

80. Distefano,R., Nigita,G., Veneziano,D., Romano,G., Croce,C.M. and Acunzo,M. (2019) isoTar: Consensus Target Prediction with Enrichment Analysis for MicroRNAs Harboring Editing Sites and Other Variations. In Laganà,A. (ed), *MicroRNA Target Identification*, Methods in Molecular Biology. Springer New York, New York, NY, Vol. 1970, pp. 211–235.

81. Vejnar,C.E. and Zdobnov,E.M. (2012) miRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic Acids Research*, **40**, 11673–11683.

82. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, **120**, 15–20.

83. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat Genet*, **39**, 1278–1284.

84. Heyne,S., Costa,F., Rose,D. and Backofen,R. (2012) GraphClust: alignment-free structural clustering of local RNA secondary structures. *Bioinformatics*, **28**, i224–i232.

85. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA Targets. *PLoS Biol*, **2**, e363.

86. Delaunay,J. (2003) Prognosis of inv(16)/t(16;16) acute myeloid leukemia (AML): a survey of 110 cases from the French AML Intergroup. *Blood*, **102**, 462–469.

87. Rubnitz,J.E. (2002) Favorable Impact of the t(9;11) in Childhood Acute Myeloid Leukemia. *Journal of Clinical Oncology*, **20**, 2302–2309.

88. Meshinchi,S., Woods,W.G., Stirewalt,D.L., Sweetser,D.A., Buckley,J.D., Tjoa,T.K., Bernstein,I.D. and Radich,J.P. (2001) Prevalence and prognostic significance of Flt3 internal tandem duplication in pediatric acute myeloid leukemia. *Blood*, **97**, 89–94.

89. den Bakker,M.A., Roden,A.C., Marx,A. and Marino,M. (2014) Histologic Classification of Thymoma: A Practical Guide for Routine Cases. *Journal of Thoracic Oncology*, **9**, S125–S130.

90. Li,Z., Qu,Z., Wang,Y., Qin,M. and Zhang,H. (2020) miR-101-3p sensitizes non-small cell lung cancer cells to irradiation. *Open Medicine*, **15**, 413–423.

91. Zhang,X., He,X., Liu,Y., Zhang,H., Chen,H., Guo,S. and Liang,Y. (2017) MiR-101-3p inhibits the growth and metastasis of non-small cell lung cancer through blocking PI3K/AKT signal pathway by targeting MALAT-1. *Biomedicine & Pharmacotherapy*, **93**, 1065–1073.

92. 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.

93. Distefano,R., Nigita,G., Veneziano,D., Romano,G., Croce,C.M. and Acunzo,M. (2019) isoTar: Consensus Target Prediction with Enrichment Analysis for MicroRNAs Harboring Editing Sites and Other Variations. In Laganà,A. (ed), *MicroRNA Target Identification*, Methods in Molecular Biology. Springer New York, New York, NY, Vol. 1970, pp. 211–235.

94. Chakrabarty,A., Tranguch,S., Daikoku,T., Jensen,K., Furneaux,H. and Dey,S.K. (2007) MicroRNA regulation of cyclooxygenase-2 during embryo implantation. *Proceedings of the National Academy of Sciences*, **104**, 15144–15149.

95. Hao,Y., Gu,X., Zhao,Y., Greene,S., Sha,W., Smoot,D.T., Califano,J., Wu,T.-C. and Pang,X. (2011) Enforced Expression of miR-101 Inhibits Prostate Cancer Cell Growth by Modulating the COX-2 Pathway *In Vivo*. *Cancer Prev Res*, **4**, 1073–1083.

96. Harati,R., Mohammad,M.G., Tlili,A., El-Awady,R.A. and Hamoudi,R. (2020) Loss of miR-101-3p Promotes Transmigration of Metastatic Breast Cancer Cells through the Brain Endothelium by Inducing COX-2/MMP1 Signaling. *Pharmaceuticals*, **13**, 144.

97. Li,C.-Y., Pang,Y.-Y., Yang,H., Li,J., Lu,H.-X., Wang,H.-L., Mo,W.-J., Huang,L.-S., Feng,Z.-B. and Chen,G. (2017) Identification of miR-101-3p targets and functional features based on bioinformatics, meta-analysis and experimental verification in hepatocellular carcinoma. *Am J Transl Res*, **9**, 2088–2105.

98. Ma,C., Zheng,C., Bai,E. and Yang,K. (2016) miR-101 inhibits glioma cell invasion via the downregulation of COX-2. *Oncology Letters*, **12**, 2538–2544.

99. Shao,Y., Li,P., Zhu,S., Yue,J., Ji,X., He,Z., Ma,D., Wang,L., Wang,Y., Zong,Y., *et al.* (2015) Cyclooxygenase-2, a Potential Therapeutic Target, Is Regulated by miR-101 in Esophageal Squamous Cell Carcinoma. *PLoS ONE*, **10**, e0140642.

100. Strillacci,A., Griffoni,C., Sansone,P., Paterini,P., Piazzi,G., Lazzarini,G., Spisni,E., Pantaleo,M.A., Biasco,G. and Tomasi,V. (2009) MiR-101 downregulation is involved in cyclooxygenase-2 overexpression in human colon cancer cells. *Experimental Cell Research*, **315**, 1439–1447.

101. Lin,X., Luo,W., Wang,H., Li,R., Huang,Y., Chen,L. and Wu,X. (2019) The Role of Prostaglandin-Endoperoxide Synthase-2 in Chemoresistance of Non-Small Cell Lung Cancer. *Front. Pharmacol.*, **10**, 836.

102. Liu,B., Qu,L. and Yan,S. (2015) Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity. *Cancer Cell Int*, **15**, 106.

103. Cao,Q., Liu,F., Ji,K., Liu,N., He,Y., Zhang,W. and Wang,L. (2017) MicroRNA-381 inhibits the metastasis of gastric cancer by targeting TMEM16A expression. *J Exp Clin Cancer Res*, **36**, 29.

104. Formosa,A., Markert,E.K., Lena,A.M., Italiano,D., Finazzi-Agro',E., Levine,A.J., Bernardini,S., Garabadgiu,A.V., Melino,G. and Candi,E. (2014) MicroRNAs, miR-154, miR-299-5p, miR-376a, miR-376c, miR-377, miR-381, miR-487b, miR-485-3p, miR-495 and miR-654-3p, mapped to the 14q32.31 locus, regulate proliferation, apoptosis, migration and invasion in metastatic prostate cancer cells. *Oncogene*, **33**, 5173–5182.

105. Li,N., He,X., Wei,Y., Wang,Y., Liu,L. and Wang,W. (2016) MiR-381 functions as a tumor suppressor in colorectal cancer by targeting Twist1. *OTT*, 10.2147/OTT.S99228.

106. Liang,Y., Zhao,Q., Fan,L., Zhang,Z., Tan,B., Liu,Y. and Li,Y. (2015) Down-regulation of MicroRNA-381 promotes cell proliferation and invasion in colon cancer through up-regulation of LRH-1. *Biomedicine & Pharmacotherapy*, **75**, 137–141.

107. Qiao,G., Li,J., Wang,J., Wang,Z. and Bian,W. (2019) miR-381 functions as a tumor suppressor by targeting ETS1 in pancreatic cancer. *Int J Mol Med*, 10.3892/ijmm.2019.4206.

108. Rui,X., Gu,T., Pan,H., Shao,S. and Shao,H. (2019) MicroRNA-381 suppresses proliferation and invasion of prostate cancer cells through downregulation of the androgen receptor. *Oncol Lett*, 10.3892/ol.2019.10471.

109. Mi,H., Wang,X., Wang,F., Li,L., Zhu,M., Wang,N., Xiong,Y. and Gu,Y. (2018) miR-381 induces sensitivity of breast cancer cells to doxorubicin by inactivation of MAPK signaling via FYN. *European Journal of Pharmacology*, **839**, 66–75.

110. Mi,H., Wang,X., Wang,F., Li,L., Zhu,M., Wang,N., Xiong,Y. and Gu,Y. (2020) SNHG15 Contributes To Cisplatin Resistance In Breast Cancer Through Sponging miR-381. *OTT*, **Volume 13**, 657–666.

111. Ming,J., Zhou,Y., Du,J., Fan,S., Pan,B., Wang,Y., Fan,L. and Jiang,J. (2015) miR-381 suppresses C/EBPα-dependent Cx43 expression in breast cancer cells. *Bioscience Reports*, **35**, e00266.

112. Mohammadi-Yeganeh,S., Hosseini,V. and Paryan,M. (2019) Wnt pathway targeting reduces triple-negative breast cancer aggressiveness through miRNA regulation in vitro and in vivo. *J Cell Physiol*, **234**, 18317–18328.

113. Wu,M., Fan,B., Guo,Q., Li,Y., Chen,R., Lv,N., Diao,Y. and Luo,Y. (2018) Knockdown of SETDB1 inhibits breast cancer progression by miR-381-3p-related regulation. *Biol Res*, **51**, 39.

114. Xue,Y., Xu,W., Zhao,W., Wang,W., Zhang,D. and Wu,P. (2017) miR-381 inhibited breast cancer cells proliferation, epithelial-to-mesenchymal transition and metastasis by targeting CXCR4. *Biomedicine & Pharmacotherapy*, **86**, 426–433.

115. Yi,D., Xu,L., Wang,R., Lu,X. and Sang,J. (2019) miR-381 overcomes cisplatin resistance in breast cancer by targeting MDR1: miR-381 overcomes cisplatin resistance. *Cell Biol Int*, **43**, 12–21.

116. Kanda,M., Shimizu,D., Tanaka,H., Tanaka,C., Kobayashi,D., Hayashi,M., Takami,H., Niwa,Y., Yamada,S., Fujii,T., *et al.* (2018) Synaptotagmin XIII expression and peritoneal metastasis in gastric cancer. *British Journal of Surgery*, **105**, 1349–1358.

117. Zhang,L., Fan,B., Zheng,Y., Lou,Y., Cui,Y., Wang,K., Zhang,T. and Tan,X. (2020) Identification SYT13 as a novel biomarker in lung adenocarcinoma. *J Cell Biochem*, **121**, 963–973.

118. Wang,Y., Xu,X., Yu,S., Jeong,K.J., Zhou,Z., Han,L., Tsang,Y.H., Li,J., Chen,H., Mangala,L.S., *et al.* (2017) Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.*, **27**, 1112–1125.

119. Katoh,T., Hojo,H. and Suzuki,T. (2015) Destabilization of microRNAs in human cells by 3′ deadenylation mediated by PARN and CUGBP1. *Nucleic Acids Res*, **43**, 7521–7534.

120. Moore,M.J., Scheel,T.K.H., Luna,J.M., Park,C.Y., Fak,J.J., Nishiuchi,E., Rice,C.M. and Darnell,R.B. (2015) miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat Commun*, **6**, 8864.

121. Van der Kwast,R.V.C.T., Woudenberg,T., Quax,P.H.A. and Nossent,A.Y. (2020) MicroRNA-411 and Its 5′-IsomiR Have Distinct Targets and Functions and Are Differentially Regulated in the Vasculature under Ischemia. *Molecular Therapy*, **28**, 157–170.

122. Huang,T., Zhou,Y., Cao,Y., Tao,J., Zhou,Z.-H. and Hang,D.-H. (2017) STK39, overexpressed in osteosarcoma, regulates osteosarcoma cell invasion and proliferation. *Oncology Letters*, **14**, 4599–4604.

