Pan-Cancer Analysis of Canonical and Modified miRNAs Enhances the Resolution of the Functional miRNAome 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 P. Marceca[7], Paolo Fadda[8], Alessandro Laganà[9], Mario Acunzo[10], Qin Ma[5], Giovanni Nigita[1], and Carlo M. Croce[1]

ABSTRACT

Epitranscriptomic studies of miRNAs have added a new layer of complexity to the cancer field. Although there is fast-growing interest in adenosine-to-inosine (A-to-I) miRNA editing and alternative cleavage that shifts miRNA isoforms, simultaneous evaluation of both modifications in cancer is still missing. Here, we concurrently profiled multiple miRNA modification types, including A-to-I miRNA editing and shifted miRNA isoforms, in >13,000 adult and pediatric tumor samples across 38 distinct cancer cohorts from The Cancer Genome Atlas and The Therapeutically Applicable Research to Generate Effective Treatments data sets. The differences between canonical miRNAs and the wider miRNAome in terms of expression, clustering, dysregulation, and prognostic standpoint were investigated. The combination of canonical miRNAs and modified miRNAs boosted the quality of clustering results, outlining unique clinicopathologic features among cohorts. Certain modified miRNAs showed opposite expression from their canonical counterparts in cancer, potentially impacting their targets and function. Finally, a shifted and edited miRNA isoform was experimentally validated to directly bind and suppress a unique target. These findings outline the importance of going beyond the well-established paradigm of one mature miRNA per miRNA arm to elucidate novel mechanisms related to cancer progression.

in gene regulation, suggesting the concurrent consideration of canonical and modified miRNAs can boost patient stratification.

The combined contribution of canonical and modified miRNAs boosts patients' clinicopathological stratification, reveals more complex gene regulation mechanisms, and provides additional potential cancer biomarkers.

Significance: Modified miRNAs may act as cancer biomarkers and function as allies or antagonists of their canonical counterparts
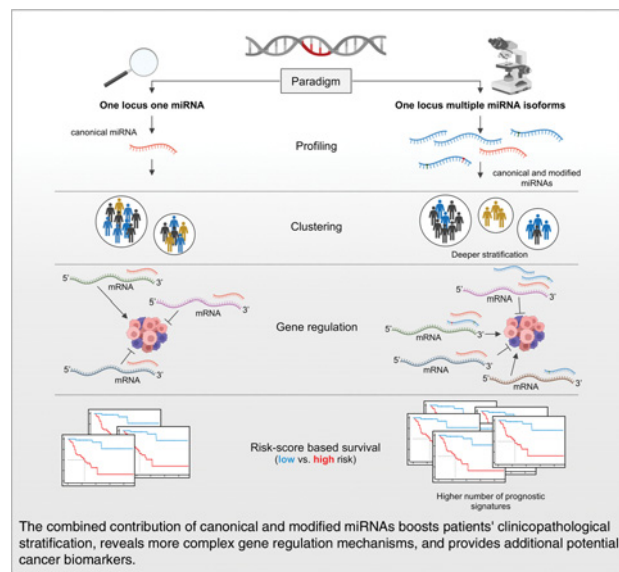
Introduction

MicroRNAs (miRNA) are small noncoding RNAs (∼21 nucleotides) that posttranscriptionally regulate gene expression (1) and whose dysregulation correlates with human diseases (2), including cancer (3). Until recently, most miRNA studies were built upon the miRNA biogenesis paradigm "one mature miRNA per miRNA

precursor arm" (1). However, the latest advancements in next-generation sequencing technologies unveiled a more intricate scenario (4), in which expressed miRNAs somehow differ from their canonical reference (5). These miRNA "variants," along with canonical miRNAs, are termed "miRNA isoforms or isomiRs" and are classified into five main categories: (i) canonical miRNAs, (ii) 5′ isomiRs, (iii) 3′ isomiRs, (iv) polymorphic isomiRs, and (v) mixed

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

isomiRs (6). 5′ isomiRs, 3′ isomiRs, and mixed isomiRs undergo sequence shifting (addition or trimming of nucleotides; refs. 7, 8), affecting the 5′, 3′, or both ends, respectively. Besides, mixed isomiRs and polymorphic isomiRs are characterized by single-nucleotide variants (SNV), specifically, DNA and RNA modifications, such as single-nucleotide polymorphisms (SNP; ref. 9), somatic mutations, and adenosine-to-inosine (A-to-I) RNA editing events, mammals' most abundant RNA editing form (10). A single A-to-I RNA editing event involving the miRNA seed region (MSR) could compromise the miRNA-mediated gene regulation process, potentially altering the miRNA function (11). Similarly, shifted isomiRs, which likely derive from alternative cleavage, may exhibit targetome differentiation (12). Although initially considered artifacts (13), recent studies have reevaluated their function (14) as they actively interact with genes (8, 12). Any change involving miRNA ends, especially the 5′-end, may diversify the molecules' targetome, leading to a more complex gene regulation machinery than previously thought (15).

In recent years, efforts to assess the biological implications of A-to-I edited miRNAs have boosted the interest in using such molecules as potential biomarkers for cancer prognosis and therapy (16, 17). Meanwhile, a surge in shifted isomiR-oriented studies led to the first pan-cancer study (18) over 32 tumor cohorts from The Cancer Genome Atlas (TCGA) data set. Although these studies have proven the importance of investigating such miRNA modifications, the concurrent profiling of these molecules in cancer is still missing.

In this work, we simultaneously profiled canonical and modified miRNAs from the most prominent and reliable cancer public resources, TCGA and The Therapeutically Applicable Research to Generate Effective Treatments (TARGET), analyzing >13,000 adult and pediatric cancer samples spread across 38 distinct cohorts. Annotated molecules were benchmarked to assess the benefits and drawbacks of using specific miRNA modification types for clustering and survival purposes. We investigated dysregulated molecules across different cohorts/tissues. Lastly, we experimentally demonstrated the phenomenon of targetome shifting among miRNA isoforms generated by the same locus in two case studies: (i) the canonical hsa-miR-101–3p and one of its shifted isomiRs (one nucleotide added at 5′-end and two nucleotides trimmed at 3′-end), in lung adenocarcinoma; and (ii) the canonical hsa-miR-381–3p and its A-to-I edited isomiR at position 4, in breast cancer.

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

Materials and Methods

The study was conducted following the Health Insurance Portability and Accountability Act (HIPAA) guidelines. The patient cohorts we used are publicly available data sets collected with patients' informed consent, generated by TCGA (https://cancer.gov/tcga) and by the Therapeutically Applicable Research to Generate Effective Treatments initiative (TARGET: http://ocg.cancer.gov/programs/target) Research Networks, both from the NCI (NIH). Furthermore, the TCGA and TARGET data were downloaded, stored, and analyzed with the approval of the NIH Data Access Committee with the following DBGap Project IDs: #11332 for TCGA and #22219 for TARGET.

Data sources

We downloaded miRNA-seq (v20) BAM files from 33 TCGA (11,082 samples) and 5 TARGET (2,378 samples) cohorts. Paired-end RNA-seq (v32) BAM files were collected for the 33 TCGA (11,123 samples) and 5 TARGET (1,127 samples) cohorts. We used four data sets of known DNA and RNA modifications: SNPs from dbSNP (v155; ref. 19), A-to-I miRNA editing sites from MiREDiBase (v1; ref. 20), and somatic mutations from COSMIC (v96; ref. 21) and TCGA/TARGET (v20). Cohorts' essential characteristics are summarized in **Table 1**. See Supplementary Information for more details.

miRNA isoform annotation

We applied an in-house workflow to annotate a wide range of miRNA isoforms, including novel molecules. Downloaded miRNA-seq BAM files were converted into FASTQ files, quality-checked, trimmed, and then annotated using miRge 2.0 (22), one of the major pipelines for canonical miRNA/modified miRNA profiling. Novel molecules were retained according to a prediction probability greater than 0.95 (computed by miRge 2.0). At the same time, annotated SNVs were inspected to ensure a minimum Phred quality score of 30, corresponding to a sequencing error probability of less than 0.1%. Samples with less than one million miRNA mapped reads were removed to improve data quality for batch correction (23).

miRNA isoform SNV filtering and annotation

SNVs were further filtered via a multistep process, which first reduced the population of unknown SNVs using the four SNV data sets. Afterward, molecules with unknown SNVs involving the first (at 5′-end) or one of the last two nucleotides (at 3′-end) were removed, potentially due to sequencing errors or flaws in the linker ligation during the assembly of the cDNA library. We calculated the SNV statistical significance via binomial test, adjusting the resulting P values via the Benjamini–Hochberg correction. Adjusted P values were used to retain molecules showing significant SNVs (adjusted $P < 0.05$) in at least 5% of the sample population (12,857). See Supplementary Information for more details.

miRNA isoform nomenclature

We adopted a human-readable way to label annotated molecules and their modifications, using a combination of five factors: canonical miRNA, pre-miRNA, 5′- and 3′-end shifting, and the "Compact Idiosyncratic Gapped Alignment Report" string, shortened CIGAR, to indicate SNVs. A label "miR-21-5p__mir-21__-1__+1__2MG21M" represents a mixed isomiR that withstands: (i) a single-nucleotide addition at both 5′- and 3′-end, and (ii) an A-to-I miRNA editing at position 3 (2M stands for two genomic matches, a guanosine (G) in place of a genomic adenosine (A), whereas 21M indicates 21 genomic matches). See Supplementary Information for more details.

