The novel microRNAs hsa-miR-nov7 and hsa-miR-nov3 are over-expressed in locally advanced breast cancer

Deepak Poduval[1], Zuzana Sichmanova[1], Anne Hege Straume[1,*], Per Eystein Lønning[1,2], Stian Knappskog[1,2,‡]

[1] Section of Oncology, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway.
[2] Department of Oncology, Haukeland University Hospital, 5021 Bergen, Norway.

* Current address: Norwegian Institute of Marine Research, 5005 Bergen, Norway.

‡ Corresponding author: Stian Knappskog, Section of Oncology, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway. E-mail: stian.knappskog@uib.no. Phone +47 55586447.

Author e-mails:

Deepak Poduval	poduval.deepak.b@gmail.com
Zuzana Sichmanova	zuzana.sichmanova@uib.no
Anne Hege Straume	anne.hege.straume@hi.no
Per Eystein Lønning	per.lonning@helse-bergen.no
Stian Knappskog	stian.knappskog@uib.no

Keywords: miRNA; in silico prediction; breast cancer; prognosis

Abstract

miRNAs are an important class of small non-coding RNAs, which play a versatile role in gene regulation at the post-transcriptional level. Expression of miRNAs is often deregulated in human cancers.

We analyzed small RNA massive parallel sequencing data from 50 locally advanced breast cancers aiming to identify novel breast cancer related miRNAs. We successfully predicted 10 novel miRNAs, out of which 2 (hsa-miR-nov3 and hsa-miR-nov7) were recurrent. Applying high sensitivity qPCR, we detected these two microRNAs in 206 and 214 out of 223 patients in the study from which the initial cohort of 50 samples were drawn. We found hsa-miR-nov3 and hsa-miR-nov7 both to be overexpressed in tumor versus normal breast tissue in a separate set of 13 patients (p=0.009 and p=0.016, respectively) from whom both tumor tissue and normal tissue were available. We observed *hsa-miR-nov3* to be expressed at higher levels in ER-positive compared to ER-negative tumors (p=0.037). Further stratifications revealed particularly low levels in the her2-like and basal-like cancers compared to other subtypes (p=0.009 and 0.040, respectively).

We predicted target genes for the 2 microRNAs and identified inversely correlated genes in mRNA expression array data available from 203 out of the 223 patients. Applying the KEGG and GO annotations to target genes revealed pathways essential to cell development, communication and homeostasis.

Although a weak association between high expression levels of *hsa-miR-nov7* and poor survival was observed, this did not reach statistical significance. *hsa-miR-nov3* expression levels had no impact on patient survival.

Introduction

miRNAs are an important class of small non-coding RNAs, playing a versatile role in the gene regulation at the post – transcriptional level [1-5]. These molecules have proven to be involved in vital cellular functions, such as development, differentiation and metabolism [6-8]. In recent years there has been increased focus on the role of miRNAs in cancer [9], and the implementation of next generation sequencing (NGS) has lead to the identification of multiple novel miRNAs as well as linked individual miRNA expression and combined signatures to tumor characteristics [10]. Currently there are 2656 distinct human miRNAs identified in the miRbase v22 [11], including more than 700 found to be deregulated in cancers [12].

Breast cancer is the most common malignancy in women. While outcome has improved significantly over the last three decades, resistance to therapy still presents a major challenge causing breast cancer related deaths [13]. As for chemoresistance in general, the underlying biological mechanisms remain poorly understood [14].

Merging evidence has indicated miRNA deregulation to play a role in breast cancer biology and outcome. Dysregulation of miRNAs may affect signal transduction pathways by targeting oncogenes and tumor suppressor genes [15], important to cancer development, progression, metastasis and potentially therapy response [16, 17]. Thus, while miR-10b, miR-125b, and miR-145 are generally downregulated, other miRNAs, like miR-21 and miR-155, are generally upregulated in breast cancer as compared to normal breast tissue [18]. Further, several miRNAs have revealed strong associations to clinical parameters [19, 20]: For example, differential expression of miR-210, miR-21, miR-106b*, miR-197, miR-let-7i, and miR-210, have been identified as a signature with prognostic value and also linked to invasiveness [21]. Moreover, miR-21 has been found linked to breast cancer metastasis and poor survival [22], while mir-29a overexpression has been shown to reduce the growth rate of breast cancer cells [23].

miRNAs are also known to be differentially regulated across different subclasses of breast cancer. E.g. while members of the mir-181 family are up regulated in breast cancer in general, miR-181c in particular is activated by the expression of HER2 gene [24]. Also, miR-140 has been found suppressed by estrogen stimulation in ERα-positive breast cancer cells, most likely due to ER response elements in the flanking element of the miR-140 promoter [25].

In the present study, we analyzed global miRNA expression in 50 locally advanced breast cancers using NGS, aiming to identify novel, potentially breast cancer specific miRNAs. We identified and validated two novel miRNAs (one not previously described and one not previously reported in breast cancer), and subsequently evaluated their expression in an extended patient series (n=223), by high sensitivity qPCR. Both were found over-expressed in breast cancer as compared to normal breast tissue. Considering different breast cancer subtypes, *hsa-miR-nov3* was expressed at particular high levels in ER-positive tumors contrasting lower levels in basal-like and Her2-like tumors. No similar patterns were observed for *hsa-miR-nov7*.

Materials and Methods

Patients

In the present work we have analyzed biopsy material from two breast cancer studies.

1) In the first study, incisional biopsies were collected before chemotherapy from 223 patients with locally advanced breast cancer in a prospective study designed to identify the response to epirubicin (n = 109) and paclitaxel (n = 114) monotherapy. Primary response to therapy as well as long-term follow up (>10 years or death) was recorded for all patients. This cohort has been described in detail previously [26].

2) In the second study, tumor breast tissue and normal breast tissue from tumor bearing and non-tumor bearing quadrants were collected from 46 anonymous breast cancer patients undergoing mastectomy, with the purpose of determining tissue estrogens. This cohort is described in detail in [27].

Using NGS, we analyzed miRNA expression in 50 patients from study 1). Next, candidate miRNAs were quantified using qPCR in all 223 patients from study 1), as well as 13 randomly selected patients from study 2), where RNA was available from tumor tissue and matching normal breast tissue (7 ER-positive and 6 ER-negative tumors). In addition, mRNA expression array data was available for 203 out of the 223 patients in study 1).

All patients provided written informed consent, and the studies conducted in accordance to national laws, regulation and ethical permissions.

Tissue Sampling and RNA extraction

Tissue samples were snap-frozen in liquid nitrogen in the operating theatre and stored in liquid nitrogen until further processing. Total RNA was extracted from the biopsies using miRvanaTM kit (ThermoFisher), according to the manufacturer's instructions. RNA integrity and concentration were determined using Bioanalyzer 2000 and Nanodrop ND2000 spectrophotometer, respectively.

miRNA-sequencing

Sample preparation and single-end sequencing were performed at the core facility of the Norwegian Genomics Consortium in Oslo, on Illumina HiSeq 2500, 1x50bp. De-multiplexing was performed using the Illumina CASAVA software. FastQC was run on all samples with the main purpose to assess sequence quality.

Novel miRNA prediction

The raw sequencing files (fastq) were processed using the novel miRNA prediction algorithm mirdeep v2.0.0.5 [10]. Potential novel miRNAs were identified using the human reference

genome (hg19) and already identified miRNAs from humans and other hominids from miRbase 20 [28]. In the mirdeep2 algorithm, filtering parameters randfold P-value less than 0.05 and scores greater than or equal to 10 were applied. Precursor structures obtained after filtering were manually identified based on the presence of 1-2 mismatches in the stem region, a loop sequence of 4–8 nt, and the presence of mature sequence in the stem region (See Suppl. Info.) [29].

cDNA synthesis and qPCR

cDNA from miRNAs was prepared using Exiqon's Universal cDNA synthesis kit II, with 20 ng of total RNA as input. qPCR was performed using Exiqon's miRCURY LNA™ Universal RT microRNA PCR system, with custom Pick-&-Mix ready to use PCR plates with an inter-plate calibrator, on a LightCycler 480 instrument (Roche). Relative expression levels for each sample were calculated by dividing the expression of the gene of interest on the average expression of two reference miRNAs: miR-16-5p and miR-30b-5p.

miRNA cloning and capillary sequencing

End products from custom miRNA specific qPCR were cloned into pCR 2.1 TOPO-TA vector (Life Technologies) by TOPO-TA cloning according to the manufacturer's instructions. The generated plasmids were amplified by transformation and cultivation of E. coli TOP10 cells (Life Technologies). The plasmids were then isolated using the Qiagen miniprep kit according to the manufacturer's instructions.

Sequencing was performed using the BigDye v.1.3 system (Applied Biosystems) and the primers following thermocycling conditions as previously described [30]. Capillary electrophoresis and data collection were performed on an automated capillary sequencer (ABI3700).

Target prediction and pathway analysis

Target prediction was performed using the offline algorithm miRanda [31, 32] and the online algorithms miRDB [33] and TargetScanHuman Custom (Release 5.2) [34].

miRanda predicts gene targets based on position specific sequence complementarity between miRNA and mRNA using weighted dynamic programming, an extension of the Smith-Waterman algorithm [35]. Also, the miRanda algorithm uses the free energy estimation between duplex of miRNA:mRNA (Vienna algorithm [36]) as an additional filter.

The miRDB is an online database of animal miRNA targets, which uses SVM (Support Vector Machine) machine-learning algorithm trained with miRNA-target binding data from already known and validated miRNA-mRNA interactions [33, 37].

TargetScanHuman Custom predicts biological miRNA targets by searching for match for the seed region of the given miRNA that is present in the conserved 8-mer and 7-mer sites [34]. It also identifies sites with conserved 3' pairing from the mismatches in the seed region [38, 39].

An in-house pan-cancer panel of 283 tumor suppressor genes was used to filter target genes of interest. The panel was generated based on the tumor suppressors within the

CGPv2/3-panels [40], Roche's Comprehensive Cancer Design as well as a manual literature search (Supporting Table 1).

Further, we used GATHER, a functional gene enrichment tool, which integrates various available biological databases to find functional molecular patterns, in order to find biological context from the target gene list [41]. With the help of GATHER, we did KEGG pathway [42], and GO (gene ontology) enrichment analyses for the common genes predicted by all three prediction algorithms.

mRNA expression

In the interest of validating miRNA targets, we analyzed inverse correlations between miRNA expression and mRNA levels. mRNA expression levels were extracted from microarray analyses performed on a Human HT-12-v4 BeadChip (Illumina) after labeling (Ambion; Aros Applied Biotechnology). Illumina BeadArray Reader (Illumina) and the Bead Scan Software (Illumina) were used to scan BeadChips. Expression signals from the beads were normalized and further processed as previously described [43]. We re-annotated the data set using illuminaHumanv4.db from AnnotationDbi package, built under Bioconductor 3.3 in R [44], to select only probes with "Perfect" annotation[45]. The probes represented 21043 identified and unique genes (13340 represented by single probe and 7703 represented by multiple probes). In the cases of multiple probes targeting the same gene, we calculated fold difference for these probes. This was done to avoid losing potentially relevant biological information if expression of one probe was significantly higher that expression of another. However, for no genes did we find a fold difference >2 fold. Therefore, the mean expression for each such gene, was calculated based on the values form each probe, weighted according to the number of beads per probe.

