

Type of the Paper (Article)

# Disparities in Lung Cancer: miRNA isoform characterization in lung adenocarcinoma

Rosario Distefano <sup>1,†</sup>, Giovanni Nigita <sup>1,†</sup>, Patricia Le <sup>2</sup>, Giulia Romano <sup>2</sup>, Mario Acunzo <sup>2,‡,\*</sup> and Patrick Nana-Sinkam <sup>2,‡,\*</sup>

<sup>1</sup> Department of Cancer Biology and Genetics and Comprehensive Cancer Center, The Ohio State University, Columbus, 43210 OH, USA.; [rosario.distefano@osumc.edu](mailto:rosario.distefano@osumc.edu) (R.D.); [giovanni.nigita@osumc.edu](mailto:giovanni.nigita@osumc.edu) (G.N.).

<sup>2</sup> Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, 23298 VA, USA. [patricia.le@vcuhealth.org](mailto:patricia.le@vcuhealth.org) (P.L.); [giulia.romano@vcuhealth.org](mailto:giulia.romano@vcuhealth.org) (G.R.); [mario.acunzo@vcuhealth.org](mailto:mario.acunzo@vcuhealth.org) (M.A.); [patrick.nana-sinkam@vcuhealth.org](mailto:patrick.nana-sinkam@vcuhealth.org) (P.N.-S.).

\* Correspondence: [mario.acunzo@vcuhealth.org](mailto:mario.acunzo@vcuhealth.org); [patrick.nana-sinkam@vcuhealth.org](mailto:patrick.nana-sinkam@vcuhealth.org).

† These authors share equal contribution.

‡ These authors share equal senior authorship.

Received: date; Accepted: date; Published: date

**Simple Summary:** Post-transcriptional modification events in miRNA molecules have revealed a more complex layer to cancer biology. Such modifications have created a novel path to developing and testing potential cancer biomarkers. Here, we concurrently profiled canonical and non-canonical miRNA molecules in White American (W) and Black or African American (B/AA) lung adenocarcinoma (LUAD) patients from The Cancer Genome Atlas (TCGA) cohort. We identified distinct potential post-transcriptional modifications in lung cancer tissues from W versus B/AA patients. Our results suggest the relevance of miRNA isoforms as potential biomarkers in lung cancer.

**Abstract:** Despite the development of targeted therapeutics, immunotherapy, and strategies for early detection, lung cancer carries a high mortality. Further, significant racial disparities in outcomes exist for which the molecular drivers have yet to be fully elucidated. The growing field of Epitranscriptomics has introduced a new layer of complexity to the molecular pathogenesis of cancer. RNA modifications can occur in coding and non-coding RNAs, such as miRNAs, possibly altering their gene regulatory function. The potential role for such modifications as clinically informative biomarkers remains largely unknown. Here, we concurrently profiled canonical miRNAs, shifted isomiRs (templated and non-templated), miRNAs with single-point modification events (RNA and DNA) in White American (W) and Black or African American (B/AA) lung adenocarcinoma (LUAD) patients. We found that while most deregulated miRNA isoforms were similar in W and B/AA LUAD tissues compared to normal adjacent tissues, there was a subgroup of isoforms with deregulation according to race. We specifically investigated an edited miRNA, miR-151a-3p with an A-to-I editing event at position 3, to determine how its altered expression may be associated with activation of divergent biological pathways between W and B/AA LUAD patients. Finally, we identified distinct race-specific miRNA isoforms that correlated with prognosis for both Ws and B/AAs. Our results suggest that concurrently profiling canonical and non-canonical miRNAs may have potential as a strategy for identifying additional distinct biological pathways and biomarkers in lung cancer.

**Keywords:** miRNA isoform, isomiR, miRNA editing, lung adenocarcinoma, race-disparities.

## 1. Introduction

In the United States, lung cancer remains the leading cause of cancer death in males and females. In 2020, the American Cancer Society reported estimates of 235,000 new cases and 130,000 deaths from lung and bronchus cancers [1]. Annual incidence and mortality rates are further segregated based on race, ethnicity, and sex [2]. Lung cancer disproportionately affects Black or African American (B/AA) males, exemplified by higher incidence and mortality rates than in other reported racial and ethnic groups [2,3]. B/AAs tend to present with later-stage disease, while those diagnosed with early-stage disease are more likely to refuse definitive surgical resection [4,5]. Furthermore, B/AAs are less likely to meet the United States Preventive Services Task Force (USPSTF) eligibility criteria for lung cancer screening due to being diagnosed with lung cancer at younger ages and with a lower pack-per-year smoking history upon diagnosis; this observation is referred to as “Smoker’s Paradox” [4,6,7]. Overall, the intersectionality of many social, environmental, economic factors and, ultimately, potential genetic factors contribute to these racial disparities in lung cancer [2,4–6,8].

The emergence of large-scale sequencing technologies for profiling the human genome has allowed for comparative analyses across patient groups. Lower frequencies of all lung cancer driver mutations, along with higher tumor mutational burden, are found in B/AAs vs. whites (Ws) [9–11]. Whole-exome sequencing of B/AA non-small cell lung cancer (NSCLC) patients without a known driver mutation revealed 47 novel and potentially damaging nonsynchronous mutations that were not found in The Cancer Genome Atlas (TCGA) [12]. This suggests that broadening current mutation profiling to encompass these novel driver genes may elucidate new targets for guiding clinical decisions [12].

Although several studies have examined the genomic landscape of lung cancers across racial groups, few studies so far have investigated for alterations in the non-coding transcriptome [7], such as miRNA isoforms [13]. Recent advances in high-throughput sequence technology have revealed an additional layer of complexity within miRNA molecules [14], paving the way for the field of “Epitranscriptomics” of ncRNAs in cancer [15]. The human genome predominantly encodes for non-coding RNAs (ncRNAs), such as microRNAs (miRNAs) [16,17]. MiRNAs are single-stranded RNAs, ~22 nucleotides in length, involved in post-transcriptional gene regulation primarily by RNA interference mechanism [14]. During miRNA biogenesis, multiple cleavage events are catalyzed by RNases (e.g., Drosha, Dicer) to process primary miRNAs (pri-miRNA) into a mature miRNA duplex [18]. High-throughput sequencing has revealed that various miRNA isoforms (isomiRs) may be derived from the same precursor miRNA (pre-miRNA) [18]. These isoforms appear from alternative processing by Drosha and Dicer, as well as post-transcriptional modifications or “RNA editing” events, which include non-templated nucleotide additions (NTA) (e.g., adenylation, uridylation) and nucleotide substitutions (e.g., cytosine-to-uridine/C-to-U, adenosine-to-inosine/A-to-I) [18–21]. Moreover, when these edits occur within the functional 5’ seed region of miRNAs (from nucleotides 2-7), miRNA-mRNA target interactions are altered [22,23], thereby creating new patterns of targeting that have been shown to affect downstream biology [18,19,21,23–26].

The spatiotemporal nature of RNA editing can provide insight into the dynamic nature of human cancers. [24,27]. Generally, human cancers display

hypo-editing of miRNAs, consequently upregulating their targets [24,26]. Expression of miRNA isoforms in triple-negative breast cancer (TNBC) tissue samples from a TCGA dataset between B/AA and W patients demonstrated differences in isoforms from loci of known oncogenic miRNAs (oncomiRs) (e.g., miR-200c, miR-21, miR-17/92 family, miR-183/96/182 family) [13]. This differential expression of isomiRs may alter the dynamics of metastatic pathways (e.g., P38, Wnt/ $\beta$ -catenin), contributing to the aggressive phenotype of TNBC [13].