Clustering analysis

We benchmarked four distinct groups (G1, G2, G3, and G4) of molecules using an in-house workflow, investigating the benefits and drawbacks of using specific miRNA modification types for clustering cancer samples across cohorts. We filtered molecules for each group according to a geometric mean >3 reads per million miRNA mapped reads (RPM), corresponding to 10.6 ± 3.1 (SD) raw read counts across cohorts. Then, the raw read count expression of the filtered molecules was corrected for batch effects using tumor purity and platform and reduced into a two-dimensional matrix via a nonlinear dimensionality reduction technique (uniform manifold approximation and projection: UMAP; ref. 24). A density-based clustering (DBSCAN; ref. 25) was finally applied to the reduced matrix to cluster cancer samples. All clustering results were filtered according to a percentage of mislabeled

Table 1. Essential characteristics of cohorts.

Cohort	Cancer type	No. of cases	Age at diagnosis (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	1,079	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

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

samples (noise) lower than 5% and evaluated via (adjusted rand index (ARI), adjusted mutual information (AMI), and Fowlkes–Mallows index scores. See Supplementary Information for more details.

Differential miRNA isoform expression analysis

We investigated dysregulated molecules across cancer cohorts, performing differential expression (DE) analyses considering a minimum of 5 samples per cohort/condition (e.g., tumor vs. normal). We filtered molecules according to a geometric mean >3 RPM in at least one of the two conditions, correcting the expression for batch effects using tumor purity and preclinical factors, such as gender, age at initial pathologic diagnosis, tumor stage, and platform. The resulting P values were adjusted via the Benjamini–Hochberg correction. We retained dysregulated molecules according to an adjusted $P < 0.05$ and |linear fold change| >1.5. See Supplementary Information for more details.

Cell lines

HEK293 (ATCC; cat. #CRL-1573), A549 (ATCC; cat. #CCL-185), H1299 (NCI-H1299; ATCC; cat. #CRL-5803), MDA-MB-231 (ATCC; cat. #HTB-26), and HCC70 (ATCC; cat. #CRL-2315) were seeded and grown in RPMI-1641 medium supplemented with 10% of FBS and penicillin–streptomycin (100 U/mL penicillin and 0.1 mg/mL streptomycin; Millipore Sigma). All cell lines were authenticated through the short-tandem repeat profiling method and tested to be free of *Mycoplasma* contamination using the Mycoplasma PCR Detection Kit (Applied Biological Materials).

Cell transfection

HEK293, A549, H1299, MDA-MB-231, and HCC70 cell lines were plated in a 6- or 12-well plate 24 hours before transfection. 100 nmol/L of mirVana miRNA mimic (Thermo Fisher Scientific) of canonical miR-101-3p (miR-101-3p__mir-101-1__0__0__21M; Assay ID: MC11414), shifted miR-101-3p (miR-101-3p__mir-101-1__-1__-2__20M; custom assay), canonical miR-381-3p (miR-381-3p__mir-381__0__0__22M; Assay ID: MC10242), and edited miR-381-3p (miR-381-3p__mir-381__0__0__3MG18M; custom assay) 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; cat. #4464058), and Anti-miR miRNA Inhibitor-negative Control #1 (Thermo Fisher Scientific; cat. #AM17010) were used as scrambled controls. After 5 hours, the transfection medium was replaced with RPMI-1641 supplemented with 10% FBS and penicillin–streptomycin and, only for H1299 cells, with mitomycin C (Millipore Sigma) at the concentration of 15 μg/mL. After 24 or 48 hours, cells were harvested and subjected to luciferase assay or RNA isolation and protein lysis. See Supplementary Information for more details.

Risk score–based prognostic signature

We applied an in-house workflow for estimating optimal prognostic signatures, benchmarking the four groups introduced in the clustering analysis via a 10-fold cross-validation strategy. Molecule expression was corrected for batch effects using tumor purity and platform. We explored both overall survival (OS; event: death; nonevent: alive) and relapse-free survival (RFS; event: relapse; nonevent: no relapse), requiring a minimum of 20 patients per event type for reliable results. We used a stratified shuffle splitting policy to ensure a proportional balance of event/nonevent patients among training (66% of patients) and validation (34% of patients) sets. The workflow exploited a combination of several strategies to narrow down the number of molecules included in each identified prognostic signature. Briefly, strategies ranged from recursive feature elimination for best-molecule selection to a combination of univariate and multivariate Cox hazard regression models for assessing the relationship between molecule expression and patients' survival. After stratifying patients into high- and low-risk groups using the patients' risk scores, we used the two groups to calculate the P value (log-rank test) and prognostic signature accuracy (area under the curve: AUC). The essential clinical characteristics of cohorts are summarized in **Table 2**. See Supplementary Information for more details.

Data availability

The source code of analyses and expression data generated in this study are available via the Zenodo repository (https://zenodo.org/record/6902726; ref. 26).

Results

miRNA isoform profiling in cancer

We investigated the miRNAome in cancer, profiling canonical and modified miRNAs at a large scale (**Fig. 1A** and **B**), obtaining ~8,300 molecules across cohorts (Supplementary Fig. S1A). On average, we identified 1,533 ±189 (SD) expressed molecules per cohort (geometric mean >3 RPM across cohort's samples), with modified miRNAs outmatching the number of canonical miRNAs by more than a factor of five (**Fig. 1C**, "columns in blue"). The 3′ isomiRs resulted in the most abundant expressed molecules (~45% per cohort), showing an expression (median of percentiles) close to that of canonical miRNAs (**Fig. 1C**, "columns in red color"). Of all cancer cohorts, TCGA-TGCT, TARGET-RT, TCGA-PCPG, TCGA-THYM, and TCGA-SKCM showed, in order, the highest number of expressed molecules (above 75th quartile; **Fig. 1C**, "square in red color") across miRNA modification types (e.g., 3′ isomiR). TCGA-GBM represented the sole exception (normal tissue only) above 75th. The TARGET-ALL-P3, TCGA-LAML, and TCGA-TGCT cohorts were characterized by the highest number of 5′ isomiRs, which could potentially extend the miRNA targetome due to MSR shifting. In terms of miRNA ends stability, the 5′-end resulted in the most stable, with shiftings mainly limited to ±1 nucleotide (Supplementary Table S1, "5′-end shifting

distribution"). In contrast, the 3′-end revealed greater mobility, with most molecules showing large trimming (up to 5) and addition (up to 2) of nucleotides (Supplementary Table S1, "3′-end shifting distribution").

As for SNVs, DNA modifications were mainly located near the 3′-end (Supplementary Table S1, "DNA SNV distribution"), whereas A-to-I miRNA editing sites were distributed within the MSR (Supplementary Table S1, "A-to-I SNV distribution"). Interestingly, the nervous tissues and thymoma showed the highest number of A-to-I edited molecules. Although most A-to-I edited miRNA molecules did not positively correlate (rho < 0.25 and P > 0.05) with ADAR1/2 gene expression (enzymes that catalyze the chemical conversion of adenosine to inosine in double-stranded RNAs; ref. 10), there were some exceptions. In the TCGA-LGG and TCGA-UVM cohorts, edited molecules positively correlated (rho > 0.25 and P < 0.05) with ADAR2, whereas TCGA-TGCT and TCGA-PCPG showed a positive correlation (rho > 0.25 and P < 0.05) with ADAR1.

Finally, we investigated the relationship between the differences in the abundance of expressed miRNA isoforms and genes across cohorts' cancer samples. For each cohort/cancer tissue, we extracted two groups of samples, one with a lower (below 25th quartile) and the other with a higher (above 75th quartile) number of expressed miRNA isoforms. We performed a gene DE analysis based on these two groups, retaining significantly dysregulated genes. Each DE analysis included, as covariates, tumor purity and preanalytical variables, such as age at initial pathologic diagnosis, tumor stage, and gender (see Supplementary Information for more details). We then performed a pathways enrichment analysis via Ingenuity Pathway Analysis software, reporting the most significant pathways enriched in at least five cohorts/cancer tissues in Supplementary Fig. S1B. Notably, the abundance of expressed miRNA isoforms correlated with the activation/deactivation of critical carcinogenesis pathways, such as the p38 MAPK, Wnt/β-catenin, HGF, and IL6-signaling pathways.

miRNA isoform–based cancer sample clustering

We benchmarked four distinct groups (G1, G2, G3, and G4) of molecules (**Fig. 2A**), investigating the benefits and drawbacks of using specific miRNA modification types for clustering cancer samples across cohorts (**Fig. 2B**). We observed a significant improvement in sample stratification when leveraging G4 (expressed canonical and modified miRNAs; **Fig. 2C**). Clustering results were examined from a clinicopathologic perspective, using available cohorts' clinicopathologic features from TCGA and TARGET (Supplementary Table S2). Clinicopathologic stratification significance was assessed via Pearson chi-squared test, reporting the most significant (Chi-square P < 0.01) and notable features from a clinicopathologic standpoint (Supplementary Table S2). Throughout the analysis, the G1-, G2-, G3-, and G4-based findings showed different behaviors in clustering cancer samples.