Statistics

Expression levels of miRNAs in tumor versus normal tissue were compared by Wilcoxon rank tests for paired samples. Inverse correlations between miRNA expression and mRNA expression were assessed by Spearman tests. The potential impact of the novel miRNAs on long-term outcome (relapse-free survival and disease-specific survival) in breast cancer patients was calculated by Log-rank tests and illustrated by Kaplan-Meier curves, using the SPSS software v.19. All p-values are reported as two-sided.

Results

Novel miRNA prediction

In order to identify novel miRNAs, 50 patients with locally advanced breast cancer (from study 1, see materials and methods) were subject to global miRNA-sequencing using massive parallel sequencing. On average, the dataset resulted in 3 million reads per sample. Using the miRNA identifier module in miRDeep2, we detected 10 novel miRNAs (Table 1). Eight out of these 10 miRNAs were detected in a single sample only, while two were expressed in two or more patients and therefore regarded as the most reliable predictions. These two miRNAs, here temporarily named *hsa-miR-nov3* and *hsa-miR-nov7*, were found in tumor samples from 2 and from 6 patients, respectively. For both of these novel miRNAs, we identified precursor structures with not more than one or two mismatches in the stem region, as well as the presence of mature miRNA sequences (Fig 1; Supporting Fig 1). Therefore, we selected these two miRNAs for further analyses. Notably, while this work was conducted, *hsa-miR-nov7* was identified by another team in lymphomas, and reported as miR-10393-3p [46].

Table 1: Novel miRNA sequences as predicted by mirdeep v2.0.0.5 from massive parallel sequencing of total miRNA in 50 locally advanced breast cancers.

miRNA	Co-ordinate	Mature sequence	Strand	Number of samples
hsa-miR-nov2	chr2:36662749..36662809	AAAAACTGCGATTACTTTTGCA	-	1
hsa-miR-nov3	chr3:186505088..186505149	AAAGCAGGATTCAGACTACAATAT	+	2
hsa-miR-nov3_2	chr3:132393169..132393224	CAAAAACTGCAATTACTTTTGC	+	1
hsa-miR-nov4	chr4:155140075..155140134	AAAAGTAATCGCTGTTTTTG	+	1
hsa-miR-nov7	chr7:138728845..138728903	AATTACAGATTGTCTCAGAGA	-	6
hsa-miR-nov8	chr8:116546693..116546762	TTAGAGCTTCAACCTCCAGTGTGA	-	1
hsa-miR-nov10	chr10:31840034..31840078	CGCGGGTGCTTACTGACCCT	+	1
hsa-miR-nov10_2	chr10:72163928..72163994	GCGGCGGCGGCGGCGGCG	+	1
hsa-miR-nov17	chr17:36760852..36760906	CCCAGCCCCACGCGTCCCCATG	-	1
hsa-miR-nov20	chr20:26189318..26189366	TGGCCGAGCGCGGCTCGTCGCC	-	1

Fig 1: Predicted novel miRNAs. Depiction of novel miRNAs (A) *hsa-miR-nov3* and (B) *hsa-miR-nov7*, identified by miRDeep2, showing (i) predicted mature and star sequences, **exp**, probabilistic model expected from Drosha/Dicer processing and **obs**, observed sequences from sequencing data (ii) density plot for read counts for mature and star sequences as well as (iii) miRNA secondary structure.

In-vitro validation of novel micro RNAs

Next, we aimed to validate our *in-silico* predictions and confirm that the sequences from which we identified *hsa-miR-nov3* and *hsa-miR-nov7* represented bona-fide novel miRNAs expressed in the patients. Utilizing total RNA from the patients found to express the two predicted novel miRNAs, we performed global poly-adenylation and cDNA synthesis followed by miRNA-specific qPCR amplification. For both miRNAs we observed positive qPCR reactions. Further, end products of the qPCRs were then ligated into carrier-plasmids and sequenced. We confirmed the resulting plasmids to contain the predicted miRNA sequences. Further, in both cases, the sequences were flanked by a poly A tail, confirming that the original molecules used as input in the poly-adenylation were present as short 22nt RNAs (Fig 2).

Fig 2: miRNA sequences. Chromatogram of capillary-sequenced qPCR products after *hsa-miR-nov3* (A) and *hsa-miR-nov7* (B) amplification. Highlighted background indicates the 22nt miRNA-sequence region (reverse complementary), followed by the Adenine homopolymer indicating *in vitro* adenylation at the expected site.

Overexpression of hsa-miR-nov7 and hsa-miR-nov3 in breast cancer

Given that the sensitivity for the novel miRNAs was better in the qPCR than in the miRNA massive parallel sequencing (MPS) analysis, we aimed to assess whether the miRNAs were expressed in a limited number of breast cancer patients only (as indicated by their detection in 2 and 6 out of 50 patients in the MPS analysis), or if they were detectable in a higher fraction of patients, when applying a more sensitive detection method. We therefore performed qPCR to quantify the expression levels of *hsa-miR-nov7 and hsa-miR-nov3* in tumor tissue across the entire cohort of patients from study 1 (n = 223). With this method, we detected *hsa-miR-nov7* and *hsa-miR-nov3* in 206 and 214 samples out of total 223 samples respectively, albeit at variable levels (Fig 3).

Interestingly, while no difference in the expression levels of *hsa-miR-nov7* was observed between breast cancer subgroups, we found a significant difference in the expression levels of *has-miR-nov3* related to estrogen receptor status. Thus, the expression levels of *has-miR-nov3* were higher in ER-positive as compared to ER-negative tumors (p=0.037; Fig 4A). Further, assessing the expression levels of the two miRNAs in mRNA-based subclasses of breast cancer according to the Perou classification [47], comparing all five classes, we observed a significant difference between the subtypes with respect to *miR-nov3* expression (p=0.041; Kruskal-Wallis test; Fig 4c). We found *hsa-miR-nov3* levels to be lower in HER2 like (p = 0.009; Mann-Whitney test) and basal-like (p = 0.04; Mann-Whitney) tumors as compared to tumors of the other classes.

Following the finding that the two miRNAs were detectable in more than 90 percent of patients, in order to assess whether the expression of these miRNAs were tumor specific we compared the levels of *hsa-miR-nov7* and *hsa-miR-nov3* expression in breast cancer tissue versus normal breast tissue. For this purpose, we randomly selected 13 patients from a study where samples of breast tumor tissue and matching normal tissue from a non-tumor bearing

quadrant of the same breast were available (study 2, see materials and methods) [27]. We detected expression of the novel miRNAs in both tumor- and normal tissue samples for all 13 patients. Notably, we found *hsa-miR-nov3* expression to be elevated in tumor compared to normal tissue in 10 out of the 13 patients (p=0.009; Wilcoxon test; Fig 5A). Similar findings were observed for *hsa-miR-nov7* with elevated expression in 10 out of 13 tumors (Wilcoxon: p = 0.016; Fig 5B). The level of overexpression (i.e. the ratio of expression levels in tumor versus normal tissue) for the two miRNAs did not correlate to each other (p>0.2; Spearman).

Notably, overexpression of *hsa-miR-nov7* in tumor versus normal tissue was observed predominantly in ER-positive tumors (overexpression in 7 out of 7 ER-positive tumors, contrasting 3 out of 6 ER-negative tumors; p=0.070; Fischer exact test).

Fig 3: Expression of novel miRNAs in breast cancer tissue. Bars indicate the relative expression of *hsa-miR-nov3* (A) and *hsa-miR-nov7* (B) in 223 breast cancer patients.

Fig 4: Expression of novel miRNAs in breast cancer tissue. Expression levels stratified by ER-status (A, B) and by expression subtypes (C, D).

Fig 5: Expression of novel miRNAs in breast cancer tissue. Bars indicate the ratio of expression in tumour tissue vs. matched normal breast tissue in 13 breast cancer patients, for *hsa-miR-nov3* (A) and *hsa-miR-nov7* (B).

hsa-miR-nov7 and hsa-miR-nov3 target prediction

Based on our finding of both novel miRNAs to be overexpressed in breast cancer, we next aimed to elucidate the functional roles for *hsa-miR-nov7 and hsa-miR-nov3* by identifying potential targets. We performed *in silico* target predictions using three different algorithms – miRanda, miRDB and TargetScanHumanCustom. miRanda, whch predicts possible targets from human transcripts in general, predicted 9200 and 12315 target genes for *hsa-miR-nov7 and hsa-miR-nov3*, respectively. miRDB, which contains curated and possible miRNA targets, predicted 570 and 530 target genes each for *hsa-miR-nov7 and hsa-miR-nov3*, respectively, while TargetScanHuman custom predicted 633 target genes for *hsa-miR-nov7,* and 282 target genes for *hsa-miR-nov3*. For increased stringency in our predictions, we restricted the potential targets to the ones called by all three algorithms (Fig 6). This left a total of 97 and 180 potential targets for *hsa-miR-nov3* and *hsa-mir-nov7*, respectively.

The two lists of 97 and 180 predicted gene targets were then used for KEGG pathway analysis and GO enrichment analysis using GATHER. The top 10 KEGG pathways and GO terms for each microRNAs are listed in Table 2. The KEGG and GO annotations for *hsa-miR-nov3* showed pathways that are important in cell development and communication. Similar analysis for *hsa-miR-nov7* unveiled pathways playing a vital role in cell functions such as communication and homeostasis. Thus, both these miRNAs implied cell functions that are vital to cancer development and progression.

In order to further substantiate these *in-silico* predictions, we performed a complete Spearman correlation analysis between the expression levels of *hsa-miR-nov7 and hsa-miR-nov3* and mRNA expression array data available for 203 out of the 233 patients in study 1. Assuming the miRNAs, in general, to execute their function by suppressing gene expression

(mRNA degradation), we restricted the analysis to genes which were negatively correlated to expression of the miRNAs. The top ranking negatively correlated genes are listed in Table 3. Notably, the only genes with Rho-values < -0.2 were *RMND5A* for *hsa-miR-nov3* and *GLUD1* and *SASH1* for *hsa-miR-nov7*. Given that the two novel miRNAs were overexpressed in breast cancer tissue, we went on to restrict the correlation analysis to an in-house list of 283-tumor suppressor genes previously described (Supporting Table 2). Among these tumor suppressors, we found 115 to be negatively correlated to *hsa-miR-nov7* and 119 to *hsa-miR-nov3*. Assessing the intersection between these negatively correlated tumor suppressor genes and the predicted targets, we obtained a list of one gene for *hsa-miR-nov3* (*ATRX*) and three genes for *hsa-miR-nov7* (*APC*, *SFRP2* and *CDH11*), but the correlations were non-significant in all 4 cases (Table 4, Fig 7).

