

10-30-2015

Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity.

Aristeidis G Telonis

Computational Medicine Center, Sidney Kimmel Medical College, Thomas Jefferson University

Phillipe Loher

Computational Medicine Center, Thomas Jefferson University

Yi Jing

Department of Pathology, Anatomy and Cell Biology, and Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania

Eric R Londin

Computational Medicine Center, Thomas Jefferson University Jefferson Medical College

Isidore Rigoutsos

Thomas Jefferson University

Follow this and additional works at: <https://jdc.jefferson.edu/tjucompmedctrfp>

 Part of the [Oncology Commons](#)

[Let us know how access to this document benefits you](#)

Recommended Citation

Telonis, Aristeidis G; Loher, Phillipe; Jing, Yi; Londin, Eric R; and Rigoutsos, Isidore, "Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity." (2015). *Computational Medicine Center Faculty Papers*. Paper 12.
<https://jdc.jefferson.edu/tjucompmedctrfp/12>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Computational Medicine Center Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity

Aristeidis G. Telonis, Phillipe Loher, Yi Jing, Eric Londin and Isidore Rigoutsos*

Computational Medicine Center, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, USA

Received April 06, 2015; Revised August 12, 2015; Accepted September 04, 2015

ABSTRACT

Here we describe our study of miRNA isoforms (isomiRs) in breast cancer (BRCA) and normal breast data sets from the Cancer Genome Atlas (TCGA) repository. We report that the full isomiR profiles, from both known and novel human-specific miRNA loci, are particularly rich in information and can distinguish tumor from normal tissue much better than the archetype miRNAs. IsomiR expression is also dependent on the patient's race, exemplified by miR-183-5p, several isomiRs of which are upregulated in triple negative BRCA in white but not black women. Additionally, we find that an isomiR's 5' endpoint and length, but not the genomic origin, are key determinants of the regulation of its expression. Overexpression of distinct miR-183-5p isomiRs in MDA-MB-231 cells followed by microarray analysis revealed that each isomiR has a distinct impact on the cellular transcriptome. Parallel integrative analysis of mRNA expression from BRCA data sets of the TCGA repository demonstrated that isomiRs can distinguish between the luminal A and luminal B subtypes and explain in more depth the molecular differences between them than the archetype molecules. In conclusion, our findings provide evidence that post-transcriptional studies of BRCA will benefit from transcending the one-locus-one-miRNA paradigm and taking into account all isoforms from each miRNA locus as well as the patient's race.

INTRODUCTION

Breast cancer (BRCA) is a heterogeneous disease and the second leading cause of cancer death among women after lung cancer (1). In 2014 in the United States there were an estimated 232 000 newly diagnosed cases, with approximately 41 000 deaths from the disease.

Over the last 30 years, in-depth molecular investigations led to the identification of histological markers that have been successfully implemented in clinical practice. These markers are used to categorize BRCA tumors and determine the most appropriate course of action in each case. Quantitation of the Estrogen Receptor (ER), the Progesterone Receptor (PR) and of the Human Epidermal Growth Factor Receptor 2 (HER2) are now routinely used to classify tumors into disease subgroups depending on the presence ("positive" or "+") or absence ("negative" or "-") of each hormone receptor (2,3). ER/PR positive tumors (ER+ only, PR+ only, or simultaneously ER+ and PR+) account for the majority (>70%) of the BRCA cases, with effective treatments targeting the ER pathway (2–5). On the other hand, 15% of the tumors are HER2+ and generally have a highly active growth factor signaling pathway with significant crosstalk with the ER cytoplasmic signaling pathway and with different tumor properties (2,3,6).

Approximately 15–20% of the BRCA cases do not express any of the three hormone receptors and are therefore characterized as "triple negative BRCA" (TNBC), or as "ER–/PR–/HER2–" (3). TNBC represents the most aggressive subtype of BRCA; in the absence of a receptor, hormone therapies cannot be used leaving chemotherapy as the only option. Extensive population studies have shown that TNBC is much more prevalent among black women (7). In particular, it is more than twice as frequent among premenopausal black women than premenopausal white women. Moreover, TNBC from black patients exhibits higher rates of proliferation, increased angiogenesis, higher grade and rates of lymph node metastasis (8–10). The resulting mortality rate in this population group is ~31%, the highest in BRCA among all racial/ethnic groups in the US.

Gene expression studies originally helped establish four basic BRCA subtypes (11,12): Luminal A (ER+ and/or PR+, HER2–, low Ki67), Luminal B (ER+ and/or PR+, HER2+, or HER2– with high Ki67), HER2-Enriched (ER–, PR–, HER2+) and Basal-like (ER–, PR–, HER2–) tumors. Two more subtypes, "normal-like"

*To whom correspondence should be addressed. Tel: +1 215 503 6152; Fax: +1 215 503 0466; Email: Isidore.Rigoutsos@jefferson.edu

and “claudin-low,” have also been described but occur at lower frequencies (8,13). The expression levels of a 50-gene panel (the “PAM50 gene signature”) is typically used to classify a surgically resected BRCA sample into one of the four intrinsic subtypes (14). It is worth noting here that the histological hormone profiles and the intrinsic subtypes are not “interchangeable”. A characteristic example is that of basal-like and TNBC tumors (15): even though the two have a significant overlap not all TNBC tumors have a basal-like phenotype. Conversely, not all basal-like cancers lack expression of the three hormone receptors.

In addition to other large-scale studies (16), the Cancer Genome Atlas Network has conducted a multi-platform integrated analysis of hundreds of BRCA samples across all hormone profiles and most intrinsic subtypes providing novel molecular insight into BRCA biology (8). Among the data types that were generated, the gene expression part of the study included the quantification of both mRNA and miRNAs.

MiRNAs are ubiquitously expressed short RNAs approximately 22 nucleotides (nt) in length. MiRNAs have been shown to regulate a large variety of molecular functions, and have been associated to a variety of diseases (17–19). In vertebrates, their biogenesis typically involves transcription of a pri-miRNA by RNA polymerase II followed by cleavage by the DGCR8/Drosha complex into a hairpin-shaped loop (pre-miRNA) that is exported to the cytoplasm. There, a pre-miRNA is processed further by Dicer, an RNase III endonuclease, leading to the formation of mature single-stranded miRNA products and their subsequent loading onto the RNA-induced silencing complex (RISC) whose main component is the Argonaute (Ago) protein (20,21). With the help of the RISC complex a miRNA controls the abundance of specific mRNAs by targeting the latter in regions that are complementary, or nearly so, to the miRNA (22,23). The miRNA profile is deregulated in BRCA, and in cancer in general, (8,24) and this subject has been systematically reviewed (25–28).

It has long been held that each arm of the pre-miRNA produces one consequential mature miRNA. This miRNA is listed in the public miRNA databases and will be referred to in what follows as the “reference” or “archetype” miRNA (29). However, the subsequent discovery of isoforms (isomiRs) that arise from the same arm as the archetype miRNA and have slightly different 5′ and/or 3′ termini is beginning to overturn the classical “one arm-one miRNA” view. As we reported recently, specific isomiRs can be substantially more abundant than the archetype miRNA (30). In fact, as we showed in recent work using lymphoblastoid cell lines derived from 452 healthy individuals isomiR profiles depend on gender, population and race (30). The implications that these isomiR findings may have for health and disease are potentially very significant. Nonetheless, with the exception of isolated efforts (31–33) the currently available knowledge remains limited.

In the current study, we extend our previous study of healthy individuals (30) to the disease context by systematically analyzing isomiR profiles in BRCA data sets from the TCGA repository. After we catalogued all the isomiRs that are abundantly present in these data sets we focused on a number of questions. Does knowledge of the isomiR pro-

files improve discrimination of normal from tumor, or of one subtype from another? Do the race-dependencies that we discovered in lymphoblastoid cell lines (30) extend to breast tissue samples? Are isomiR abundances correlated? Is the genomic locus or the hairpin arm, i.e. the arm from which the isomiRs arise, a deciding factor of isomiR abundance? Can we discern any evidence of additional mechanisms that affect isomiR levels? Do distinct isomiRs from the same hairpin arm affect the cellular transcriptome differently?

MATERIALS AND METHODS

IsomiR nomenclature

To discriminate among isomiRs from the same locus, we use the naming rules we introduced previously (30). In brief, for each isomiR we note the unique ID assigned to the locus (e.g. MIMAT0000076) as well as its most common name (e.g. miR-21-5p), followed by two numbers separated by vertical bars, e.g. miR-21-5p|+2|-11. The first number indicates the relative position of the isomiR’s 5′ terminus with respect to the archetype’s 5′ end, whereas the second number indicates the analogous relationship for the isomiR’s and the archetype’s 3′ termini: a positive sign (+) indicates that the isomiR’s terminus is downstream from the archetype terminus (in 5′→3′ direction); a negative sign (–) indicates that the isomiR’s terminus is upstream of the archetype’s terminus. Using this notation, miR-21-5p|0|0 denotes the archetype miRNA that arises from the left (5p) arm of the mir-21 precursor.

Sample source, read mapping, isomiR identification, normalization and filtering

Deep sequencing data for the 316 samples included in this study were obtained from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>). The samples were chosen in a way that almost all hormone profiles were included. Publicly available high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) data for three BRCA model cell lines were also downloaded (34). Mapping was performed on the GRCh37 human genome assembly, as described previously (30). No insertions or deletions were allowed. To focus on the isomiRs of the data set, we used the reference coordinates of the miR-Base Rel. 20 miRNAs (29) as well as the 3494 miRNA loci that we recently reported (35) (we will refer to these loci as “novel”) and expanded them by 6 nt at each side. We kept the molecules that are fully in this window and have a total length of 16–28 nt, inclusive. To exclude lowly expressed isomiRs, we considered for statistical analysis only isomiRs that had more than 100 reads in at least one sample. As the average sequencing depth (number of total reads) for the BRCA samples was 5.5 M reads, this threshold corresponds to a threshold of about 18 transcripts per million (TPM). This filtering process left us with 1958 isomiRs in the final expression matrix. In order to exclude sequencing depth biases, we normalized the expression of each isomiR (the number of reads) by dividing it with the number of uniquely mapped reads for each data set. Data for mRNA

(Level 3-RNASeqV2) expression for the same TCGA samples were downloaded from the TCGA data portal (<https://tcga-data.nci.nih.gov>). Genes with a mean expression lower than 100 reads per million (RPM) were excluded from the analysis. This left us with 12 657 genes in the expression matrix that was scaled (standardized) per gene before the statistical analysis.

Cell culture, RNA transfection and microarray analysis

MDA-MB-231 cells were grown in DMEM medium (Fisher Scientific) supplemented with 10% fetal bovine serum (Life Technologies), 1% Penicillin and Strep (Fisher Scientific), and 1% glutamine (Fisher Scientific), at 37°C in a humidified atmosphere containing 5% CO₂. The mirVana miRNA mimic for miR-183-5p, the Negative Control #1, along with the two customized isomiRs (miR-183-5p|–2|–2l, 5'-TGTATGGCACTGGTAGAATTCA-3' and miR-183-5p|+2|+2l, 5'-TGGCACTGGTAGAATTCACTGT-3') were purchased from Life Technologies and were transfected at a final concentration of 50 nM using the Lipofectamine RNAiMAX transfection Reagent (Life Technologies). Transfections were done in triplicate. After 48 h, RNA was extracted from the cells using the TRizol reagent (Life Technologies).

RNA was quantified on a Nanodrop ND-100 spectrophotometer, followed by RNA quality assessment analyzing on an Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA). Fragmented biotin-labeled cDNA (from 100 ng of RNA) was prepared using the GeneChip WT Plus Kit (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Affymetrix gene chips, Human Transcriptome Array 2.0 (Affymetrix, Santa Clara, CA), were hybridized with 5 µg fragmented and biotin-labeled cDNA in 200 µl of hybridization cocktail. Target denaturation was performed at 99°C for 5 min. and then 45°C for 5 min, followed by hybridization with rotation 60 rpm for 16 h at 45°C. Arrays were then washed and stained using Gene chip Fluidic Station 450, using Affymetrix GeneChip hybridization wash and stain kit. Chips were scanned on an Affymetrix Gene Chip Scanner 3000, using Command Console Software. Quality Control of the experiment was performed by Expression Console Software v1.4.1. The SST-RMA algorithm was used for data processing and normalization; only the top 80% expressed protein-coding genes (15 581 in number) from ENSEMBL75 (36) were considered for further statistical analysis.