The conventional approach to studying miRNA expression is to cluster miRNA isoforms within a shifting range of a few nucleotides (miRNA-arms). Unlike this approach, we concurrently profiled canonical miRNAs, shifted isomiRs (templated and non-templated), miRNAs with single-point modification events (e.g., A-to-I RNA editing, SNPs, and nucleotide insertions) in lung adenocarcinoma (LUAD) samples from TCGA data sets. In particular, given the racial disparities prevalent in lung cancer, we assessed the differential expression of canonical and non-canonical miRNAs between B/AA and W from the LUAD-TCGA cohort. Among the deregulated miRNA isoforms, we elected to investigate the edited form of miR-151a-3p, with an A-to-I editing event in the seed region at position three, in W versus B/AA LUAD tissues. Here, we demonstrate, based on pathway enrichment analysis, that edited miR-151a-3p has a potential differential biological effect in B/AA versus W LUAD lung cancers.

Moreover, we identified additional miRNA isoform expression patterns that significantly and distinctly stratify patients in overall survival (OS) and relapse-free survival (RFS) in B/AA versus W patients with lung adenocarcinoma. These results suggest that racial intragroup and intergroup differences in miRNA isoform expression may serve as additional classifiers in the development of clinically meaningful biomarkers.

## 2. Results

### 2.1. Systematic miRNA isoform characterization in Black or African and White American lung adenocarcinoma patients

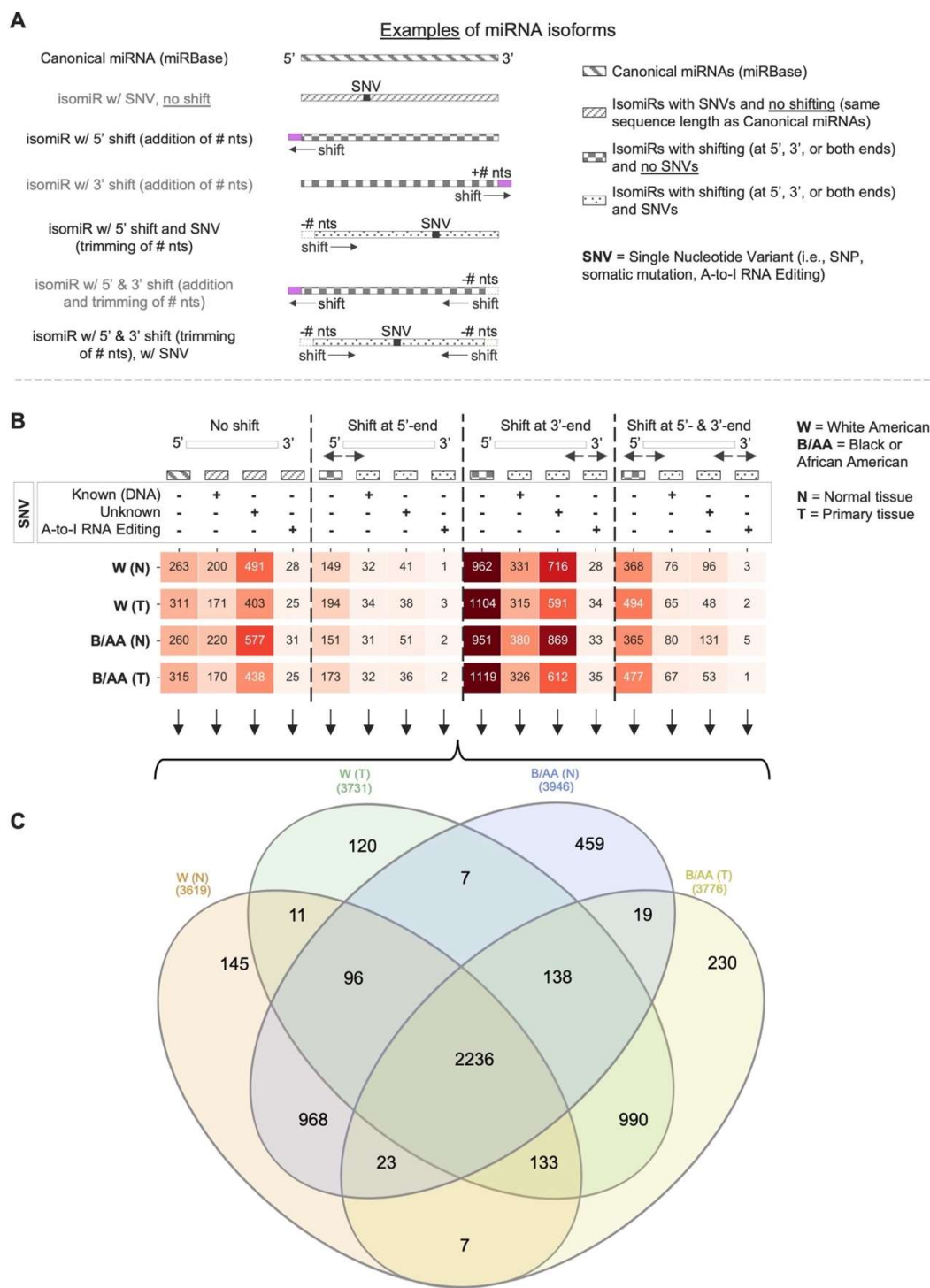
To investigate expression patterns of miRNA isoforms in both B/AA and W lung adenocarcinoma patients, we analyzed small-RNA sequencing (sRNA-seq) sample data from the LUAD-TCGA cohort. In particular, we examined data from 485 tissues, precisely 58 B/AA (52 primary solid tumors and six adjacent normal tissues) and 427 W (387 primary solid tumors and 40 adjacent normal tissues) samples. The cohort characteristics are summarized in Table S1.

In order to concurrently detect and quantify canonical miRNAs, miRNAs with single nucleotide variant (SNV) events (e.g., Single Nucleotide Polymorphisms (SNPs), A-to-I RNA editing, and nucleotide insertions), and templated/non-templated shifted isomiRs (see Figure 1A), we performed an in-house workflow (see Figure S1 and Materials and Methods section) which implements the miRge 2.0 pipeline [28]. We also created a specific nomenclature for the detected miRNA isoforms based on the miRGe3 [29] format (see Figure S2 and Materials and Methods for more details). SNVs occurring in miRNA sequences were further annotated with A-to-I RNA editing sites (MiREDiBase database [30]), somatic mutations (COSMIC database [31], and TCGA and TARGET cohorts), and known SNPs (dbSNPs) [32]. Furthermore, we employed a filtering phase, discarding miRNA isoforms with a not-annotated SNV in the first or last two nucleotides (see

Materials and Methods for more details). Finally, we identified a total of 5582 expressed miRNA isoforms (see Figure 1B, Table S2 and Materials and Methods section for more details) in the LUAD-TCGA cohort (3785 in W normal tissues, 3832 in W tumor tissues, 4137 in B/AA normal tissues, and 3881 in B/AA tumor tissues). Noteworthy, one of the expressed molecules is a novel miRNA, labeled as *miR-n47-3p* (Genomic coordinates: chr19, 13836509-13836528, *negative* strand).

In all four sample groups, the miRNA isoforms with a shift in the 3'end with no SNVs are the most representative ones, which are >3.5 times more than the canonical miRNA molecules (Figure 1B). Interestingly, considering the expression of each of 16 miRNA isoform categories, the class of the miRNA isoforms with a 3'end shifting with no SNVs had an expression comparable to the expression of canonical molecules (see Figure S3 and Table S2), rendering these isoforms attractive from a functional standpoint.

Furthermore, intersecting the expressed miRNA isoforms identified in each group sample, we found >2.2K miRNA isoforms in common (Figure 1C). Moreover, we identified some miRNA isoforms which are specifically expressed in each group: 145 in W(N), 120 in W(T), 459 in B/AA(N), and 230 in B/AA(T). Interestingly, our results show that B/AA samples have more miRNA isoforms distinctly expressed than W samples, highlighting a disparity between the two races, which, interestingly, is not evident if we take into consideration only canonical molecules (see Figure S4).