In order to get a broader overview of potential biological function, we selected the 100 gene transcripts with the strongest positive and the top 100 gene transcripts with the strongest negative correlation to the two miRNAs (independent of previous target-predictions) and performed gene ontology analyses. We detected no cancer related pathways or cellular functions to be significantly associated with *hsa-miR-nov7* (Supporting Table S3). However, for *hsa-miR-nov3*, KEGG analysis of the negatively correlated genes revealed associations to Hepatorcellular carcinoma as well as several pathways related to drug metabolism (Supporting Table S4).

Fig 6: Target genes predicted. Venn-diagrams illustrating the number of target genes predicted by TargetScan, mirDB and Miranda for the two novel miRNAs *hsa-mir-nov3* (A) and *hsa-mir-nov7* (B).

Fig 7: Correlations to tumor suppressor genes. Scatter plots showing correlation of target tumor suppressors with A) *hsa-miR-nov3* and B) *hsa-miR-nov7*.

Table 2: Top 10 GO and KEGG annotation.

A) GO annotation - *hsa-miR-nov3*

#	Annotation	ln(Bayes factor)	neg ln(p value)	FE: neg ln(p value)	FE: neg ln(FDR)
1	GO:0009653 [3]: morphogenesis	94.88	7.98	100.5	92.91
2	GO:0007275 [2]: development	87.32	7.58	92.89	85.99
3	GO:0007154 [3]: cell communication	85.41	7.46	90.99	84.5
4	GO:0009887 [4]: organogenesis	74.65	6.99	80.24	74.26
5	GO:0048513 [3]: organ development	74.65	6.99	80.24	74.26
6	GO:0007165 [4]: signal transduction	74.2	6.97	79.77	73.97
7	GO:0007242 [5]: intracellular signaling cascade	66.52	6.59	72.18	66.53
8	GO:0007010 [6]: cytoskeleton organization and biogenesis	55.54	6.04	61.15	55.63
9	GO:0009790 [3]: embryonic development	48.63	5.7	54.28	48.88
10	GO:0006928 [4]: cell motility	47.82	5.65	53.49	48.2

B) KEGG annotation - *hsa-miR-nov3*

#	Annotation	Total Genes With Ann	ln(Bayes factor)	neg ln(p value)	FE: neg ln(p value)	FE: neg ln(FDR)
1	path:hsa04810: Regulation of actin cytoskeleton	35	9.03	4.07	13.96	9.57
2	path:hsa04010: MAPK signaling pathway	36	6.93	3.75	11.8	8.1
3	path:hsa04510: Focal adhesion	32	4.15	3.26	8.94	5.94
4	path:hsa04110: Cell cycle	18	4.1	3.25	9.08	5.95
5	path:hsa04060: Cytokine-cytokine receptor interaction	33	3.23	3.07	7.97	5.24
6	path:hsa04620: Toll-like receptor signaling pathway	17	2.97	3.01	7.92	5.24
7	path:hsa04210: Apoptosis	16	2.13	2.82	7.06	4.55
8	path:hsa04512: ECM-receptor interaction	14	1.17	2.55	6.09	3.72
9	path:hsa04630: Jak-STAT signaling pathway	21	1.01	2.51	5.77	3.52
10	path:hsa05050: Dentatorubropallidoluysian atrophy (DRPLA)	5	0.7	2.41	5.9	3.59

C) GO annotation - *hsa-miR-nov7*

#	Annotation	ln(Bayes factor)	neg ln(p value)	FE: neg ln(p value)	FE: neg ln(FDR)
1	GO:0007154 [3]: cell communication	60.17	6.3	65.79	58.14
2	GO:0007275 [2]: development	54.83	6	60.38	53.43
3	GO:0007165 [4]: signal transduction	50.84	5.81	56.44	49.89
4	GO:0009653 [3]: morphogenesis	48.96	5.72	54.56	48.3
5	GO:0050794 [3]: regulation of cellular process	41.31	5.3	46.94	40.9
6	GO:0009987 [2]: cellular process	40.56	5.26	46.33	40.48
7	GO:0009887 [4]: organogenesis	40.37	5.25	45.94	40.24
8	GO:0048513 [3]: organ development	39.98	5.23	45.54	40.09
9	GO:0007242 [5]: intracellular signaling cascade	39.87	5.22	45.58	40.09
10	GO:0050789 [2]: regulation of biological process	39.18	5.18	44.62	39.27

D) KEGG annotation - *hsa-miR-nov3*

#	Annotation	Total Genes With Ann	ln(Bayes factor)	neg ln(p value)	FE: neg ln(p value)	FE: neg ln(FDR)
1	path:hsa04630: Jak-STAT signaling pathway	27	5.53	3.5	10.48	6.6
2	path:hsa04350: TGF-beta signaling pathway	18	5.22	3.45	10.3	6.6
3	path:hsa04010: MAPK signaling pathway	33	3.15	3.04	7.94	4.57
4	path:hsa04210: Apoptosis	17	2.51	2.91	7.49	4.27
5	path:hsa04620: Toll-like receptor signaling pathway	17	2.28	2.85	7.25	4.24
6	path:hsa04020: Calcium signaling pathway	4	2.23	2.84	0	0
7	path:hsa00471: D-Glutamine and D-glutamate metabolism	3	1.12	2.54	6.48	3.7
8	path:hsa04510: Focal adhesion	29	0.96	2.49	5.64	3.23
9	path:hsa05030: Amyotrophic lateral sclerosis (ALS)	5	-0.17	0	5.04	2.78
10	path:hsa04512: ECM-receptor interaction	13	-0.28	0	4.61	2.39

Table 3: Spearman correlation table for *hsa-miR-nov3* and *hsa-miR-nov7* and their top 25 target genes (ranked by inverse correlation).

A) *hsa-miR-nov3*

Gene Symbol	Estimate	P.value	Expression (mean)
RMND5A	-0.2018	0.0038	14.0750
YES1	-0.1649	0.0184	17.0218
PALM2-AKAP2	-0.1455	0.0378	13.0997
SLC7A1	-0.1224	0.0811	16.8650
RAPGEF5	-0.1208	0.0853	14.9652
CTDSPL2	-0.1196	0.0885	15.4945
SLC4A5	-0.1077	0.1251	15.0101
HIPK1	-0.1046	0.1366	13.3737
ABHD12	-0.0998	0.1555	16.2313
FMNL2	-0.0982	0.1624	16.0939
POU4F1	-0.0933	0.1844	13.4684
RPS6KA3	-0.0905	0.1981	14.6430
LARP1	-0.0890	0.2054	15.0210
WIPI2	-0.0702	0.3184	14.7316
MTCH1	-0.0575	0.4139	18.6604
DIAPH1	-0.0528	0.4530	16.7109
MARCKS	-0.0481	0.4946	18.6286
LUZP1	-0.0453	0.5200	17.1097
DNAJC8	-0.0449	0.5238	18.2152
CLOCK	-0.0436	0.5354	15.7894
SLAMF6	-0.0415	0.5557	15.4277
CDAN1	-0.0405	0.5655	16.6394
PCDH11X	-0.0359	0.6104	13.4661
RYBP	-0.0346	0.6234	16.9184
FGF1	-0.0344	0.6249	13.9423

491 B) *hsa-miR-nov7*
492

Gene Symbol	Estimate	P.value	Expression (Mean)
GLUD1	-0.2274	0.0011	18.0399
SASH1	-0.2095	0.0026	16.9164
MARK1	-0.1883	0.0070	15.0356
ARID5B	-0.1877	0.0072	17.7569
ELOVL5	-0.1854	0.0079	17.5656
PUM1	-0.1707	0.0147	17.8295
PNRC2	-0.1599	0.0224	15.4674
UNC13B	-0.1583	0.0238	15.5633
FLRT2	-0.1581	0.0239	15.7323
ZFHX4	-0.1482	0.0344	14.7383
CHIC1	-0.1479	0.0348	13.5807
MAN1A1	-0.1457	0.0375	15.4956
CPEB2	-0.1387	0.0478	14.6995
PDE4D	-0.1377	0.0495	13.9823
TMED7	-0.1366	0.0514	17.1083
NDFIP1	-0.1280	0.0680	16.1458
CSMD1	-0.1269	0.0704	13.8158
MITF	-0.1187	0.0908	14.0482
ITSN1	-0.1185	0.0915	14.8011
CTDSPL2	-0.1178	0.0932	15.4945
ATAD2B	-0.1178	0.0932	14.9892
SFRP2	-0.1129	0.1080	18.4511
DPP10	-0.1119	0.1110	13.4306
BMPR2	-0.1107	0.1149	17.1664
EIF5A2	-0.1100	0.1174	14.5450

493
494
495

Table 4: List of intersection between correlated tumour suppressor genes and the predicted targets of *hsa-miR-nov3* and *hsa-miR-nov7*.

hsa-miR-nov3	***hsa-miR-nov7***
ATRX	APC
	CDH11
	SFRP2

Expression of hsa-miR-nov7 and hsa-miR-nov3 and clinical outcome in breast cancer

Since both *hsa-miR-nov7* and *hsa-miR-nov3* were overexpressed in the tumor tissue of breast cancer patients, we assessed whether any of the two novel miRNAs were associated to clinical outcomes in study 1 (223 breast cancer patients). Given that these patients were enrolled in a prospective study specifically designed to assess response to primary chemotherapy administered as epirubicin or paclitaxel monotherapy in a neoadjuvant setting [26, 48], we assessed the association of *hsa-miR-nov7 and hsa-miR-nov3* levels with primary therapy response and with long term survival (10-years).

We found no association between any of the two novel miRNAs and primary response to either epirubicin or paclitaxel (data not shown). Regarding survival, we observed a weak association between high levels of *hsa-miR-nov7* and poor survival in the paclitaxel treated arm of the study, with the strongest associations observed for relapse free survival, however, none of these associations reached statistical significance (Fig 8). No effect was observed in the epirubicin treated arm. Further, for *hsa-miR-nov3*, no significant correlation to outcome was recorded.

Given the skewed expression levels between breast cancer subtypes for *hsa-miR-nov3*, we performed survival analyses stratified for ER-status and subtypes. These analyses revealed no significant associations to survival (data not shown).

Fig 8: miRNAs and breast cancer survival. Kaplan-Meier curves showing (i) disease-specific and (ii) relapse-free survival of locally advanced breast cancer patients treated with epirubicin or paclitaxel monotherapy in the neoadjuvant setting (study 1), with respect to expression levels of (A) hsa-miR-nov3 and (B) hsa-miR-nov7 on all samples.