Multivariate statistical analysis

Statistical analysis was performed in R (<http://www.R-project.org/>) (37). The packages *lawstat* (38), *amap* (39), *DiscrMiner* (40), *samr* (41), *VennDiagram* (42) and *dendextend* (43) were used for Levene's test, Hierarchical Clustering (HCL) (44), Partial Least Squares–Discriminant Analysis (PLS-DA) (45), Significance Analysis of Microarrays (SAM) (41) and visualization of the Venn diagrams and the dendrograms, respectively. Principal Component Analysis (PCA) (44) was performed with the *prcomp* function. For the construction of the Venn diagram of the number of isomiRs present in each cell line from (34), the normalized expression values were averaged across replicates of the

same cell line and an isomiR was considered to be present if it belonged to the upper 90% quantile of the distribution of the number of reads per isomiR. DAVID (46,47) was used to identify enriched gene ontology (GO) Biological Process (BP) terms with an FDR threshold of 10%. The background list for the DAVID analysis comprised the expressed genes in the respective experiments.

Target prediction and miRNA-mRNA integration

Targets of isomiRs were predicted using *rma22* (48). The full list of the protein-coding ENSEMBL75 (36) transcripts was used as targets and the targeting site was allowed to be anywhere in the 5'- and 3'-UTR as well as the coding sequence. We defined nucleotides 2–7 as the seed region, enforced the presence at least 12 paired bases in the heteroduplex, including G:U pairs, while allowing at most one unpaired base (mismatch or bulge) within the seed region. An additional filtering for *P*-value (threshold of 0.01) of the specific loci being a miRNA response element (MRE) was done during the integration of isomiR and mRNA data for the comparison of the luminal A and luminal B BRCA subtypes. An isomiR was considered to target a gene if it had at least one predicted target site in at least one of the gene's mRNA transcripts satisfying the above criteria. The isomiR-gene pairs were visualized using Cytoscape (49).

RESULTS

In this study, we considered a total of 5353 human miRNA precursors: these included the 1859 precursors contained in Release 20 of miRBase (29) and the 3494 novel precursors that we reported recently (35). For each miRNA hairpin, we considered each arm separately. We analyzed 294 tumors and 22 normal samples from the TCGA/BRCA repository (8) and processed the data as previously described (30). For each miRNA hairpin arm, we considered an “active window” that extended six nucleotides (nt) beyond the boundaries of the “archetype” miRNA that is currently listed in miRBase or that was reported by us (35). All reads of each analyzed data set were mapped on the entire genome: only reads that mapped within each arm's active window and whose length ranged from 16 to 28 nt inclusive were kept and processed further. We refer to all the distinct molecules that arise from a single miRNA precursor arm as “isomiRs”. Naturally, the “archetype” or “reference” miRNA listed in miRBase for a locus is included among the locus' isomiRs.

The archetype miRNA is but one of several products arising from the locus at hand

We first counted the number of hairpin arms that produce one or more miRNA molecules. We found a total of 405 hairpin arms producing 1958 distinct isomiRs (the full data set is included as Supplementary Table S1). Of the 405 hairpin arms, 32 are novel (35) and produce 105 distinct isomiRs. On average there are five distinct isomiRs that arise from each locus. However, this number exhibits significant variation as can be seen from Figure 1A: 103 (25.4%; 16 of which are novel) hairpin arms produce a single miRNA

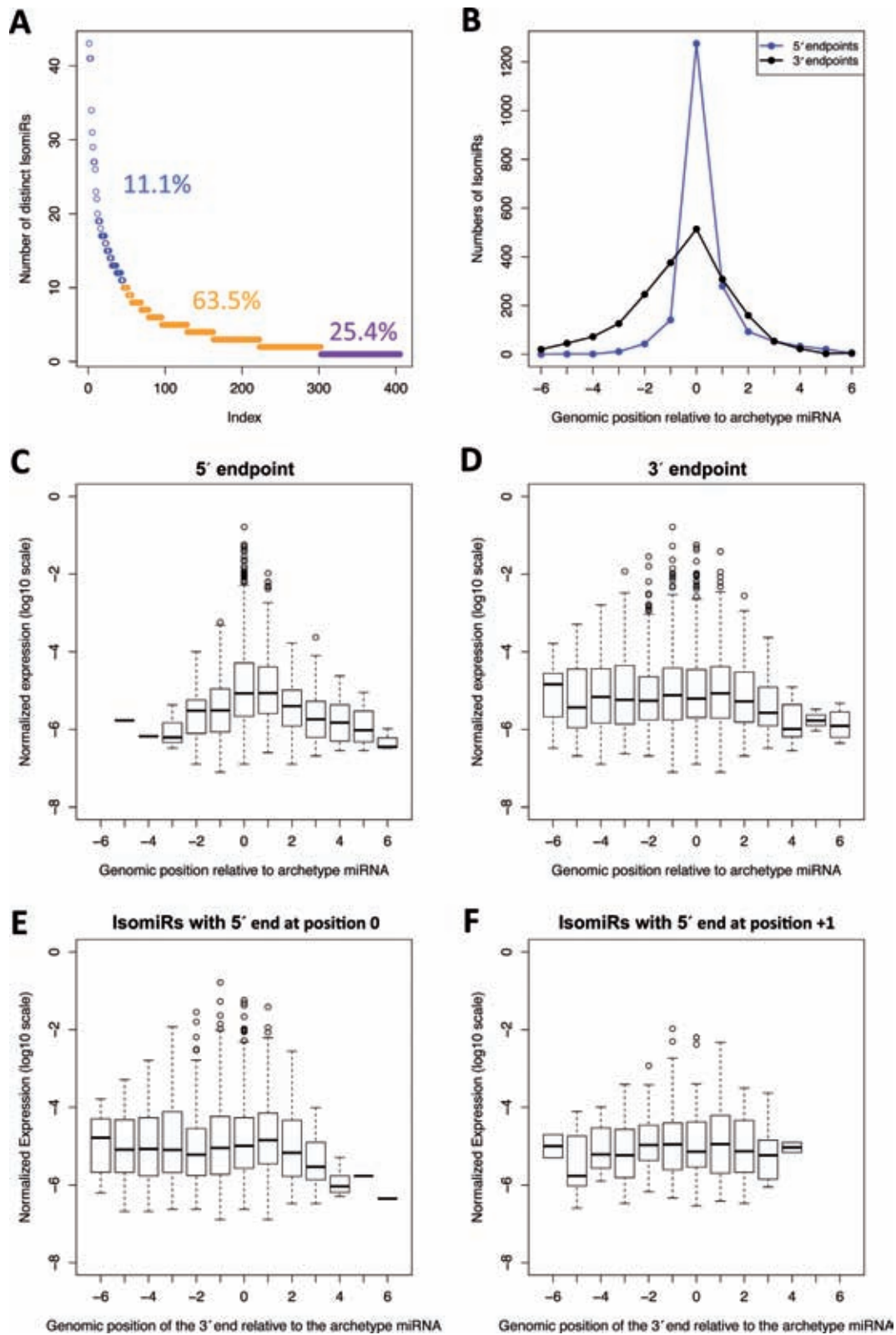


Figure 1. General characteristics of the isomiRs. (A) The number of distinct isomiRs per locus. (B) Distribution of the number of isomiR endpoints at each genomic position relative to the archetype's coordinates. (C–D) Relative expression of the isomiRs starting (C) or ending (D) at each genomic position relative to the archetype's coordinates. (E–F) Relative expression of the isomiRs with the 5' terminus at the reference (E) or at the +1 position (F) and a varying 3' terminus (x axis). For (C–F) the expression values were calculated using normal breast samples from white women.

molecule whereas the majority of the arms (257 of 405, or 63.5%; 14 of which are novel) produce between 2 and 10 isomiRs. For 45 arms (11.1%; two of which are novel) we observed more than 10 isomiRs generated from each. Among these 45 arms, miR-21-5p had the most isomiRs (43 isomiRs) followed by miR-10a-5p (41 isomiRs) and miR-183-5p (41 isomiRs). It is noteworthy that 90 arms do not produce the archetype miRNA and that the majority of them (53 out of 90) produce only one other (non-archetype) isomiR. We also note that out of the 32 novel miRNA loci found to be expressed, half of them are human-specific, 14 are primate-specific (not found in the mouse genome) and two are found in mice but not in *Drosophila melanogaster* nor in *Saccharomyces cerevisiae*. Interestingly, the average expression of these arms was not correlated with evolutionary conservation (Supplementary Figure S1A).

As the 5' and the 3' isomiRs are expected to affect the miRNA locus' properties in different ways, we plotted the distribution of the 5' and 3' termini of the isomiRs around the termini of the archetype miRNA (Figure 1B). Levene's test indicated that the variances in the two distributions differ in a statistically significant manner (P -value $< 10^{-4}$) with the 3' termini of the isomiRs exhibiting more diversity than the 5' termini (30). We also examined how the abundance of isomiRs changes as a function of the location of each isomiR's termini relatively to those of the archetype. To this end, we calculated the median expression of each isomiR in normal samples from white females and plotted it as a function of the deviation from the archetype's 5' or the 3' termini (Figure 1C and D). For the 5' terminus, we observed an overall reduction in the abundance of the isomiRs, by as much as an order of magnitude, as the distance from the reference 5' position increased (Figure 1C). However, this was not the case for the 3' termini: the abundance of isomiRs with termini ending as many as 6 nt away (either -6 or $+6$) from the reference position was roughly unchanged compared to that of the archetype.

To dissect the relationships of the two endpoints, we first decomposed the curve of the 3' termini in Figure 1B to its 5'-specific counterparts (Supplementary Figure S1B). We found that the majority of the isomiRs starting at the reference position also end at the reference 3' position. However, when we plotted the average expression of the isomiRs with the same 5' end but varying 3' ends, we observed that the expression levels remain on average similar (Figure 1E–F) independently of (i) the starting and ending positions (Figure 1E–F), although with a larger dynamic range for isomiRs ending at the reference position and of (ii) the absolute numbers of isomiRs ending at each position (Supplementary Figure S1B). It should be noted that these data describe the totality of the isomiRs irrespective of their genomic origin and do not reflect individual loci or hairpin arms. As exemplified in Supplementary Figure S1C–E, and also shown in our previous work (30), there are specific hairpin arms where the archetype miRNA is not the most abundant molecule.

Collectively, the above observations mirror our earlier findings in a different tissue (30) and indicate that the reference molecules (i.e. archetype miRNA) capture only a portion of a diverse repertoire of miRNA products that can arise from a given precursor arm.

IsomiRs improve the discrimination between normal and BRCA samples

In previous work we analyzed samples from healthy individuals (30) and showed that isomiRs are constitutively present across like samples while also exhibiting differences in a gender-, population- and race-dependent manner. Here we extend the earlier analysis to include disease samples. As an example we focus on normal breast and on ER+/PR+/HER2– BRCA samples, all from white patients, from the TCGA cohort. We used PCA to project the datapoints (samples) from their original multi-dimensional space onto a 2-D space and also used HCL to assess the similarity/dissimilarity of the analyzed data sets (44). Figure 2 shows the results of PCA and HCL when only the archetype miRNAs (panels A and B) and when the complete collection of isomiRs is considered (panels C and D). In the first case, the normal and tumor samples are part of one big cluster and mixed (Figure 2A and B). However, when using the full set of isomiRs the two groups of samples formed two distinct clusters with a very small cross-talk (Figure 2C and D). The improvement in the ability to discriminate in an unsupervised way between the normal and the tumor data sets is very clear. We obtained similar results for the rest of normal versus BRCA-subtype comparisons; see, e.g. Supplementary Figure S2. Exclusion of our novel miRNAs had minimal effect on the discrimination patterns described (data not shown).