In TCGA-ESCA, G1, G2, and G4 classified patients according to their clinical stages and histologic subtypes (adeno vs. squamous cell carcinoma), whereas G3 failed to identify any notable clinicopathologic feature. We observed a similar trend in TCGA-TGCT, in which G3 could not cluster cancer samples according to their clinical stage, in stark contrast with what was obtained with G1, G2, and G4. These results suggest that, at least in these settings, canonical miRNAs play a fundamental, predictive role.

Nonetheless, adding additional molecules boosted the capability of identifying notable clinicopathologic features rather than relying solely on canonical miRNAs. In TCGA-KIRC, G2 was able to separate

Table 2. Clinical characteristics of cohorts.

Cohort	# Cases	Stages (I/II/III/IV/NA)	OS			RFS		
			# 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	1,079	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

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

late-stage (IV stage) cancer samples from the rest. In TCGA-BRCA, G2 exclusively clustered cancer samples according to the expression of the hormone receptors. It then stratified, similarly to G3 and G4, samples according to their histologic characteristics (ductal versus lobular). The TCGA-BRCA cohort represents an example in which G1 failed to identify significant clinical features. We observed the same behavior in TCGA-STAD, with the sole G2, G3, and G4 dividing cancer samples into clusters characterized by different clinical stages.

In other cohorts, G3 and G4 showed similar capabilities in stratifying cancer samples according to their clinicopathologic features. In particular, G3 and G4 classified cancer samples at different stages of the disease in TCGA-LUSC, whereas in TCGA-HNSC, they recognized well-differentiated tumors (g1 neoplasm histologic grade).

We identified cases in which G4 outperformed the other groups. In TCGA-CESC, G4 clustered cancer samples according to their histologic subtypes (adeno vs. squamous cell carcinoma). In TCGA-READ, G4 identified low-grade tubular adenocarcinomas among all the others.

Altogether, our results highlighted a substantial improvement in clustering, particularly when not limiting the analysis to the sole canonical miRNAs. In this context, G4 performed better than all the other benchmarked groups.

Differentially abundant miRNA isoforms across cancer tissues

To determine whether miRNA isoforms were dysregulated across cohorts/cancer tissues, we performed a DE analysis comparing primary solid, recurrent solid, metastatic, and normal tissues. Each DE analysis included, as covariates, tumor purity and preanalytical variables, such as age at initial pathologic diagnosis, stage of disease, gender, and platform (see Supplementary Information for more details). Dysregulated molecules were retained according to an adjusted $P < 0.05$ and |linear fold change| >1.5 (27).
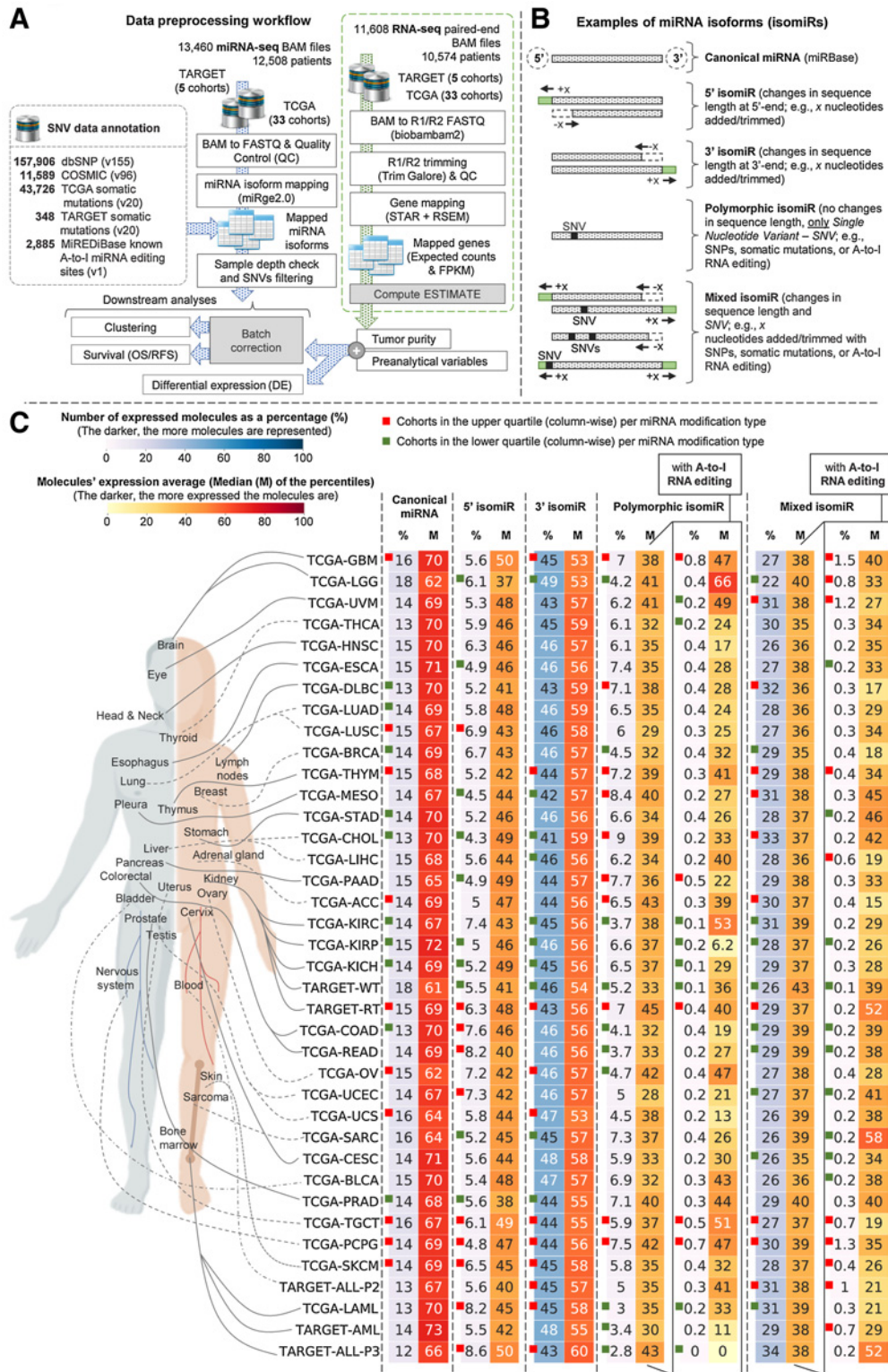
Figure 1.
Data preprocessing workflow, miRNA isoforms classification, and miRNA modification types distribution. **A–C,** In-house data preprocessing workflow and data sources (**A**), examples of annotated miRNA isoforms (**B**), and distribution of expressed molecules (geometric mean >3 RPM) across cohorts and miRNA modification types (**C**). Panel **C** reports, for each cohort, the number of molecules per miRNA modification type as a percentage (%), along with the median (M) of the percentiles computed using molecules' expression average. Squares in green and red indicate (column-wise) whether a cohort belongs to the lower (25th) or upper quartile (75th), respectively.
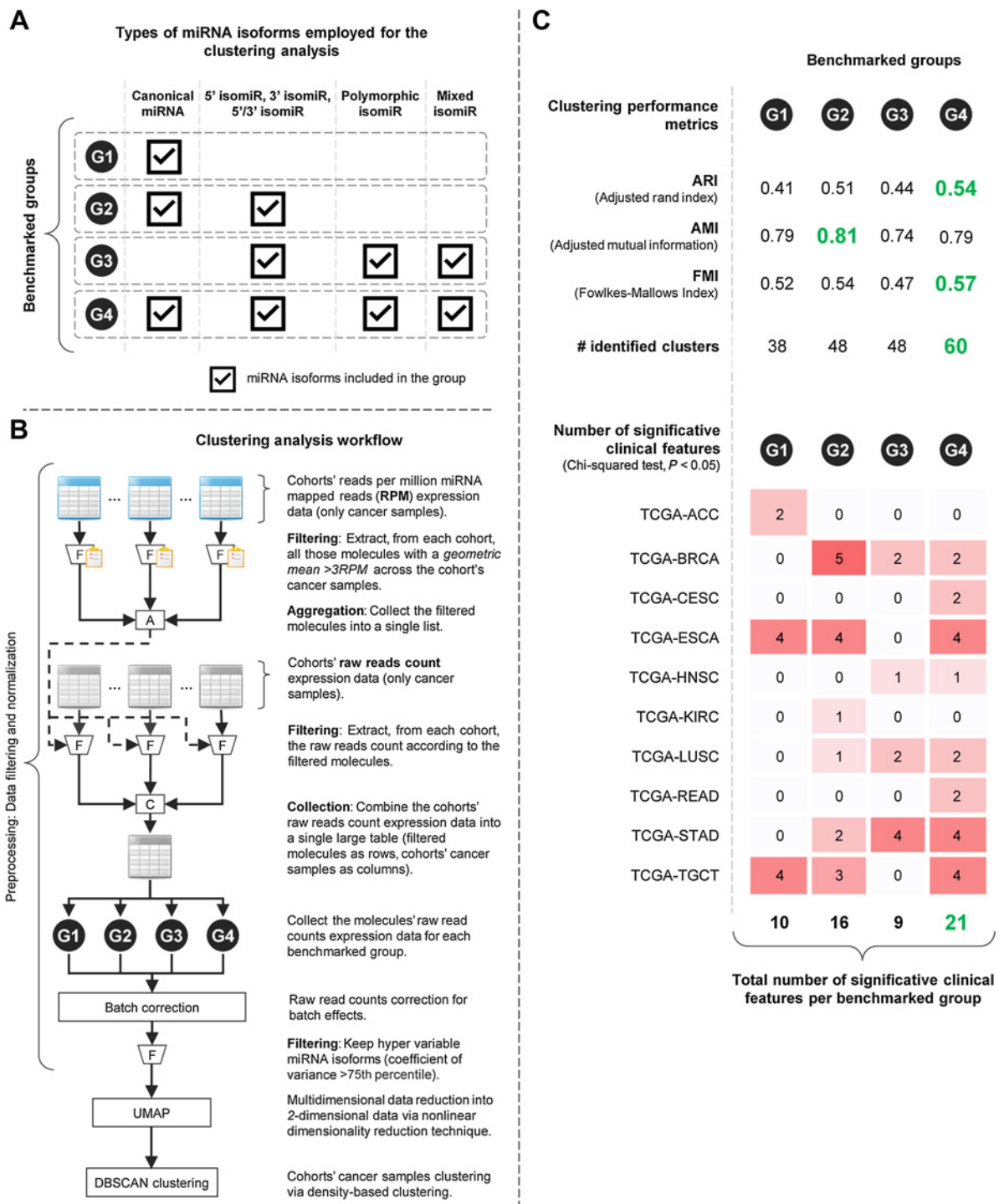
Figure 2.
miRNA isoform-based clustering better delineates clinicopathologic stratification. **A–C,** Type of miRNA isoforms used by the four benchmarked groups of molecules (**A**), workflow used to benchmark the four groups (**B**), and clustering results for the four groups (**C**). Panel **C** reports quality scores, the number of identified clusters, and the number of unique significant (Chi-square $P < 0.01$) clinicopathologic features identified for each benchmarked group. See Supplementary Table S2 for the list of specified clinicopathologic features, with the most prominent and significant features highlighted in green. See Supplementary Information for more details.