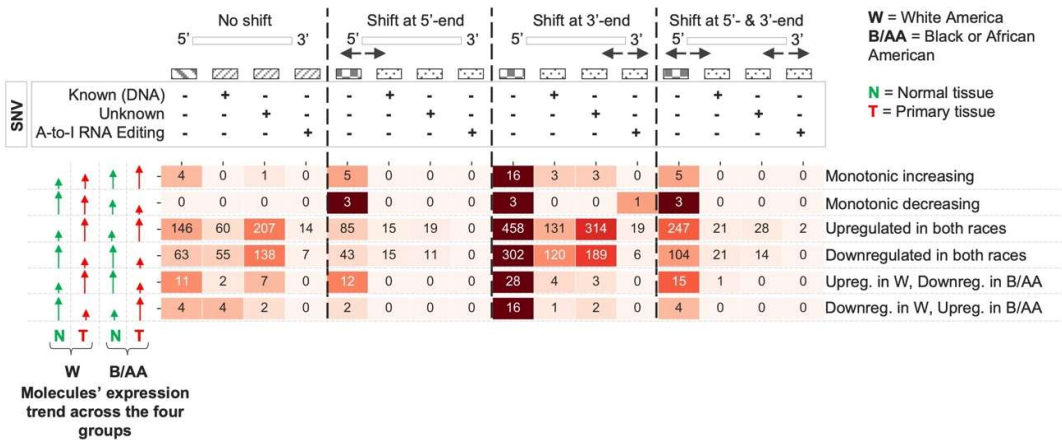
**Figure 1.** Canonical and non-canonical miRNAome profiling in lung adenocarcinoma. **A)** Examples of miRNA isoforms detected from sRNA-seq data. **B)** Distribution of expressed miRNA isoforms in primary solid tumor (T) and normal (N) tissues from both W and B/AA LUAD patient samples. The heatmap shows the number of expressed miRNA isoforms (canonical and non-canonical ones) for each of the four miRNA isoform categories (no shift, shift at 5'-end, shift at 3'-end, and shift at 5'- & 3'-end). **C)** Venn diagram indicates the commonly expressed miRNA isoforms between the four group samples: W (N), W(T), B/AA (N), and B/AA (T).



2.2. miRNA isoform dysregulation in lung adenocarcinoma race disparities

To identify potential dysregulated miRNA isoforms in lung adenocarcinoma lung across races, we performed a multivariate analysis comparing primary solid tumor (T) and normal adjacent (N) tissues from both W and B/AA LUAD patient samples. The raw read counts of the expressed miRNA isoforms ([RPM geometric mean]>1 Reads Per Million across the four sets of samples) were first normalized, and then we performed a multivariate analysis (see Materials and Methods for more details), considering four distinct sample groups: I) W Normal, II) W Tumor, III) B/AA Normal, and IV) B/AA Tumor tissue samples.

We obtained a total of 2935 deregulated miRNA isoforms (FDR<.05), which we subdivided into six different trends, based on the trend in the four groups mentioned above (see Figure 2 and Table S3): I) *monotonic increasing*, meaning that the average (specifically, the geometric mean) miRNA isoforms increase along four tissue types (from W Normal to B/AA Tumor tissue samples); II) *monotonic decreasing*, indicating that the average of miRNA isoforms decrease along the four tissue types; III) *upregulated in both races*, that are the miRNA isoforms upregulated in both W and B/AA patients, and are not *monotonic increasing*; IV) *downregulated in both races*, namely the miRNA isoforms downregulated in both W and B/AA patients, and they are not *monotonic decreasing*; V) *Upreg. in W, Downreg. in B/AA*, considering the miRNA isoforms upregulated in W patients and, at the same time, downregulated in B/AA patients; VI) *Downreg. in W, Upreg. in B/AA*, representing the miRNA isoforms downregulated in W patients and, at the same time, upregulated in B/AA patients. Our results show that upregulated and downregulated miRNA molecules in both races are the most representative trends (>94%). Nevertheless, a few miRNA molecules (~6%, primarily non-canonical ones) fall into one of the four remaining trends (monotonic or opposite trends), showing specific differences between W and B/AA.



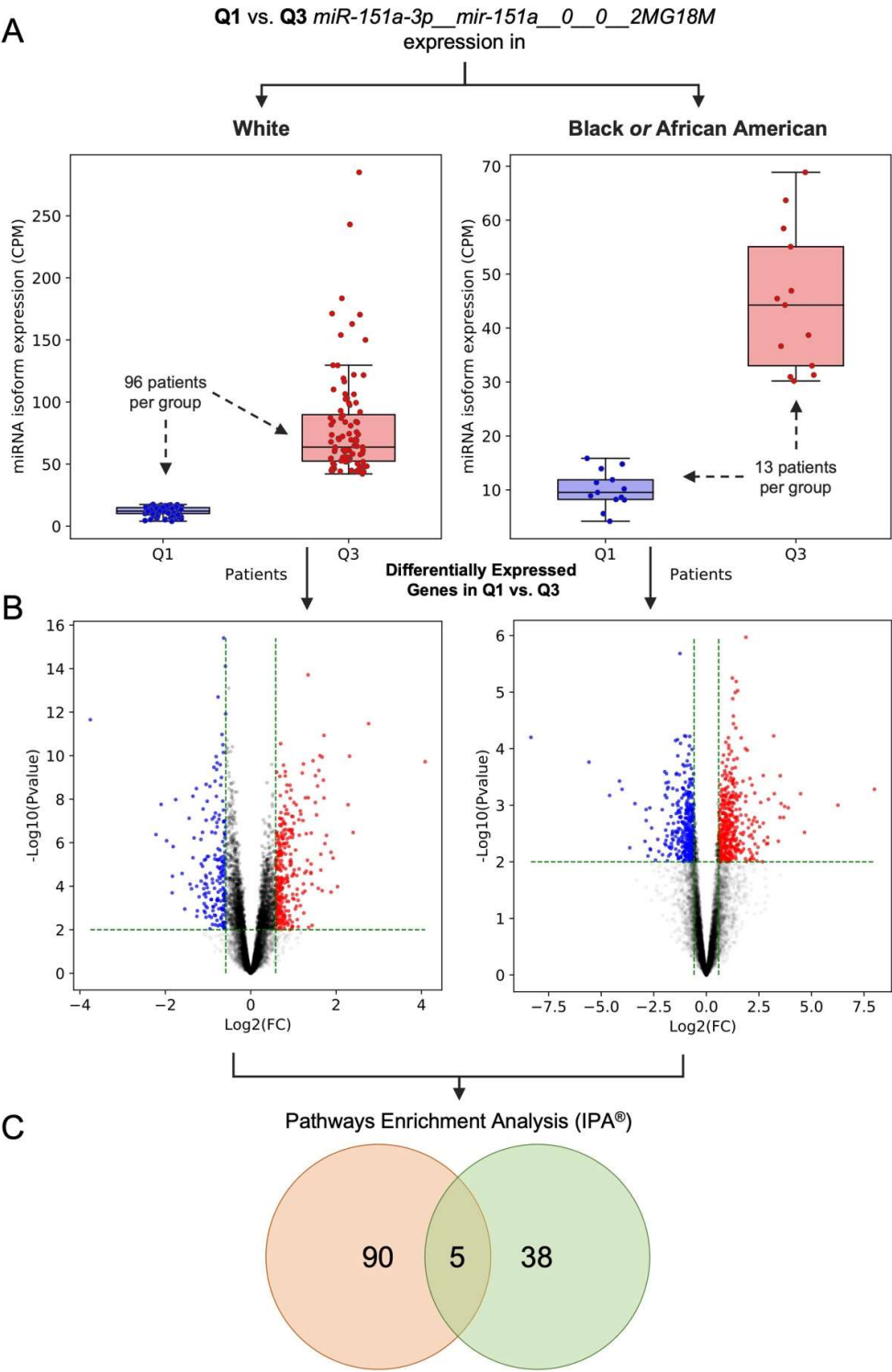
**Figure 2.** Deregulated miRNA isoforms in both W and B/AA lung adenocarcinoma patient samples. The heatmap shows the number of deregulated miRNA isoforms (canonical and non-canonical ones) for each of the four miRNA isoform categories (no shift, shift at 5'-end, shift at 3'-end, and shift at 5'- & 3'-end). Each trend is defined based on the geometric mean of each group.