123. Li,C., Wang,A., Chen,Y., Liu,Y., Zhang,H. and Zhou,J. (2020) MicroRNA-299-5p inhibits cell metastasis in breast cancer by directly targeting serine/threonine kinase 39. *Oncol Rep*, 10.3892/or.2020.7486.

124. Li,Q., Zhang,S., Hu,M., Xu,M. and Jiang,X. (2019) Silencing of synaptotagmin 13 inhibits tumor growth through suppressing proliferation and promoting apoptosis of colorectal cancer cells. *Int J Mol Med*, 10.3892/ijmm.2019.4412.

125. Li,Z., Zhu,W., Xiong,L., Yu,X., Chen,X. and Lin,Q. (2016) Role of high expression levels of STK39 in the growth, migration and invasion of non-small cell type lung cancer cells. *Oncotarget*, **7**, 61366–61377.

126. Zhang,C., Wang,X., Fang,D., Xu,P., Mo,X., Hu,C., Abdelatty,A., Wang,M., Xu,H., Sun,Q., *et al.* (2021) STK39 is a novel kinase contributing to the progression of hepatocellular carcinoma by the PLK1/ERK signaling pathway. *Theranostics*, **11**, 2108–2122.

127. Zhao,Q., Zhu,Y., Liu,L., Wang,H., Jiang,S., Hu,X. and Guo,J. (2018) STK39 blockage by RNA interference inhibits the proliferation and induces the apoptosis of renal cell carcinoma. *OTT*, **Volume 11**, 1511–1519.

128. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E., *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, **47**, D941–D947.

129. Sherry,S.T. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, **29**, 308–311.

## TABLE AND FIGURES LEGENDS

**Figure 1. Data preprocessing workflow, isomiRs classification, and modification types distribution**

(A-C) In-house data preprocessing workflow (A), examples of annotated isomiRs (B), and distribution of expressed molecules across cohorts and modification types (C). See Supplementary Table 2 for the complete list of A-to-I RNA editing sites employed by the workflow. See Supplementary Table 3 for more detailed information regarding the distribution of modification types of expressed molecules across 5p and 3p arms, the shifting amount at 5'- and 3'-ends, along with the number of molecules affected by SNPs, somatic mutations, and A-to-I RNA editing sites.

**Figure 2. MiRNA isoform-based clustering better delineates clinical-pathological stratification**

(A-B) Clustering benchmarks results related to three different sets (*CAN*, *ISO_wo_SNV*, and *ISO*) of molecules (A) and a comparison between *the three* sets to highlight their ability to separate cohorts' cancer samples (B). Panel A reports quality scores (*Adjusted Rand Index* - ARI, *Adjusted Mutual Information* - AMI, and *Fowlkes-Mallows Index* - FMI) and the number of identified clusters for each set. Panel B compares clustering based on *CAN*, *ISO_wo_SNV*, and *ISO* sets, highlighting cancer sample separation. See Supplementary Figure 2A-B for more detailed information on how we defined the three sets (A) and the designed workflow (B) for data visualization and benchmarking. See Supplementary Figure 3A-C for a complete comparison between *CAN*-, *ISO_wo_SNV*-, and *ISO*-based clustering. See Supplementary Table 5 for more detailed information on the most prominent and significant clinical-pathological features taken into account for clustering-based clinical-pathological analysis.

**Figure 3. MiRNA isoforms dysregulated across cohorts and tissues**

(A-B) Distribution of dysregulated molecules per cohort/comparison and modification type (A), and most prominent modification types (5'- and 3'-end shifting, SNPs/somatic mutations, and A-to-I RNA editing sites) (B). See Supplementary Table 6 for the complete information on dysregulated molecules across cohorts/comparisons, and Supplementary Table 7 for detailed information on all modification types affecting dysregulated molecules (5'- and 3'-end shifting, SNPs/somatic mutations, and A-to-I RNA editing sites).

**Figure 4. MiRNA isoforms experimental gene targeting validation**

(A-E) *miR-101-3p* and *miR-101-3p (-1|-2)* experimental targeting validation in lung cancer cells. Expression of both miR-101-3p miRNA isoforms in normal and tumor samples in TCGA-LUAD cohort (A). Graphical representation of the binding site between the PTGS2 3' UTR and the two miRNA isoforms, miR-101-3p (canonical miRNA) and miR-101-3p (-1|-2) (isomiR) (B). Luciferase assay for psiCHECK-2-PTGS2 3' UTR WT construct co-transfected with mirVana™ miRNA mimics for miR-101-3p, miR-101-3p (-1|-2), and negative scramble miRNA control (SCR) in HEK-293 cells performed 24 hours after the transfection (C). Histograms report the expression of miR-101-3p and miR-101-3p

(-1|-2) in the A549 lung cancer cell line after 48 hours from the transfection with specific mirVana™ miRNA mimics for both miRNA isoforms and the negative scramble miRNA control (D). Western blotting depicts the downregulation of PTGS2 protein in A549 cells after miR-101-3p overexpression (E). Densitometric quantification of western blotting signals from three independent experiments (Figure 4E) was performed using ImageJ (U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997–2018) (E). (F-J) *miR-381-3p* and *miR-381-3p_4_A_G* experimental targeting validation in Triple-Negative (TN) breast cancer cells. Expression of both miR-381-3p miRNA isoforms in normal and breast cancer samples in TCGA-BRCA cohort (F). Graphical representation of STY13 3' UTR binding with miR-381-3p_4_A_G (edited miRNA) (G). Luciferase assay for psiCHECK-2-SYT13 3' UTR WT construct (H) co-transfected with mirVana™ miRNA mimics for miR-381-3p, miR-381-3p_4_A_G, and negative scramble miRNA control (SCR) in HEK-293 cells performed 24 hours after the transfection. qRT-PCR shows the increment of miR-381-3p and miR-381-3p_4_A_G after specific mirVana™ miRNA mimics transfection compared with negative scramble miRNA control (SCR) in MDA-MB-231 breast cancer cells (I). Western blotting represents the downregulation of SYT13 protein (J) in MDA-MB-231 cells after miR-381-3p_4_A_G upregulation via mirVana miRNA mimic transfection. The histogram reports densitometric quantification of western blotting signals from three independent experiments (J), were performed using ImageJ (U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997–2018). Pictures are representative of at least three experiments. The fold of increase in the graphics is the mean values of 3 replicates. $P-value<0.05$ was considered statistically significant. Annotations for * $0.01 \leq p-value<0.05$ , ** $0.001 \leq p-value<0.01$ , and *** $p-value<0.001$ are provided accordingly. Error bars indicate the standard deviation (SD) for the three biological replicates. See Supplementary Table 9 for more details.

**Figure 5. Overall and Relapse Free Survival risk score-based signature**

(A-B) Overview of risk score-based signatures for Overall Survival (A) and Relapse Free Survival (B), supplied with the number of molecules in each signature, the corresponding p-value (Log Rank test), and the area under the curve (AUC) score. See Supplementary Figure 4, Supplementary Figure 5, and Supplementary Table 10 for more detailed information regarding the workflow employed for results generation, the complete list of Kaplan-Meier curves for both OS and RFS signatures, and the list of molecules for each signature, respectively.

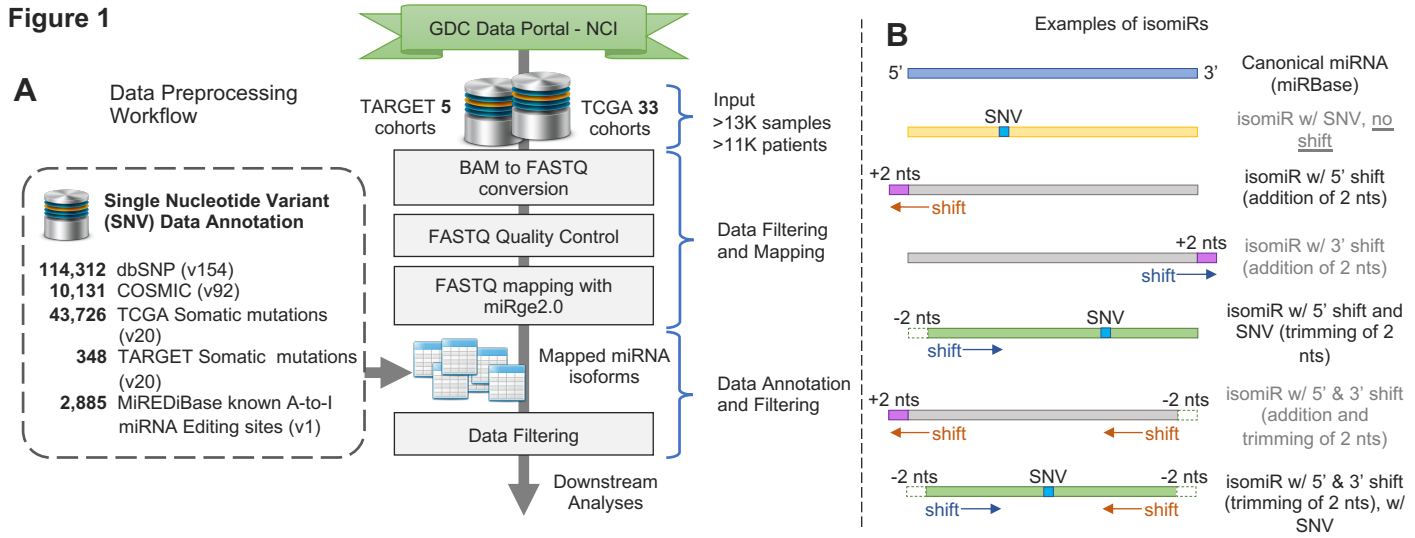
**Table 1. TCGA/TARGET cohorts basic characteristics**