Overall, dysregulated molecules (**Fig. 3**) followed a pattern comparable to what we reported in **Fig. 1C**, with 3′ isomiRs being the most abundant molecules across cohorts/tissues. Dysregulated molecules were similarly distributed across miRNA arms (5p and 3p), with A-to-I edited molecules being the sole exception (∼4 times more abundant for the 3p arm). Considering the negligible difference between the two miRNA arms, we focused on the sole distribution of dysregulated molecules across miRNA modification types. Our results confirmed the higher stability of the 5′-end, with most dysregulated molecules characterized by shifting of ±2 nucleotides, whereas molecules with 3′-end shifting showed wider additions (up to +3) and trimmings (up to −6). DNA modifications were mainly located near the 3′-end, whereas A-to-I miRNA editing sites were distributed within the MSR.

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

Among differentially expressed molecules (adjusted $P < 0.05$ and | linear fold change| >1.5), we identified 85 dysregulated canonical miRNAs (out of 392) showing opposite expression trends compared with their modified miRNAs (canonical miRNA downregulated and modified miRNAs upregulated, or vice versa; Supplementary Table S3, "miRs with opposite expr. trend"). Out of the 85 molecules, we selected the canonical miR-101–3p (among the top-five dysregulated canonical miRNAs with opposite expression trends), whose expression resulted in being downregulated in four cohorts (TCGA-LUAD, TCGA-LIHC, TCGA-HNSC, and TCGA-CHOL), as the candidate to assess the potential targeting shifting among different miRNA isoforms. In TCGA-LUAD, the downregulated canonical miR-101–3p presented an upregulated shifted miRNA isoform (with no SNVs), termed miR-101–3p (−1|−2), characterized by shifting at 5′-end (addition of one nucleotide) and 3′-end (trimming of two nucleotides). The 5′-end shifting facilitated the research of exclusive putative targets using predictor tools due to the different MSR, whereas 3′-end shifting made the design of a custom TaqMan probe for real-time PCR analysis specific for the miRNA isoform easier. Moreover, we verified that both molecules were correctly loaded by AGO2 (Supplementary Fig. S2A–B). Finally, the interest in such a canonical miRNA was corroborated by a recent work in which authors assessed one shifted miRNA isoform of canonical miR-101–3p in the human brain (28).

As shown in **Fig. 4A** and **B**, the two molecules were characterized by an opposite expression trend, with the canonical miRNA downregulated in TCGA-LUAD cancer samples. Looking for putative and exclusive targets, we retained dysregulated genes (Supplementary Table S4) based on their significance (see Supplementary Information) and opposite expression trends (i.e., miRNA isoform up and genes down, or vice versa; **Fig. 4C** and **D**). Finally, we intersected dysregulated genes with the list of predicted targets generated by isoTar (Supplementary Table S4; ref. 29), requiring a minimum consensus of two prediction tools. Out of the reduced set of genes, we elected *prostaglandin-endoperoxide synthase 2* (*PTGS2* or *COX2*), an oncogene studied in cancers (30, 31), which is also a validated target for miR-101–3p and overexpressed in lung cancer (32), and *desmocollin-2* (*DSC2*), a putative target for the miR-101–3p (−1|−2), that was recently characterized as an oncogene in lung cancer (33). In line with the literature, we confirmed the direct binding (Supplementary Fig. S2C) through luciferase assay (**Fig. 4E**) between *PTGS2* and the downregulated canonical miR-101–3p. After miR-101–3p ectopic overexpression in HEK293 cells, we observed a ∼40% reduction of luciferase activity compared with the scramble-negative control (SCR; **Fig. 4E**). In contrast, miR-101–3p (−1|−2) overexpression resulted in a minor reduction in luciferase activity (**Fig. 4E**). The

direct binding miRNA gene was confirmed by using a mutated psiCHECK-2-PTGS2 3′ UTR vector, holding the deletion of the binding sites for the canonical miR-101–3p. In this case, the assay showed no decrease in luciferase activity after overexpressing the two miR-101–3p molecules (Supplementary Fig. S2D).

After transfecting the canonical miR-101-3p (Supplementary Fig. S2E), Western blotting experiments highlighted a significant downregulation of endogenous PTGS2 in the A549 and H1299 lung cancer cell lines, whereas no PTGS2 protein level variation was observed transfecting miR-101-3p (-1|-2; **Fig. 4F** and **G**). Our findings demonstrated that of the two molecules, only the downregulated canonical miR-101–3p exclusively targeted *PTGS2* (upregulated in lung cancer).

DSC2 is a protein implicated in mediating cell adhesion, epithelial cell proliferation, and tumorigenesis (33). Likewise PTGS2, we demonstrated via luciferase assay, using both wild-type (WT) and mutated psicheck2-DSC2 3′UTR (**Fig. 4H**; Supplementary Fig. S2D), and Western blotting (**Fig. 4I** and **J**), the exclusive targeting of DSC2 by miR-101-3p (−1|−2).

Although the two miRNA isoforms arise from the same locus, our experiments demonstrated that shifting a few nucleotides led to a distinct targetome for these two molecules in lung cancer.

Dysregulated A-to-I edited miRNA isoforms in cancer

We measured the A-to-I RNA editing abundance across cancer cohorts/tissues, detecting 59 unique dysregulated A-to-I edited miRNA isoforms (Supplementary Table S3, "Edited miRs dysreg. in cancer") that originated from 18 distinct miRNA arms. The edited miR-381–3p (A-to-I RNA editing at position 4) resulted in one of the most diffused dysregulated edited molecules. Its downregulation interested 7 of 26 cohorts/tissues, a trend confirmed by previous studies (16, 17) and observed in several tumors, including breast cancer (34).