Discussion

We investigated whether we could detect novel, previously undescribed miRNAs and, if so, address their potential association to other defined biological parameters and to outcome in a cohort of locally advanced breast cancer. We successfully predicted 10 new miRNAs, out of

which 2 were deemed reliable because of their detected presence in more than one patient. Although these two novel miRNAs (preliminary termed *hsa-miR-nov7* and *hsa-miR-nov3*) were only predicted from 8 samples among the 50 initially sequenced biopsies, we found them to be expressed in all patients by highly sensitive qPCR at varying levels. In addition to our *in vitro* validations, the qPCR detection validated the initial NGS based analysis, detecting these two miRNAs.

Since expression of the two miRNAs was confirmed in breast tumor tissue from the majority of patients analyzed, we went on to assess the relative expression levels in tumor versus matched normal breast tissue, collected from a non-tumor bearing quadrant. Our finding that both novel miRNAs had higher expression levels in tumor than in normal tissue indicates a potential functional role in breast cancer. However, although being overexpressed, the biological role of these two miRNAs in cancer should be interpreted with caution. The expression levels are very low, and it is therefore uncertain whether they will have a major impact on cellular functions. However, when assessing the potential functional roles of these microRNAs by *in silico* prediction of targets followed by validation using correlation to mRNA-array data, the KEGG and GO annotations for these targets revealed cellular functions of potential importance in development and progression of cancer. As such, our present findings may warrant further investigations into the functions of the two miRNAs. Notably, regarding *hsa-miR-nov3*, it was of particular interest that this miRNA was significantly higher expressed in ER-positive as compared to ER-negative breast cancers. Accordingly, we found relatively high expression levels of *hsa-miR-nov3* in tumors of the luminal and normal-like subtypes, contrasting low expression levels in basal-like and her2-like tumors [49, 50]. This finding may indicate a potential role for *hsa-miR-nov3* restricted to ER-positive tumors.

Regarding potential specific targets, we narrowed these down by first assessing the intersect of three different target prediction algorithms, and then the intersect of this result with a predefined list of tumor suppressors. Although none of the remaining genes after this filtering had a statistically significant inverse correlation with the miRNAs, we identified some potentially interesting connections: For *hsa-miR-nov3,* we propose *ATRX* as a target. This is a gene in the SWI/SNF family, involved in chromatin remodelling, and it has previously been found subject to loss of heterozygosity (LOH) in breast cancer [51]. Importantly, we recently reported mutations in the SWI/SNF family genes to be enriched in relapsed breast cancer as compared to primary cancers [52]. Thus, this supports the hypothesis of a breast cancer promoting function for *hsa-miR-nov3*. For *hsa-miR-nov7*, we propose *APC*, *SFRP2*, and *CDH11* as potential targets. Interestingly, the two former are involved in regulation of the Wnt-signalling pathway [53-55] and both have previously been reported as targets for several miRNAs in breast cancer [56-58]. Taken together, this may imply a role for *hsa-miR-nov7* in Wnt signaling. Notably, during our work with the present project, *hsa-miR-nov7*, was identified by Lim and colleagues and coined miR-10393-3p [46]. They found this miRNA to target genes involved in chromatin modifications associated with pathogenesis of Diffuse large B-cell lymphoma (DLBCL). While this differs from our present finding, it may likely be explained by tissue specific effects of the miRNA.

Regarding any predictive or prognostic role for the two investigated miRNAs, we found no significant impact on survival. While we recorded a non-significant trend towards an association between miRnov7 expression and overall survival in the paclitaxel arm, further studies on larger patient cohorts are warranted to clarify this issue. Alternatively, the miRNAs could play a role in tumorigenesis but not later tumor progression. As such, the observed

overexpression in tumor tissue compared to normal breast tissue may be a remaining signal from tumorigenesis.

Whether cancer related overexpression of the two miRNAs described here is merely consequences of other molecular mechanisms in cancer cells or whether the two miRNAs may be involved in tumorigenesis, but not subsequent cancer progression, remains unknown.

Acknowledgements

We thank Beryl Leirvaag and Gjertrud T. Iversen for technical assistance.

Funding

This work was performed in the Mohn Cancer Research Laboratory. The project was funded by grants from the Trond Mohn Research Foundation, The Norwegian Cancer Society, The Norwegian Research Council and the Norwegian Health Region West.

Conflict of interest

The authors declare no conflict of interest.

References

1. Beezhold KJ, Castranova V, Chen F. Microprocessor of microRNAs: regulation and potential for therapeutic intervention. Mol Cancer. 2010;9:134. Epub 2010/06/03. doi: 10.1186/1476-4598-9-134. PubMed PMID: 20515486; PubMed Central PMCID: PMCPmc2887798.
2. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-54. Epub 1993/12/03. PubMed PMID: 8252621.
3. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. The EMBO journal. 2002;21(17):4663-70. Epub 2002/08/29. PubMed PMID: 12198168; PubMed Central PMCID: PMCPMC126204.
4. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901-6. Epub 2000/03/08. doi: 10.1038/35002607. PubMed PMID: 10706289.
5. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science (New York, NY). 2001;294(5543):853-8. Epub 2001/10/27. doi: 10.1126/science.1064921. PubMed PMID: 11679670.
6. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, et al. A uniform system for microRNA annotation. RNA (New York, NY). 2003;9(3):277-9. Epub 2003/02/20. PubMed PMID: 12592000; PubMed Central PMCID: PMCPmc1370393.
7. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. MicroRNAs and other tiny endogenous RNAs in C. elegans. Current biology : CB. 2003;13(10):807-18. Epub 2003/05/16. PubMed PMID: 12747828.
8. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, et al. The small RNA profile during Drosophila melanogaster development. Developmental cell. 2003;5(2):337-50. Epub 2003/08/16. PubMed PMID: 12919683.

9. Mishra S, Yadav T, Rani V. Exploring miRNA based approaches in cancer diagnostics and therapeutics. Critical reviews in oncology/hematology. 2016;98:12-23. Epub 2015/10/21. doi: 10.1016/j.critrevonc.2015.10.003. PubMed PMID: 26481951.
10. Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol. 2008;26(4):407-15. Epub 2008/04/09. doi: 10.1038/nbt1394. PubMed PMID: 18392026.
11. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73. Epub 2013/11/28. doi: 10.1093/nar/gkt1181. PubMed PMID: 24275495; PubMed Central PMCID: PMCPmc3965103.
12. Wang Y, Yu Y, Tsuyada A, Ren X, Wu X, Stubblefield K, et al. Transforming growth factor-beta regulates the sphere-initiating stem cell-like feature in breast cancer through miRNA-181 and ATM. Oncogene. 2011;30(12):1470-80. Epub 2010/11/26. doi: 10.1038/onc.2010.531. PubMed PMID: 21102523; PubMed Central PMCID: PMCPmc3063856.
13. Lonning PE, Knappskog S, Staalesen V, Chrisanthar R, Lillehaug JR. Breast cancer prognostication and prediction in the postgenomic era. Ann Oncol. 2007;18(8):1293-306. Epub 2007/02/24. doi: mdm013 [pii]
10.1093/annonc/mdm013. PubMed PMID: 17317675.
14. Lonning PE, Knappskog S. Mapping genetic alterations causing chemoresistance in cancer: identifying the roads by tracking the drivers. Oncogene. 2013;32:5315 - 30. doi: 10.1038/onc.2013.48. PubMed PMID: 23474753.
15. Gotte M, Mohr C, Koo CY, Stock C, Vaske AK, Viola M, et al. miR-145-dependent targeting of junctional adhesion molecule A and modulation of fascin expression are associated with reduced breast cancer cell motility and invasiveness. Oncogene. 2010;29(50):6569-80. Epub 2010/09/08. doi: 10.1038/onc.2010.386. PubMed PMID: 20818426.
16. Rask L, Balslev E, Jorgensen S, Eriksen J, Flyger H, Moller S, et al. High expression of miR-21 in tumor stroma correlates with increased cancer cell proliferation in human breast cancer. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica. 2011;119(10):663-73. Epub 2011/09/16. doi: 10.1111/j.1600-0463.2011.02782.x. PubMed PMID: 21917003.
17. Harquail J, Benzina S, Robichaud GA. MicroRNAs and breast cancer malignancy: an overview of miRNA-regulated cancer processes leading to metastasis. Cancer biomarkers : section A of Disease markers. 2012;11(6):269-80. Epub 2012/12/19. doi: 10.3233/cbm-120291. PubMed PMID: 23248185.
18. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;65(16):7065-70. Epub 2005/08/17. doi: 10.1158/0008-5472.Can-05-1783. PubMed PMID: 16103053.
19. Nassar FJ, Nasr R, Talhouk R. MicroRNAs as biomarkers for early breast cancer diagnosis, prognosis and therapy prediction. Pharmacology & Therapeutics. 2017;172:34-49. doi: https://doi.org/10.1016/j.pharmthera.2016.11.012.
20. Goh JN, Loo SY, Datta A, Siveen KS, Yap WN, Cai W, et al. microRNAs in breast cancer: regulatory roles governing the hallmarks of cancer. Biological Reviews. 2016;91(2):409-28. doi: doi:10.1111/brv.12176.
21. Volinia S, Galasso M, Sana ME, Wise TF, Palatini J, Huebner K, et al. Breast cancer signatures for invasiveness and prognosis defined by deep sequencing of microRNA. Proc Natl Acad Sci U S A. 2012;109(8):3024-9. doi: 10.1073/pnas.1200010109. PubMed PMID: 22315424; PubMed Central PMCID: PMCPMC3286983.
22. Yan LX, Huang XF, Shao Q, Huang MAY, Deng L, Wu QL, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA (New York, NY). 2008;14(11):2348-60. doi: 10.1261/rna.1034808. PubMed PMID: 18812439; PubMed Central PMCID: PMCPMC2578865.
23. Wu JZ, Yang TJ, Lu P, Ma W. Analysis of signaling pathways in recurrent breast cancer. Genetics and Molecular Research. 2014;13(4):10097-104. doi: 10.4238/2014.December.4.4. PubMed PMID: WOS:000350229200042.