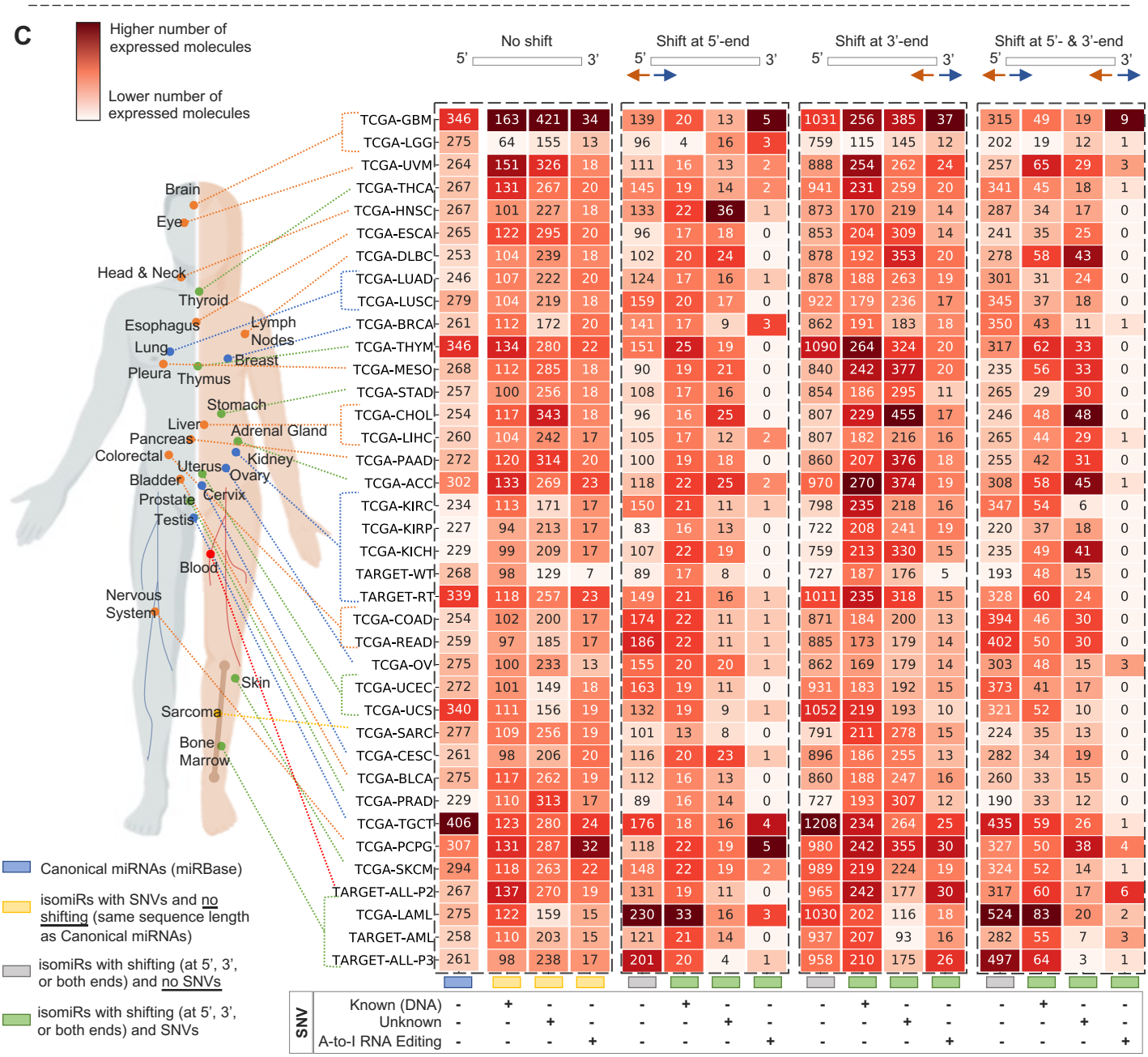
The table reports cohorts' essential characteristics, including the number of cases, age at diagnoses, gender, and race.

**Table 2. TCGA/TARGET cohorts clinical characteristics**

The table shows cohorts' clinical characteristics for both Overall Survival (OS) and Relapse Free Survival (RFS), including the number of cases, stages, number of events/no events (censored).

**Figure 1**




| | No shift | | | | Shift at 5'-end | | | | Shift at 3'-end | | | | Shift at 5'- & 3'-end | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TCGA-GBM | 346 | 163 | 421 | 34 | 139 | 20 | 13 | 5 | 1031 | 256 | 385 | 37 | 315 | 49 | 19 | 9 |
| TCGA-LGG | 275 | 64 | 155 | 13 | 96 | 4 | 16 | 3 | 759 | 115 | 145 | 12 | 202 | 19 | 12 | 1 |
| TCGA-UVM | 264 | 151 | 326 | 18 | 111 | 16 | 13 | 2 | 888 | 254 | 262 | 24 | 257 | 65 | 29 | 3 |
| TCGA-THCA | 267 | 131 | 267 | 20 | 145 | 19 | 14 | 2 | 941 | 231 | 259 | 20 | 341 | 45 | 18 | 1 |
| TCGA-HNSC | 267 | 101 | 227 | 18 | 133 | 22 | 36 | 1 | 873 | 170 | 219 | 14 | 287 | 34 | 17 | 0 |
| TCGA-ESCA | 265 | 122 | 295 | 20 | 96 | 17 | 18 | 0 | 853 | 204 | 309 | 14 | 241 | 35 | 25 | 0 |
| TCGA-DLBC | 253 | 104 | 239 | 18 | 102 | 20 | 24 | 0 | 878 | 192 | 353 | 20 | 278 | 58 | 43 | 0 |
| TCGA-LUAD | 246 | 107 | 222 | 20 | 124 | 17 | 16 | 1 | 878 | 188 | 263 | 19 | 301 | 31 | 24 | 0 |
| TCGA-LUSC | 279 | 104 | 219 | 18 | 159 | 20 | 17 | 0 | 922 | 179 | 236 | 17 | 345 | 37 | 18 | 0 |
| TCGA-BRCA | 261 | 112 | 172 | 20 | 141 | 17 | 9 | 3 | 862 | 191 | 183 | 18 | 350 | 43 | 11 | 1 |
| TCGA-THYM | 346 | 134 | 280 | 22 | 151 | 25 | 19 | 0 | 1090 | 264 | 324 | 20 | 317 | 62 | 33 | 0 |
| TCGA-MESO | 268 | 112 | 285 | 18 | 90 | 19 | 21 | 0 | 840 | 242 | 377 | 20 | 235 | 56 | 33 | 0 |
| TCGA-STAD | 257 | 100 | 256 | 18 | 108 | 17 | 16 | 0 | 854 | 186 | 295 | 11 | 265 | 29 | 30 | 0 |
| TCGA-CHOL | 254 | 117 | 343 | 18 | 96 | 16 | 25 | 0 | 807 | 229 | 455 | 17 | 246 | 48 | 48 | 0 |
| TCGA-LIHC | 260 | 104 | 242 | 17 | 105 | 17 | 12 | 2 | 807 | 182 | 216 | 16 | 265 | 44 | 29 | 1 |
| TCGA-PAAD | 272 | 120 | 314 | 20 | 100 | 19 | 18 | 0 | 860 | 207 | 376 | 18 | 255 | 42 | 31 | 0 |
| TCGA-ACC | 302 | 133 | 269 | 23 | 118 | 22 | 25 | 2 | 970 | 270 | 374 | 19 | 308 | 58 | 45 | 1 |
| TCGA-KIRC | 234 | 113 | 171 | 17 | 150 | 21 | 11 | 1 | 798 | 235 | 218 | 16 | 347 | 54 | 6 | 0 |
| TCGA-KIRP | 227 | 94 | 213 | 17 | 83 | 16 | 13 | 0 | 722 | 208 | 241 | 19 | 220 | 37 | 18 | 0 |
| TCGA-KICH | 229 | 99 | 209 | 17 | 107 | 22 | 19 | 0 | 759 | 213 | 330 | 15 | 235 | 49 | 41 | 0 |
| TARGET-WT | 268 | 98 | 129 | 7 | 89 | 17 | 8 | 0 | 727 | 187 | 176 | 5 | 193 | 48 | 15 | 0 |
| TARGET-RT | 339 | 118 | 257 | 23 | 149 | 21 | 16 | 1 | 1011 | 235 | 318 | 15 | 328 | 60 | 24 | 0 |
| TCGA-COAD | 254 | 102 | 200 | 17 | 174 | 22 | 11 | 1 | 871 | 184 | 200 | 13 | 394 | 46 | 30 | 0 |
| TCGA-READ | 259 | 97 | 185 | 17 | 186 | 22 | 11 | 1 | 885 | 173 | 179 | 14 | 402 | 50 | 30 | 0 |
| TCGA-OV | 275 | 100 | 233 | 13 | 155 | 20 | 20 | 1 | 862 | 169 | 179 | 14 | 303 | 48 | 15 | 3 |
| TCGA-UCEC | 272 | 101 | 149 | 18 | 163 | 19 | 11 | 0 | 931 | 183 | 192 | 15 | 373 | 41 | 17 | 0 |
| TCGA-UCS | 340 | 111 | 156 | 19 | 132 | 19 | 9 | 1 | 1052 | 219 | 193 | 10 | 321 | 52 | 10 | 0 |
| TCGA-SARC | 277 | 109 | 256 | 19 | 101 | 13 | 8 | 0 | 791 | 211 | 278 | 15 | 224 | 35 | 13 | 0 |
| TCGA-CESC | 261 | 98 | 206 | 20 | 116 | 20 | 23 | 1 | 896 | 186 | 255 | 13 | 282 | 34 | 19 | 0 |
| TCGA-BLCA | 275 | 117 | 262 | 19 | 112 | 16 | 13 | 0 | 860 | 188 | 247 | 16 | 260 | 33 | 15 | 0 |
| TCGA-PRAD | 229 | 110 | 313 | 17 | 89 | 16 | 14 | 0 | 727 | 193 | 307 | 12 | 190 | 33 | 12 | 0 |
| TCGA-TGCT | 406 | 123 | 280 | 24 | 176 | 18 | 16 | 4 | 1208 | 234 | 264 | 25 | 435 | 59 | 26 | 1 |
| TCGA-PCPG | 307 | 131 | 287 | 32 | 118 | 22 | 19 | 5 | 980 | 242 | 355 | 30 | 327 | 50 | 38 | 4 |
| TCGA-SKCM | 294 | 118 | 263 | 22 | 148 | 22 | 19 | 2 | 989 | 219 | 224 | 19 | 324 | 52 | 14 | 1 |
| TARGET-ALL-P2 | 267 | 137 | 270 | 19 | 131 | 19 | 11 | 0 | 965 | 242 | 177 | 30 | 317 | 60 | 17 | 6 |
| TCGA-LAML | 275 | 122 | 159 | 15 | 230 | 33 | 16 | 3 | 1030 | 202 | 116 | 18 | 524 | 83 | 20 | 2 |
| TARGET-AML | 258 | 110 | 203 | 15 | 121 | 21 | 14 | 0 | 937 | 207 | 93 | 16 | 282 | 55 | 7 | 3 |
| TARGET-ALL-P3 | 261 | 98 | 238 | 17 | 201 | 20 | 4 | 1 | 958 | 210 | 175 | 26 | 497 | 64 | 3 | 1 |

| SNV | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Known (DNA) | - | + | - | - | - | + | - | - | - | + | - | - | - | + | - | - |
| Unknown | - | - | + | - | - | - | + | - | - | - | + | - | - | - | + | - |
| A-to-I RNA Editing | - | - | - | + | - | - | - | + | - | - | - | + | - | - | - | + |