We investigated the canonical miR-381–3p and the above-mentioned edited forms (shortened miR-381-3p_4_A_G) in the breast cancer cohort (TCGA-BRCA). The expression of the two molecules exhibited a significant downregulation in cancer samples (**Fig. 4K** and **L**). In line with the previous section, we applied a similar workflow (see Supplementary Information) to assess potential target variability between the two molecules. After retaining significantly dysregulated genes (Supplementary Table S4), characterized by an opposite expression trend (miRNA isoform *down*, genes *up*; **Fig. 4M** and **N**), we crossed them with the list of gene targets predicted by isoTar (Supplementary Table S4). Out of the reduced set of potential direct targets for miR-381–3p, we elected *ubiquitin conjugating enzyme E2 C* (*UBE2C*), a gene that promotes breast cancer proliferation, migration, and invasion, whose overexpression correlates with poor clinical outcomes (35). Using luciferase reporter vectors containing the WT or mutated 3′ UTR of the gene, the latter lacking the miR-381–3p binding site, together with the two miRNA molecules (canonical and edited miRNAs) in HEK293 cells, we demonstrated the direct binding between *UBE2C* and miR-381–3p (**Fig. 4O**; Supplementary Fig. S2D). Following the miR-381–3p overexpression (Supplementary Fig. S2E), Western blotting experiments in triple-negative breast cancer (TNBC) cell lines MDA-MB-231 and HCC70 corroborated our findings, demonstrating a significant downregulation of UBE2C (∼50%), as depicted by densitometry (**Fig. 4P** and **Q**). At the same time, among miR-381-3p_4_A_G targets, we selected synaptotagmin 13 (SYT13), an oncogene involved in different cancers (36, 37). We validated the direct binding (Supplementary Fig. S2C) between *SYT13* and miR-381-3p_4_A_G through luciferase assay in HEK293 cells, with
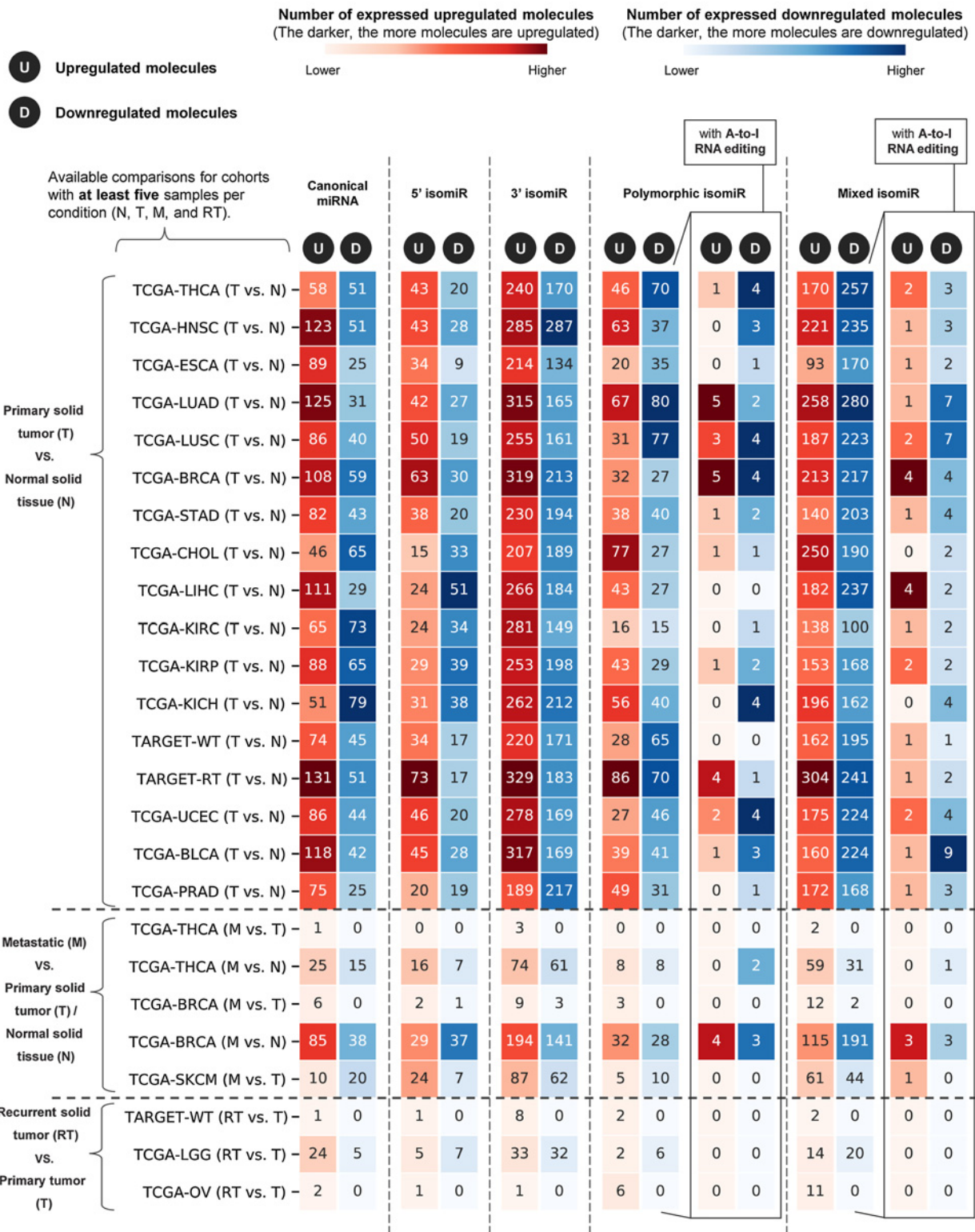
Column-wise gradient color.

Number of expressed upregulated molecules
(The darker, the more molecules are upregulated)

Number of expressed downregulated molecules
(The darker, the more molecules are downregulated)

U Upregulated molecules

D Downregulated molecules

Lower — Higher   Lower — Higher

Available comparisons for cohorts with **at least five** samples per condition (N, T, M, and RT).	Canonical miRNA U	Canonical miRNA D	5' isomiR U	5' isomiR D	3' isomiR U	3' isomiR D	Polymorphic isomiR U	Polymorphic isomiR D	with A-to-I RNA editing U	with A-to-I RNA editing D	Mixed isomiR U	Mixed isomiR D	with A-to-I RNA editing U	with A-to-I RNA editing D
Primary solid tumor (T) vs. Normal solid tissue (N)														
TCGA-THCA (T vs. N)	58	51	43	20	240	170	46	70	1	4	170	257	2	3
TCGA-HNSC (T vs. N)	123	51	43	28	285	287	63	37	0	3	221	235	1	3
TCGA-ESCA (T vs. N)	89	25	34	9	214	134	20	35	0	1	93	170	1	2
TCGA-LUAD (T vs. N)	125	31	42	27	315	165	67	80	5	2	258	280	1	7
TCGA-LUSC (T vs. N)	86	40	50	19	255	161	31	77	3	4	187	223	2	7
TCGA-BRCA (T vs. N)	108	59	63	30	319	213	32	27	5	4	213	217	4	4
TCGA-STAD (T vs. N)	82	43	38	20	230	194	38	40	1	2	140	203	1	4
TCGA-CHOL (T vs. N)	46	65	15	33	207	189	77	27	1	1	250	190	0	2
TCGA-LIHC (T vs. N)	111	29	24	51	266	184	43	27	0	0	182	237	4	2
TCGA-KIRC (T vs. N)	65	73	24	34	281	149	16	15	0	1	138	100	1	2
TCGA-KIRP (T vs. N)	88	65	29	39	253	198	43	29	1	2	153	168	2	2
TCGA-KICH (T vs. N)	51	79	31	38	262	212	56	40	0	4	196	162	0	4
TARGET-WT (T vs. N)	74	45	34	17	220	171	28	65	0	0	162	195	1	1
TARGET-RT (T vs. N)	131	51	73	17	329	183	86	70	4	1	304	241	1	2
TCGA-UCEC (T vs. N)	86	44	46	20	278	169	27	46	2	4	175	224	2	4
TCGA-BLCA (T vs. N)	118	42	45	28	317	169	39	41	1	3	160	224	1	9
TCGA-PRAD (T vs. N)	75	25	20	19	189	217	49	31	0	1	172	168	1	3
Metastatic (M) vs. Primary solid tumor (T) / Normal solid tissue (N)														
TCGA-THCA (M vs. T)	1	0	0	0	3	0	0	0	0	0	2	0	0	0
TCGA-THCA (M vs. N)	25	15	16	7	74	61	8	8	0	2	59	31	0	1
TCGA-BRCA (M vs. T)	6	0	2	1	9	3	3	0	0	0	12	2	0	0
TCGA-BRCA (M vs. N)	85	38	29	37	194	141	32	28	4	3	115	191	3	3
TCGA-SKCM (M vs. T)	10	20	24	7	87	62	5	10	0	0	61	44	1	0
Recurrent solid tumor (RT) vs. Primary tumor (T)														
TARGET-WT (RT vs. T)	1	0	1	0	8	0	2	0	0	0	2	0	0	0
TCGA-LGG (RT vs. T)	24	5	5	7	33	32	2	6	0	0	14	20	0	0
TCGA-OV (RT vs. T)	2	0	1	0	1	0	6	0	0	0	11	0	0	0