The findings suggest that, as a group, the isomiRs are better descriptors of the tissue at hand (both normal and tumor states). To evaluate alterations of the isomiR biogenesis at a global level, we plotted the graph of Figure 1E–F but for the ER+/PR+/HER2– BRCA samples (Supplementary Figure S3). We did not observe any global patterns of isomiR production, e.g. a global preponderance of 3' shortening or 5' extensions in the hormone positive tumors as compared to the normal samples, indicating that the observed differences are probably caused by alterations in the regulation of specific miRNA loci. To statistically capture such differences, we used the non-parametric method for multivariate significance analysis known as “Significance Analysis of Microarrays” or SAM (41). For the normal versus ER+/PR+/HER2– comparison, SAM identified 224 isomiRs with statistically significant differential abundances (DA); 221 of the 224 are downregulated in the ER+/PR+/HER2– data sets (Supplementary Table S2 and Figure S4). The 224 DA isomiRs originate from 61 distinct hairpin arms, four of which are novel. The 5p arm of miR-10b generated the highest number of DA isomiRs: 22 DA isomiRs out of 29 isomiRs expressed. The 5p arm of miR-21 generated 43 isomiRs, the highest number of distinct isomiRs across all transcribed miRNA arms; however, only three of these isomiRs were DA between the normal and ER+/PR+/HER2– data sets. We note that the three DA miR-21-5p isomiRs were the only isomiRs that were found to be more abundant in the 101 analyzed ER+/PR+/HER2– BRCA tumors from the TCGA cohort (the archetype miRNA was among the three).

The 3p arm of miR-143 is another noteworthy source of isomiRs. This arm produced a total of 27 isomiRs that include the archetype miRNA. Of these 27

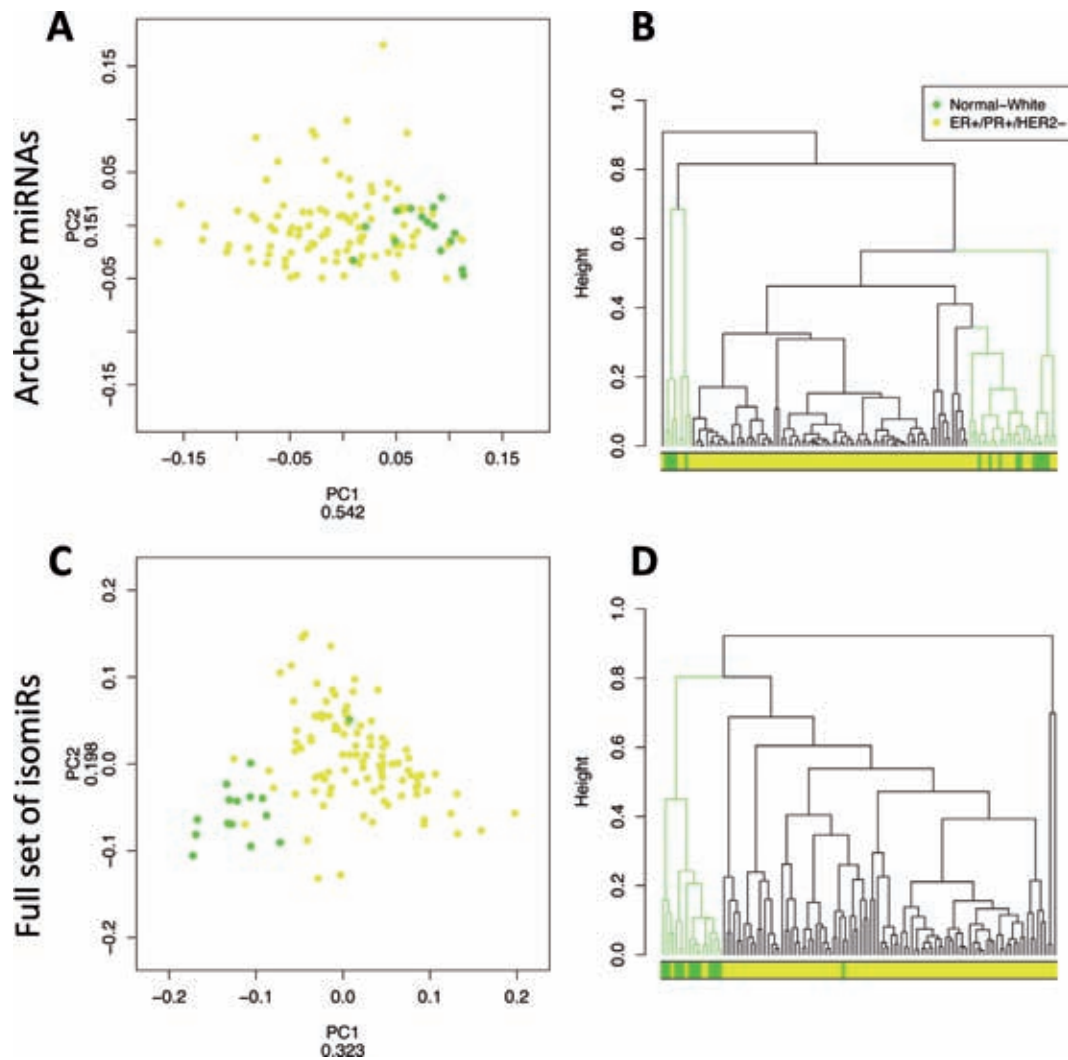


Figure 2. The full set of isomiRs better separated the normal from the tumor samples than the archetype miRNAs. Principal Component Analysis (PCA) (A and C) and Hierarchical Clustering (HCL; metric: Pearson Correlation) (B and D) of the ER+/PR+/HER2- BRCA (yellow) and normal (green) samples. (A–B) PCA and HCL with only the archetype miRNAs. No clear separation of the tumor and normal groups can be observed. (C–D) PCA and HCL using the full set of isomiRs showing a clear discrimination of the two groups with little overlap. The numbers on the axes labels of PCA indicate the variance explained by each principal component (PC).

isomiRs, 14 exhibited significantly lower abundance in the ER+/PR+/HER2- data sets compared to the normal data sets. Interestingly, and unlike the case of miR-21-5p above, the miR-143 archetype miRNA was not among the DA isomiRs.

Looking across the 61 hairpin arms, we found that the archetype miRNA was the sole DA isomiR between normal and ER+/PR+/HER2- in only 14 cases. In an additional 28 cases, the DA isomiRs included the archetype miRNA and at least one more isoform. Finally, for the remaining 19 arms, the archetype miRNA was not among the DA isomiRs. In other words, 77% (47 of 61) of the miRNA-producing hairpins contributed isomiRs that were DA abundant between normal and ER+/PR+/HER2- and whose significance in the BRCA context will be missed if one looks only at the archetype miRNA. This is in agreement with the results summarized in Figure 2 where inclu-

sion of the full isomiR profile led to an improved discrimination between the two tissue states.

The isomiR profiles in BRCA samples depend on race

We recently showed in a different biological context that the isomiR profiles in healthy individuals differ between human races (white and black) and also between populations belonging to the same race (30). Racial differences among breast cancer patients have been documented previously, with the case of TNBC providing the most characteristic such example (10,50). However, there are many open questions regarding the molecular basis of these phenotypic differences, especially at the miRNA/isomiR level.

In order to examine the possibility of race-dependencies we studied the isomiR profiles in 51 white TNBC patients and 16 black TNBC subjects. We also studied the isomiRs in normal samples from 15 white and 6 black subjects. We

first carried out a normal versus tumor comparison and did so separately for each race. In both instances, unsupervised analyses indicated that the isomiR profile could clearly discriminate between normal and tumor samples from same-race subjects (Figure 3A-D), a strong indication that there are significant differences in the isomiR profiles between the normal and tumor tissue states. For the samples from the white subjects, the observed cross-talk comprised only three data sets (Figure 3A and B). In the case of black subjects, there was no cross-talk as the normal and tumor data sets clustered distinctly (Figure 3C and D).

In addition to PCA and HCL, we also used the non-parametric SAM (see above), separately for each race in order to identify which race-specific isomiRs are deregulated in cancer. For the normal and TNBC data sets from the white subjects, SAM identified 305 isomiRs that are DA in tumor compared to the normal samples (Supplementary Table S3 and Figure S5A). Specifically 156 of these 305 isomiRs exhibited lower abundance in the tumor samples; the remaining 149 isomiRs exhibited higher abundance in tumor. Repeating the analysis for the TNBC and normal data sets from the black subjects we found 162 isomiRs that were DA between normal and tumor. All 162 of these isomiRs exhibited lower abundance in tumor (Supplementary Table S3 and Figure S5B). Interestingly, for both races, non-archetype isomiRs from the 5p arm of miR-10b were at the top in terms of relative significance in the normal versus TNBC comparisons.

In all, 113 hairpin arms (from both miRBase and novel genomic loci) gave rise to DA isomiRs in the TNBC versus normal comparison in white subjects. The number dropped to 47 hairpin arms when data sets from only black subjects were analyzed. Only 30 hairpin arms were common to both collections. Nonetheless we note that even though a hairpin arm may be a source of DA isomiRs in both races, the identity of the DA isomiRs from that arm is not the same for both races; e.g. different miR-143-3p isomiRs may be DA between normal and tumor in white subjects than in black subjects. Next, we compared the identities of the two collections of DA isomiRs (305 and 162, respectively—see above) found that only 89 isomiRs were common to both races and DA in the TNBC versus normal comparisons: 216 DA isomiRs are responsible for molecular events that are exclusively present in the white subjects and 73 DA isomiRs are responsible for molecular events that are exclusively present in the black subjects. In addition, we examined the TNBC data sets to determine whether racial molecular differences could be identified between white and black subjects. In recognition of the fact that TNBC is a diverse cancer subtype (3), we used a supervised analysis method (Partial Least Squares—Discriminant Analysis or PLS-DA) (51). We were able to discriminate the TNBC data sets along the race boundary (Figure 3E). Using SAM, we identified 21 isomiRs (originating from 8 hairpin arms) that were DA between white and black TNBC patients (Supplementary Table S3 and Figure S5C). Interestingly, 10 of these DA isomiRs arise from TJU_CMC_MD2.ID00121, a novel human-specific miRNA locus that we recently discovered (35). It is reasonable to assume that the observed differences underlie differences in the pathobiology of TNBC.

Finally, we examined the normal data sets alone to determine whether differences existed between white and black subjects. For consistency, we again used PLS-DA and were able to easily discriminate the normal data sets of different races (Figure 3F). Using SAM, we identified 22 DA isomiRs (from 14 distinct hairpin arms) all of which have lower abundance in the normal breast samples of white subjects (Supplementary Table S3 and Figure S5D). The overlap among the identified DA isomiRs is shown in Supplementary Figure S5E. These results indicate that normal breast tissue physiology includes a race-specific molecular component. These characteristics potentially provide a different dynamic biological background, beyond the genome, that may explain race-dependent contributions to the events that guide cancer development.

Several isomiRs have correlated abundances that depend on their endpoint choices and not on the locus of origin

In a top-down approach, one miRNA gene gives rise to two arms, and each arm to multiple isomiRs. Therefore, the isomiR profile is the result of transcriptional choices of the unit of origin and of post-transcriptional events that operate on the products of the two arms. From a genome-centric view, the genomic origin should dominate the post-transcriptional biogenesis in regulating the abundance of the isomiRs encoded from different genes but from a biogenesis perspective the choice on the abundance of functionally related isomiRs should be mainly determined by post-transcriptional processing mechanisms. To investigate these hypotheses, we first considered the genomic loci producing miR-29a-3p and miR-29c-3p. The sequences of these two archetype miRNAs share the same seed (5'-NAGCACC-3'), have similar mRNA targeting profiles (52,53) but are produced from different chromosomes. The sequences of the archetype miRNAs also differ only in one base in the middle of the miRNAs' span. Using HCL (Pearson correlation as a metric) on the expression of products of miR-29a-3p and miR-29b-3p arms in all samples in the BRCA data set, we clustered the isomiRs based on their abundance profile: two isomiRs, whether from the same locus or not, will end-up on the same sub-tree if they have correlated expression across samples. We observed that the primary determinant of the clustering was not the isomiRs' genomic origin but rather their primary characteristics, i.e. the specifics of their 5' and 3' endpoints (Figure 4A). Looking at Figure 4A, we can recognize two main clusters being formed: a "gold" cluster containing mainly isomiRs that have a shortened 3' end and consequently short lengths, and also a "black" cluster containing the archetype miRNAs and isomiRs with smaller deviations on their 5' and 3' endpoints.