Figure 2

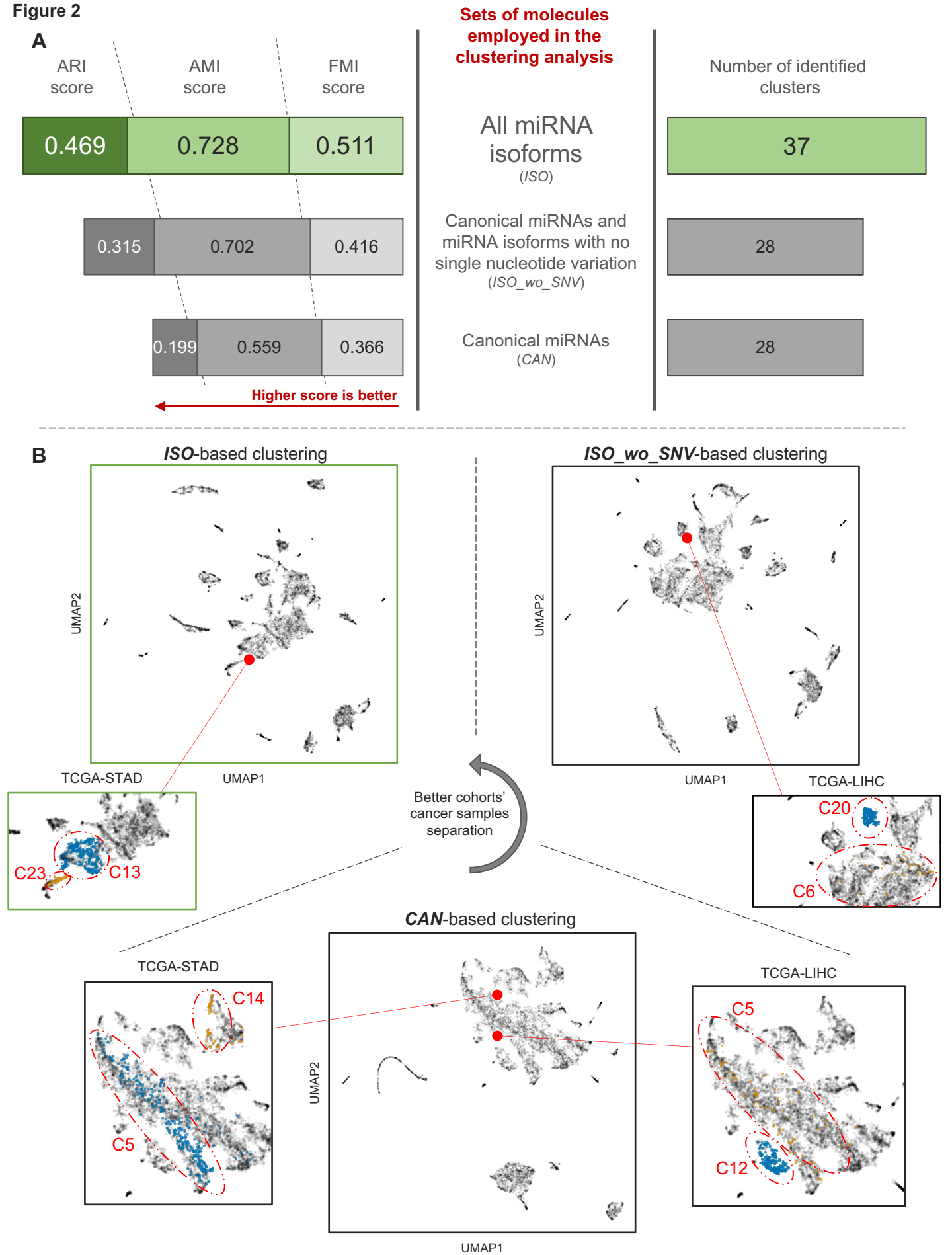

A
ARI score
AMI score
FMI score
Sets of molecules employed in the clustering analysis
Number of identified clusters
0.469
0.728
0.511
All miRNA isoforms
(ISO)
37
0.315
0.702
0.416
Canonical miRNAs and miRNA isoforms with no single nucleotide variation
(ISO_wo_SNV)
28
0.199
0.559
0.366
Canonical miRNAs
(CAN)
28
Higher score is better
B
ISO-based clustering
ISO_wo_SNV-based clustering
UMAP2
UMAP1
TCGA-STAD
C23
C13
TCGA-LIHC
C20
C6
Better cohorts' cancer samples separation
CAN-based clustering
TCGA-STAD
C14
C5
TCGA-LIHC
C5
C12

**Figure 3**

**A**

**M:** Metastatic
**N:** Solid Tissue Normal
**RT:** Recurrent Solid Tumor
**T:** Primary Solid Tumor

Canonical miRNAs (miRBase)

isomiRs with SNVs and no shifting (same sequence length as Canonical miRNAs)

isomiRs with shifting (at 5', 3', or both ends) and no SNVs

isomiRs with shifting (at 5', 3', or both ends) and SNVs

**SNV** = Single Nucleotide Variant (i.e., SNP, somatic mutation, A-to-I RNA Editing)

| | | No shift (5' – 3') | | | | Shift at 5'-end (5' ← → 3') | | | | Shift at 3'-end (5' ← → 3') | | | | Shift at 5'- & 3'-end (5' ← → 3') | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T vs. N** | TCGA-THCA (T vs. N) | 100 | 70 | 130 | 11 | 61 | 10 | 10 | 1 | 383 | 111 | 137 | 14 | 127 | 19 | 18 | 0 |
| | TCGA-HNSC (T vs. N) | 159 | 72 | 140 | 11 | 72 | 13 | 17 | 0 | 505 | 111 | 163 | 12 | 188 | 25 | 15 | 0 |
| | TCGA-ESCA (T vs. N) | 107 | 51 | 108 | 7 | 42 | 11 | 5 | 0 | 380 | 105 | 137 | 7 | 118 | 20 | 2 | 0 |
| | TCGA-LUAD (T vs. N) | 145 | 68 | 177 | 15 | 74 | 12 | 12 | 0 | 523 | 135 | 218 | 14 | 198 | 22 | 23 | 0 |
| | TCGA-LUSC (T vs. N) | 167 | 58 | 114 | 13 | 118 | 14 | 11 | 0 | 582 | 113 | 152 | 14 | 244 | 24 | 10 | 0 |
| | TCGA-BRCA (T vs. N) | 155 | 72 | 108 | 14 | 85 | 10 | 4 | 3 | 508 | 131 | 143 | 17 | 237 | 28 | 9 | 1 |
| | TCGA-STAD (T vs. N) | 136 | 66 | 152 | 11 | 67 | 10 | 6 | 0 | 467 | 116 | 183 | 9 | 153 | 19 | 15 | 0 |
| | TCGA-CHOL (T vs. N) | 112 | 76 | 240 | 14 | 50 | 12 | 15 | 0 | 448 | 150 | 311 | 14 | 137 | 32 | 24 | 0 |
| | TCGA-LIHC (T vs. N) | 143 | 62 | 145 | 11 | 61 | 12 | 7 | 2 | 436 | 110 | 137 | 13 | 165 | 25 | 20 | 1 |
| | TCGA-KIRC (T vs. N) | 125 | 60 | 74 | 9 | 91 | 13 | 8 | 1 | 476 | 134 | 143 | 13 | 209 | 27 | 4 | 0 |
| | TCGA-KIRP (T vs. N) | 117 | 51 | 104 | 7 | 44 | 9 | 6 | 0 | 354 | 113 | 141 | 13 | 116 | 16 | 7 | 0 |
| | TCGA-KICH (T vs. N) | 109 | 57 | 152 | 9 | 53 | 7 | 11 | 0 | 406 | 120 | 222 | 7 | 141 | 22 | 31 | 0 |
| | TARGET-WT (T vs. N) | 166 | 66 | 91 | 5 | 52 | 11 | 4 | 0 | 504 | 114 | 106 | 4 | 110 | 34 | 10 | 0 |
| | TARGET-RT (T vs. N) | 215 | 66 | 165 | 11 | 100 | 17 | 11 | 0 | 613 | 152 | 210 | 8 | 225 | 43 | 10 | 0 |
| | TCGA-COAD (T vs. N) | 211 | 76 | 173 | 15 | 147 | 11 | 6 | 1 | 717 | 154 | 181 | 12 | 341 | 39 | 27 | 0 |
| | TCGA-UCEC (T vs. N) | 165 | 54 | 58 | 6 | 122 | 12 | 8 | 0 | 625 | 142 | 133 | 8 | 299 | 28 | 12 | 0 |
| | TCGA-BLCA (T vs. N) | 172 | 68 | 157 | 14 | 77 | 10 | 6 | 0 | 523 | 127 | 161 | 14 | 182 | 22 | 11 | 0 |
| | TCGA-PRAD (T vs. N) | 106 | 67 | 218 | 8 | 48 | 9 | 8 | 0 | 421 | 123 | 225 | 7 | 123 | 17 | 7 | 0 |
| **M vs. N/T** | TCGA-THCA (M vs. N) | 73 | 52 | 81 | 8 | 35 | 8 | 4 | 0 | 241 | 74 | 85 | 7 | 67 | 16 | 12 | 1 |
| | TCGA-BRCA (M vs. N) | 125 | 53 | 57 | 12 | 53 | 13 | 5 | 0 | 326 | 122 | 122 | 9 | 160 | 35 | 9 | 1 |
| | TCGA-SKCM (M vs. T) | 38 | 22 | 69 | 4 | 38 | 1 | 3 | 0 | 193 | 39 | 39 | 1 | 84 | 11 | 3 | 1 |
| **RT vs. N/T** | TARGET-WT (RT vs. N) | 110 | 45 | 115 | 3 | 43 | 7 | 3 | 1 | 387 | 95 | 113 | 1 | 101 | 22 | 13 | 0 |
| | TCGA-LGG (RT vs. T) | 8 | 3 | 2 | 0 | 3 | 0 | 0 | 0 | 21 | 4 | 4 | 0 | 4 | 2 | 1 | 0 |
| **SNV** | Known (DNA) | - | + | - | - | - | + | - | - | - | + | - | - | - | + | - | - |
| | Unknown | - | - | + | - | - | - | + | - | - | - | + | - | - | - | + | - |
| | A-to-I RNA Editing | - | - | - | + | - | - | - | + | - | - | - | + | - | - | - | + |

Higher number of dysregulated molecules

Lower number of dysregulated molecules

**B**

**T:** Primary Solid Tumor

Lower number of dysregulated molecules — % — Higher number of dysregulated molecules

**UP:** Upregulated
**DOWN:** Downregulated

## Nucleotide shifting distribution at 5'- and 3'-end

(values displayed in percentage)

| | 5'-end | | | | | | | | | | | | 3'-end | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| T (%) | 0 | 1 | 5 | 80 | 10 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 7 | 11 | 24 | 32 | 11 | 4 | 1 | 0 | 0 | 0 | 0 | UP |
| | 0 | 2 | 6 | 80 | 11 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 5 | 5 | 5 | 9 | 15 | 29 | 17 | 7 | 3 | 1 | 0 | 0 | 0 | DOWN |