### 2.3. Dysregulation of edited miR-151a-3p in Black or African and White American lung adenocarcinoma patients: a case of study

To study the potential impact of deregulated non-canonical miRNAs on downstream targets, we examined the edited form of miR-151a-3p with an A-to-I editing event in position 3. MiR-151-3p is highly expressed in all four distinct tissue types, compared to the other miRNA isoforms (W(N)~89<sup>th</sup> percentile, W(T)~87<sup>th</sup> perc., B/AA(N)~90<sup>th</sup> perc., and B/AA(T)~89<sup>th</sup> perc.). Moreover, as shown in Table S3 and Figure S5, it was down-regulated in both W and B/AA tumor samples, although more marked in B/AA tumor samples (based on the geometric mean of each group). We sought to determine if the edited miRNA had a differential impact, even though the trend in expression was similar in both races. Interestingly, downregulation of edited miR-151a-3p correlated with the downregulation of the adenosine deaminase RNA specific B1 (ADARB1 or ADAR2) gene, but not with the ADAR (ADAR1) gene (see Figure S5), which is consistent with previous observations that ADAR2 deregulation could drive such editing events in miR-151a-3p [26,33,34]. In addition, the edited miRNA molecule has an A-to-I editing event in position 3 (in the miRNA seed region - MSR), which could result in a downstream targetome distinct from that of the conical miR-151a-3p, as is the case in other miRNAs that we and others we have demonstrated in the past [22,23,35].

In order to uncover the potential downstream impact of the edited miR-151a-3p in both W and B/AA patient samples, we first subdivided the tumor samples (for both races) into two distinct groups based on the low (first quartile, Q1) and high (third quartile, Q3) expression of the miRNA molecule (see Figure 3A). We then employed a differential expression analysis of genes, comparing Q1 vs. Q3 groups (see Figure 3B and Table S4). As we were interested in the impact on gene expression of the miRNA molecule, we considered all those significant D.E. genes ( $p\text{-value} < .01$  and  $|\text{linear fold change}| > 1.5$ , see Materials and Methods section for more details), which potentially are direct or indirect targets. Subsequently, we performed a pathway enrichment analysis using the Ingenuity® Pathway Analysis (IPA) software (see Figure 3C and Table S4) based on the D.E. genes.

In summary, we demonstrate that the expression of edited miR-151a-3p is associated with different canonical pathways comparing the two patient groups (W and B/AA LUAD patients). Indeed, only a few canonical pathways are shared between the two population groups (see Figure 3C).



**Figure 3.** Different gene and pathway dysregulation related to A-to-I (in position 3) edited miR-151a-3p expression in LUAD race disparities. **A)** Box plots indicate the expression of edited miR-151a-3p in both W and B/AA tumor samples with low (first quartile, Q1) and high (third quartile, Q3) expression of the miRNA molecule. **B)** Scatter plots show the D.E. genes ( $p\text{-value} < .01$  and  $|linear\ fold\ change| > 1.5$ ) related to the expression edited miRNA (Q1 vs. Q3). **C)** Venn diagram indicates the common significant enriched pathways (based on the D.E. genes) between W and B/AA patient samples.



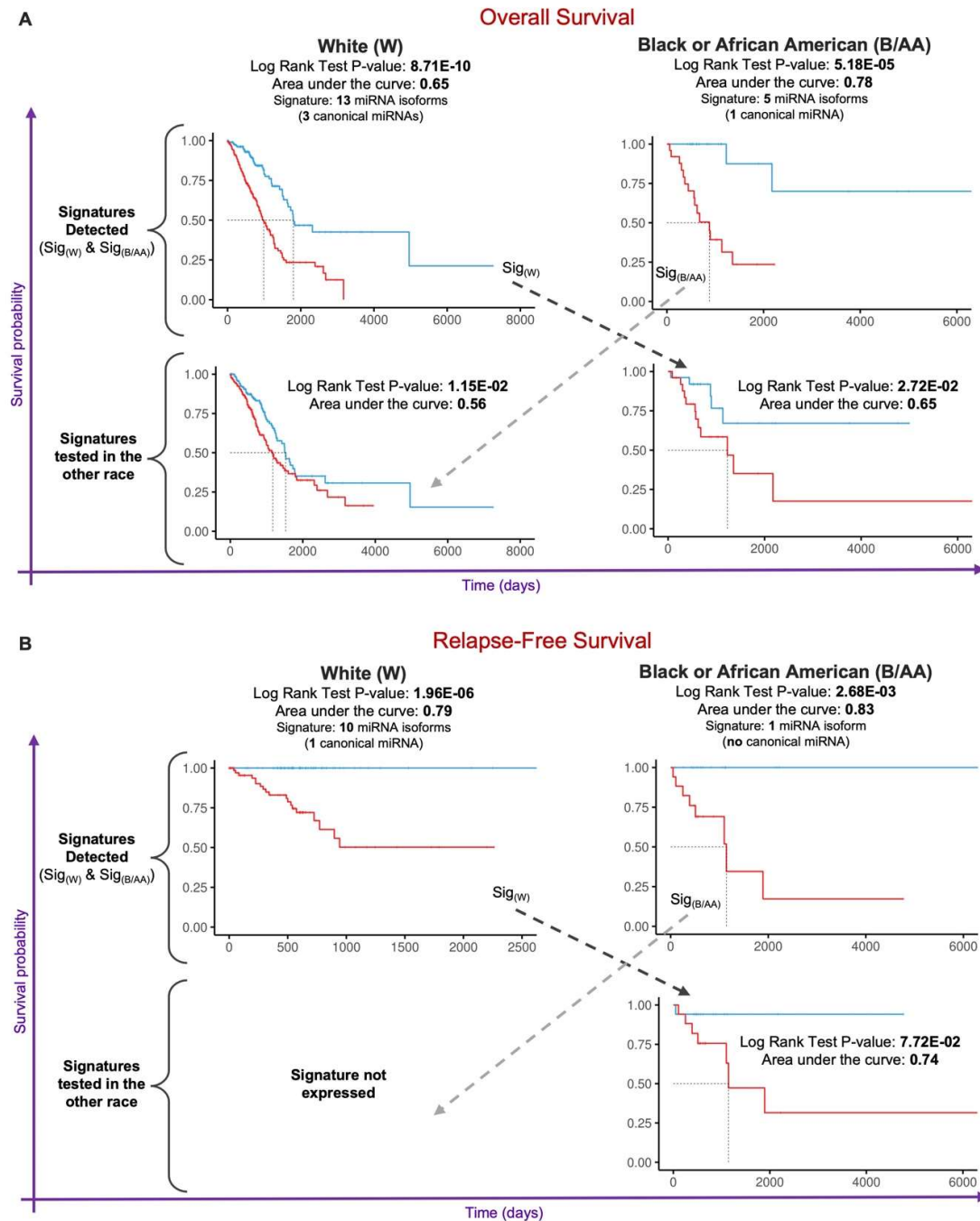
#### 2.4 Prognostic miRNA isoform signature in lung adenocarcinoma race disparities

Considering the significant number of expressed miRNA isoforms identified through the concurrent proofing of canonical and non-canonical miRNAs (see Figure 1B), we decided to investigate the potential of such miRNA isoforms as clinical biomarkers. To accomplish this goal, we designed a two-stage workflow (see Figure S6 and Materials and Methods) to identify prognostic signatures for Overall (OS) and Relapse-Free (RFS) survival.