689 24. Tashkandi H, Shah N, Patel Y, Chen H. Identification of new miRNA biomarkers
690 associated with HER2-positive breast cancers. Oncoscience. 2015;2(11):924-9. Epub
691 2015/12/24. PubMed PMID: 26697527; PubMed Central PMCID: PMCPmc4675790.
692 25. Güllü G. Clinical significance of miR-140-5p and miR-193b expression in patients.
693 2015;38(1):21-9. doi: 10.1590/s1415-475738120140167. PubMed PMID: 25983620; PubMed
694 Central PMCID: PMCPMC4415571.
695 26. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S, et al. CHEK2
696 mutations affecting kinase activity together with mutations in TP53 indicate a functional
697 pathway associated with resistance to epirubicin in primary breast cancer. PLoS ONE.
698 2008;3(8):e3062. Epub 2008/08/30. doi: 10.1371/journal.pone.0003062. PubMed PMID:
699 18725978.
700 27. Lonning PE, Helle H, Duong NK, Ekse D, Aas T, Geisler J. Tissue estradiol is selectively
701 elevated in receptor positive breast cancers while tumour estrone is reduced independent of
702 receptor status. J Steroid Biochem Mol Biol. 2009;117(1-3):31-41. Epub 2009/07/14. doi:
703 10.1016/j.jsbmb.2009.06.005. PubMed PMID: 19591931.
704 28. Sasidharan V, Lu YC, Bansal D, Dasari P, Poduval D, Seshasayee A, et al. Identification of
705 neoblast- and regeneration-specific miRNAs in the planarian Schmidtea mediterranea. RNA
706 (New York, NY). 2013;19(10):1394-404. Epub 2013/08/27. doi: 10.1261/rna.038653.113.
707 PubMed PMID: 23974438; PubMed Central PMCID: PMCPMC3854530.
708 29. Krishna S, Nair A, Cheedipudi S, Poduval D, Dhawan J, Palakodeti D, et al. Deep
709 sequencing reveals unique small RNA repertoire that is regulated during head regeneration in
710 Hydra magnipapillata. Nucleic Acids Res. 2013;41(1):599-616. Epub 2012/11/21. doi:
711 10.1093/nar/gks1020. PubMed PMID: 23166307; PubMed Central PMCID: PMCPMC3592418.
712 30. Knappskog S, Chrisanthar R, Lokkevik E, Anker G, Ostenstad B, Lundgren S, et al. Low
713 expression levels of ATM may substitute for CHEK2 /TP53 mutations predicting resistance
714 towards anthracycline and mitomycin chemotherapy in breast cancer. Breast Cancer Res.
715 2012;14(2):R47. doi: 10.1186/bcr3147. PubMed PMID: 22420423; PubMed Central PMCID:
716 PMC3446381.
717 31. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila.
718 Genome biology. 2003;5(1):R1. Epub 2004/01/08. doi: 10.1186/gb-2003-5-1-r1. PubMed PMID:
719 14709173; PubMed Central PMCID: PMCPmc395733.
720 32. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets.
721 PLoS biology. 2004;2(11):e363. Epub 2004/10/27. doi: 10.1371/journal.pbio.0020363. PubMed
722 PMID: 15502875; PubMed Central PMCID: PMCPmc521178.
723 33. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and
724 functional annotations. Nucleic Acids Res. 2015;43(Database issue):D146-52. Epub 2014/11/08.
725 doi: 10.1093/nar/gku1104. PubMed PMID: 25378301; PubMed Central PMCID:
726 PMCPmc4383922.
727 34. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines,
728 indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20. Epub
729 2005/01/18. doi: 10.1016/j.cell.2004.12.035. PubMed PMID: 15652477.
730 35. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol.
731 1981;147(1):195-7. Epub 1981/03/25. PubMed PMID: 7265238.
732 36. Wuchty S, Fontana W, Hofacker IL, Schuster P. Complete suboptimal folding of RNA and
733 the stability of secondary structures. Biopolymers. 1999;49(2):145-65. Epub 1999/03/10. doi:
734 10.1002/(sici)1097-0282(199902)49:2<145::aid-bip4>3.0.co;2-g. PubMed PMID: 10070264.
735 37. Wang X, El Naqa IM. Prediction of both conserved and nonconserved microRNA targets
736 in animals. Bioinformatics. 2008;24(3):325-32. Epub 2007/12/01. doi:
737 10.1093/bioinformatics/btm595. PubMed PMID: 18048393.
738 38. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved
739 targets of microRNAs. Genome Res. 2009;19(1):92-105. Epub 2008/10/29. doi:
740 10.1101/gr.082701.108. PubMed PMID: 18955434; PubMed Central PMCID: PMCPMC2612969.
741 39. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA
742 targeting specificity in mammals: determinants beyond seed pairing. Molecular cell.

2007;27(1):91-105. Epub 2007/07/07. doi: 10.1016/j.molcel.2007.06.017. PubMed PMID: 17612493; PubMed Central PMCID: PMCPMC3800283.

40. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat Med. 2015;21(7):751-9. doi: 10.1038/nm.3886. PubMed PMID: 26099045; PubMed Central PMCID: PMC4500826.

41. Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics. 2006;22(23):2926-33. Epub 2006/09/27. doi: 10.1093/bioinformatics/btl483. PubMed PMID: 17000751.

42. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30. Epub 1999/12/11. PubMed PMID: 10592173; PubMed Central PMCID: PMCPMC102409.

43. Curtis C. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. 486(7403):346-52. doi: 10.1038/nature10983. PubMed PMID: 22522925; PubMed Central PMCID: PMCPMC3440846.

44. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

45. Barbosa-Morais NL, Dunning MJ, Samarajiwa SA, Darot JF, Ritchie ME, Lynch AG, et al. A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. Nucleic Acids Res. 2010;38(3):e17. Epub 2009/11/20. doi: 10.1093/nar/gkp942. PubMed PMID: 19923232; PubMed Central PMCID: PMCPMC2817484.

46. Lim EL, Trinh DL, Scott DW, Chu A, Krzywinski M, Zhao Y, et al. Comprehensive miRNA sequence analysis reveals survival differences in diffuse large B-cell lymphoma patients. Genome biology. 2015;16:18. Epub 2015/02/28. doi: 10.1186/s13059-014-0568-y. PubMed PMID: 25723320; PubMed Central PMCID: PMCPMC4308918.

47. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52. doi: http://www.nature.com/nature/journal/v406/n6797/suppinfo/406747a0_S1.html.

48. Chrisanthar R, Knappskog S, lokkevik E, Anker G, Ostenstad B, Lundgren S, et al. Predictive and prognostic impact of TP53 mutations and MDM2 promoter genotype in primary breast cancer patients treated with epirubicin or paclitaxel. PLoS ONE. 2011;6(4):e19249. doi: doi:10.1371/journal.pone.0019249.

49. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52. doi: 10.1038/35021093. PubMed PMID: 10963602.

50. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74. doi: 10.1073/pnas.191367098. PubMed PMID: 11553815; PubMed Central PMCID: PMC58566.

51. Roy D, Guida P, Zhou G, Echiburu-Chau C, Calaf GM. Gene expression profiling of breast cells induced by X-rays and heavy ions. International journal of molecular medicine. 2008;21(5):627-36. Epub 2008/04/22. PubMed PMID: 18425356.

52. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32(2):169-84.e7. doi: 10.1016/j.ccell.2017.07.005. PubMed PMID: 28810143; PubMed Central PMCID: PMCPMC5559645.

53. von Marschall Z, Fisher LW. Secreted Frizzled-Related Protein-2 (sFRP2) Augments Canonical Wnt3a-induced Signaling. Biochem Biophys Res Commun. 2010;400(3):299-304. doi: 10.1016/j.bbrc.2010.08.043. PubMed PMID: 20723538; PubMed Central PMCID: PMCPMC2952323.

54. Rattner A, Hsieh JC, Smallwood PM, Gilbert DJ, Copeland NG, Jenkins NA, et al. A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of

797 frizzled receptors. Proc Natl Acad Sci U S A. 1997;94(7):2859-63. PubMed PMID: 9096311;
798 PubMed Central PMCID: PMCPMC20287.
799 55. Hankey W, Frankel WL, Groden J. Functions of the APC tumor suppressor protein
800 dependent and independent of canonical WNT signaling: Implications for therapeutic targeting.
801 Cancer metastasis reviews. 2018;37(1):159-72. doi: 10.1007/s10555-017-9725-6. PubMed
802 PMID: PMC5803335.
803 56. Isobe T, Hisamori S, Hogan DJ, Zabala M, Hendrickson DG, Dalerba P, et al. miR-142
804 regulates the tumorigenicity of human breast cancer stem cells through the canonical WNT
805 signaling pathway. eLife. 2014;3. Epub 2014/11/19. doi: 10.7554/eLife.01977. PubMed PMID:
806 25406066; PubMed Central PMCID: PMCPMC4235011.
807 57. Tan Z, Zheng H, Liu X, Zhang W, Zhu J, Wu G, et al. MicroRNA-1229 overexpression
808 promotes cell proliferation and tumorigenicity and activates Wnt/beta-catenin signaling in
809 breast cancer. Oncotarget. 2016;7(17):24076-87. Epub 2016/03/19. doi:
810 10.18632/oncotarget.8119. PubMed PMID: 26992223; PubMed Central PMCID:
811 PMCPMC5029685.
812 58. Liu S, Wang Z, Liu Z, Shi S, Zhang Z, Zhang J, et al. miR-221/222 activate the Wnt/β-
813 catenin signaling to promote triple negative breast cancer. Journal of Molecular Cell Biology.
814 2018:mjy041-mjy. doi: 10.1093/jmcb/mjy041.
815
816
817

818 Supporting information

819

820 **S1 Fig. Predicted novel miRNAs.** Table on the upper left shows miRDeep2 scores and
821 read counts. RNA secondary structure for miRNA on the top right. Color code for depiction
822 as follows mature sequence in red, loop sequence in yellow and purple for star sequences.
823 Density plot in the middle shows distribution of reads in precursor reads predicted. Dotted
824 lines illustrate alignment and mm, number of mismatches. Exp, is potential precursor model
825 predicted by algorithm with taking accounts of stability based on free energy, position and
826 read frequencies according to Dicer/Drosha processing of miRNA. Obs, is postion and reads
827 found from deep sequencing data. (A) *hsa-miR-nov3* and (B) *hsa-miR-nov7*.

828

829 **S1 Table**. **In-house pan-cancer panel of 283 tumor suppressor genes**. Panel generated
830 based on CGPv2/3-panels [40], Roche's Comprehensive Cancer Design along with manual
831 literature search, to filter target genes of interest.

832

833 **S2 Table. Correlation miRNAs and tumour suppressor genes.** Spearman correlation
834 table for *hsa-miR-nov3* (A) and *hsa-miR-nov7* (B) inversely correlated tumor suppressor
835 genes.