← Shifting →  ← Shifting →

## Top 5 known (DNA) SNVs distribution

(values displayed in percentage)

| | | Sequence nucleotides | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref-Mod | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| T (%) | G-T | 0 | 6 | 10 | 28 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 6 | 9 | 15 | 8 | 5 | 1 | 0 | 1 | 1 | |
| | C-T | 4 | 3 | 2 | 9 | 3 | 1 | 1 | 0 | 3 | 2 | 0 | 2 | 6 | 4 | 1 | 1 | 1 | 1 | 2 | 3 | 8 | 16 | 19 | 4 | 2 | 2 | 0 | 1 | 0 | |
| | G-A | 3 | 1 | 2 | 0 | 0 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 1 | 2 | 4 | 5 | 0 | 5 | 15 | 30 | 11 | 3 | 1 | 0 | 1 | 0 | UP |
| | C-A | 13 | 1 | 9 | 8 | 3 | 0 | 0 | 1 | 3 | 0 | 0 | 1 | 2 | 3 | 0 | 1 | 1 | 2 | 5 | 1 | 15 | 7 | 21 | 3 | 0 | 1 | 0 | 0 | 0 | |
| | A-G | 0 | 4 | 2 | 2 | 4 | 4 | 2 | 4 | 9 | 2 | 4 | 0 | 4 | 0 | 4 | 2 | 0 | 0 | 0 | 0 | 2 | 18 | 22 | 7 | 0 | 0 | 0 | 0 | 0 | |
| | C-T | 2 | 0 | 0 | 0 | 0 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 | 4 | 3 | 1 | 2 | 1 | 8 | 0 | 4 | 18 | 32 | 9 | 1 | 0 | 0 | 0 | 0 | |
| | C-A | 3 | 3 | 9 | 4 | 3 | 0 | 0 | 0 | 2 | 4 | 2 | 3 | 11 | 5 | 1 | 1 | 2 | 2 | 3 | 3 | 8 | 15 | 14 | 0 | 1 | 2 | 0 | 0 | 0 | |
| | G-T | 6 | 0 | 2 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 1 | 1 | 4 | 1 | 33 | 10 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | DOWN |
| | C-A | 0 | 13 | 6 | 11 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 7 | 13 | 22 | 13 | 1 | 0 | 0 | 0 | 0 | |
| | G-A | 0 | 2 | 2 | 2 | 0 | 0 | 7 | 2 | 20 | 2 | 5 | 2 | 2 | 0 | 5 | 0 | 0 | 0 | 0 | 2 | 10 | 17 | 15 | 2 | 0 | 0 | 0 | 0 | 0 | |

## A-to-I RNA sites distribution

(values displayed in percentage)

| | | Sequence nucleotides | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref-Mod | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | |
| T (%) | A-G | 5 | 10 | 8 | 11 | 15 | 6 | 5 | 0 | 2 | 8 | 0 | 6 | 0 | 0 | 2 | 8 | 3 | 3 | 0 | 0 | 2 | 3 | 0 | 3 | UP |
| | | 11 | 11 | 8 | 14 | 6 | 3 | 0 | 0 | 3 | 6 | 0 | 6 | 0 | 0 | 3 | 8 | 3 | 3 | 0 | 8 | 6 | 3 | 0 | 0 | DOWN |

Figure 4

Shifted isomiR

Edited miRNA

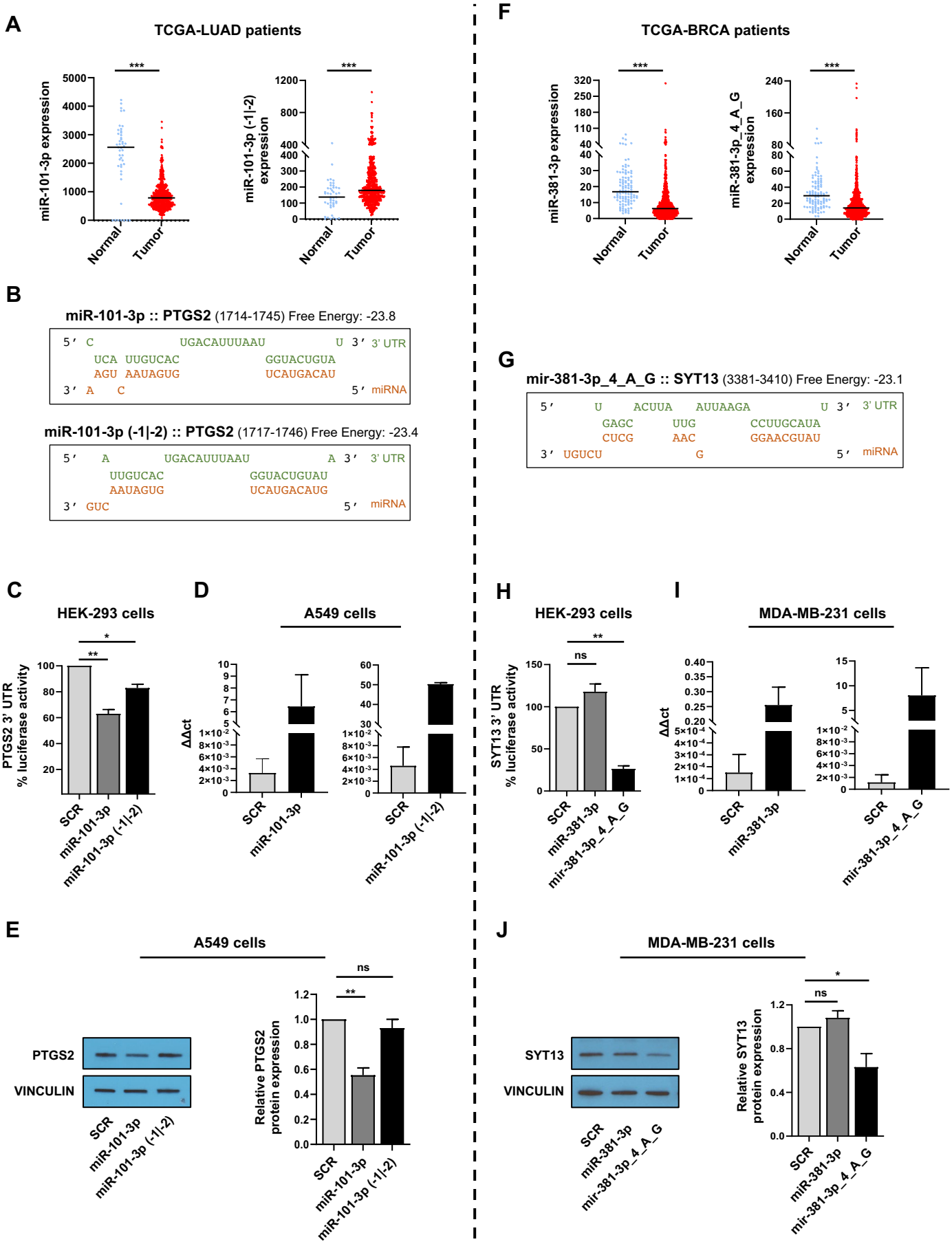
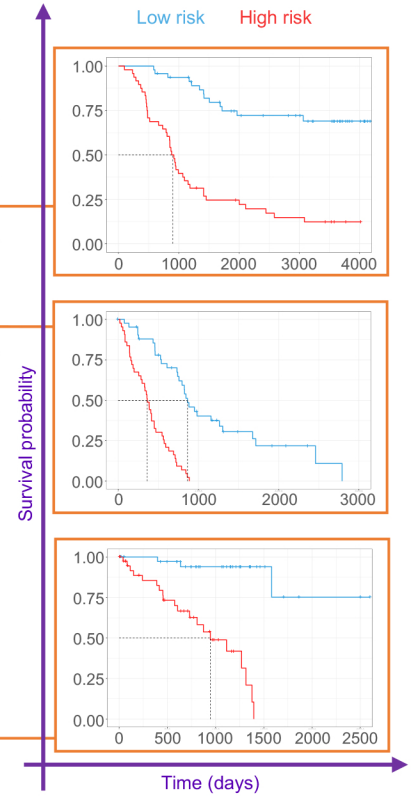


Figure 5

A

## Overall Survival
(12 signatures)

| Signature | Canonical miRNAs | miRNA isoforms | P-value (Log Rank Test) | Area Under the Curve |
|---|---|---|---|---|
| TCGA-LIHC (T) | 1 | 16 | 1.79e-11 | 0.70 |
| TCGA-LGG (T) | 6 | 11 | 3.44e-16 | 0.71 |
| TARGET-ALL-P2 (PBDB) | 1 | 10 | 3.16e-10 | 0.79 |
| TCGA-LAML (PBDP) | | 9 | 8.31e-12 | 0.72 |
| TCGA-MESO (T) | | 9 | 3.56e-10 | 0.79 |
| TCGA-ACC (T) | 1 | 7 | 7.04e-07 | 0.78 |
| TARGET-ALL-P2 (RBDB) | | 8 | 1.69e-07 | 0.85 |
| TARGET-AML (PBDP) | 1 | 5 | 2.47e-07 | 0.71 |
| TCGA-PRAD (T) | 1 | 3 | 1.21e-02 | 0.70 |
| TCGA-THCA (T) | | 4 | 2.21e-05 | 0.76 |
| TARGET-ALL-P2 (PBDP) | | 4 | 1.71e-06 | 0.87 |
| TCGA-UVM (T) | | 2 | 1.36e-07 | 0.76 |




**T**: Primary Solid Tumor
**RBDB**: Recurrent Blood Derived Cancer – Bone Marrow
**PBDP**: Primary Blood Derived Cancer – Peripheral blood
**PBDB**: Primary Blood Derived Cancer – Bone Marrow

B

## Relapse-Free Survival
(9 signatures)

| Signature | Canonical miRNAs | miRNA isoforms | P-value (Log Rank Test) | Area Under the Curve |
|---|---|---|---|---|
| TCGA-HNSC (T) | 2 | 10 | 7.16e-12 | 0.73 |
| TCGA-STAD (T) | | 5 | 8.83e-06 | 0.70 |
| TCGA-ESCA (T) | 1 | 2 | 1.52e-04 | 0.72 |
| TCGA-KIRP (T) | 1 | 2 | 8.87e-05 | 0.78 |
| TCGA-PCPG (T) | | 2 | 3.65e-04 | 0.74 |
| TCGA-KICH (T) | | 1 | 9.27e-04 | 0.80 |
| TCGA-UVM (T) | | 1 | 4.44e-06 | 0.78 |
| TCGA-SKCM (T) | | 1 | 3.00e-05 | 0.77 |
| TCGA-ACC (T) | | 1 | 1.55e-05 | 0.71 |