As shown in Figure 4A, we identified two different signatures for OS, with 13 miRNA isoforms (including three canonical miRNAs) for W LUAD patients and five miRNA isoforms (including one canonical miRNA) for B/AA LUAD patients (see Table S5). We have also tested these prognostic signatures in the other race group, finding that the 13-miRNA-isoform signature identified in W LUAD patients carried similar accuracy for B/AA LUAD samples (the AUC is 0.65 for both race groups); in contrast, the 5-miRNA-isoform signature identified in B/AA LUAD patient appeared to be more accurate for B/AAs (AUC=0.78) compared to Ws (AUC=0.56).

Similar to the OS analyses, we identified two distinct miRNA isoform signatures for RFS for W and B/AA LUAD patients (see Figure 4B). In particular, an RFS signature harboring ten miRNA isoforms (including a canonical miRNA) for W LUAD patients and a signature for B/AA LUAD (see Table S5). Testing the RFS signatures in the other group, we determined that the ten-miRNA-isoform signature for W LUAD patients (AUC=0.79) had similar accuracy for B/AA LUAD patients (AUC=0.74); however, the miRNA isoform signature for B/AA LUAD patients was not expressed within W LUAD patient tissues and thus, was not testable as a potential biomarker.

In conclusion, through the analysis of both canonical and non-canonical miRNA molecules, we identified distinct patterns of expression of isoforms between B/AA and W LUAD patients that harbor clinical significance. These results suggest that investigation of miRNA isoforms may serve as an additional avenue for identifying clinically informative biomarkers and, more importantly, elucidating novel cancer biology.



**Figure 4.** miRNA isoform prognostic signatures in lung adenocarcinoma across race, supplied with significance (Log Rank test pvalue) and the area under the curve (AUC) score. **A)** miRNA isoforms signatures for overall survival in W and B/AA patient samples. **B)** miRNA isoform signatures for relapse-free survival in W and B/AA patient samples. The bottom part of each panel shows the results of the signatures test in the other race.

### 3. Discussion

Disparities in lung cancer diagnosis and treatment have contributed to widening the gap in care and outcomes [2]. In particular, Black or African American (B/AA) males are disproportionately burdened by this disease, as evidenced by higher incidence and mortality rates [2,3]. The current USPSTF criteria for lung cancer screening is based on age and smoking history, but B/AAs are less

likely to meet the criteria [4–7]. B/AAs are generally diagnosed at a later stage disease, at a younger age, and with a lower pack-per-year smoking history when compared to White Americans (W) [4–7]. Although cigarette smoking is a major risk factor for the development of lung cancer, it is clear that socioeconomic and epigenetics factors contribute to disparities associated with this disease. Therefore, we and others propose that a more comprehensive approach needs to be developed for lung cancer risk stratification, detection, and treatment in underrepresented populations.

The advent of high throughput sequencing platforms has driven the discovery of non-coding RNAs (ncRNAs), such as microRNAs (miRNAs) [16,17], as both drivers of cancer biology and clinical biomarkers. The dysregulation of miRNA expression has been well-documented in all human cancers. Previous studies have mainly focused on evaluating canonical miRNAs, which represent excellent diagnostic cancer biomarkers if they are considered as single molecules instead of clusters. One limitation of this approach is that only a few hundred canonical miRNAs are expressed. In light of this, we concurrently profiled both canonical and non-canonical miRNA molecules to significantly expand the number of potential cancer biomarkers from several hundred to thousands. The inclusion of non-canonical miRNAs could significantly improve the resolution of potential bio-classifiers. This will be particularly valuable for elucidating subtle molecular differences that may contribute to lung cancer disparities.

Here, we profiled canonical miRNAs, shifted isomiRs (templated and non-templated), and miRNAs with single nucleotide variant (SNV) events (e.g., single nucleotide polymorphisms (SNPs), A-to-I RNA editing, nucleotide insertions) in lung adenocarcinoma (LUAD) samples coming from the TCGA (see Figure 1). We conducted differential expression analysis in order to compare the levels of miRNA isoforms between B/AA and W LUAD patients (see Figure 2). We elected to study the function of edited miR-151a-3p, which undergoes an A-to-I editing site in the seed region at position 3. This approach was based on similar studies by us and others who have previously investigated the functional consequences of miRNAs that possess an editing site within the seed region [22,23,35]. Our results here demonstrate that the differential expression of a single miRNA isoform (edited miR-151a-3p) in B/AAs and Ws may be associated with a distinct set of pathways (see Figure 3). However, not all sequence shifts observed between miRNA isoforms will alter the original function. Instead, these isomiRs can serve as candidate classifiers. Therefore, we sought to identify miRNA isoforms that classify patients in terms of overall survival (OS) and relapse-free survival (RFS). Intriguingly, we have found two different prognostic signatures of OS and RFS that could distinguish B/AAs but not Ws LUAD patients (see Figure 4), illustrating the importance of considering racial differences in miRNA isoform expression when developing clinical classifiers.

Altogether, the results obtained from these studies demonstrate the potential of profiling a wider miRNAome. We conclude that concurrently profiling canonical and non-canonical miRNAs may provide additional insights into cancer biology and clinically useful biomarkers. We recognize the limitations of the current study, including, most importantly, the relatively low number of cases of B/AA lung cancer within the TCGA database, the possibility that observed differences in isoforms may be driven by factors other than race, and the need for independent validation of any clinical classifier. Nevertheless, the encouraging findings from our study suggest that a more comprehensive approach to miRNA profiling that incorporates both canonical and non-canonical forms may prove valuable in identifying additional molecular mechanisms of lung cancer.

## 4. Materials and Methods

### 4.1 Data acquisition and preparation

sRNA-seq sequencing BAM files of 485 lung adenocarcinoma tissue samples (TCGA-LUAD project, with 439 primary solid tumors and 46 adjacent normal tissues) were downloaded via the GDC data portal (<https://gdc-portal.nci.nih.gov>). Before the detection analysis, we converted BAM files into FASTQ ones (which contain raw, unaligned reads with quality scores) by using the *bamtofastq* function from *BEDTools* (v2.25.0) package [36]. Finally, all the raw sequence data were quality-filtered using the *ConDeTri* tool (v2.3) [37], with the following parameters: *-pb=fq -lq=20 -hq=30 -minlen=15 -sc=33*.

We also downloaded raw read count and FPKM values (where FPKM means *Fragments Per Kilobase Million*) of genes detected from RNA-seq samples (matched with the sRNA-seq samples described above) from the TCGA-LUAD project via the GDC data portal.

#### 4.2 MiRNA Isoforms Mapping, Annotation, and Filtering

Quality-filtered sequences were aligned to the human genome (hg38) and, subsequently, annotated via an in-house designed workflow (Figure S1). The mapping and annotation phases leveraged on miRge 2.0 [28] tool, a pipeline for canonical miRNAs/miRNA isoforms annotation based on miRGeneDB 2.0 [38] and the miRBase (v22) [39]. The miRge 2.0 pipeline was executed using the default parameters, performing the *prediction* module to discover potential novel miRNA molecules. As illustrated in Figure 1A, annotated miRNA molecules include canonical miRNAs, shifted miRNAs, and miRNA with single nucleotide variants (SNVs), such as A-to-I RNA Editing sites and Single Nucleotide Polymorphisms (SNPs), and somatic mutations. After the annotation phase, the annotated miRNA molecules were enriched with additional information, such as known SNPs (dbSNP v154 [32]), somatic mutations (COSMIC database v92 [31], and TCGA and TARGET cohorts), and A-to-I RNA Editing sites from MiREDiBase (v1.0) database [30]. Moreover, to eliminate potential imperfection in the linker ligation during the cDNA library construction or sequencing errors [40–42], we discarded all those miRNA isoforms with at least a not-annotated SNV in the first or last two nucleotides.