Having shown that independently transcribed isomiRs of nearly-identical archetypes do not cluster based on the transcript of origin, we examined isomiRs that arise from a polycistronic transcript as is the case with the miR-17/92 cluster (28). This cluster comprises six miRNAs (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92a-1) whose pri-miRNAs are transcribed as a single RNA; the DGCR8/Drosha complex identifies and cleaves each characteristic hairpin structure (28), but with different efficien-

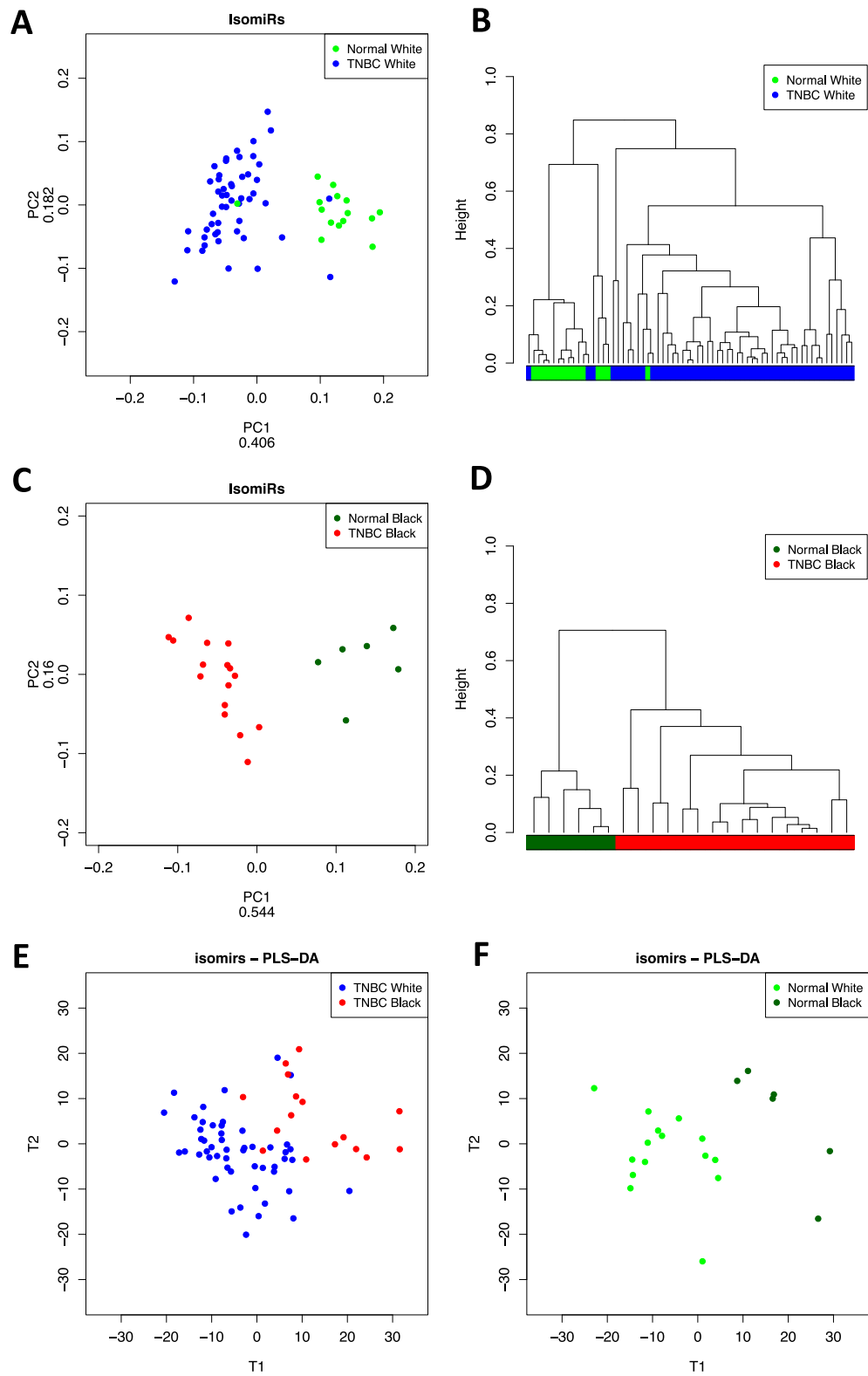


Figure 3. Race-specific isomiR profiles in TNBC and normal tissue. Principal Component Analysis (PCA) (A and C) and Hierarchical Clustering (HCL; metric: Pearson Correlation) (B and D) of the normal and TNBC samples in white (A and B) and black (C and D) women. The separation of the tumor from the normal samples is evident in all of the four analyses. (E–F) Partial Least Squares—Discriminant Analysis (PLS-DA) for the white and black normal tissues (E) and tumor samples (F). In all these analyses, the separation of the samples from black and white subjects is evident. The numbers on the axes labels of PCA indicate the variance explained by each principal component (PC).

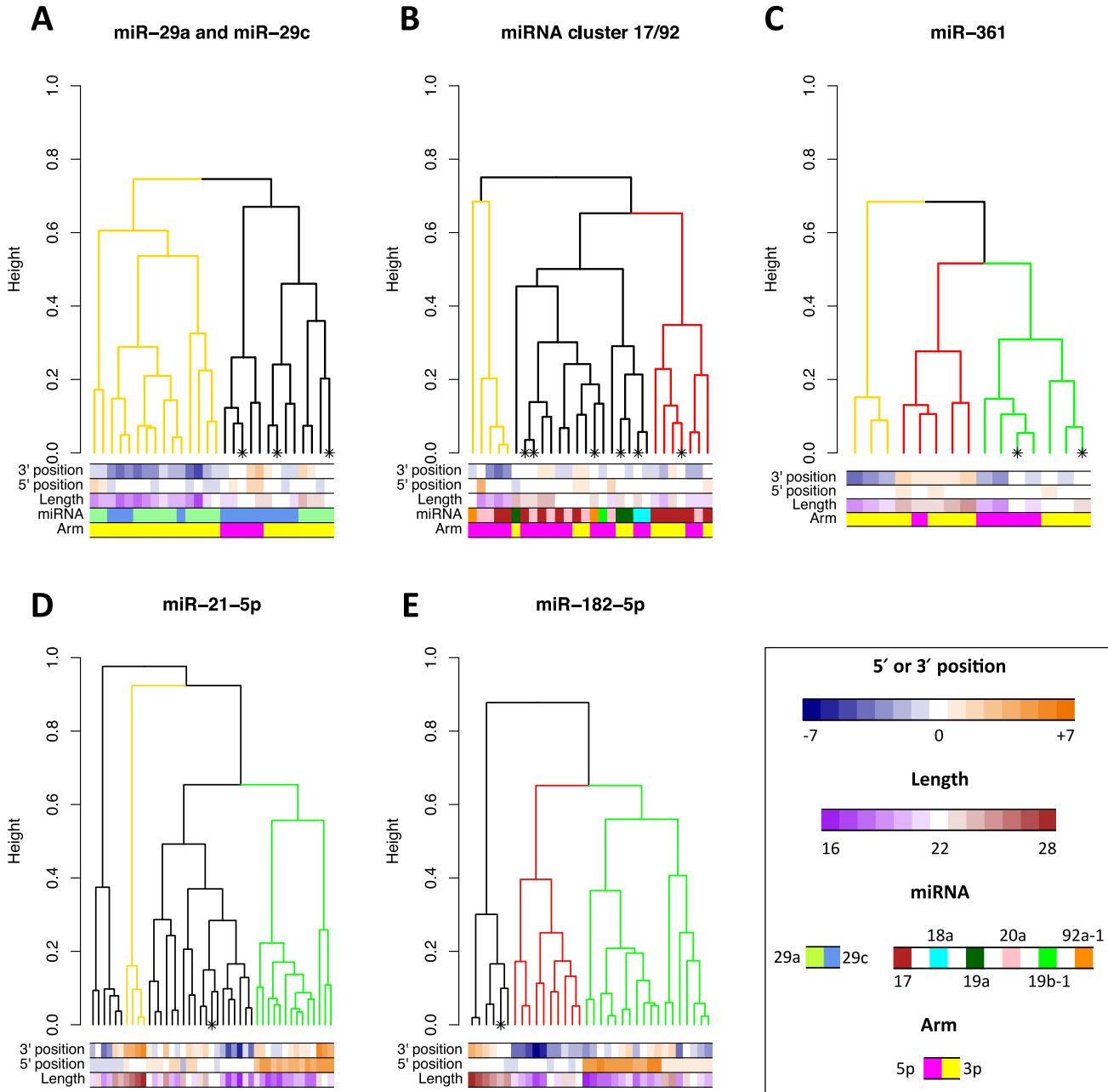


Figure 4. Correlation patterns among isomiRs. HCL analyses (metric: Pearson Correlation) on the isomiRs of selected loci. (A) HCL of the isomiRs of two functionally related miRNA loci, miR-29a and miR-29c. The absence of absolute clustering based on the transcript of origin indicates that the transcription is not the main determinant of the isomiR relative abundance. (B) HCL of the isomiRs from the miRNAs of the 17/92 cluster. IsomiRs do not cluster based on the hairpin of origin suggesting stronger downstream processing mechanisms. (C) HCL of the isomiRs from the two arms of one common hairpin. The choice on the arm of the hairpin has a mild effect on determining the correlation patterns of the isomiRs that originate from each arm. (D–E) HCL of the isomiRs from distinct loci: miR-21-5p (D) and miR-182-5p (E). IsomiRs are clustered based on specific characteristics including the relative 5' and 3' endpoints and the length. For (A–E), branches colored yellow, green and red represent clusters of isomiRs with specific common characteristics that are described in the main text. Asterisks mark the archetype miRNA.

cies for each hairpin (54,55). HCL analysis of the isomiRs' abundances revealed the absence of clusters based on the respective miRNA precursor of origin as well as the absence of clusters of miRNA family members that share the same seed (Figure 4B). Additionally, the arm of origin did not seem to be a contributor to determining the relative abundances of the isomiRs. However, we found that isomiRs clustered with respect to other characteristics, such as their length: the "golden" and "red" clusters contain mainly isomiRs that are shorter than the respective archetype miRNA (Figure 4B).

We also explored whether the identity of the source hairpin arm impacts on isomiR abundance. We focused on mir-361, a pre-miRNA that produces isomiRs from both of its arms. As seen in Figure 4C, HCL analysis revealed that isomiRs clustered together if they were 3' shortened ("gold" group). Analogously, isomiRs from both precursor arms that were 3' lengthened clustered together ("red" group). We note that the red group contains all of the 3' lengthened isomiRs. Lastly, the two archetype miRNAs (indicated by the asterisks in Figure 4C) as well as isomiRs from both the 5p and 3p arms that had only slightly modified lengths clustered together forming the "green" group on the right part of the dendrogram. The lack of clearly defined clusters that are based on the source arm of this miRNA precursor indicates that arm identity, in addition to the transcript and hairpin source, is not the main determinant of isomiR abundance.

Having excluded genomic origin, co-transcription and arm identity from the list of major modulators of the isomiRs' relative abundances, we focused on and analyzed separately the isomiR profiles from the 5p arms of two specific pre-miRNAs, mir-21 and mir-182. We selected these two miRNA loci because they produce a significant number of isomiRs, 43 and 34 respectively, that we can use to interpret the correlation patterns. As can be seen from Figure 4D–E, HCL revealed that in both cases the isomiRs with a shortened 5' end had correlated abundances and clustered together: for both miR-21-5p and miR-182-5p the "green" group comprised most of the 5'-shortened isomiRs and only those. Specifically for the isomiRs of miR-182-5p, the green cluster comprises two subgroups: one subgroup contained short isomiRs that were shortened largely on their 3' ends; the second group contained short isomiRs whose 5' termini were very substantially shifted to the right (in the 5'→3' orientation) of the archetype's 5' terminus.

The above results suggest a complex set of events that regulate the abundance of isomiRs from a given arm. We emphasize that these results are not in conflict with the observation that there was no global change in the isomiR production in tumor as compared to the normal tissue (Supplementary Figure S3). In this section we focused on the processing of specific loci and hairpin arms, while the results of Figure 1D–E and Supplementary Figure S3 describe the totality of isomiRs without considering possible mechanisms that regulate the abundance of isomiRs from specific loci/hairpin arms. The absence of clusters that comprise isomiRs with the same transcriptional history (Figure 4A–B) emphasizes that the regulation is post-transcriptional in nature and operates independently on each hairpin and each hairpin arm (Figure 4A–C). Nonetheless, the selection

of a relatively extreme 5' endpoint appears to have a higher priority than that of the 3' end (Figure 4D–E).