Figure 3.
miRNA isoforms dysregulation across cohorts and tissues. Distribution of dysregulated molecules per cohort/tissues and miRNA modification type. The figure reports the number of significant (adjusted $P < 0.05$) upregulated (U; linear fold change >1.5) and downregulated (D; linear fold change <-1.5) molecules, including A-to-I edited miRNA isoforms.
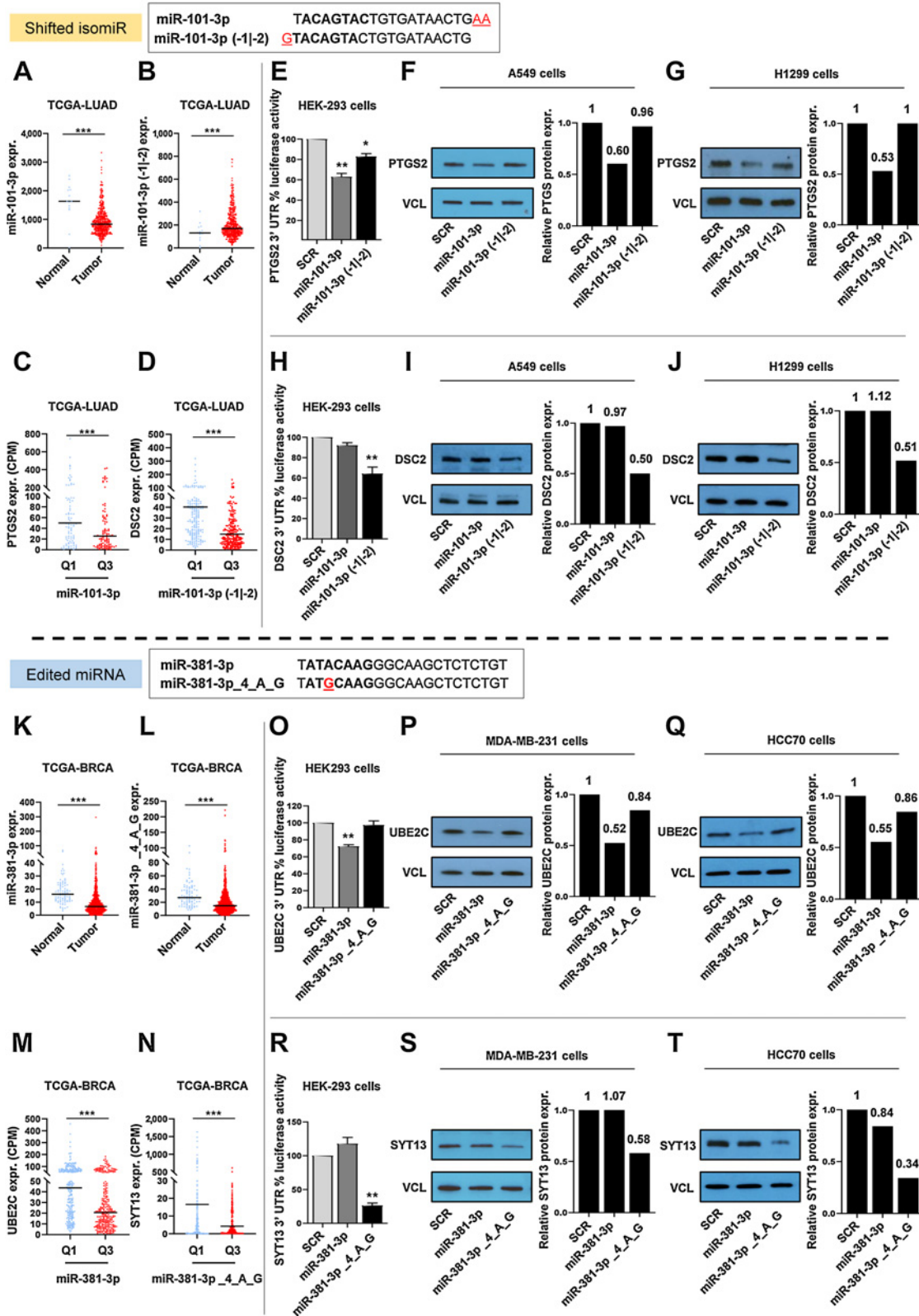
a reduced luciferase activity of ~80% (**Fig. 4R**). Then, we confirmed these data by repeating the luciferase assay using a mutated psicheck2-SYT12 3′ UTR without the miR-381-3p_4_A_G MSR, observing an unvaried luciferase activity even after the overexpression of the edited miRNA (Supplementary Fig. S2D). Finally, after miR-381-3p_4_A_G overexpression (Supplementary Fig. S2E), Western blotting experiments in MDA-MB-231 and HCC70 cell lines confirmed the downregulation of SYT13 as shown by densitometry (**Fig. 4S** and **T**).

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 explored OS and RFS (Supplementary Fig. S3), benchmarking the four groups (G1, G2, G3, and G4) introduced in the clustering analysis (**Fig. 2A**). Overall, we observed higher performance when including modified miRNAs. Without relying solely on canonical miRNAs (G1), we were able to achieve a greater number of significant prognostic signatures (log-rank test-based $P < 0.05$ and AUC ≥ 0.7; **Fig. 5A**). Notably, we identified 2 (G1), 4 (G2), 3 (G3), and 6 (G4) OS prognostic signatures, whereas the RFS ones were 1 (G1), 2 (G2), 3 (G3), and 4 (G4). The complete list of identified signatures and additional details are reported in Supplementary Table S5.

It would be fair to think that adding more molecules might increase the number of identified signatures. To test such a hypothesis, we randomly extracted expressed molecules (geometric mean >3 RPM) from G4, sampling them into five groups of different sizes (10%, 25%, 50%, 75%, and 100% of all expressed molecules in G4), and repeating the process ten times per group size. We applied the repeated-measures ANOVA and Wilcoxon rank-sum tests, which ultimately rejected the initial hypothesis, confirming neither statistically significant differences across group sizes nor moving from one group size to the next in order (e.g., from 10% to 25%; **Fig. 5B**).

Discussion

Several scientific contributions have magnified our understanding of the so-called miRNA Epitranscriptome, igniting the interest in miRNA modifications such as A-to-I RNA editing (10, 16, 17, 38–40) and alternative miRNA cleavage (7, 14, 18, 41–45). However, the concurrent occurrence of A-to-I miRNA editing, DNA modifications (SNPs or somatic mutations), and shifted isomiRs has yet to be widely explored.

In this work, we simultaneously characterized a broader set of microRNA modification types, processing miRNA-seq data from >13,000 adult and pediatric cancer samples across 38 distinct TCGA and TARGET cohorts. Overall, the number of expressed modified miRNAs exceeded by 5-fold the canonical counterpart. The 3′ isomiR resulted in the most predominant miRNA modification type among identified ones, which similarly affected both 5p and 3p arms. The 3′-end showed less stability than the 5′-end (8), with larger trimming of nucleotides, whereas the absence of broader additions (>3 nts) could be due to degradation processes carried out by some enzymes that remove the exceeding part spurting out the RISC complex (46). As known, 5′- and 3′-end regions perform different functionalities, with the first carrying out the miRNA::mRNA partial base-pairing through MSR (1) and the latter playing a role in miRNA::mRNA interaction stabilization (47, 48), especially in the presence of mismatches or bubbles (49). Changes at the 5′-end alter the MSR and may affect the molecule's targetome (50). Meanwhile, the higher number of expressed 3′-end shifted molecules could be attributed to the cell's attempt to modulate miRNAs activity, perhaps trying to overcome the weakness of specific miRNA::mRNA bindings under particular conditions (49).

The A-to-I RNA editing phenomenon may represent another way to alter the MSR and, consequently, the molecules' biological function (10). Noteworthy, 76% of dysregulated edited molecules reported at least one A-to-I editing site within the MSR. The TCGA-LGG cohort presented the highest number of expressed edited miRNA isoforms, which is in line with the literature, as it is well known that such a phenomenon is abundantly present in the nervous system (51).

By contrast, the distribution of the most representative known DNA variant forms (top-five) observed across dysregulated miRNA isoforms mostly falls near the 3′-end (between the 21st and 24th nucleotides), potentially influencing either the miRNA lifespan or the targeting stability (46, 52).

We then explored the underlying differences in the abundance of expressed miRNA isoforms in each cohort/cancer tissue from a functional standpoint. We compared cancer samples with a lower (first quartile) and higher (third quartile) number of expressed miRNA isoforms. A functional analysis displayed the activation/deactivation of several critical pathways involved in proliferation, metastasization, tumor immune escape, invasion, and angiogenesis, such as the Wnt/β-catenin, p38 MAPK, IL6, and HGF signaling pathways. Our results confirm the hypothesis that a different abundance of miRNA isoforms could be associated with the enrichment of, but not limited to, cancer-related pathways as a consequence of a not-necessarily direct dysregulation of crucial genes.

We explored molecules' ability to cluster cancer samples across cohorts, benchmarking four different groups of molecules according to