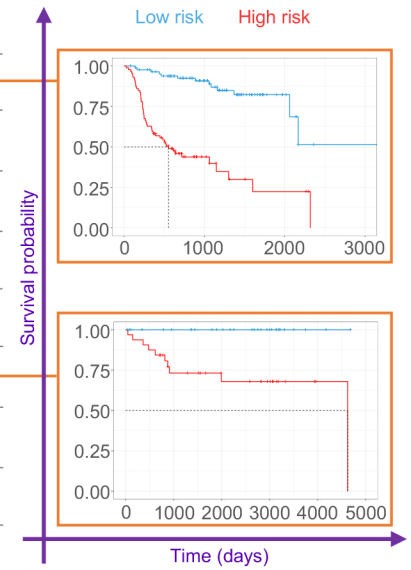


**Table 1. TCGA/TARGET cohorts basic characteristics**

| Cohort | Cancer type | No. of Cases | Age at diagnoses (Mean ± SD /NA) | Gender (M/F/NA) | Race (White/AA/Other/NA) |
|---|---|---|---|---|---|
| TARGET-ALL-P2 | Acute Lymphoblastic Leukemia - Phase II | 191 | 6.8 ± 5.3 /0 | 101/90/0 | 133/20/7/31 |
| TARGET-ALL-P3 | Acute Lymphoblastic Leukemia - Phase III | 38 | 8.4 ± 5.3 /0 | 21/17/0 | 1/1/0/36 |
| TARGET-AML | Acute Myeloid Leukemia | 701 | 9.1 ± 6.1 /22 | 344/335/22 | 502/77/33/89 |
| TARGET-RT | Rhabdoid tumors | 66 | 1.2 ± 2.2 /0 | 35/31/0 | 49/8/0/9 |
| TARGET-WT | Wilms tumor | 127 | 4.1 ± 2.8 /0 | 53/74/0 | 95/18/0/14 |
| TCGA-ACC | Adrenocortical carcinoma | 80 | 46.4 ± 15.9 /0 | 31/49/0 | 67/1/1/11 |
| TCGA-BLCA | Bladder Urothelial Carcinoma | 409 | 68.0 ± 10.6 /1 | 302/107/0 | 324/23/44/18 |
| TCGA-BRCA | Breast invasive carcinoma | 1079 | 58.6 ± 13.2 /17 | 12/1066/1 | 746/182/62/89 |
| TCGA-CESC | Cervical squamous cell carcinoma and endocervical adenocarcinoma | 307 | 48.2 ± 13.8 /2 | 0/307/0 | 211/30/30/36 |
| TCGA-CHOL | Cholangiocarcinoma | 36 | 63.0 ± 12.8 /0 | 16/20/0 | 31/2/3/0 |
| TCGA-COAD | Colon adenocarcinoma | 444 | 66.8 ± 13.1 /4 | 231/211/2 | 213/59/12/160 |
| TCGA-DLBC | Lymphoid Neoplasm Diffuse Large B-cell Lymphoma | 47 | 56.3 ± 14.1 /0 | 22/25/0 | 28/1/18/0 |
| TCGA-ESCA | Esophageal carcinoma | 184 | 62.5 ± 11.9 /0 | 157/27/0 | 114/5/45/20 |
| TCGA-GBM | Glioblastoma multiforme | 5 | 0 ± 0 /5 | 0/0/5 | 0/0/0/5 |
| TCGA-HNSC | Head and Neck squamous cell carcinoma | 524 | 60.9 ± 11.9 /1 | 383/141/0 | 449/47/13/15 |
| TCGA-KICH | Kidney Chromophobe | 66 | 51.5 ± 14.3 /0 | 39/27/0 | 58/4/2/2 |
| TCGA-KIRC | Kidney renal clear cell carcinoma | 516 | 60.5 ± 12.1 /0 | 335/181/0 | 445/56/8/7 |
| TCGA-KIRP | Kidney renal papillary cell carcinoma | 291 | 61.5 ± 12.1 /5 | 214/77/0 | 207/61/8/15 |
| TCGA-LAML | Acute Myeloid Leukemia | 188 | 54.9 ± 16.2 /0 | 101/87/0 | 171/13/2/2 |
| TCGA-LGG | Brain Lower Grade Glioma | 512 | 43.0 ± 13.4 /2 | 281/230/1 | 471/21/9/11 |
| TCGA-LIHC | Liver hepatocellular carcinoma | 373 | 59.3 ± 13.4 /4 | 254/119/0 | 183/17/163/10 |
| TCGA-LUAD | Lung adenocarcinoma | 513 | 65.3 ± 9.9 /30 | 239/274/0 | 387/52/8/66 |
| TCGA-LUSC | Lung squamous cell carcinoma | 478 | 67.4 ± 8.6 /9 | 354/124/0 | 333/30/9/106 |
| TCGA-MESO | Mesothelioma | 87 | 63.0 ± 9.8 /0 | 71/16/0 | 85/1/1/0 |
| TCGA-OV | Ovarian serous cystadenocarcinoma | 489 | 59.9 ± 11.5 /11 | 0/486/3 | 422/32/18/17 |
| TCGA-PAAD | Pancreatic adenocarcinoma | 178 | 64.6 ± 10.9 /0 | 98/80/0 | 157/6/11/4 |
| TCGA-PCPG | Pheochromocytoma and Paraganglioma | 179 | 47.3 ± 15.1 /0 | 78/101/0 | 148/20/7/4 |
| TCGA-PRAD | Prostate adenocarcinoma | 494 | 61.0 ± 6.8 /11 | 494/0/0 | 146/7/2/339 |
| TCGA-READ | Rectum adenocarcinoma | 161 | 64.2 ± 11.8 /1 | 86/74/1 | 81/6/1/73 |
| TCGA-SARC | Sarcoma | 259 | 60.8 ± 14.7 /1 | 119/140/0 | 227/18/6/8 |
| TCGA-SKCM | Skin Cutaneous Melanoma | 448 | 58.1 ± 15.6 /8 | 276/172/0 | 425/1/12/10 |
| TCGA-STAD | Stomach adenocarcinoma | 436 | 65.7 ± 10.7 /9 | 281/155/0 | 273/13/88/62 |

| | | | | | |
|---|---|---|---|---|---|
| TCGA-TGCT | Testicular Germ Cell Tumors | 150 | 32.0 ± 9.3 /16 | 134/0/16 | 119/6/4/21 |
| TCGA-THCA | Thymoma | 506 | 47.3 ± 15.8 /0 | 136/370/0 | 334/27/53/92 |
| TCGA-THYM | Thyroid carcinoma | 124 | 58.2 ± 13.0 /1 | 64/60/0 | 103/6/13/2 |
| TCGA-UCEC | Uterine Carcinosarcoma | 550 | 63.9 ± 11.2 /15 | 0/539/11 | 367/107/33/43 |
| TCGA-UCS | Uterine Corpus Endometrial Carcinoma | 57 | 69.7 ± 9.2 /0 | 0/57/0 | 44/9/3/1 |
| TCGA-UVM | Uveal Melanoma | 80 | 61.6 ± 13.9 /0 | 45/35/0 | 55/0/0/25 |

List of cohorts and their essential characteristics, including the number of cases, age at diagnoses, gender, and race (White, African American (AA), others).

**Table 2. TCGA/TARGET cohorts clinical characteristics**

| | | | Overall Survival (OS) | | | Relapse-Free Survival (RFS) | | |
|---|---|---|---|---|---|---|---|---|
| Cohort | # Cases | Stages (I/II/III/IV/NA) | # events | # censored | Median Follow-UP (Months) | # events | # censored | Median Follow-UP (Months) |
| TARGET-ALL-P2 | 191 | 0/0/0/0/191 | 80 | 68 | 56.05 | 0 | 0 | 0.00 |
| TARGET-ALL-P3 | 38 | 0/0/0/0/38 | 12 | 17 | 29.33 | 0 | 0 | 0.00 |
| TARGET-AML | 701 | 0/0/0/0/701 | 265 | 413 | 59.30 | 0 | 0 | 0.00 |
| TARGET-RT | 66 | 2/13/27/0/24 | 32 | 26 | 7.98 | 0 | 0 | 0.00 |
| TARGET-WT | 127 | 17/49/41/14/6 | 52 | 75 | 54.27 | 0 | 0 | 0.00 |
| TCGA-ACC | 80 | 9/37/16/16/2 | 29 | 51 | 39.42 | 36 | 39 | 27.40 |
| TCGA-BLCA | 409 | 2/131/139/135/2 | 179 | 229 | 17.87 | 79 | 232 | 16.50 |
| TCGA-BRCA | 1079 | 182/609/244/20/24 | 149 | 929 | 27.57 | 31 | 363 | 33.27 |
| TCGA-CESC | 307 | 163/70/46/21/7 | 71 | 236 | 21.27 | 26 | 172 | 21.12 |
| TCGA-CHOL | 36 | 19/9/1/7/0 | 18 | 18 | 21.50 | 17 | 13 | 10.72 |
| TCGA-COAD | 444 | 73/168/125/65/13 | 101 | 340 | 22.30 | 30 | 1 | 21.80 |
| TCGA-DLBC | 47 | 7/17/5/12/6 | 9 | 38 | 26.37 | 6 | 21 | 26.37 |
| TCGA-ESCA | 184 | 18/82/62/16/6 | 77 | 107 | 13.35 | 43 | 126 | 12.60 |
| TCGA-GBM | 5 | 0/0/0/0/5 | 0 | 0 | 0.00 | 0 | 0 | 0.00 |
| TCGA-HNSC | 524 | 27/85/92/320/0 | 223 | 300 | 21.50 | 64 | 107 | 23.77 |
| TCGA-KICH | 66 | 21/25/14/6/0 | 9 | 56 | 74.93 | 10 | 54 | 73.67 |
| TCGA-KIRC | 516 | 253/55/123/82/3 | 172 | 344 | 39.38 | 2 | 32 | 13.60 |
| TCGA-KIRP | 291 | 180/25/52/16/18 | 44 | 246 | 25.62 | 18 | 140 | 20.40 |
| TCGA-LAML | 188 | 0/0/0/0/188 | 114 | 63 | 12.17 | 0 | 0 | 0.00 |
| TCGA-LGG | 512 | 0/0/0/0/512 | 124 | 385 | 22.60 | 65 | 186 | 20.13 |
| TCGA-LIHC | 373 | 173/86/85/5/24 | 129 | 243 | 19.83 | 102 | 174 | 13.78 |
| TCGA-LUAD | 513 | 277/121/84/24/7 | 182 | 322 | 21.88 | 31 | 156 | 18.40 |
| TCGA-LUSC | 478 | 230/158/80/6/4 | 199 | 273 | 22.25 | 30 | 139 | 17.00 |
| TCGA-MESO | 87 | 10/16/45/16/0 | 73 | 13 | 17.10 | 48 | 32 | 11.80 |
| TCGA-OV | 489 | 1/27/374/80/7 | 308 | 177 | 34.87 | 60 | 0 | 18.43 |
| TCGA-PAAD | 178 | 21/147/3/4/3 | 93 | 85 | 15.48 | 56 | 106 | 13.08 |
| TCGA-PCPG | 179 | 0/0/0/0/179 | 6 | 173 | 25.17 | 15 | 163 | 23.48 |
| TCGA-PRAD | 494 | 0/0/0/0/494 | 10 | 484 | 30.80 | 60 | 44 | 24.87 |
| TCGA-READ | 161 | 29/48/50/24/10 | 26 | 134 | 20.58 | 10 | 0 | 27.15 |
| TCGA-SARC | 259 | 0/0/0/0/259 | 98 | 161 | 31.57 | 91 | 141 | 22.52 |
| TCGA-SKCM | 448 | 74/128/164/23/59 | 210 | 229 | 37.47 | 221 | 215 | 27.58 |
| TCGA-STAD | 436 | 58/128/180/43/27 | 168 | 263 | 14.07 | 45 | 177 | 13.42 |
| TCGA-TGCT | 150 | 59/14/14/0/63 | 4 | 130 | 42.03 | 30 | 99 | 28.80 |
| TCGA-THCA | 506 | 284/52/113/55/2 | 16 | 490 | 31.50 | 25 | 355 | 33.98 |
| TCGA-THYM | 124 | 0/0/0/0/124 | 9 | 114 | 41.77 | 16 | 107 | 38.13 |
| TCGA-UCEC | 550 | 339/49/123/28/11 | 88 | 450 | 30.32 | 28 | 165 | 34.47 |
| TCGA-UCS | 57 | 22/5/20/10/0 | 35 | 22 | 20.37 | 29 | 25 | 12.97 |
| TCGA-UVM | 80 | 0/39/37/4/0 | 23 | 57 | 26.13 | 17 | 62 | 22.33 |

Cohorts' clinical characteristics, including the number of cases, stage, number of events/no events (censored) for Overall and Relapse Free Survival.