The Ago loading of isomiRs is cell-type specific

In our previous work, we reported that isomiRs are loaded on Ago (30). To test the same hypothesis in the BRCA context, we used publicly available Ago HITS-CLIP from three different BRCA model cell-lines, MCF-7 (ER+/PR+/HER2-), MDA-MB-231 (ER-/PR-/HER2-) and BT474 (ER+/PR+/HER2+) (34). The number of distinct isomiRs found to be loaded on Ago in each cell line was largely the same in all three cases, as more than 75% of the isomiRs had significant expression in all three cell lines (Figure 5A; the full data set is provided in Supplementary Table S4). However, in terms of their relative Ago-loaded abundance, the three cell lines exhibited distinct profiles, as PCA (Figure 5B) and HCL (Figure 5C) revealed that the replicates of each cell line clustered together and distinctly from other cell lines. As each cell line corresponds to a different molecular BRCA subtype, these data suggest that different cell lines exploit the miRNA/isomiR machinery in different ways and therefore differentially regulating gene expression.

IsomiRs from the same hairpin arm differentially regulate the cellular transcriptome

In light of the differences that we observed in both normal and breast tumor samples and the different Ago-loaded isomiR profiles in cell lines, we sought to explore how different isomiRs affect the mRNA transcript profile of a cell. Currently, there is very limited knowledge regarding the targets of isomiRs (31–33) especially for isomiRs that differ on their 5' termini, which are expected to have differences on the seed region of the corresponding isoform. The seed, which comprises positions 2 through 7 from the 5' end of a miRNA, is known to be a key determinant of a miRNA's targetome (56–58). We focused on the isomiRs of the miR-183-5p locus, as the locus' role in BRCA, and cancer in general, has been previously demonstrated (59–61). However, the isomiRs' functional roles remain unexplored with no, to the best of our knowledge, previously reported systemic studies. In TNBC, 10 isomiRs of this locus are upregulated in white subjects as compared to the adjacent normal tissue but not in black subjects (Supplementary Table S3).

We selected the archetype miRNA and two more isomiRs of the same length that are found expressed in the BRCA data set (Supplementary Table S1). The two isomiRs were shifted two base pairs upstream (miR-183-5p1-2l-2l) or downstream (miR-183-5p1+2l+2l). We transfected MDA-MB-231 cells with a miRNA mimic for each isomiR and also with a negative control RNA mimic and analyzed the extracted RNA with microarrays. PCA and HCL on the top 80% expressed mRNA transcripts indicated that each treatment had a distinct effect on the transcriptome of the cells (Figure 6A–B). The transcript profiles of the samples are well correlated (the low values of the Y axis of Figure 6B and Supplementary Figure S6A); however, the three replicates in each treatment formed clear tight clusters (Figure 6A–B). We used SAM to find DA genes af-

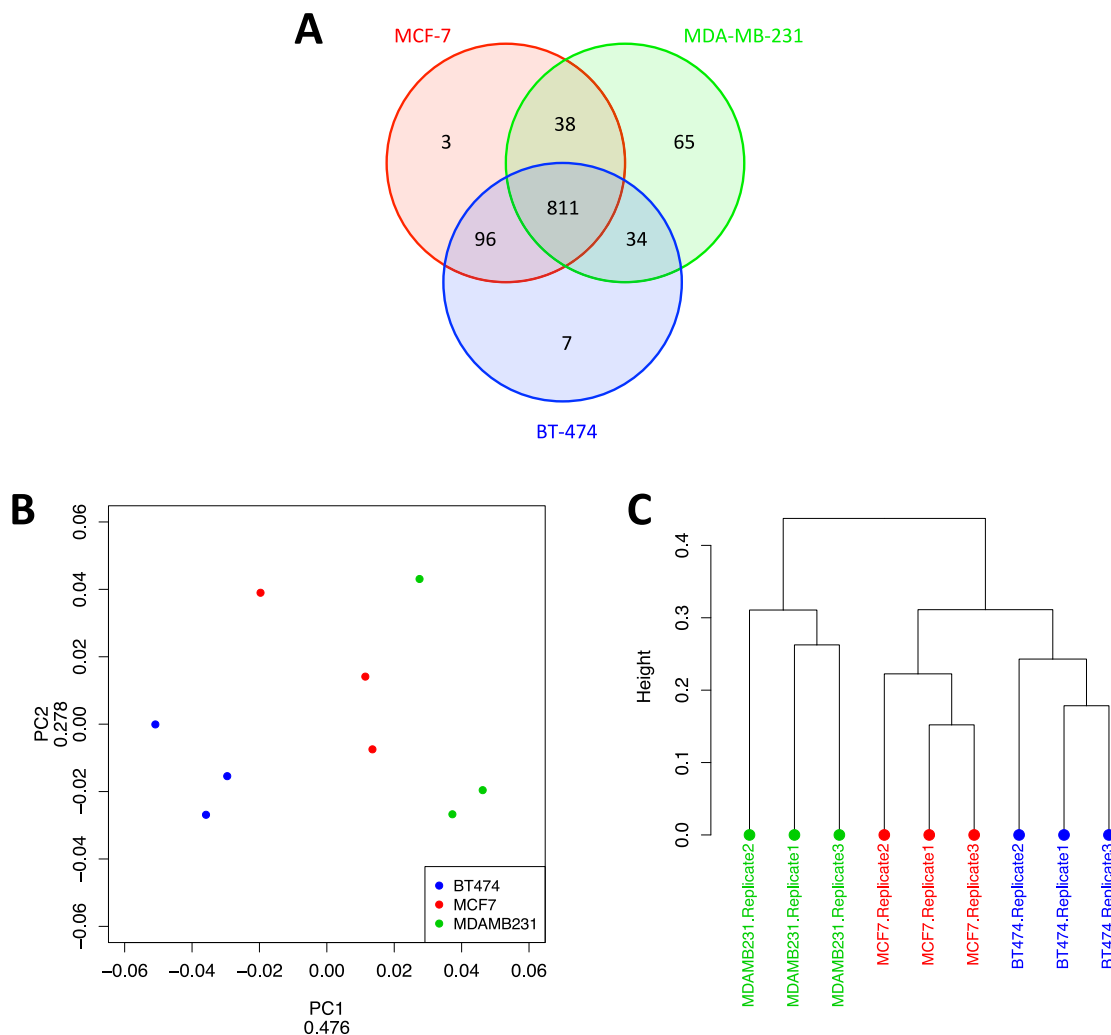


Figure 5. Ago-loaded isomiR profiles are different among BRCA model cell lines. Venn diagram showing how many distinct isomiRs are found in the Ago HITS-CLIP of each cell line (A). Principal Component Analysis (PCA) (B) and Hierarchical Clustering (HCL; metric: Kendall's tau coefficient) (C) indicated that the Ago-loaded isomiR profile is characteristic for each cell line and consistent among replicates. The numbers on the axes labels of PCA indicate the variance explained by each principal component (PC).

ter each isomiR treatment (Supplementary Figure S6B–6D and Supplementary Table S5) as compared to the control treatment. Figure 6C shows the number of DA mRNA transcripts after the treatment with each isomiR as compared to the control one. It is evident that the isomiRs regulate the cell transcriptome differently as there are fewer than 15% of all the mRNAs that are common to all three cases (Figure 6C). A similar trend was also obvious for the upregulated transcripts (Figure 6C). We ran DAVID (46,47) in order to identify significantly enriched GO BP terms in the transcripts with significantly differentiated abundance, separately for each isomiR and direction (up or downregulated) (Supplementary Table S6) and found that no enriched biological process was common to the three groups for either the upregulated or downregulated mRNAs (Supplementary Figure S6E–6F). For example, the archetype miRNA led to the decrease in abundance of mRNAs encoding proteins involved in the regulation of kinase activity and phosphorus metabolism whereas the miR-183-5p|+2l+2l isomiR de-

creased the abundance of genes involved in Ras signaling and ribonucleotide monophosphate metabolic processes (Supplementary Table S6). Interestingly, we found a set of genes that are regulated in opposite ways by different isomiRs (Figure 6D). For example, the transcript of the *EGFR* gene was upregulated after overexpression of the archetype miR-183-5p miRNA but downregulated after the transfection with the miR-183-5p|–2l–2l mimic, as compared to the control treatment (Figure 6D). The opposite trend was observed for *NRAS*. No specific BP term was found enriched in this group of genes (data not shown). Examples of the expression of DA transcripts are shown in Supplementary Figure S7.

To determine whether each negatively significant transcript is a potential direct target of the respective isomiR, we used rna22 (48) to predict target sites on the mRNA sequence. Supplementary Table S7 contains all the predictions for the three isomiRs with which we experimented; the predictions are summarized in Supplementary Figure S8.

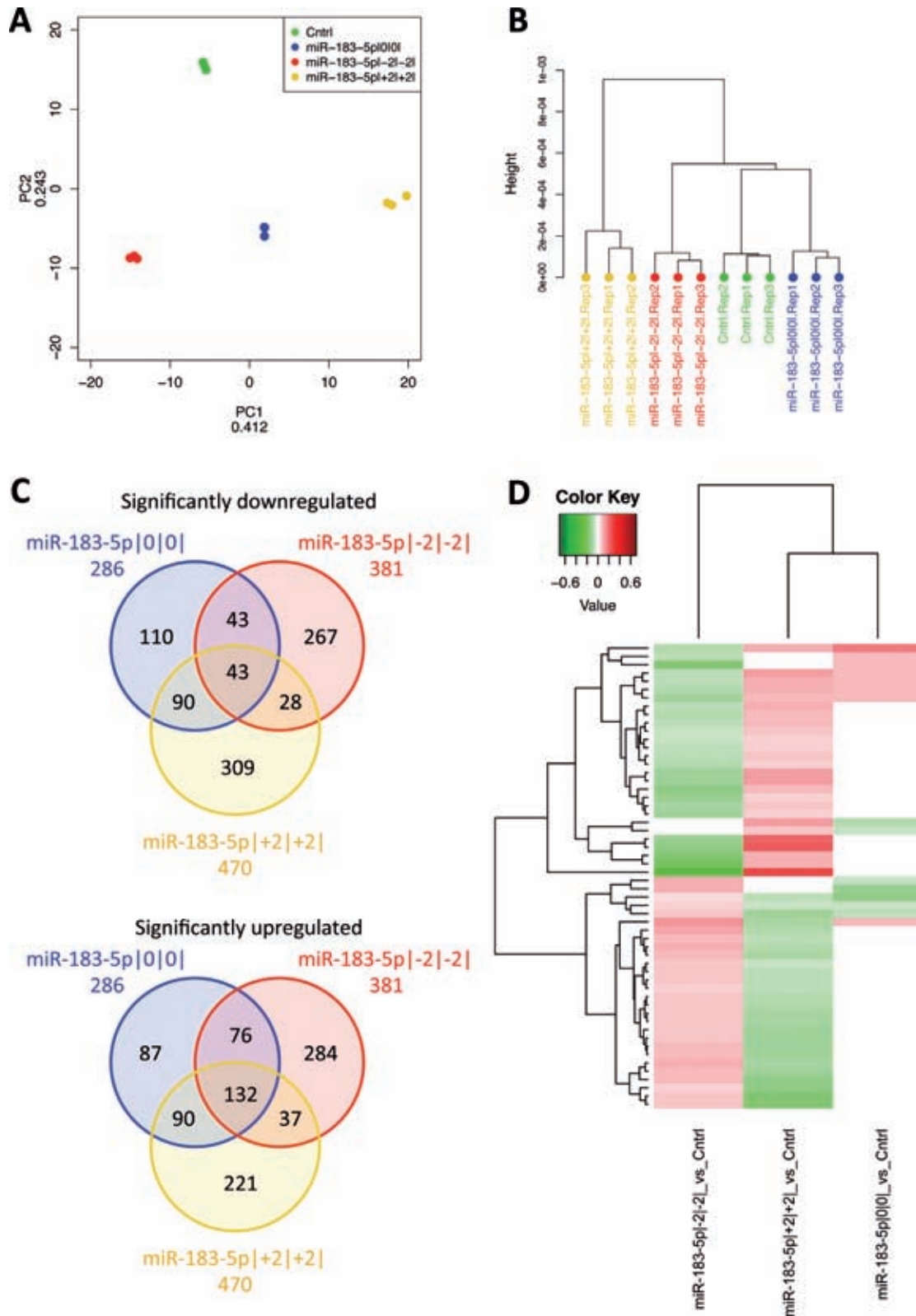


Figure 6. Different isomiRs from the same locus have different effects on the cell transcriptome. Principal Component Analysis (PCA) (A) and Hierarchical Clustering (HCL; metric: Pearson correlation) (B) on the mRNA transcripts of the four treatment groups quantified by microarrays. (C) Venn diagram of the common down—(top graph) and upregulated (bottom graph) mRNAs among the treatments with the specific isomiR as compared to the control treatment. (D) Heatmap of the differentially abundant mRNAs but in opposite directions in at least two comparisons. The numbers on the axes labels of PCA indicate the variance explained by each principal component (PC).