The miRNA isoform nomenclature uses the information from the mirGFF3 format [29], which was extended with other features. More details about the miRNA isoform nomenclature are illustrated in Figure S2.

Finally, data were collected into tab-separated text files, considering those miRNA isoforms with an expression of [RPM geometric mean]>1 (where RPM means *Reads Per Million*) for the downstream analyses.

#### 4.3 miRNA isoform quantification and multivariate analysis in LUAD race disparities

Raw read counts of the expressed miRNA isoforms were normalized via the *trimmed mean of M-values* (TMM) [43] method by employing the *calcNormFactors* function available in *edgeR* (v3.32.1) [44,45] R (v3.4.4) package from *BioConductor* (v3.6) [46].

In the determination of the expressed miRNA isoforms for each group (W (N), W (T), B/AA (N), and B/AA (T)), the read counts of the miRNA isoforms were normalized with the TMM method separately for each group (see Figure 1B and 1C).

Concerning the multivariate analysis, using the normalized data (normalizing all the four sample groups together), we performed a Kruskal-Wallis test comparing the four LUAD sample groups (see Figure 2). Finally, each pvalue was adjusted by Benjamini-Hochberg's correction method [47].

#### 4.4 Differential gene expression and functional enrichment analyses

Firstly, for both W and B/AA LUAD patients, we grouped samples according to the first (Q1) and third (Q3) quartile of edited miR-151a-3p to identify deregulated genes related to the expression edited miRNA molecule (Q1 vs. Q3) considered in our study. Before the normalization step, low expressed genes were filtered out, retaining those genes with a minimum expression of [FPKM geometric mean]>1. Subsequently, the set of expression-filtered genes was used to extract the corresponding raw read counts and then normalized via TMM [43]. Finally, for both W and B/AA LUAD patients, the differential expression analyses were performed by using *edgeR*. All those dysregulated expressed (D.E.) genes with  $p\text{-value} < .01$  and  $|\text{linear fold change}| > 1.5$  were considered for the functional enrichment analyses (see Table S4). Later, a pathway enrichment analysis was performed employing the Ingenuity® Pathway Analysis (IPA®) software (v01-16) based on the D.E. genes. All those pathways with a  $p\text{-value} < .01$  were retained (see Table S4).

#### 4.5 Prognostic signature identification

The right-censored data were considered for the patient's survival time for overall (OS) and relapse-free (RFS) survival analyses. The prognostic miRNA isoform signature workflow consists of two main stages, as shown in Figure S6.

In the first stage, the patients' miRNA isoforms expression data are assembled, starting by whether the sample group (either W or B/AA patients) has at least *ten* patients per event type: *dead* and *alive* for the OS and *relapse* or *not-relapse* for RFS. Subsequently, the raw read counts are taken for the selected patients and then normalized, with the method described in Section 4.4. Finally, a correlation matrix is calculated out of the normalized data.

The second stage starts with removing highly correlated miRNA isoforms, retaining those miRNA molecules with a correlation lower than a specific *threshold*, ranging from 0.6 to 0.8 (with incremental steps of 0.01). At this point, a first step generates a univariate Cox proportional hazard regression model, assessing the relationship between miRNA isoforms expression and patients' survival. Each *Cox p-value* is adjusted using Benjamini-Hochberg's method. Subsequently, a multivariate Cox proportional hazard regression model is performed, based on the most significant miRNA isoforms (*Cox FDR* < .05) coming from the univariate model. Both univariate and multivariate Cox regression models leverage the *CoxPHFitter* function from the *lifelines* (v0.25.4) Python package. The list of miRNA isoforms present in this stage is narrowed down by keeping those significant miRNA molecules (with a *Cox p-value* < .05 according to the multivariate model) and then applying a feature selection strategy. The selection strategy consists of the Recursive Feature Elimination method from the *scikit-learn* (v0.22.1) Python package, using a Logistic Regression Model used as the estimator. Successively, for each patient, a *Risk Score* is computed by linearly combining the expression value and the regression coefficient (univariate model) associated with the reduced set of miRNA isoforms (selection strategy), as defined below:

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

In which *n* is the number of reduced miRNA isoforms, *exp* indicates the expression, and  $\beta$  represents the regression coefficient (from the univariate Cox model). Patients are then separated into high- and low-risk groups using the median of Risk Scores as a cutoff. Finally, the last step ends the Kaplan-Meier plot is generated, calculating the p-value (Log Rank Test) and a prognostic signature accuracy (area under the curve - AUC) based on the two risk groups. The Kaplan-Meier plots were generated using the *survfit* and *ggsurvplot* functions from the *survival* (v3.2-3) R package. The prognostic signature with the highest AUC is selected at the end of the workflow.

#### 4.6 Statistical analyses and graphs

Venn diagrams in Figures 1B, 3C, and S3 were created by using *InteractiVenn* webtool [48]. The heatmaps present in Figures 1 and 2 were generated using *matplotlib* and *seaborn* Python libraries. Regarding the heatmaps, a miRNA molecule could be counted more than one time if it has more types of SNVs (e.g., known DNA variant, A-to-I RNA editing event, or unknown SNV).

## 5. Conclusions

In summary, we profiled miRNA isoforms in 485 lung adenocarcinoma (LUAD) samples, specifically from 427 White American and 58 Black or African American patient donors, to investigate the role of miRNA isoform in lung cancer race disparities. Unlike conventional methods, we concurrently profiled canonical and non-canonical miRNA isoforms, discovering more miRNA isoforms expressed explicitly in B/AA samples compared with Ws. Most of the deregulated miRNA isoforms deregulated LUAD samples conserve the same trend (up-or down-regulation) between the two race groups, and just a few miRNA molecules have an exclusive race-related trend. We have also analyzed an expressed edited miRNA as a case study, showing how its expression related to different pathways between W and B/AA LUAD patients.

Moreover, based on miRNA isoform molecules, we have discovered two distinctive overall survival signatures, one W and one for B/AA LUAD patients, and two different relapse-free survival



signatures, one for both race-group patients. In particular, the prognostic signatures are unique to B/AA LUAD patients, suggesting that these miRNA isoforms require further examination and validation as potential race-specific prognostic biomarkers in lung cancer.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: Data processing workflow of LUAD TCGA sRNA-seq samples; Figure S2: miRNA isoform nomenclature; Figure S3: miRNA isoform expression distribution across the four group samples, namely W (N), W(T), B/AA (N), and B/AA (T); Figure S4: Expressed canonical miRNA in the four group samples; Figure S5: Comparison of expression between the edited miR-151a-3p and ADARs genes in LUAD-TCGA data; Figure S6: Two-stage workflow to identify prognostic miRNA isoform signatures; Table S1: Clinical information of TCGA-LUAD patients; Table S2: List of the expressed miRNA isoforms across the four group samples; Table S3: List of deregulated miRNA isoforms across the four group samples; Table S4: List of deregulated transcripts and significant enriched canonical pathways associated to the expression of the edited miR-151a-3p; Table S5: Potential prognostic miRNA isoform signatures identified in both W and B/AA LUAD patients.