836

837 **S3 Table. Correlations mir7 and gene ontology.**

838

839 **S4 Table. Correlations mir3 and gene ontology.**

840

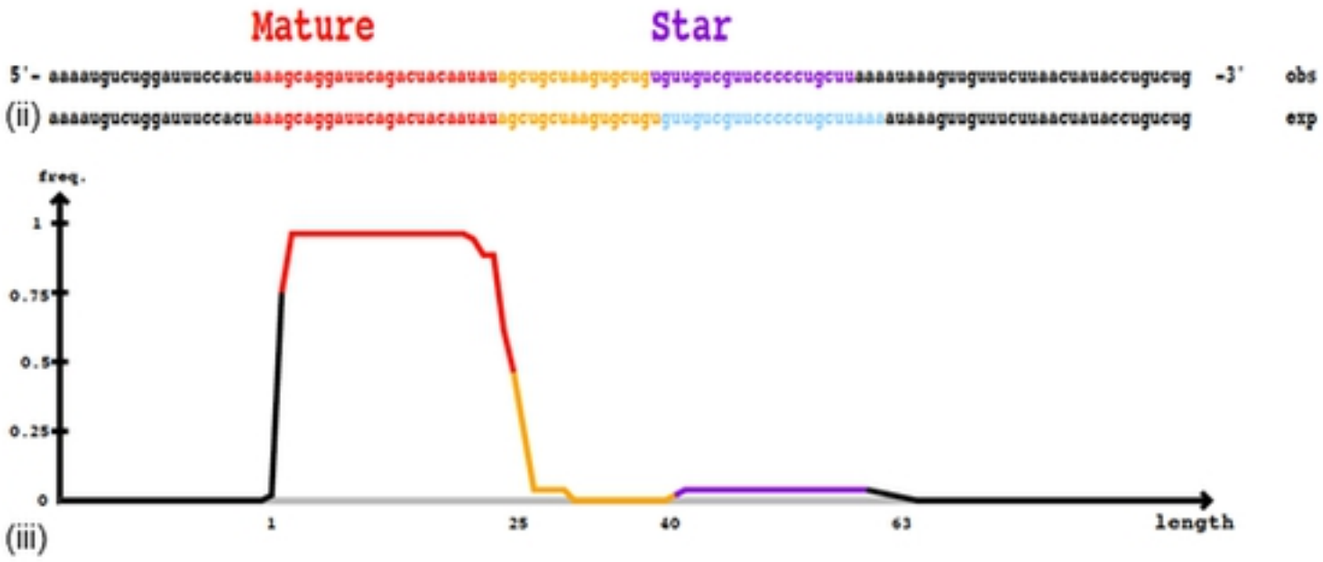
841 **S1 File. Supporting information mirDeep.**

842

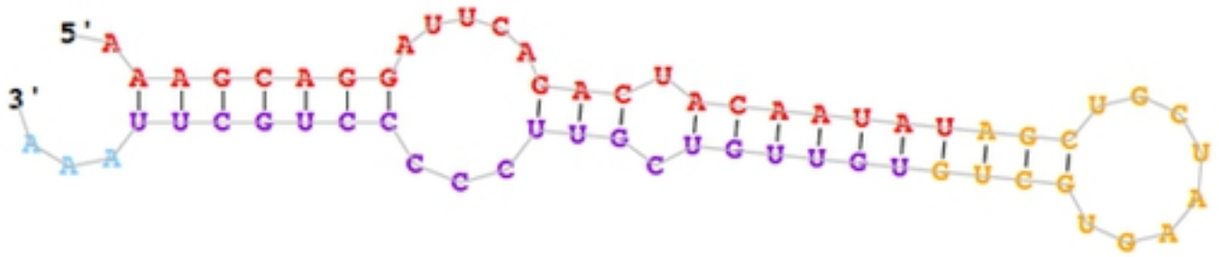
Figure 1

A. ***hsa-miR-nov3***

(i)



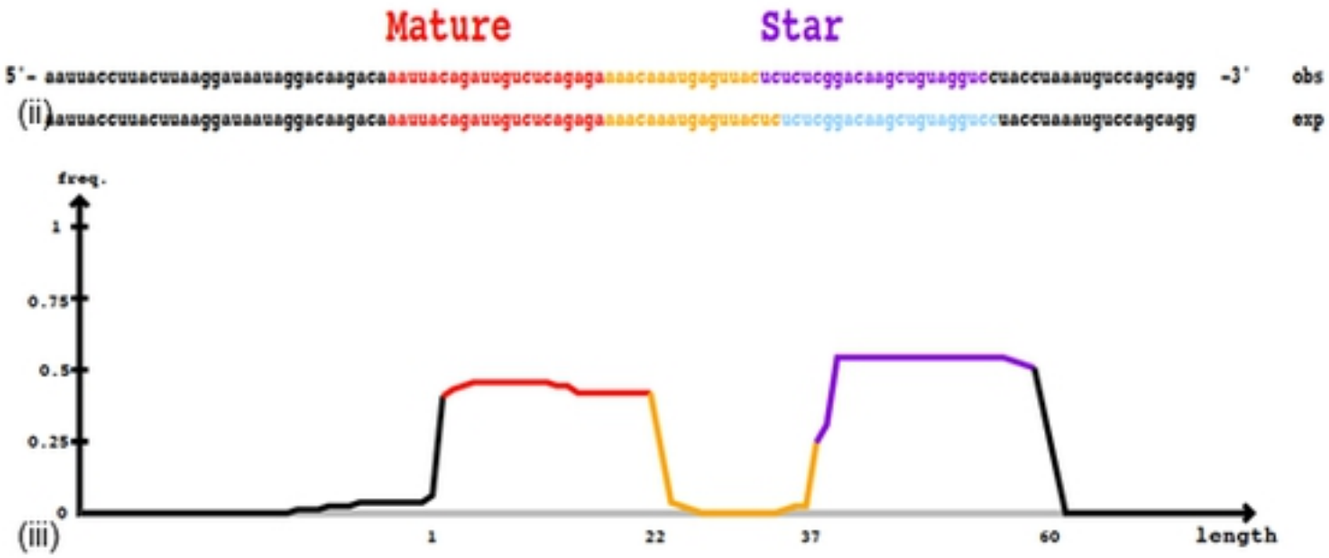
(iii)

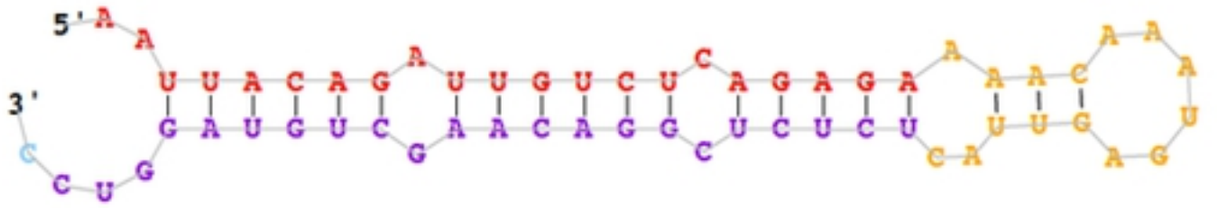


Figure

B. ***hsa-miR-nov7***

(i)

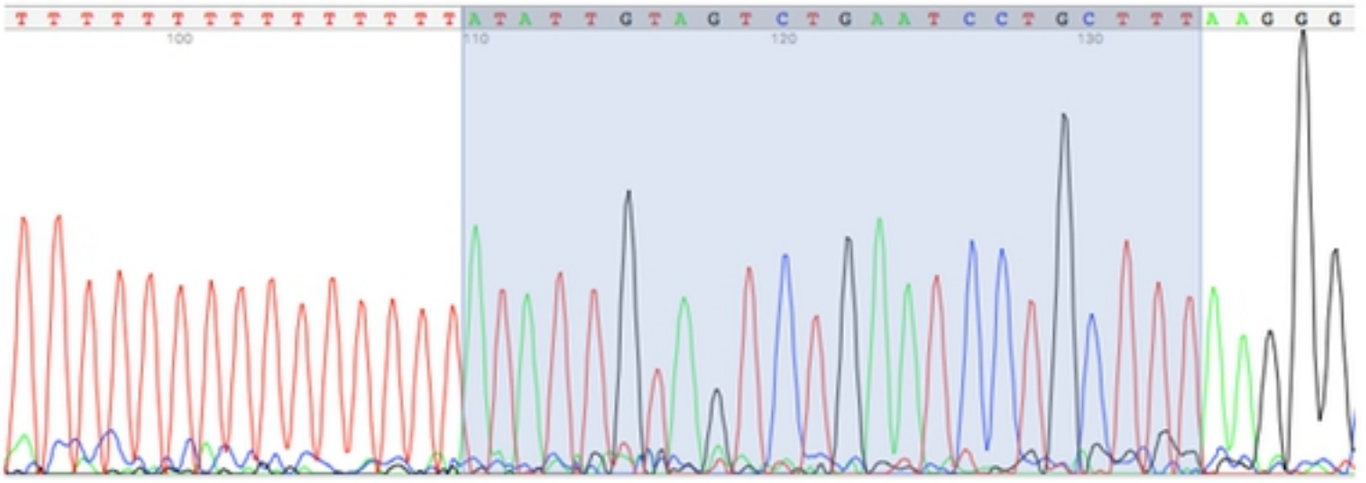




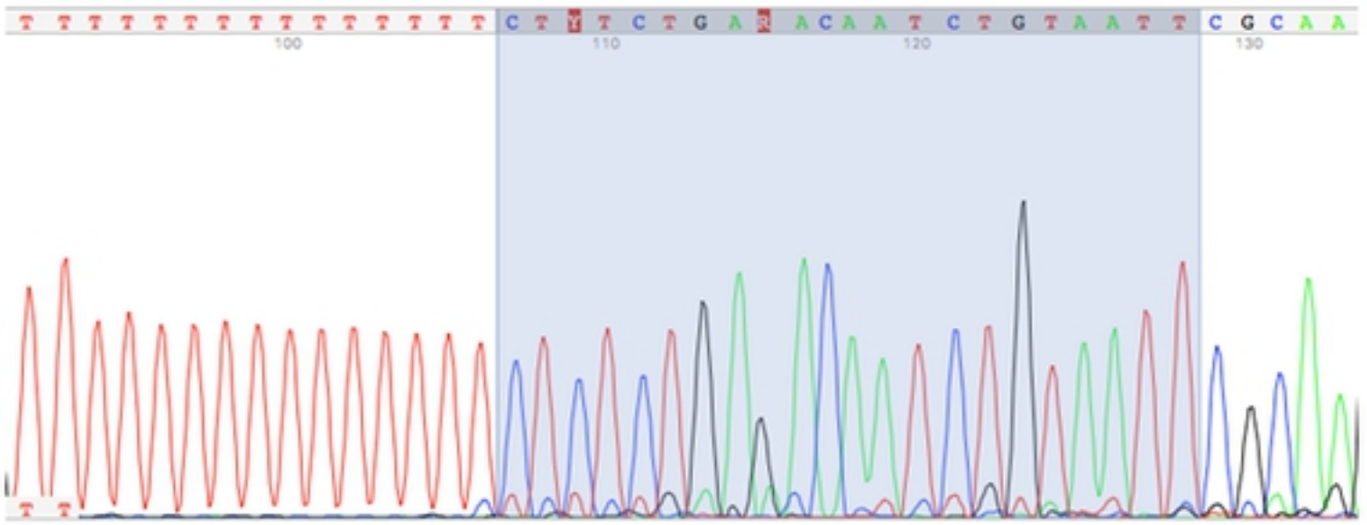
Figure

Figure 2

A)