We observed that rna22 can predict that more than 70% of the negatively significant and DA mRNAs can potentially be a direct target of each isomiR (Supplementary Figure S8).

In summary, these findings provide experimental evidence that in BRCA each of the miR-183-5p isomiRs has a different targetome and the effect of each isomiR on the transcriptome can change drastically even when there is a shift by two nucleotides with respect to the archetype miRNA. The findings also underline the importance of focusing on and separately studying specific isomiRs when trying to gauge the regulatory role that a specific miRNA hairpin arm can have on expressed mRNAs.

IsomiRs can distinguish between the luminal A and luminal B subtypes

The luminal A and luminal B subtypes were established as the most common subtypes early on based on gene expression analyses (8,12). As reported in the original TCGA BRCA analysis, use of the PAM50 gene signature classified the samples in the intrinsic subtypes, including luminal A and luminal B, but a miRNA-based analysis could not distinguish these two subtypes (8). Because of the recognized underlying heterogeneity and intra-group variability, we opted to use a supervised method (PLS-DA) to specifically examine whether the ER+/PR+/HER2– luminal A samples could be distinguished from the luminal B ones using the full isomiR profiles. For this analysis, we only considered BRCA samples from white women. As shown in Figure 7A, the isomiRs can easily separate the two luminal subtypes without any overlap between the resulting clusters. From the PLS-DA model we extracted the isomiRs that contribute the most to the observed discrimination and list them in Table 1. Additionally, we used SAM to identify DA isomiRs between the luminal A and the luminal B subtype (Figure 7B, Table 1 and Supplementary Table S8 and Figure S9A). Both approaches showed concordance in the identified isomiRs. Of note are isomiRs from the miR-141, miR-200c and miR-191-5p loci which have been implicated previously in BRCA biology (see Discussion).

Using the mRNA expression data from the TCGA data portal (8), we performed PCA and HCL on the ER+/PR+/HER2– luminal A and luminal B data sets (Supplementary Figure S9B and C). Both of these unsupervised methods showed a good separation of the two BRCA subtypes based on mRNA expression levels. Using the SAM method we identified 397 differentially expressed mRNAs (Supplementary Table S9 and Figure S9D): specifically, 125 were downregulated and 272 were upregulated in luminal B compared to luminal A. DAVID analysis of the upregulated mRNAs identified a signature involved in cell proliferation and cell cycle (Supplementary Table S10). Analogously, similar analysis of the downregulated mRNAs revealed a signature related to interactions with the extracellular matrix and cell adhesion (Supplementary Table S10).

We proceeded to integrate our analysis of the mRNA expression data with our isomiR analyses. We used the rna22 tool (48) to predict targets among the differentially abundant mRNA (Supplementary Table S11) and deter-

Table 1. Significant isomiRs for the discrimination of luminal A and luminal B ER+/PR+/HER2– BRCA miRNA profiles

IsomiR	VIP score ¹	Change B versus A ²
MIMAT0004598 hsa-miR-141-5p 0 +1	2.437	↑
MIMAT0003258 hsa-miR-590-5p 0 0	2.404	↑
MIMAT0000617 hsa-miR-200c-3p -1 -1	2.289	↑
MIMAT0004586 hsa-miR-15b-3p 0 -1	2.272	↑
MIMAT0004586 hsa-miR-15b-3p 0 +1	2.261	↑
MIMAT0000440 hsa-miR-191-5p 0 +1	2.257	↑
MIMAT0000432 hsa-miR-141-3p 0 +1	2.23	↑
MIMAT0000440 hsa-miR-191-5p 0 0	2.199	↑
MIMAT0004586 hsa-miR-15b-3p 0 0	2.195	↑
MIMAT0000456 hsa-miR-186-5p 0 +2	2.193	↑
MIMAT0000417 hsa-miR-15b-5p 0 0	2.186	↑
MIMAT0000753 hsa-miR-342-3p 0 +1	2.174	↑
MIMAT0004951 hsa-miR-887-3p 0 +1	2.155	↑
MIMAT0000227 hsa-miR-197-3p 0 -1	2.148	↑
MIMAT0000617 hsa-miR-200c-3p -1 -2	2.122	↑
MIMAT0000759 hsa-miR-148b-3p 0 +2	2.109	↑
MIMAT0000068 hsa-miR-15a-5p 0 +2	2.076	↑
MIMAT0000617 hsa-miR-200c-3p 0 0	2.057	↑
TJU_CMC_MD2.ID02126.3p-miR 0 0	2.049	↑
MIMAT0004586 hsa-miR-15b-3p 0 -2	2.048	↑
MIMAT0000691 hsa-miR-130b-3p 0 0	2.047	↑
MIMAT0005951 hsa-miR-1307-3p 0 -4	2.038	↑
MIMAT0004598 hsa-miR-141-5p 0 +2	2.037	↑
MIMAT0000093 hsa-miR-93-5p 0 0	2.014	↑
MIMAT0022727 hsa-miR-1307-5p 0 +2	2.014	↑
MIMAT0000757 hsa-miR-151a-3p -2 +1	2.013	↑
MIMAT0004509 hsa-miR-93-3p 0 0	2.01	↑

¹ VIP score as calculated by the PLS-DA model of Figure 7A

² Direction of change of each isomiR as identified by SAM (Supplementary Table S8) in luminal B as compared to luminal A samples

mine isomiR-mRNA pairs that are anti-correlated: i.e. the isomiR is more and the mRNA is less abundant in luminal B as compared to luminal A, or *vice versa*. We found that 111 out of the 125 downregulated genes are direct predicted targets of at least one upregulated isomiR (Figure 7C; Supplementary Table S11). Moreover, 86 of the 272 upregulated genes are direct predicted targets of at least one downregulated isomiR. The number of predicted interactions among the isomiR profile and the genes is significantly larger when all the isomiRs are considered compared to the interactions mediated by the archetype miRNAs only (Figure 7C). We note that the anti-correlations described here comprise both abundant and less abundant isomiRs (Supplementary Figure S9E).

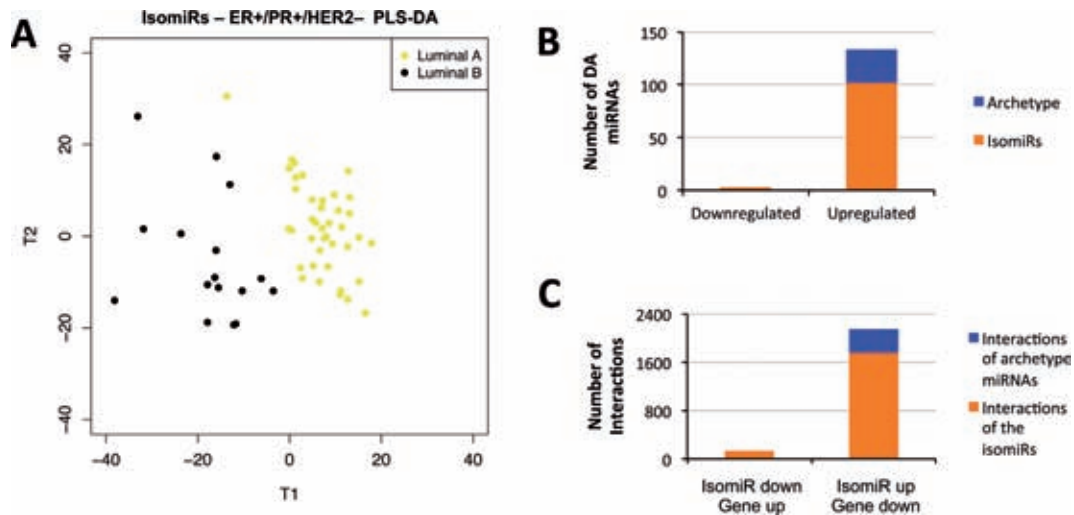


Figure 7. Luminal A and Luminal B intrinsic subtypes have distinct isomiR profiles. (A) Partial Least Squares—Discriminant Analysis (PLS-DA) of the isomiR profiles of ER+/PR+/HER2- BRCA cases from white women. (B) Number of differentially abundant (DA) isomiRs in the luminal B as compared to the luminal A subtype. Each bar divided into the number of DA archetype miRNAs (blue part) and the number of the rest DA isomiRs (orange part). (C) Number of interactions of isomiR-gene anti-pairs. The blue part of each bar represents interactions of archetype miRNAs; the orange of non-archetype ones.

To visualize the isomiR:gene anti-correlated pairs, we constructed a network for each type of pairs with specific biological processes highlighted (Supplementary Figure S10A–B). From this integrative analysis, we observed the evident role of two let-7 isomiRs in inhibiting proliferation and DNA repair/replication (62); these isomiRs have distinct 5' termini and thus distinct seed regions. The archetype let-7c-5p miRNA was *not* found to be DA. Two of the most important isomiRs for the discrimination of the two subtypes, miR-141-5p|0|+1| and miR-200c-3p|–1|–1| had nine and six interactions, respectively. The two networks are highly intrac connected indicating that combinations of isomiRs cooperatively regulate the cell transcriptome. These results strengthen the functional importance of isomiRs in the BRCA context and further support the argument that their deregulation is of significance in the study of the different BRCA subtypes and racial disparities (30).

DISCUSSION

In this study, we expanded our previous work with isomiRs extending it to the disease context and in particular breast cancer. We first catalogued all possible isomiRs that are detectable by RNA-seq and in virtually all cases of known miRNAs we found a clear deviation from the “one arm-one miRNA” dogma. For example, miR-21-5p produces as many as 43 distinct isomiRs. However, the resulting distribution of isomiRs that a locus can produce is skewed with an average of five isomiRs per locus. With regard to the isomiRs' endpoints, and consistent with previous studies by others and us (30,33,63), we observed a variation at the position of each terminus. Specifically, the 3' end was found to be more variable than the 5' end not only in terms of the number of isomiRs ending at each position but also their expression levels. DGCR8/Drosha and Dicer are responsible for the cleavage events during miRNA biogenesis and are commonly considered to be the enzymes choosing

the pri-miRNA and isomiR endpoints (64). However, non-templated nucleotide additions have been reported that can affect miRNA stability, target identification and targeting power (64,65). Lastly, the last few nucleotides of a miRNA have the potential to act as a localization signal for the isoform (66).

Given this increased complexity, we next examined whether considering the full isomiR profile versus only the archetype miRNAs provides any additional discriminatory power when comparing normal and tumor samples. We focused on two common BRCA subtypes, ER+/PR+/HER2- and ER-/PR-/HER2+ and were able to demonstrate in both cases that the full isomiR profile is a more powerful descriptor (Figure 2 and Supplementary Figure S2). In the case of ER+/PR+/HER2-, significance analysis identified isomiRs from many loci as being differentially expressed between the normal and tumor samples. MiR-10b (67) and miR-21 (68), two miRNAs with known and previously documented roles in BRCA, contributed multiple isomiRs to our results. In fact, many of the isomiRs arising from these two loci have modified 5' endpoints and, thus, distinct targetomes in comparison to the archetype miRNA. This is a strong argument for studying *all of the isoforms* that arise from a genomic locus and not only the archetype miRNA.