**Author Contributions:** Conceptualization, G.N., M.A. and P.N.-S.; methodology, R.D. and G.N.; software, R.D. and G.N.; formal analysis, R.D. and G.N.; data curation, R.D.; writing – original draft preparation, G.N. and P.L.; writing – review and editing, all authors; visualization, R.D. and G.N.; supervision, M.A. and P.N.-S.; project administration, M.A. and P.N.-S.; funding acquisition, M.A. and P.N.-S. All authors have read and agreed to the final version of the manuscript.

**Funding:** This work was supported by the following National Institutes of Health Grants: NCI 5U01CA213330, 1P20CA252717-01A1, and NCATS 5KL2TR002648.

**Data Availability Statement:** All data are available in Tables S1-5. Raw sequence data were retrieved via the GDC data portal after obtaining authorization from the data access committee (DBGap Project ID: 11332)

**Acknowledgments:** The results shown here are based on data generated by the TCGA Research Network (<http://cancergenome.nih.gov/>). We want to thank the Cancer IT Operation Group of The Ohio State University, in particular, Mr. Thomas Moore, for his special technical assistance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *Ca Cancer J Clin* **2021**, *71*, 7–33, doi:10.3322/caac.21654.
2. Rivera, M.P.; Katki, H.A.; Tanner, N.T.; Triplett, M.; Sakoda, L.C.; Wiener, R.S.; Cardarelli, R.; Carter-Harris, L.; Crothers, K.; Fathi, J.T.; et al. Addressing Disparities in Lung Cancer Screening Eligibility and Healthcare Access. An Official American Thoracic Society Statement. *Am J Resp Crit Care* **2020**, *202*, e95–e112, doi:10.1164/rccm.202008-3053st.
3. Thomas, P.L.; Madubata, C.J.; Aldrich, M.C.; Lee, M.M.; Owonikoko, T.K.; Minna, J.D.; Rudin, C.M.; Sage, J.; Lovly, C.M. A Call to Action: Dismantling Racial Injustices in Preclinical Research and Clinical Care of Black Patients Living with Small Cell Lung Cancer. *Cancer Discov* **2021**, *11*, 240–244, doi:10.1158/2159-8290.cd-20-1592.
4. Tanner, N.T.; Gebregziabher, M.; Halbert, C.H.; Payne, E.; Egede, L.E.; Silvestri, G.A. Racial Differences in Outcomes within the National Lung Screening Trial. Implications for Widespread Implementation. *Am J Resp Crit Care* **2015**, *192*, 200–208, doi:10.1164/rccm.201502-0259oc.
5. Ezer, N.; Mhango, G.; Bagiella, E.; Goodman, E.; Flores, R.; Wisnivesky, J.P. Racial Disparities in Resection of Early Stage Non-Small Cell Lung Cancer. *Med Care* **2020**, *58*, 392–398, doi:10.1097/mlr.0000000000001280.



6. Kitts, A.K.B. The Patient Perspective on Lung Cancer Screening and Health Disparities. *J Am Coll Radiol* **2019**, *16*, 601–606, doi:10.1016/j.jacr.2018.12.028.
7. Mitchell, K.A.; Zingone, A.; Toulabi, L.; Boeckelman, J.; Ryan, B.M. Comparative Transcriptome Profiling Reveals Coding and Noncoding RNA Differences in NSCLC from African Americans and European Americans. *Clin Cancer Res* **2017**, *23*, 7412–7425, doi:10.1158/1078-0432.ccr-17-0527.
8. Hastert, T.A.; McDougall, J.A.; Strayhorn, S.M.; Nair, M.; Beebe-Dimmer, J.L.; Schwartz, A.G. Social Needs and Health-related Quality of Life among African American Cancer Survivors: Results from the Detroit Research on Cancer Survivors Study. *Cancer* **2021**, *127*, 467–475, doi:10.1002/cncr.33286.
9. Araujo, L.H.; Lammers, P.E.; Matthews-Smith, V.; Eisenberg, R.; Gonzalez, A.; Schwartz, A.G.; Timmers, C.; Shilo, K.; Zhao, W.; Natarajan, T.G.; et al. Somatic Mutation Spectrum of Non–Small-Cell Lung Cancer in African Americans: A Pooled Analysis. *J Thorac Oncol* **2015**, *10*, 1430–1436, doi:10.1097/jto.0000000000000650.
10. Araujo, L.H.; Timmers, C.; Bell, E.H.; Shilo, K.; Lammers, P.E.; Zhao, W.; Natarajan, T.G.; Miller, C.J.; Zhang, J.; Yilmaz, A.S.; et al. Genomic Characterization of Non–Small-Cell Lung Cancer in African Americans by Targeted Massively Parallel Sequencing. *J Clin Oncol* **2015**, *33*, 1966–1973, doi:10.1200/jco.2014.59.2444.
11. Choudhury, N.J.; Egtesad, M.; Kadri, S.; Cursio, J.; Ritterhouse, L.; Segal, J.; Husain, A.; Patel, J.D. Fewer Actionable Mutations but Higher Tumor Mutational Burden Characterizes NSCLC in Black Patients at an Urban Academic Medical Center. *Oncotarget* **2019**, *10*, 5817–5823, doi:10.18632/oncotarget.27212.
12. Lusk, C.M.; Watza, D.; Dyson, G.; Craig, D.; Ratliff, V.; Wenzlaff, A.S.; Lonardo, F.; Bollig-Fischer, A.; Bepler, G.; Purrington, K.; et al. Profiling the Mutational Landscape in Known Driver Genes and Novel Genes in African American Non–Small Cell Lung Cancer Patients. *Clin Cancer Res* **2019**, *25*, 4300–4308, doi:10.1158/1078-0432.ccr-18-2439.
13. Telonis, A.G.; Rigoutsos, I. Race Disparities in the Contribution of MiRNA Isoforms and TRNA-Derived Fragments to Triple-Negative Breast Cancer. *Cancer Res* **2017**, *78*, canres.1947.2017, doi:10.1158/0008-5472.can-17-1947.
14. Lee, L.W.; Zhang, S.; Etheridge, A.; Ma, L.; Martin, D.; Galas, D.; Wang, K. Complexity of the MicroRNA Repertoire Revealed by Next-Generation Sequencing. *RNA (New York, N.Y.)* **2010**, *16*, 2170–2180, doi:10.1261/rna.2225110.
15. Esteller, M.; Pandolfi, P.P. The Epitranscriptome of Noncoding RNAs in Cancer. *Cancer Discov* **2017**, *7*, 359–368, doi:10.1158/2159-8290.cd-16-1292.
16. Consortium, I.H.G.S. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921, doi:10.1038/35057062.
17. Ling, H.; Fabbri, M.; Calin, G.A. MicroRNAs and Other Non-Coding RNAs as Targets for Anticancer Drug Development. *Nat Rev Drug Discov* **2013**, *12*, 847–865, doi:10.1038/nrd4140.
18. Glogovitis, I.; Yahubyan, G.; Würdinger, T.; Koppers-Lalic, D.; Baev, V. IsomiRs–Hidden Soldiers in the MiRNA Regulatory Army, and How to Find Them? *Biomol* **2020**, *11*, 41, doi:10.3390/biom11010041.

19. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry* **2010**, *79*, 321–349, doi:10.1146/annurev-biochem-060208-105251.
20. Nishikura, K. A-to-I Editing of Coding and Non-Coding RNAs by ADARs. *Nature reviews. Molecular cell biology* **2016**, *17*, 83–96, doi:10.1038/nrm.2015.4.
21. Nigita, G.; Marceca, G.P.; Tomasello, L.; Distefano, R.; Calore, F.; Veneziano, D.; Romano, G.; Nana-Sinkam, S.P.; Acunzo, M.; Croce, C.M. NcRNA Editing: Functional Characterization and Computational Resources. *Methods Mol Biology* **2019**, *1912*, 133–174, doi:10.1007/978-1-4939-8982-9\_6.
22. Kawahara, Y.; Zinshteyn, B.; Sethupathy, P.; Iizasa, H.; Hatzigeorgiou, A.G.; Nishikura, K. Redirection of Silencing Targets by Adenosine-to-Inosine Editing of MiRNAs. *Science* **2007**, *315*, 1137–1140, doi:10.1126/science.1138050.
23. Nigita, G.; Acunzo, M.; Romano, G.; Veneziano, D.; Laganà, A.; Vitiello, M.; Wernicke, D.; Ferro, A.; Croce, C.M. MicroRNA Editing in Seed Region Aligns with Cellular Changes in Hypoxic Conditions. *Nucleic acids research* **2016**, *44*, 6298–6308, doi:10.1093/nar/gkw532.
24. Pinto, Y.; Buchumenski, I.; Levanon, E.Y.; Eisenberg, E. Human Cancer Tissues Exhibit Reduced A-to-I Editing of MiRNAs Coupled with Elevated Editing of Their Targets. *Nucleic acids research* **2018**, *46*, 71–82, doi:10.1093/nar/gkx1176.
25. Xu, X.; Wang, Y.; Mojumdar, K.; Zhou, Z.; Jeong, K.J.; Mangala, L.S.; Yu, S.; Tsang, Y.H.; Rodriguez-Aguayo, C.; Lu, Y.; et al. A-to-I-Edited MiRNA-379-5p Inhibits Cancer Cell Proliferation through CD97-Induced Apoptosis. *Journal of Clinical Investigation* **2019**, *57*, 1798–5356, doi:10.1172/jci123396.
26. 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. Systematic Characterization of A-to-I RNA Editing Hotspots in MicroRNAs across Human Cancers. *Genome Research* **2017**, *27*, 1112–1125, doi:10.1101/gr.219741.116.
27. Baysal, B.E.; Sharma, S.; Hashemikhabir, S.; Janga, S.C. RNA Editing in Pathogenesis of Cancer. *Cancer Research* **2017**, *77*, 3733–3739, doi:10.1158/0008-5472.can-17-0520.
28. Lu, Y.; Baras, A.S.; Halushka, M.K. MiRge 2.0 for Comprehensive Analysis of MicroRNA Sequencing Data. *BMC Bioinformatics* **2018**, *19*, 275, doi:10.1186/s12859-018-2287-y.
29. Desvignes, T.; Loher, P.; Eilbeck, K.; Ma, J.; Urgese, G.; Fromm, B.; Sydes, J.; Aparicio-Puerta, E.; Barrera, V.; Espín, R.; et al. Unification of MiRNA and IsomiR Research: The MirGFF3 Format and the Mirtop API. *Bioinformatics* **2019**, *36*, 698–703, doi:10.1093/bioinformatics/btz675.
30. Marceca, G.P.; Distefano, R.; Tomasello, L.; Lagana, A.; Russo, F.; Calore, F.; Romano, G.; Bagnoli, M.; Gasparini, P.; Ferro, A.; et al. MiREDiBase, a Manually Curated Database of Validated and Putative Editing Events in MicroRNAs. *Sci Data* **2021**, *8*, 199, doi:10.1038/s41597-021-00979-8.

31. 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. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2019**, *47*, D941–D947, doi:10.1093/nar/gky1015.
32. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI Database of Genetic Variation. *Nucleic acids research* **2001**, *29*, 308–311.
33. Alon, S.; Mor, E.; Vigneault, F.; Church, G.M.; Locatelli, F.; Galeano, F.; Gallo, A.; Shomron, N.; Eisenberg, E. Systematic Identification of Edited MicroRNAs in the Human Brain. *Genome Research* **2012**, *22*, 1533–1540, doi:10.1101/gr.131573.111.
34. Tomaselli, S.; Galeano, F.; Alon, S.; Raho, S.; Galardi, S.; Polito, V.A.; Presutti, C.; Vincenti, S.; Eisenberg, E.; Locatelli, F.; et al. Modulation of MicroRNA Editing, Expression and Processing by ADAR2 Deaminase in Glioblastoma. *Genome Biology* **2015**, *16*, 5, doi:10.1186/s13059-014-0575-z.
35. Yang, W.; Chendrimada, T.P.; Wang, Q.; Higuchi, M.; Seeburg, P.H.; Shiekhattar, R.; Nishikura, K. Modulation of MicroRNA Processing and Expression through RNA Editing by ADAR Deaminases. *Nature Structural & Molecular Biology* **2006**, *13*, 13–21, doi:10.1038/nsmb1041.
36. Quinlan, A.R.; Hall, I.M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26*, 841–842, doi:10.1093/bioinformatics/btq033.
37. Smeds, L.; Künstner, A. ConDeTri--a Content Dependent Read Trimmer for Illumina Data. *PLoS ONE* **2011**, *6*, e26314, doi:10.1371/journal.pone.0026314.
38. Fromm, B.; Domanska, D.; Høye, E.; Ovchinnikov, V.; Kang, W.; Aparicio-Puerta, E.; Johansen, M.; Flatmark, K.; Mathelier, A.; Hovig, E.; et al. MirGeneDB 2.0: The Metazoan MicroRNA Complement. *Nucleic acids research* **2020**, *48*, D132–D141, doi:10.1093/nar/gkz885.
39. Kozomara, A.; Griffiths-Jones, S. MiRBase: Annotating High Confidence MicroRNAs Using Deep Sequencing Data. *Nucleic acids research* **2014**, *42*, D68–D73, doi:10.1093/nar/gkt1181.
40. Hoon, M.J.L. de; Taft, R.J.; Hashimoto, T.; Kanamori-Katayama, M.; Kawaji, H.; Kawano, M.; Kishima, M.; Lassmann, T.; Faulkner, G.J.; Mattick, J.S.; et al. Cross-Mapping and the Identification of Editing Sites in Mature MicroRNAs in High-Throughput Sequencing Libraries. *Genome Research* **2010**, *20*, 257–264, doi:10.1101/gr.095273.109.
41. Jones, M.R.; Quinton, L.J.; Blahna, M.T.; Neilson, J.R.; Fu, S.; Ivanov, A.R.; Wolf, D.A.; Mizgerd, J.P. Zcchc11-Dependent Uridylation of MicroRNA Directs Cytokine Expression. *Nat Cell Biol* **2009**, *11*, 1157–1163, doi:10.1038/ncb1931.
42. Katoh, T.; Sakaguchi, Y.; Miyauchi, K.; Suzuki, T.; Kashiwabara, S.; Baba, T.; Suzuki, T. Selective Stabilization of Mammalian MicroRNAs by 3' Adenylation Mediated by the Cytoplasmic Poly(A) Polymerase GLD-2. *Gene Dev* **2009**, *23*, 433–438, doi:10.1101/gad.1761509.

43. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol* **2010**, *11*, R25–R25, doi:10.1186/gb-2010-11-3-r25.
44. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140, doi:10.1093/bioinformatics/btp616.
45. McCarthy, D.J.; Chen, Y.; Smyth, G.K. Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation. *Nucleic Acids Res* **2012**, *40*, 4288–4297, doi:10.1093/nar/gks042.
46. 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. Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biology* **2004**, *5*, R80, doi:10.1186/gb-2004-5-10-r80.
47. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B* **1995**, *57*, 289–300, doi:10.2307/2346101.
48. Heberle, H.; Meirelles, G.V.; Silva, F.R. da; Telles, G.P.; Minghim, R. InteractiVenn: A Web-Based Tool for the Analysis of Sets through Venn Diagrams. *BMC Bioinformatics* **2015**, *16*, 169, doi:10.1186/s12859-015-0611-3.