Figure 4.
miRNA isoforms experimental gene targeting validation. miR-101-3p (**A, C,** and **E–G**) and miR-101-3p (−1|−2) (**B, D,** and **H–J**) experimental targeting validation in lung cancer cells. Expression of miR-101-3p (**A**) and miR-101-3p (−1|−2) (**B**) in normal and tumor samples in TCGA-LUAD cohort. *PTGS2* (**C**) and *DSC2* (**D**) expression in TCGA-LUAD samples in miR-101-3p/miR-101-3p (−1|−2) first (Q1) and third (Q3) quartile. Luciferase assay for psiCHECK-2-PTGS2 3′ UTR WT (**E**) and psiCHECK-2-DSC2 3′ UTR WT (**H**) constructs cotransfected with mirVana miRNA mimics for miR-101-3p, miR-101-3p (−1|−2), and negative scramble miRNA control (SCR) in HEK293 cells performed 24 hours after the transfection. Western blotting depicts the downregulation of PTGS2 (**F** and **G**) or DSC2 (**I** and **J**) proteins in A549 and H1299 cells, respectively, after miR-101-3p and miR-101-3p (−1|−2) overexpression. Densitometric quantification of Western blotting signals (**F, G, I** and **J**) was performed using ImageJ (NIH; https://imagej.nih.gov/ij/, 1997–2018). miR-381-3p (**K, M,** and **O–Q**) and miR-381-3p_4_A_G (**L, N,** and **R–T**) experimental targeting validation in TNBC cells. Expression of both miR-381-3p miRNA isoforms in normal and breast cancer samples in TCGA-BRCA cohort (**K** and **L**). Luciferase assay for psiCHECK-2-UBE2C 3′ UTR WT (**O**) and psiCHECK-2-SYT13 3′ UTR WT (**R**) constructs cotransfected with mirVana miRNA mimics for miR-381-3p, miR-381-3p_4_A_G, and negative scramble miRNA control (SCR) in HEK293 cells performed 24 hours after the transfection. Western blotting represents the downregulation of UBE2C (**P** and **Q**) and SYT13 (**S** and **T**) proteins in MDA-MB-231 and HCC70 cells after miR-381-3p and miR-381-3p_4_A_G upregulation via mirVana miRNA mimic transfection. The histogram reports densitometric quantification of Western blotting signals (**P, Q, S,** and **T**) performed using ImageJ (NIH; 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 value of 3 replicates. $P < 0.05$ was considered statistically significant. Annotations for *, $0.01 \leq P < 0.05$; **, $0.001 \leq P < 0.01$; ***, $P < 0.001$ are provided accordingly. Error bars indicate the three biological replicates' SD. The horizontal bar in each violin-like plot indicates the median. In Western blot experiments. VCL, VINCULIN.
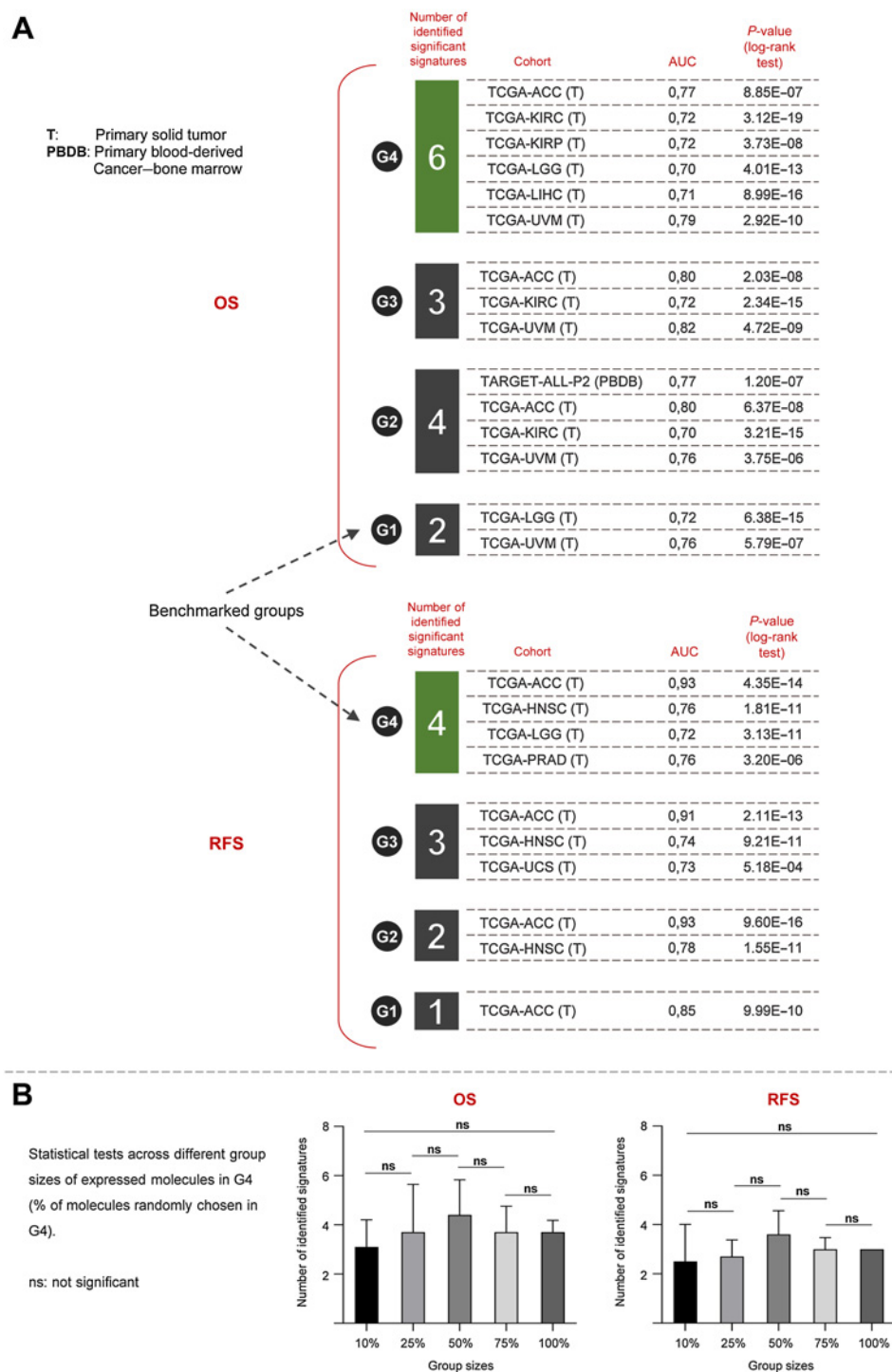
A

B

Statistical tests across different group sizes of expressed molecules in G4 (% of molecules randomly chosen in G4).

ns: not significant

Figure 5.
Overall and RFS risk score–based signatures. **A** and **B,** Overview of risk score–based signatures for OS and RFS (**A**) and results of statistical tests performed across groups of molecules of different sizes (**B**). Panel **A** reports the number of identified signatures per benchmarked group, the AUC, and the P (log-rank test). See Supplementary Fig. S3 and Supplementary Table S5 for more detailed information regarding the workflow used and the list of identified signatures.

specific miRNA modification types. Going from canonical miRNAs (G1) to all expressed canonical and modified miRNAs (G4) allowed us to gain a higher cluster fragmentation that reflected an in-depth clinicopathologic stratification. Altogether, our results depicted a more complex scenario in which canonical and modified miRNAs work together to uncover the underlying histopathologic differences among cancers. Thus, excluding one of the two may substantially limit our understanding of tumor heterogeneity.

Canonical and modified miRNAs were significantly dysregulated across cohorts/tissues. We identified 85 dysregulated canonical miRNAs (out of 392) showing opposite expression trends compared with their modified miRNAs. Among these 85 molecules, we elected to study miR-101-3p and its upregulated shifted isomiR, with one nucleotide added at 5′-end and two trimmed at 3′-end (shortened miR-101–3p ($-1|-2$)), investigating both molecules in TCGA-LUAD. As known, the most robust miRNA::mRNA binding occurs when the

miRNA 5′-end starts with uracil (U), and the opposite nucleotide in the mRNA is adenine (A; known as t1A; ref. 53). Although the shifted isomiR started with guanosine (G) as opposed to the canonical miR-101–3p, which began with U, we demonstrated that both molecules were correctly loaded by AGO2 (Supplementary Fig. S2A and S2B) and targeted unique mRNAs. Our findings are in line with a previous study (28).

Aiming to assess differences in targeting efficiency, we examined dysregulated and predicted gene targets for the two molecules. We chose *PTGS2* (*COX-2*), an oncogene in lung cancer (32), which is a validated canonical miR-101–3p target gene in different cancers (30, 31). Although the predicted binding sites for both miRNA molecules on *PTGS2*'s 3′ UTR were comparable in terms of binding free energy, we experimentally proved that the sole canonical miR-101–3p targeted PTGS2, highlighting the specificity of the miRNA–target gene match. At the same time, DSC2, a protein involved in cell adhesion and recently characterized as a metastasis mediator in lung cancer (33), is exclusively targeted by miR-101–3p (−1|−2). Although the two molecules show opposite expression trends, one could speculate that this does not necessarily translate into different biological functions, as they both target unique oncogenic genes. It might represent a hidden "backup system" of the cancer cells to sidestep the loss of function given the canonical miR-101–3p downregulation.

In the second case study, we assessed the targetome shifting between 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. Although the role of the canonical miR-381–3p is broadly acknowledged as a tumor suppressor (54), particularly in breast cancer (55), very little is known about the edited form in cancer (16, 17). Among the potential targets of the edited miR-381–3p, we elected to investigate *SYT13* in breast cancer (TCGA-BRCA), a well-known oncogene in different tumor models (36, 37). 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 regulate the SYT13 expression exclusively. Afterward, we experimentally validated a unique target (UBE2C) for miR-381-3p to point out the difference in targeting for the two molecules.

Finally, the OS and RFS results highlighted the importance of including all miRNA isoforms (G4) over other groups benchmarked in our study (G1, G2, and G3), which allowed us to identify 6 (OS) and 4 (RFS) significant prognostic signatures. We tested the correlation between using different amounts of molecules and the number of identified signatures. Interestingly, our results ultimately show no significant improvements in varying the number of molecules used, confirming that modified miRNAs may carry biological information.