Having established the importance of the full isomiR profile, we also extended our previous work on population and race differences to the disease context (30). The samples in the BRCA data set allowed us to compare white and black patients with triple negative (ER-/PR-/HER2-) BRCA. Within each race, multivariate statistical analyses clearly separated the normal from the tumor samples (Figure 3). Notably, the isomiRs responsible for differentiating the normal from the disease state in one race were largely different from the isomiRs that differentiated the normal from the disease state in the other race. Further analyses of the tumor samples from both races iden-

tified specific isomiRs that exhibited significantly different expression profiles in each race. Interestingly, isomiRs from TJU_CMC_MD2.ID00121, a novel human-specific miRNA locus we recently reported (35), were among the most significantly differentiated isomiRs between the two races. The genomic locus for TJU_CMC_MD2.ID00121 partially overlaps with the *BMP8B* gene, which has been linked with pancreatic (69) and gastric cancer (70), but is located on the opposite strand from *BMP8B*. The implications of this genomic co-localization are currently unknown. Another miRNA locus, miR-125a, gave rise to isomiRs in a race-dependent manner; this miRNA was investigated previously in BRCA (71) but not in the context of TNBC patients from different races. Lastly, we were also able to show a distinction in the isomiR profile of normal tissues from white and black women extending our previously identified race-dependent differences to another tissue type (30).

Given the evident importance of isomiRs for tumor biology, we proceeded to investigate whether isomiRs have correlated expression. First, we examined isomiRs from the miR-29a and miR-29c loci that lie on different chromosomes but are known to be functionally related. We found a lack of correlation among miR-29 isomiRs from the same genomic locus suggesting the existence of post-transcriptional events that weigh in and establish the abundance of isomiRs (Figure 4A). Along these lines, analysis of the polycistronic transcript that corresponds to the miR-17/92 cluster (28) showed that neither the hairpin nor the arm of origin were the main determinants of the isomiRs' abundance: indeed, isomiRs from different hairpins and different arms formed clusters based on other characteristics, namely the position of the 5' endpoint relative to the reference coordinate and the relative position of the 3' endpoint (Figure 4B). These results indicate that a multitude of molecular and biochemical events and mechanisms, both known (e.g. Droscha and Dicer cleavage (64), non-templated nucleotide additions (65) etc.) and unknown, likely coordinate the expression and abundance of the produced isomiRs. In addition, the machinery that would ultimately determine the collection of isomiRs depends on race, gender, tissue type and disease subtype (Figure 3) (30). Therefore, one would have to account for all these variables when elucidating the biogenesis mechanisms.

As previous work has shown that isomiRs are loaded on Ago (30,33,63), another element that we explored is the loading of the miRNA isoforms on Ago proteins and the targeting of mRNAs. To this end we analyzed the isomiR loading on Ago in three BRCA model cell lines and observed that the overlap of isomiRs among the three cell lines is *qualitatively* high (Figure 5). However, from a *quantitative* standpoint each cell line was characterized by its own specific Ago-loaded isomiR footprint. This could simply be the result of differing transcriptional events among the cell lines but in the absence of the corresponding RNA-seq profiles for these cell lines no firm conclusions can be drawn. Alternatively, these quantitative differences could be the result of currently-not-understood biological mechanisms regulating the relative amount of isomiR loading that is cell-type-specific following the cell-type-specific expression of isomiRs. As different molecular BRCA subtypes are de-

scribed by each of the analyzed cell lines, the results suggest that isomiRs are an integral and important part of the post-transcriptional regulatory layer of gene expression.

As far as targeting is concerned, previous work has shown that isomiRs of the same miRNA locus may target different mRNAs but these mRNAs (or genes) were considered to belong to the same pathways (63). Since, as we showed above, isomiRs with differing 5' ends are also produced, we investigated the effect of the overexpression of three different miR-183-5p isomiRs on the transcriptome of a BRCA model cell line, MDA-MB-231. We used microarrays to globally study how protein-coding gene expression is affected following treatment with each isomiR as compared to a control treatment. Our choice of locus in this case was driven by the fact that miR-183 has been shown to be an important miRNA in many disorders (61,72–74) including breast cancer (59,75,76), and isomiRs from this locus were found to be upregulated in TNBC from white subjects compared to the normal breast tissue, but not in black subjects (Supplementary Table S3). The microarray analysis showed that the transcript profile was clearly distinct for each treatment with a limited overlap on the number of differentially abundant genes among the three isomiR treatments as compared to the control cells (Figure 6). Among the genes that were DA, notable examples include *PDCD4* (61,77), that was downregulated after overexpression of two of the three isomiRs, and *GPRC5A* (78,79), that was upregulated in all isomiR treatments. DAVID analysis revealed that these differentially abundant genes code for proteins that are part of different biological processes (Supplementary Figure S6). A functional overlap is likely to exist, as e.g. the Ras pathway (downregulated after miR-183-5p+2l+2l treatment; Supplementary Table S6) is linked with cell proliferation (downregulated after treatment with the archetype miR-183-5p miRNA; Supplementary Table S6), but a similar physiological impact of the isomiRs cannot be concluded from microarray and gene expression studies. It is also likely that isomiRs work synergistically to direct the MDA-MB-231 cells toward a more aggressive phenotype, as the miR-183-5p locus is considered to work as an oncomiR (59). Interestingly, there was a group of genes that were differentially abundant in two or more treatments with an isomiR as compared to the control treatment but in opposite directions. Importantly, there are significant genes in cancer that fall in this category, like *EGFR* (80), *NRAS* (81), *IL8* and *IL6* (82). From this perspective, an isomiR that will lead to the downregulation of an oncogene will be considered as a tumor suppressor miRNA, although a distinct isomiR from the same hairpin arm could be acting as an oncomiR. However, the interplay among these products, how they interact and how they affect mRNA abundance in cases of co-expression are currently open questions.

The most common BRCA cases are those of the hormone receptor positive (HR+) cancers that are usually classified *molecularly* as luminal A and luminal B (8,12,83). We performed an integrated analysis of the isomiRs and the mRNAs in the luminal A and luminal B ER+/PR+/HER2– tumors and found that the *isomiR profile* can easily discriminate the two subtypes (Figure 7, Table 1 and Supplementary Table S9), which suggests currently not understood roles for the isomiRs in this context. Among the isomiRs that are

most important for the discrimination of the two subtypes were miR-141 and miR-200c, two loci with known links to BRCA progression and aggressiveness (84–86). Through integration with the mRNA data, we show that the full set of isomiRs explains in more detail the mRNA differences between the luminal A and luminal B subtypes (Figure 7 and Supplementary Figure S10).

From an evolutionary perspective, Figure 3F and our previous results (30) suggest that different races, taxa below the rank of human species, have distinct isomiR profiles, as far as normal breast tissue and lymphoblastoid cell lines are concerned. Additionally, we find isomiRs expressed from human- and primate-specific miRNA hairpin arms but no isomiRs from these loci were found to be DA among races (Supplementary Table S3). It was previously proposed that a preference for a dominant 5' isomiR exists and that it can change in different organisms through a mechanism called 'seed shifting' (87,88) resulting in different dominant 5' isomiRs in different species (33). Collectively, these data suggest a potentially dynamic role of isomiRs in evolution.

In conclusion, our results show that by considering all miRNAs isoforms from the transcribed miRNA loci instead of only the archetype miRNAs can help provide a deeper understanding of the post-transcriptional regulatory events that are at play in BRCA. Incorporation of isomiRs in such studies enhances the ability to discriminate normal tissue from tumor samples and to discriminate among different breast cancer subtypes. From a mechanistic standpoint our finding that the isomiR endpoints are not randomly distributed but are in fact correlated with the isomiR expression pattern in ways that are largely unrelated to the miRNA genomic location, miRNA hairpin, and miRNA arm of origin is noteworthy and warrants additional studies. Lastly, our findings show that overexpression of isomiRs from the same arm change the mRNA profile in distinct ways and affect multiple and different biological processes, which in turn highlights the need for future studies of post-transcriptional regulation to be isomiR-centric and not archetype-centric.

ACCESSION NUMBERS

The accession number for the microarray reported in this paper is GEO: GSE71950.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENT

Research reported in this publication was supported by the Kimmel Cancer Center Genomics Shared Resource of Thomas Jefferson University, Philadelphia.

FUNDING

W. M. Keck Foundation grant (IR) [in part]. Funding for open access charge: The Computational Medicine Center of Thomas Jefferson University.

Conflict of interest statement. None declared.

REFERENCES

- DeSantis, C., Ma, J., Bryan, L. and Jemal, A. (2014) Breast cancer statistics, 2013. *CA: Cancer J. Clin.*, **64**, 52–62.
- Thakkar, J.P. and Mehta, D.G. (2011) A review of an unfavorable subset of breast cancer: estrogen receptor positive progesterone receptor negative. *Oncologist*, **16**, 276–285.
- Dawson, S.J., Provenzano, E. and Caldas, C. (2009) Triple negative breast cancers: clinical and prognostic implications. *Eur. J. Cancer*, **45**(Suppl. 1), 27–40.
- Huang, B., Warner, M. and Gustafsson, J.A. (2014) Estrogen receptors in breast carcinogenesis and endocrine therapy. *Mol. Cell. Endocrinol.*, doi:10.1016/j.mce.2014.11.015.
- Jordan, V.C. and Brodie, A.M. (2007) Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids*, **72**, 7–25.
- Vici, P., Pizzuti, L., Natoli, C., Gamucci, T., Di Lauro, L., Barba, M., Sergi, D., Botti, C., Michelotti, A., Moscetti, L. *et al.* (2015) Triple positive breast cancer: A distinct subtype? *Cancer Treat. Rev.*, **41**, 69–76.
- Ademuyiwa, F.O., Edge, S.B., Erwin, D.O., Orom, H., Ambrosone, C.B. and Underwood, W. 3rd (2011) Breast cancer racial disparities: unanswered questions. *Cancer Res.*, **71**, 640–644.
- Cancer Genome Atlas, N. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.
- Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S. *et al.* (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*, **295**, 2492–2502.
- Dietze, E.C., Sistrunk, C., Miranda-Carboni, G., O'Regan, R. and Seewaldt, V.L. (2015) Triple-negative breast cancer in African-American women: disparities versus biology. *Nat. Rev. Cancer*, **14**, 248–254.
- Sorlie, T. (2007) Molecular classification of breast tumors: toward improved diagnostics and treatments. *Methods Mol. Biol.*, **360**, 91–114.
- Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10869–10874.
- Prat, A. and Perou, C.M. (2011) Deconstructing the molecular portraits of breast cancer. *Mol. Oncol.*, **5**, 5–23.
- Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z. *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.*, **27**, 1160–1167.
- Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R. *et al.* (2011) Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod. Pathol.*, **24**, 157–167.
- Dawson, S.J., Rueda, O.M., Aparicio, S. and Caldas, C. (2013) A new genome-driven integrated classification of breast cancer and its implications. *EMBO J.*, **32**, 617–628.
- Eroles, P., Tormo, E., Pineda, B., Espin, E. and Lluch, A. (2015) MicroRNAs in breast cancer: one More Turn in Regulation. *Curr. Drug Targets*, **16**.
- Kusenda, B., Mraz, M., Mayer, J. and Pospisilova, S. (2006) MicroRNA biogenesis, functionality and cancer relevance. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech. Repub.*, **150**, 205–215.
- Hata, A. and Lieberman, J. (2015) Dysregulation of microRNA biogenesis and gene silencing in cancer. *Sci. Signal.*, **8**, re3.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Winter, J., Jung, S., Keller, S., Gregory, R.I. and Diederichs, S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.*, **11**, 228–234.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Schirle, N.T., Sheu-Gruttadauria, J. and MacRae, I.J. (2014) Structural basis for microRNA targeting. *Science*, **346**, 608–613.
- Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. *et al.*