# A concurrent canonical and modified miRNAome pan-cancer study on TCGA and TARGET cohorts leads to an enhanced resolution in cancer

Rosario Distefano[1,*], Luisa Tomasello[1], Gian Luca Rampioni Vinciguerra[1,2], Pierluigi Gasparini[1,3,4], Yujia Xiang[5], Marina Bagnoli[6], Gioacchino Paolo Marceca[7], Paolo Fadda[8], Alessandro Laganà[9], Mario Acunzo[10], Qin Ma[5], Giovanni Nigita[1,*,†], and Carlo M. Croce[1,*,†]

[1] Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210, USA.
[2] Faculty of Medicine and Psychology, Department of Clinical and Molecular Medicine, University of Rome "Sapienza," Santo Andrea Hospital, Rome, RM, 00189, Italy.
[3] School of Biomedical Sciences and Pharmacy, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, NSW 2308, Australia.
[4] Hunter Medical Research Institute, New Lambton Heights, NSW 2305, Australia.
[5] Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, 43210, USA.
[6] Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, MI, 20133, Italy.
[7] Department of Clinical and Experimental Medicine, University of Catania, Catania, CT, 95100, Italy.
[8] Genomics Shared Resource, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210, USA.
[9] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.
[10] Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, 23298, USA.

* To whom correspondence should be addressed. Tel: +1 614 292 4930; Fax: +1 614 292 3558; Email: Carlo.Croce@osumc.edu
Correspondence may also be addressed to Rosario Distefano. Tel.: +1 614 292 7278; E-mail: Rosario.Distefano@osumc.edu
Correspondence may also be addressed to Giovanni Nigita. Tel.: +1 614 688 1805; E-mail: Giovanni.Nigita @osumc.edu
† These authors contributed equally to this work.

## SUPPLEMENTARY FIGURE AND TABLE LEGENDS

**Supplementary Figure 1. Significant pathways enriched in at least five cohorts/cancer tissues**

The figure shows significant ($|z-score| \geq 2$ and $p-value < 0.01$) pathways enriched in at least five cohorts/cancer tissues. Pathways and cohorts/cancer tissues are clustered according to the Bray-Curtis distance metric and Complete Linkage method.

**Supplementary Figure 2. Sample visualization and clustering workflow**

(A-B) The figure shows how we defined the three sets to benchmark (A) and the designed workflow used for data visualization and benchmarking (B). The three sets (*CAN*, *ISO_wo_SNV*, and *ISO*) encompassed expressed molecules based on specific modification types. In particular, the "*CAN*" set considered only canonical miRNAs (miRBase v22); the "*ISO_wo_SNV*" one used both canonical miRNAs and shifted isomiRs without SNVs; the "*ISO*" set employed all expressed canonical miRNAs and isomiRs, including the shifted ones. The workflow extracts expressed molecules from every cohort, condensing the information into a single massive table (expressed molecules as rows, cohorts' samples as columns). A nonlinear dimensionality-reduction technique (Uniform Manifold Approximation and Projection - UMAP) (McInnes et al., 2018) is applied to reduce high-dimensional data into a two-dimensional matrix for data visualization and evaluation. The reduced two-dimensional matrix is then used to perform an unsupervised clustering by leveraging the DBSCAN algorithm (Ester et al., 1996) to test the sets' clustering capability.

**Supplementary Figure 3. Dataset-based clustering comparison**

(A-C) Comparison between *CAN*- (A), *ISO_wo_SNV*- (B), and *ISO*-based cancer samples clustering (C). Panels (A-C) display only clusters related to the most prominent and significant clinical-pathological features we considered (Supplementary Table 4). Each panel shows common and unique cohorts identified by the three sets. Clusters and their IDs are highlighted throughout the figure. Coloring is used to highlight clusters within the same set/cohort.

**Supplementary Figure 4. Risk Score-Based Prognostic Signature Discovery workflow**

The figure shows the 2-stages workflow designed for prognostic signature discovery for Overall Survival (OS) and Relapse Free Survival (RFS).

**Supplementary Figure 5. Risk Score-Based Prognostic Signatures**

The figure shows the complete list of Overall Survival (OS) and Relapse Free Survival (RFS) Kaplan-Meier curves, each one associated with a specific risk score-based signature.

**Supplementary Table 1. Primers list for cloning and sequencing of target genes 3' UTRs**

List of primers used for cloning and sequencing experiments. See Materials and Methods for more details.

**Supplementary Table 2. A-to-I RNA editing sites**

The table reports the list of known A-to-I RNA Editing sites available in MiREDiBase (v1).

**Supplementary Table 3. Expressed canonical miRNAs/miRNA isoforms distribution over modification types and miRNA arms**

The table shows the distribution of modification types of expressed molecules over 5p and 3p arms, the shifting extent at 5'- and 3'-ends, along with the number of molecules subjected to SNPs, somatic mutations, and A-to-I RNA editing sites.

**Supplementary Table 4. Enriched pathways across cohorts/cancer tissues**

The table reports significant pathways enriched in at least one cohort/cancer tissue, retained according to $|z-score| \geq 2$ and $p-value < 0.01$.

**Supplementary Table 5. Clustering clinical-pathological features**

The table reports the most prominent and significant clinical-pathological features per cohort considered to investigate clustering results from a clinical standpoint. Features are grouped according to each benchmarked set of molecules (*CAN*, *ISO_wo_SNV*, and *ISO*).

**Supplementary Table 6. Dysregulated miRNA isoforms across cohort/cancer tissues**

The table reports the complete list of dysregulated molecules across cohorts/cancer tissues, retained according to a $|linear\ fold\ change| > 1.5$ and an $FDR\ adjusted\ p-value < 0.05$.

**Supplementary Table 7. Distribution of dysregulated miRNA isoform across modification types**

The table outlines the distribution of dysregulated molecules and modification types across 5p and 3p arms, filtered according to a $|linear\ fold\ change| > 1.5$ and an $FDR\ adjusted\ p-value < 0.05$.

The table provides additional information on the distribution of dysregulated molecules with 5'- and 3'-end shifting, SNPs/somatic mutations, and A-to-I RNA editing sites.

**Supplementary Table 8. Dysregulated miRNA isoforms with opposite trends**

The table reports canonical miRNAs characterized by an opposite expression trend than their miRNA isoforms across cohorts/comparisons. Molecules are retained according to a $|linear\ fold\ change| > 1.5$ and an $FDR\ adjusted\ p-value < 0.05$.

**Supplementary Table 9. Dysregulated genes supplied with predicted targets for the selected case studies**

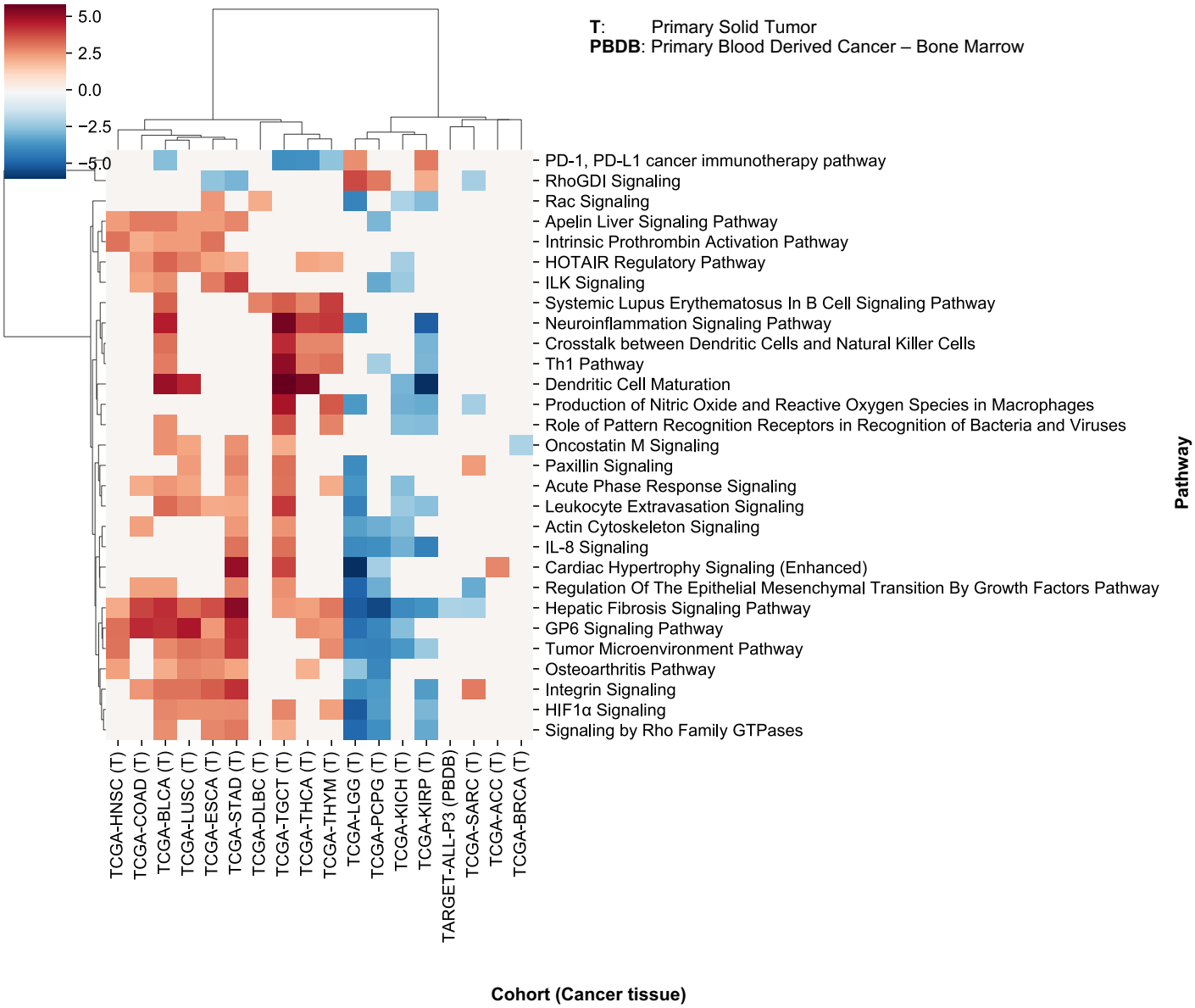
The table shows the dysregulated genes between the first (Q1) and third (Q3) quartiles of each case study: canonical miRNAs *miR-101-3p* and *miR-381-3p*, isomiR miR-101-3p (-1|-2), and edited miRNA

miR-381-3p_4_A_G. Whether available, each dysregulated gene is supplied with a predicted target consensus generated via isoTar, based on five prediction tools: PITA, RNAhybrid, TargetScan, miRanda, and miRmap.