In conclusion, we can consider modified miRNAs as independent functional molecules that may work in support or contrast to their canonical counterpart. The potential of such molecules as novel diagnostic and prognostic cancer biomarkers is indisputable, making the characterization of the wider miRNAome essential to help uncover miRNA-related mechanisms in cancer.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

R. Distefano: Conceptualization, resources, data curation, software, formal analysis, validation, visualization, methodology, writing–original draft, writing–review and editing. **L. Tomasello:** Resources, validation, investigation, writing–original draft, writing–review and editing. **G.L. Rampioni Vinciguerra:** Validation, writing–original draft, writing–review and editing. **P. Gasparini:** Validation, writing–original draft, writing–review and editing. **Y. Xiang:** Writing–review and editing. **M. Bagnoli:** Writing–review and editing. **G.P. Marceca:** Writing–review and editing. **P. Fadda:** Writing–review and editing. **A. Laganà:** Writing–review and editing. **M. Acunzo:** Writing–review and editing. **Q. Ma:** Writing–review and editing. **G. Nigita:** Conceptualization, resources, formal analysis, supervision, validation, investigation, methodology, writing–original draft, writing–review and editing. **C.M. Croce:** Supervision, funding acquisition, project administration, writing–review and editing.

Acknowledgments

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281–97.
2. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. Cell 2012;148:1172–87.
3. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet 2009;10:704–14.
4. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA 2010;16:2170–80.
5. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucl Acids Res 2014;42:D68–73.
6. Tomasello L, Distefano R, Nigita G, Croce CM. The MicroRNA family gets wider: the IsomiRs classification and role. Front Cell Dev Biol 2021;9:668648.
7. Bofill-De Ros X, Kasprzak WK, Bhandari Y, Fan L, Cavanaugh Q, Jiang M, et al. Structural differences between Pri-miRNA paralogs promote alternative drosha cleavage and expand target repertoires. Cell Rep 2019;26:447–59.
8. Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, et al. 5′ isomiR variation is of functional and evolutionary importance. Nucleic Acids Res 2014;42:9424–35.
9. Saunders MA, Liang H, Li W-H. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A 2007;104:3300–5.
10. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol 2016;17:83–96.
11. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 2007;315:1137–40.

12. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol 2011;12:R126.

13. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007;129:1401–14.

14. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl trans-ferases contributes to miRNA transcriptome complexity. Genome Res 2011;21: 1450–61.

15. Salem O, Erdem N, Jung J, Münstermann E, Wörner A, Wilhelm H, et al. 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 2016;17:566.

16. Wang Y, Xu X, Yu S, Jeong KJ, Zhou Z, Han L, et al. Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. Genome Res 2017;27:1112–25.

17. Pinto Y, Buchumenski I, Levanon EY, Eisenberg E. Human cancer tissues exhibit reduced A-to-I editing of miRNAs coupled with elevated editing of their targets. Nucleic Acids Res 2018;46:71–82.

18. Telonis AG, Magee R, Loher P, Chervoneva I, Londin E, Rigoutsos I. Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. Nucleic Acids Res 2017;45:2973–85.

19. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.

20. Marceca GP, Distefano R, Tomasello L, Lagana A, Russo F, Calore F, et al. MiREDiBase, a manually curated database of validated and putative editing events in microRNAs. Sci Data 2021;8:199.

21. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7.

22. Lu Y, Baras AS, Halushka MK. miRge 2.0 for comprehensive analysis of microRNA sequencing data. BMC Bioinf 2018;19:275.

23. Ibing S, Michels BE, Mosdzien M, Meyer HR, Feuerbach L, Körner C. On the impact of batch effect correction in TCGA isomiR expression data. NAR Cancer 2021;3:zcab101.

24. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:180203426. 2018;

25. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; 1996, pp. 226–31.

26. Distefano R, Nigita G, Croce CM. A concurrent canonical and modified miRNAome pan-cancer study on TCGA and TARGET cohorts leads to an enhanced resolution in cancer. Zenodo; 2022; Available from: https://zenodo.org/record/6902726.

27. Distefano R, Nigita G, Croce CM. Dyregulated miRNA isoforms across TCGA and TARGET cohorts. Zenodo; 2022. Available from: https://zenodo.org/record/6726643.

28. Llorens F, Bañez-Coronel M, Pantano L, del Río JA, Ferrer I, Estivill X, et al. A highly expressed miR-101 isomiR is a functional silencing small RNA. BMC Genomics 2013;14:104.

29. Distefano R, Nigita G, Veneziano D, Romano G, Croce CM, Acunzo M. isoTar: consensus target prediction with enrichment analysis for MicroRNAs harboring editing sites and other variations. In: Laganà A, editor. MicroRNA target identification. New York, NY: Springer New York; 2019. Pp. 211–35.

30. Chakrabarty A, Tranguch S, Daikoku T, Jensen K, Furneaux H, Dey SK. MicroRNA regulation of cyclooxygenase-2 during embryo implantation. Proc Natl Acad Sci U S A 2007;104:15144–9.

31. Hao Y, Gu X, Zhao Y, Greene S, Sha W, Smoot DT, et al. Enforced expression of miR-101 inhibits prostate cancer cell growth by modulating the COX-2 pathway in vivo. Cancer Prev Res 2011;4:1073–83.

32. Liu B, Qu L, Yan S. Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity. Cancer Cell Int 2015;15:106.

33. Li K, Wu R, Zhou M, Tong H, Luo KQ. Desmosomal proteins of DSC2 and PKP1 promote cancer cells survival and metastasis by increasing cluster formation in circulatory system. Sci Adv 2021;7:eabg7265.

34. Xue Y, Xu W, Zhao W, Wang W, Zhang D, Wu P. miR-381 inhibited breast cancer cells proliferation, epithelial-to-mesenchymal transition and metastasis by targeting CXCR4. Biomed Pharmacother 2017;86:426–33.

35. Mo C, Gao L, Zhu X, Wei K, Zeng J, Chen G, et al. The clinicopathological significance of UBE2C in breast cancer: a study based on immunohistochemistry, microarray and RNA-sequencing data. Cancer Cell Int 2017;17:83.

36. Kanda M, Shimizu D, Tanaka H, Tanaka C, Kobayashi D, Hayashi M, et al. Synaptotagmin XIII expression and peritoneal metastasis in gastric cancer. Br J Surg 2018;105:1349–58.

37. Zhang L, Fan B, Zheng Y, Lou Y, Cui Y, Wang K, et al. Identification SYT13 as a novel biomarker in lung adenocarcinoma. J Cell Biochem 2020;121:963–73.

38. Cesarini V, Silvestris DA, Tassinari V, Tomaselli S, Alon S, Eisenberg E, et al. ADAR2/miR-589–3p axis controls glioblastoma cell migration/invasion. Nucleic Acids Res 2018;46:2045–59.

39. Nigita G, Veneziano D, Ferro A. A-to-I RNA editing: current knowledge sources and computational approaches with special emphasis on non-coding RNA molecules. Front Bioeng Biotechnol 2015;3:37.

40. Nigita G, Distefano R, Veneziano D, Romano G, Rahman M, Wang K, et al. Tissue and exosomal miRNA editing in non-small cell lung cancer. Sci Rep 2018; 8:10222.

41. Guo L, Li Y, Cirillo KM, Marick RA, Su Z, Yin X, et al. mi-IsoNet: systems-scale microRNA landscape reveals rampant isoform-mediated gain of target interaction diversity and signaling specificity. Brief Bioinform 2021;22: bbab091.

42. Loher P, Londin ER, Rigoutsos I. IsomiR expression profiles in human lympho-blastoid cell lines exhibit population and gender dependencies. Oncotarget 2014; 5:8790–802.

43. Loher P, Karathanasis N, Londin E, Bray P, Pliatsika V, Telonis AG, et al. IsoMiRmap—fast, deterministic, and exhaustive mining of isomiRs from short RNA-seq datasets. Bioinformatics 2021;37:1828–38.

44. Telonis AG, Loher P, Jing Y, Londin E, Rigoutsos I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. Nucleic Acids Res 2015;43:9158–75.

45. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs – the overlooked repertoire in the dynamic microRNAome. Trends Genet 2012;28:544–9.

46. Katoh T, Hojo H, Suzuki T. Destabilization of microRNAs in human cells by 3′ deadenylation mediated by PARN and CUGBP1. Nucleic Acids Res 2015;43: 7521–34.

47. Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, et al. Differential regulation of microRNA stability. RNA 2010;16:1032–9.

48. Brodersen P, Voinnet O. Revisiting the principles of microRNA target recog-nition and mode of action. Nat Rev Mol Cell Biol 2009;10:141–8.

49. Moore MJ, Scheel TKH, Luna JM, Park CY, Fak JJ, Nishiuchi E, et al. miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nat Commun 2015;6:8864.

50. Van der Kwast RVCT, Woudenberg T, Quax PHA, Nossent AY. MicroRNA-411 and its 5′-IsomiR have distinct targets and functions and are differentially regulated in the vasculature under ischemia. Mol Ther 2020;28:157–70.

51. Yang Y, Okada S, Sakurai M. Adenosine-to-inosine RNA editing in neurological development and disease. RNA Biol 2021;18:999–1013.

52. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S-i, Baba T, et al. Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes Dev 2009;23:433–8.

53. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. Nat Rev Mol Cell Biol 2019;20:21–37.

54. Qiao G, Li J, Wang J, Wang Z, Bian W. miR⊠381 functions as a tumor suppressor by targeting ETS1 in pancreatic cancer. Int J Mol Med 2019;2:593–607.

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