- (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.
25. Berindan-Neagoe, I. and Calin, G.A. (2014) Molecular pathways: microRNAs, cancer cells, and microenvironment. *Clin. Cancer Res.*, **20**, 6247–6253.
 26. Di Leva, G., Garofalo, M. and Croce, C.M. (2014) MicroRNAs in cancer. *Annu. Rev. Pathol.*, **9**, 287–314.
 27. López-Camarillo, C.S. and Marchat, L.A. (2013) *MicroRNAs in cancer*. CRC Press/Taylor & Francis Group, Boca Raton.
 28. Mogilyansky, E. and Rigoutsos, I. (2013) The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ.*, **20**, 1603–1614.
 29. Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
 30. Loher, P., Londin, E.R. and Rigoutsos, I. (2014) IsomiR Expression Profiles in Human Lymphoblastoid Cell Lines Exhibit Population and Gender Dependencies. *Oncotarget*, **5**, 8790–8802.
 31. Vickers, K.C., Sethupathy, P., Baran-Gale, J. and Remaley, A.T. (2013) Complexity of microRNA function and the role of isomiRs in lipid homeostasis. *J. Lipid Res.*, **54**, 1182–1191.
 32. Guo, L. and Chen, F. (2014) A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene*, **544**, 1–7.
 33. Tan, G.C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I.M., Robinson, S., Zhang, S., Ellis, P., Langford, C.F. et al. (2014) 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.*, **42**, 9424–9435.
 34. Pillai, M.M., Gillen, A.E., Yamamoto, T.M., Kline, E., Brown, J., Flory, K., Hesselberth, J.R. and Kabos, P. (2014) HITS-CLIP reveals key regulators of nuclear receptor signaling in breast cancer. *Breast Cancer Res. Treat.*, **146**, 85–97.
 35. Londin, E., Loher, P., Telonis, A.G., Quann, K., Clark, P., Jing, Y., Hatzimichael, E., Kirino, Y., Honda, S., Lally, M. et al. (2015) Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1106–E1115.
 36. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. et al. (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.
 37. Core Team, R. (2014) R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria*, <http://www.R-project.org/>.
 38. Hui, W., Gel, Y.R. and Gastwirth, J.L. (2008) lawstat: An R Package for Law, Public Policy and Biostatistics. *J. Stat. Softw.*, **28**, 1–26.
 39. Lucas, A. (2014) amap: Another Multidimensional Analysis Package. *R Package Version 0.8–14*, <http://CRAN.R-project.org/package=amap>.
 40. Sanchez, G. (2013) Discriminer: Tools of the Trade for Discriminant Analysis. *R Package Version 0.1–29*, <http://CRAN.R-project.org/package=Discriminer>.
 41. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
 42. Chen, H. and Boutros, P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, **12**, 35.
 43. Galili, T. (2014) dendextend: Extending R's dendrogram functionality. *R Package Version 0.18.3*, <http://CRAN.R-project.org/package=dendextend>.
 44. Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.*, **2**, 418–427.
 45. Barker, M. and Rayens, W. (2003) Partial least squares for discrimination. *J. Chemometrics*, **17**, 166–173.
 46. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
 47. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
 48. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B. and Rigoutsos, I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
 49. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
 50. Yancy, H.F., Mason, J.A., Peters, S., Thompson, C.E. 3rd, Littleton, G.K., Jett, M. and Day, A.A. (2007) Metastatic progression and gene expression between breast cancer cell lines from African American and Caucasian women. *J. Carcinog.*, **6**, 8.
 51. Mungall, A.J., Palmer, S.A., Sims, S.K., Edwards, C.A., Ashurst, J.L., Wilming, L., Jones, M.C., Horton, R., Hunt, S.E., Scott, C.E. et al. (2003) The DNA sequence and analysis of human chromosome 6. *Nature*, **425**, 805–811.
 52. Kriegel, A.J., Liu, Y., Fang, Y., Ding, X. and Liang, M. (2012) The miR-29 family: genomics, cell biology, and relevance to renal and cardiovascular injury. *Physiol. Genomics*, **44**, 237–244.
 53. Slusarz, A. and Pulakat, L. (2015) The two faces of miR-29. *J. Cardiovas. Med.*, **16**, 480–490.
 54. Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T. et al. (2011) Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol.*, **8**, 1105–1114.
 55. Chaulk, S.G., Xu, Z., Glover, M.J. and Fahlman, R.P. (2014) MicroRNA miR-92a-1 biogenesis and mRNA targeting is modulated by a tertiary contact within the miR-17~92 microRNA cluster. *Nucleic Acids Res.*, **42**, 5234–5244.
 56. Clark, P.M., Loher, P., Quann, K., Brody, J., Londin, E.R. and Rigoutsos, I. (2014) Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci. Rep.*, **4**, 5947.
 57. Rigoutsos, I. and Tsirigos, A. (2010) In: Slack, F. (ed). *MicroRNAs in Development and Cancer*. Imperial College Press, pp. 237–259.
 58. Xia, Z., Clark, P., Huynh, T., Loher, P., Zhao, Y., Chen, H.W., Ren, P., Rigoutsos, I. and Zhou, R. (2012) Molecular dynamics simulations of Ago silencing complexes reveal a large repertoire of admissible 'seed-less' targets. *Sci. Rep.*, **2**, 569.
 59. Li, P., Sheng, C., Huang, L., Zhang, H., Huang, L., Cheng, Z. and Zhu, Q. (2014) MiR-183/-96/-182 cluster is up-regulated in most breast cancers and increases cell proliferation and migration. *Breast Cancer Res.*, **16**, 473.
 60. Ueno, K., Hirata, H., Shahryari, V., Deng, G., Tanaka, Y., Tabatabai, Z.L., Hinoda, Y. and Dahiya, R. (2013) microRNA-183 is an oncogene targeting Dkk-3 and SMAD4 in prostate cancer. *Br. J. Cancer*, **108**, 1659–1667.
 61. Dambal, S., Shah, M., Mihelich, B. and Nonn, L. (2015) The microRNA-183 cluster: the family that plays together stays together. *Nucleic Acids Res.*, **43**, 7173–7188.
 62. Bussing, I., Slack, F.J. and Grosshans, H. (2008) let-7 microRNAs in development, stem cells and cancer. *Trends Mol. Med.*, **14**, 400–409.
 63. Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heeter, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E. et al. (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.*, **12**, R126.
 64. Starega-Roslan, J., Koscianska, E., Kozlowski, P. and Krzyzosiak, W.J. (2011) The role of the precursor structure in the biogenesis of microRNA. *Cell. Mol. Life Sci.*, **68**, 2859–2871.
 65. Koppers-Lalic, D., Hackenberg, M., Bijnsdorp, I.V., van Eijndhoven, M.A., Sadek, P., Sie, D., Zini, N., Middeldorp, J.M., Ylstra, B., de Menezes, R.X. et al. (2014) Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell Rep.*, **8**, 1649–1658.
 66. Hwang, H.W., Wentzel, E.A. and Mendell, J.T. (2007) A hexanucleotide element directs microRNA nuclear import. *Science*, **315**, 97–100.
 67. Biagioni, F., Bossel Ben-Moshe, N., Fontemaggi, G., Yarden, Y., Domany, E. and Blandino, G. (2013) The locus of microRNA-10b: a critical target for breast cancer insurgence and dissemination. *Cell Cycle*, **12**, 2371–2375.
 68. Chen, J. and Wang, X. (2014) MicroRNA-21 in breast cancer: diagnostic and prognostic potential. *Clin. Transl. Oncol.*, **16**, 225–233.
 69. Cheng, Z., Cui, W., Ding, Y., Liu, T., Liu, W., Qin, Y., Xia, W., Xu, J., Zhang, Y. and Zou, X. (2014) BMP8B mediates the survival of pancreatic cancer cells and regulates the progression of pancreatic cancer. *Oncol. Rep.*, **32**, 1861–1866.
 70. Mima, K., Fukagawa, T., Kurashige, J., Takano, Y., Uchi, R., Ueo, H., Matsumura, T., Ishibashi, M., Sawada, G., Takahashi, Y. et al. (2013) Gene expression of bone morphogenic protein 8B in the primary site,

- peripheral blood and bone marrow of patients with gastric cancer. *Oncol. Lett.*, **6**, 387–392.
71. Lehmann, T.P., Korski, K., Ibbs, M., Zawierucha, P., Grodecka-Gazdecka, S. and Jagodzinski, P.P. (2013) rs12976445 variant in the pri-miR-125a correlates with a lower level of hsa-miR-125a and ERBB2 overexpression in breast cancer patients. *Oncol. Lett.*, **5**, 569–573.
 72. Donatelli, S.S., Zhou, J.M., Gilvary, D.L., Eksioglu, E.A., Chen, X., Cress, W.D., Haura, E.B., Schabath, M.B., Coppola, D., Wei, S. *et al.* (2014) TGF-beta-inducible microRNA-183 silences tumor-associated natural killer cells. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 4203–4208.
 73. Li, X., Kroin, J.S., Kc, R., Gibson, G., Chen, D., Corbett, G.T., Pahan, K., Fayyaz, S., Kim, J.S., van Wijnen, A.J. *et al.* (2013) Altered spinal microRNA-146a and the microRNA-183 cluster contribute to osteoarthritic pain in knee joints. *J. Bone Miner. Res.*, **28**, 2512–2522.
 74. Rizos, E., Sifakas, N., Koumariou, A., Katsantoni, E., Filippopoulou, A., Ntounas, P., Touloumis, C., Kastania, A. and Zoumpourlis, V. (2012) miR-183 as a molecular and protective biomarker for cancer in schizophrenic subjects. *Oncol. Rep.*, **28**, 2200–2204.
 75. Chang, Y.Y., Kuo, W.H., Hung, J.H., Lee, C.Y., Lee, Y.H., Chang, Y.C., Lin, W.C., Shen, C.Y., Huang, C.S., Hsieh, F.J. *et al.* (2015) Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing. *Mol. Cancer*, **14**, 36.
 76. Lowery, A.J., Miller, N., Dwyer, R.M. and Kerin, M.J. (2010) Dysregulated miR-183 inhibits migration in breast cancer cells. *BMC Cancer*, **10**, 502.
 77. Li, J., Fu, H., Xu, C., Tie, Y., Xing, R., Zhu, J., Qin, Y., Sun, Z. and Zheng, X. (2010) miR-183 inhibits TGF-beta1-induced apoptosis by downregulation of PDCD4 expression in human hepatocellular carcinoma cells. *BMC Cancer*, **10**, 354.
 78. Zhou, H. and Rigoutsos, I. (2014) MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA*, **20**, 1431–1439.
 79. Zhou, H. and Rigoutsos, I. (2014) The emerging roles of GPRC5A in diseases. *Oncoscience*, **1**, 765–776.
 80. Masuda, H., Zhang, D., Bartholomeusz, C., Doihara, H., Hortobagyi, G.N. and Ueno, N.T. (2012) Role of epidermal growth factor receptor in breast cancer. *Breast Cancer Res. Treat.*, **136**, 331–345.
 81. Samatar, A.A. and Poulidakos, P.I. (2014) Targeting RAS-ERK signalling in cancer: promises and challenges. *Nat. Rev. Drug Discov.*, **13**, 928–942.
 82. Hartman, Z.C., Poage, G.M., den Hollander, P., Tsimelzon, A., Hill, J., Panupinthu, N., Zhang, Y., Mazumdar, A., Hilsenbeck, S.G., Mills, G.B. *et al.* (2013) Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8. *Cancer Res.*, **73**, 3470–3480.
 83. Ades, F., Zardavas, D., Bozovic-Spasojevic, I., Pugliano, L., Fumagalli, D., de Azambuja, E., Viale, G., Sotiriou, C. and Piccart, M. (2014) Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *J. Clin. Oncol.*, **32**, 2794–2803.
 84. Finlay-Schultz, J., Cittelly, D.M., Hendricks, P., Patel, P., Kabos, P., Jacobsen, B.M., Richer, J.K. and Sartorius, C.A. (2014) Progesterone downregulation of miR-141 contributes to expansion of stem-like breast cancer cells through maintenance of progesterone receptor and Stat5a. *Oncogene*, **34**, 3676–3687.
 85. Le, M.T., Hamar, P., Guo, C., Basar, E., Perdigao-Henriques, R., Balaj, L. and Lieberman, J. (2014) miR-200-containing extracellular vesicles promote breast cancer cell metastasis. *J. Clin. Invest.*, **124**, 5109–5128.
 86. Tuomarila, M., Luostari, K., Soini, Y., Kataja, V., Kosma, V.M. and Mannermaa, A. (2014) Overexpression of microRNA-200c predicts poor outcome in patients with PR-negative breast cancer. *PLoS One*, **9**, e109508.
 87. Berezhikov, E. (2011) Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.*, **12**, 846–860.
 88. Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S. and Peterson, K.J. (2009) The deep evolution of metazoan microRNAs. *Evol. Dev.*, **11**, 50–68.