**Supplementary Table 10. Risk score-based signatures list**

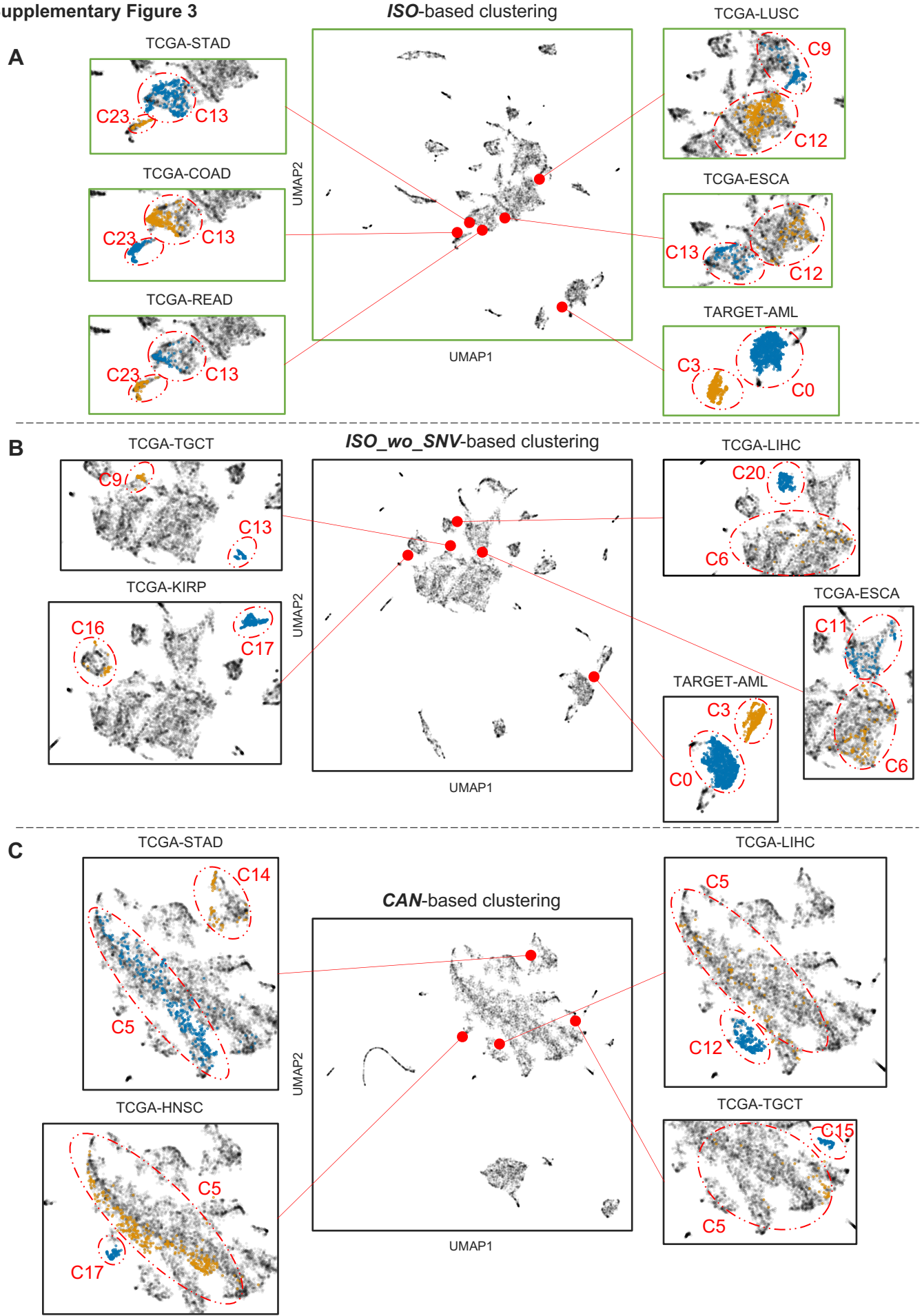
The table reports the most prominent and significant risk score-based signatures for Overall Survival (OS) and Relapse Free Survival (RFS), each one supplied with the molecules list, the area under the curve score (AUC), and p-value. Canonical miRNAs are highlighted in grey.

**Supplementary Figure 1**

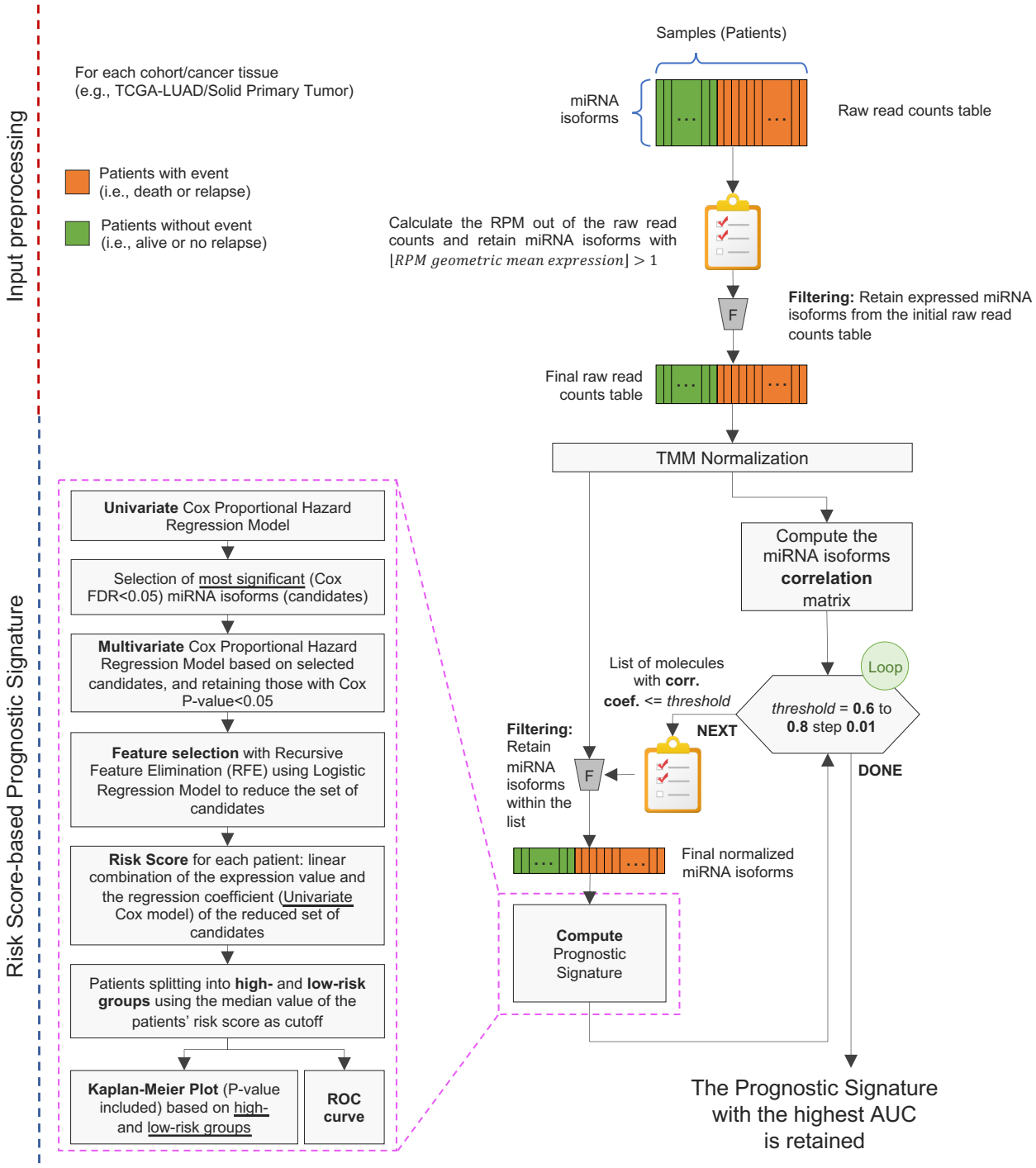


**Supplementary Figure 2**

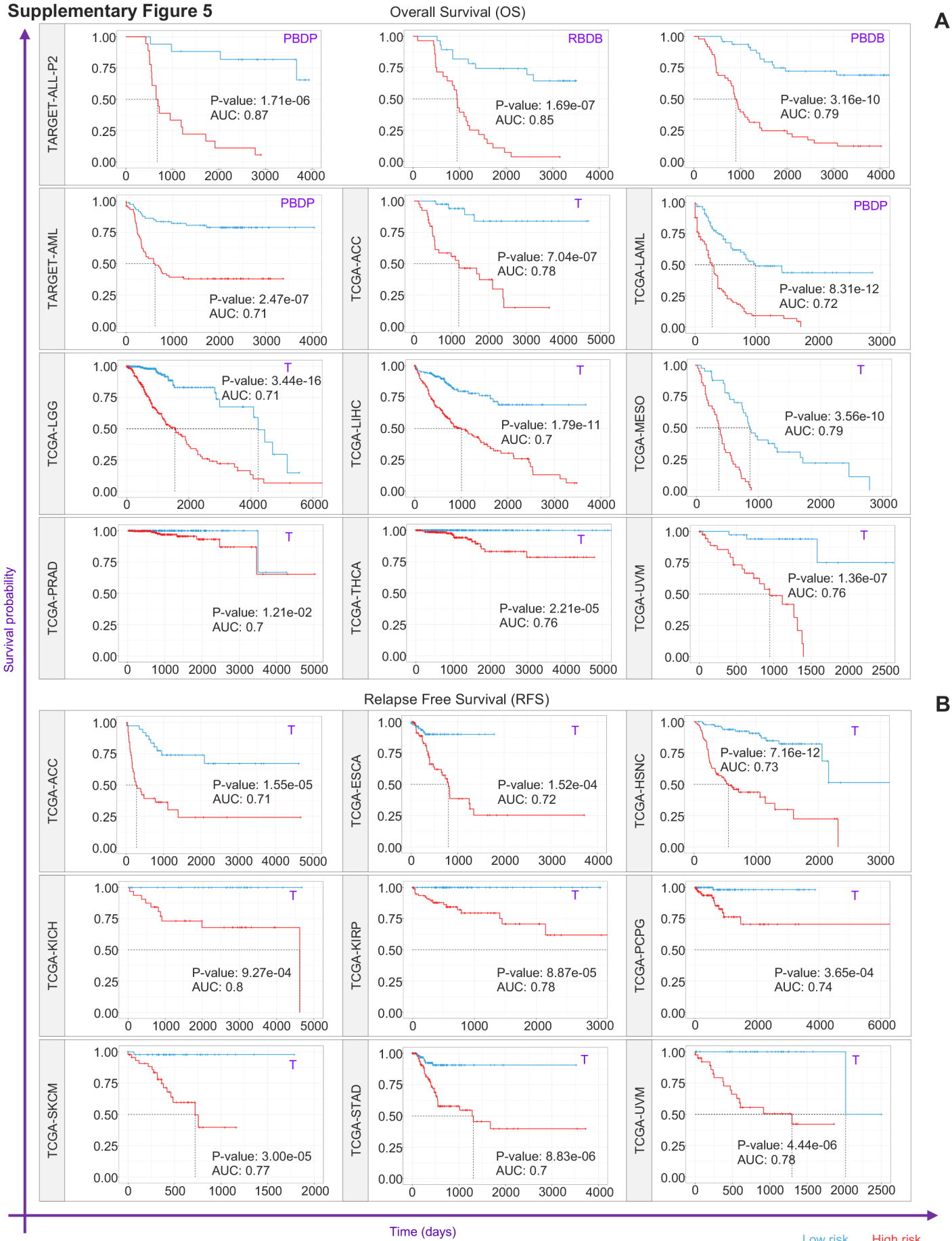
**Supplementary Figure 3**



**Supplementary Figure 4**



**Supplementary Figure 5**




Cancer tissues (GDC nomenclature):
**T**: Primary Solid Tumor
**RBDB**: Recurrent Blood Derived Cancer – Bone Marrow
**PBDP**: Primary Blood Derived Cancer – Peripheral blood
**PBDB**: Primary Blood Derived Cancer – Bone Marrow