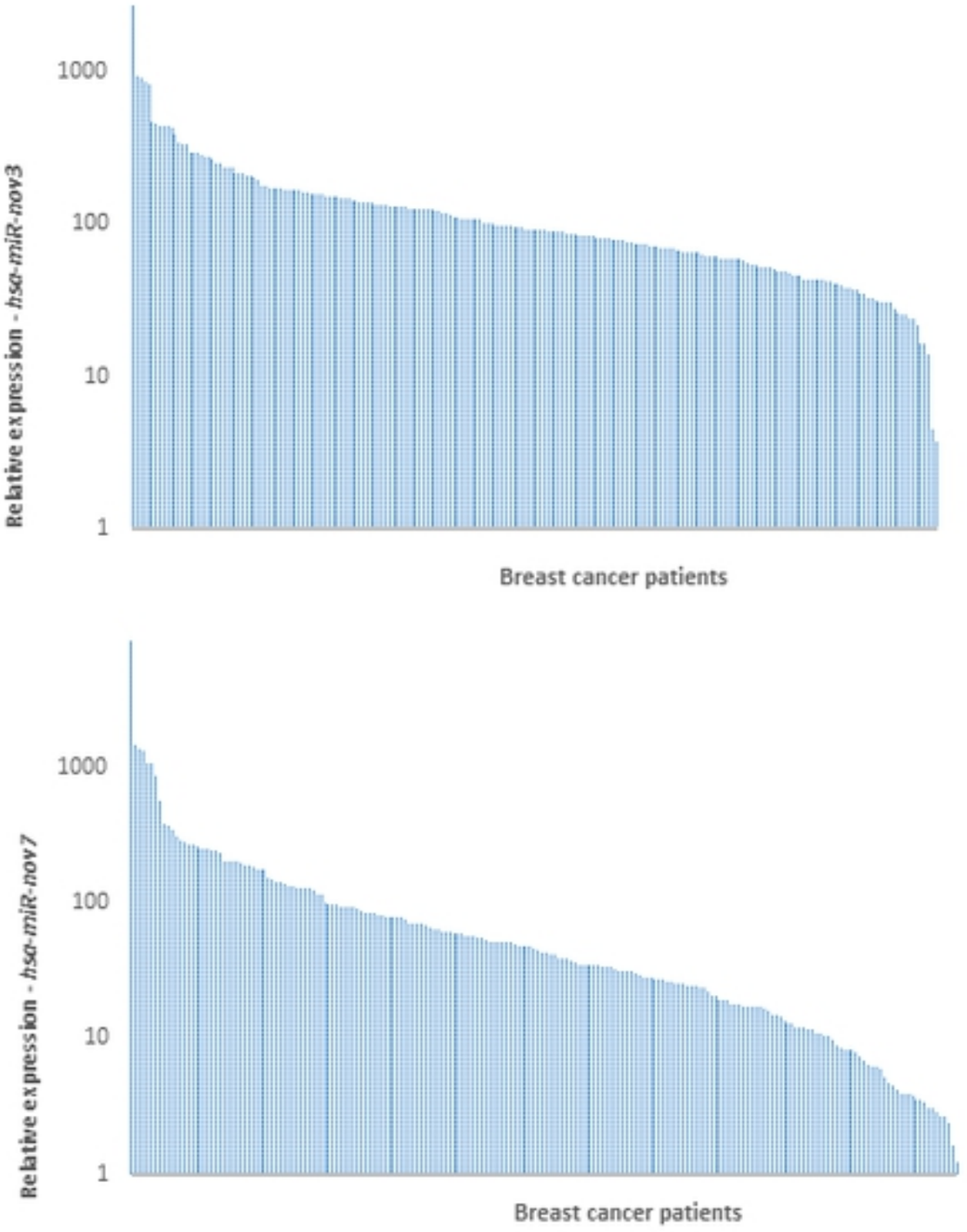


B)



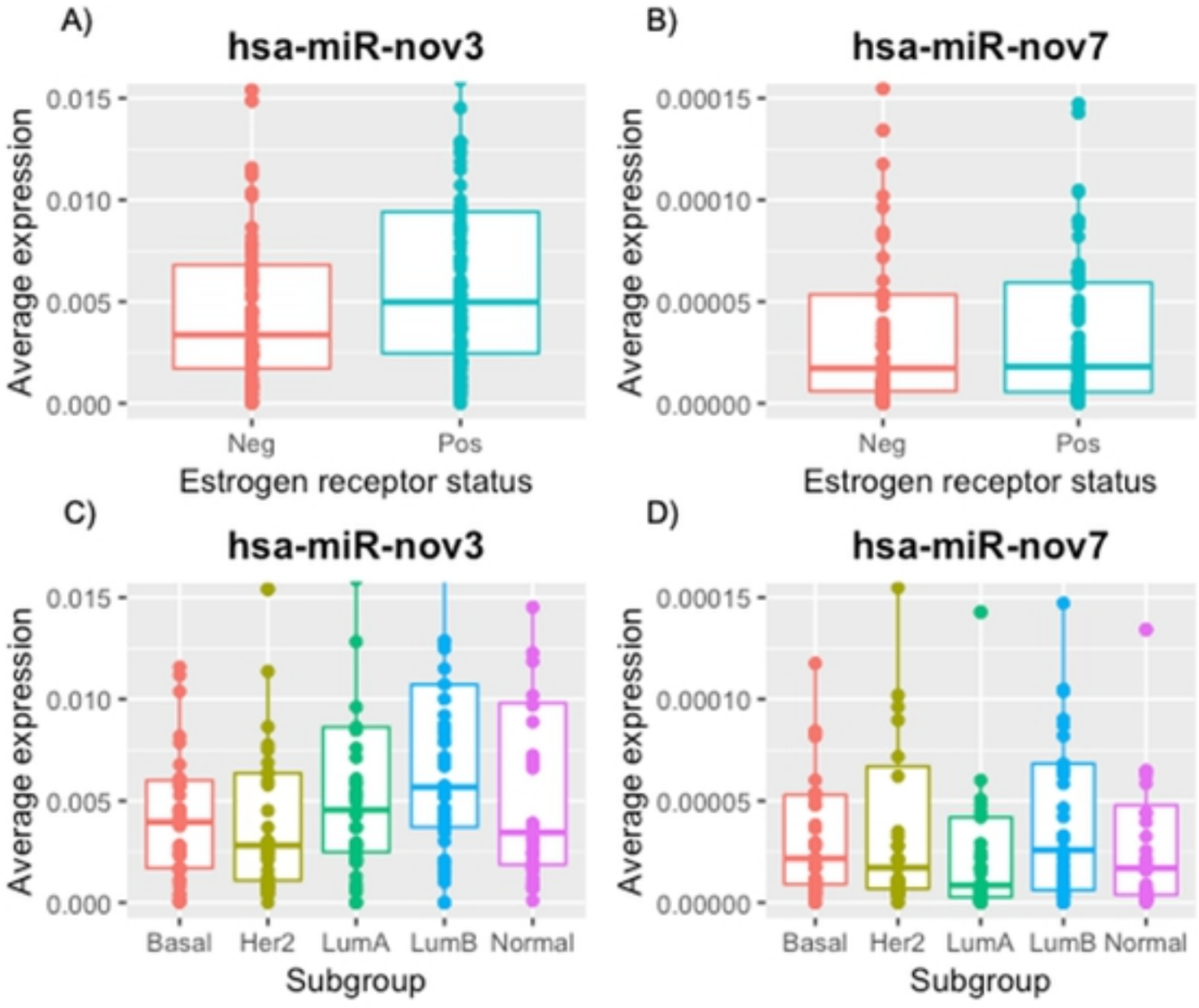
Figure

Figure 3



Figure

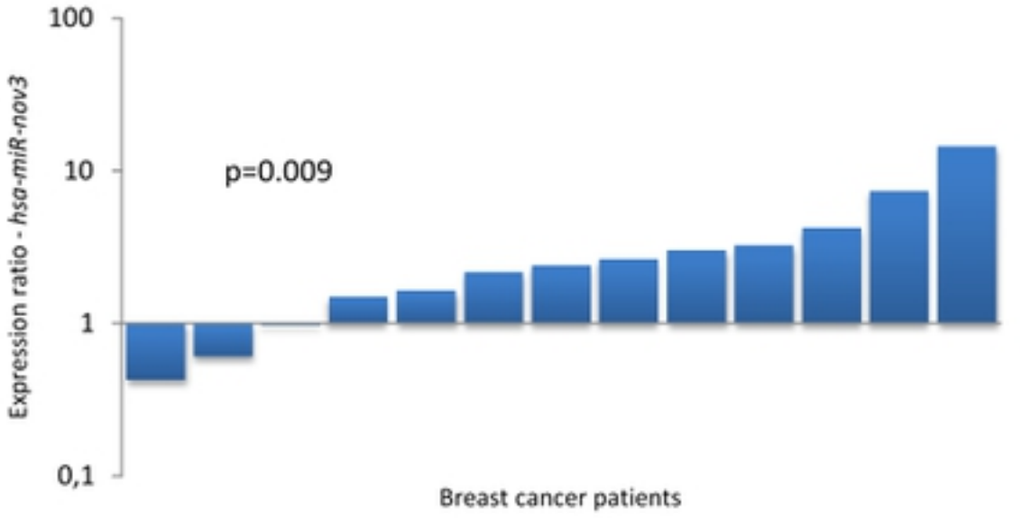
Figure 4



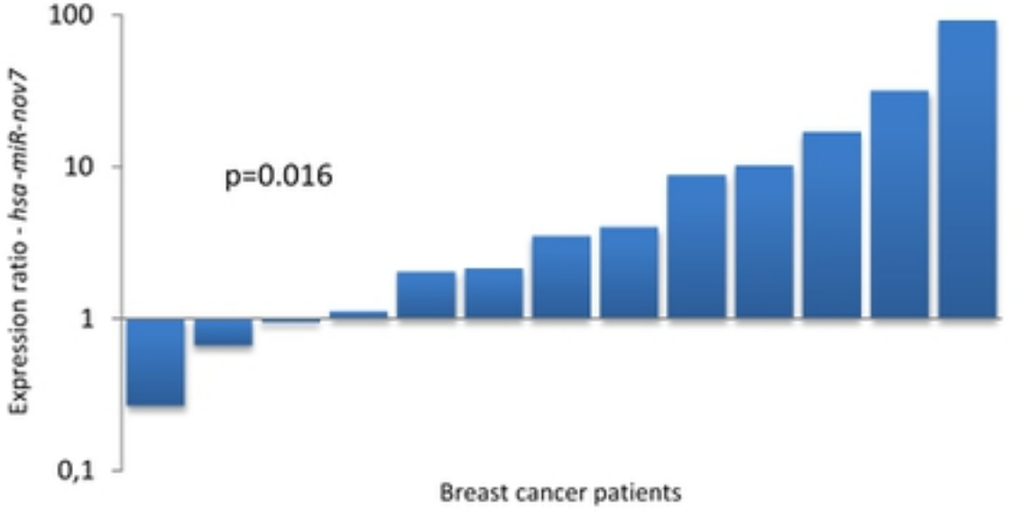
Figure

Figure 5

A)



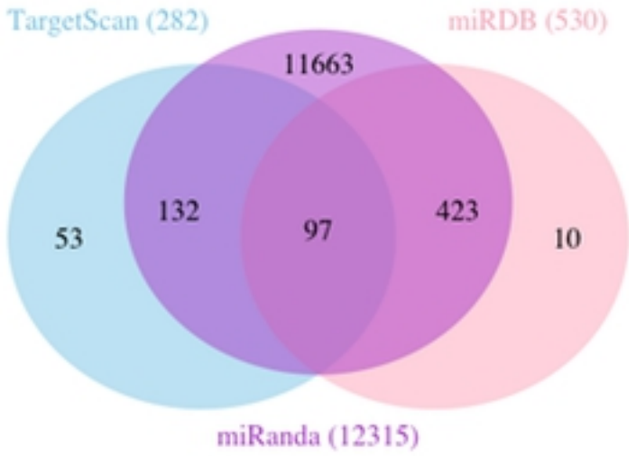
B)



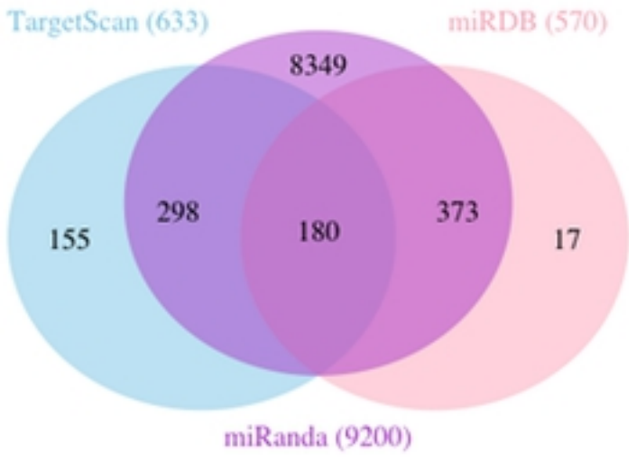
Figure

Figure 6

A)
hsa-mir-nov3



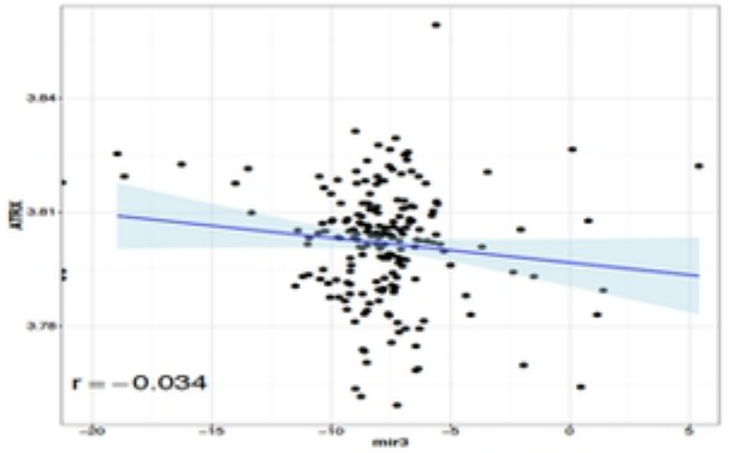
B)
hsa-mir-nov7



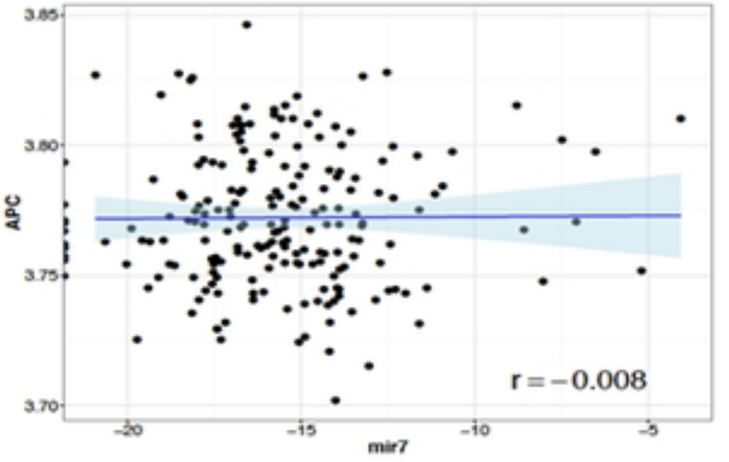
Figure

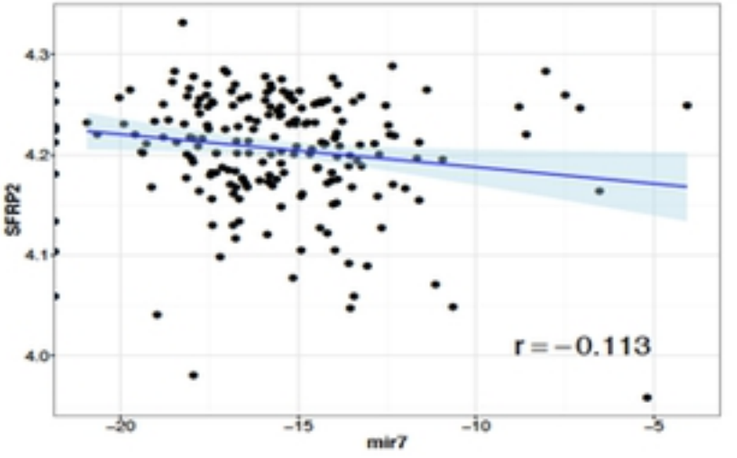
Figure 7

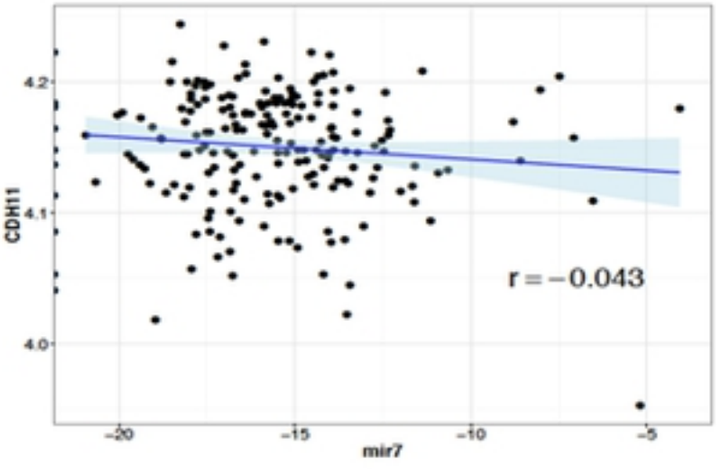
A)
hsa-miR-nov3



B)
hsa-miR-nov7

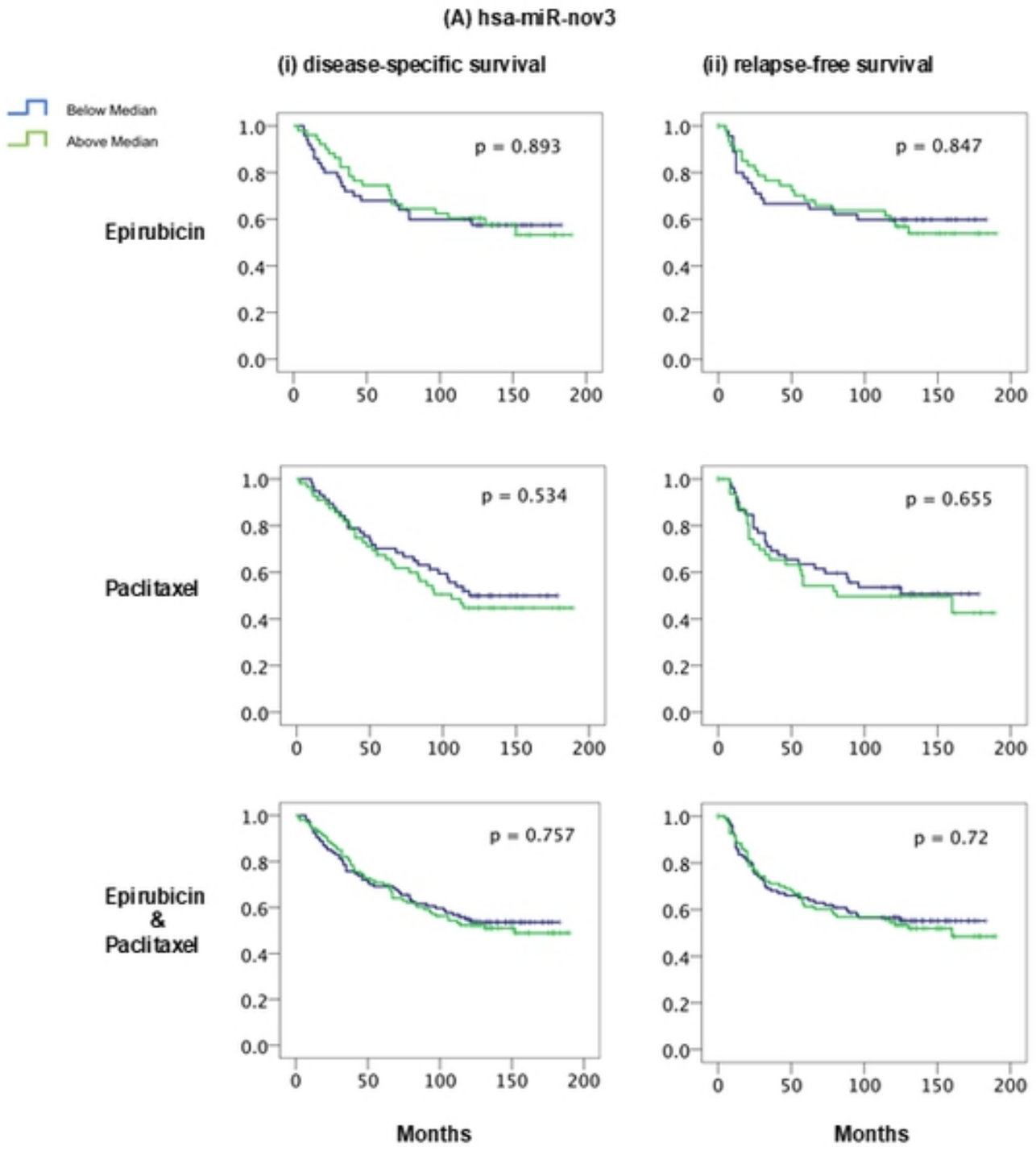




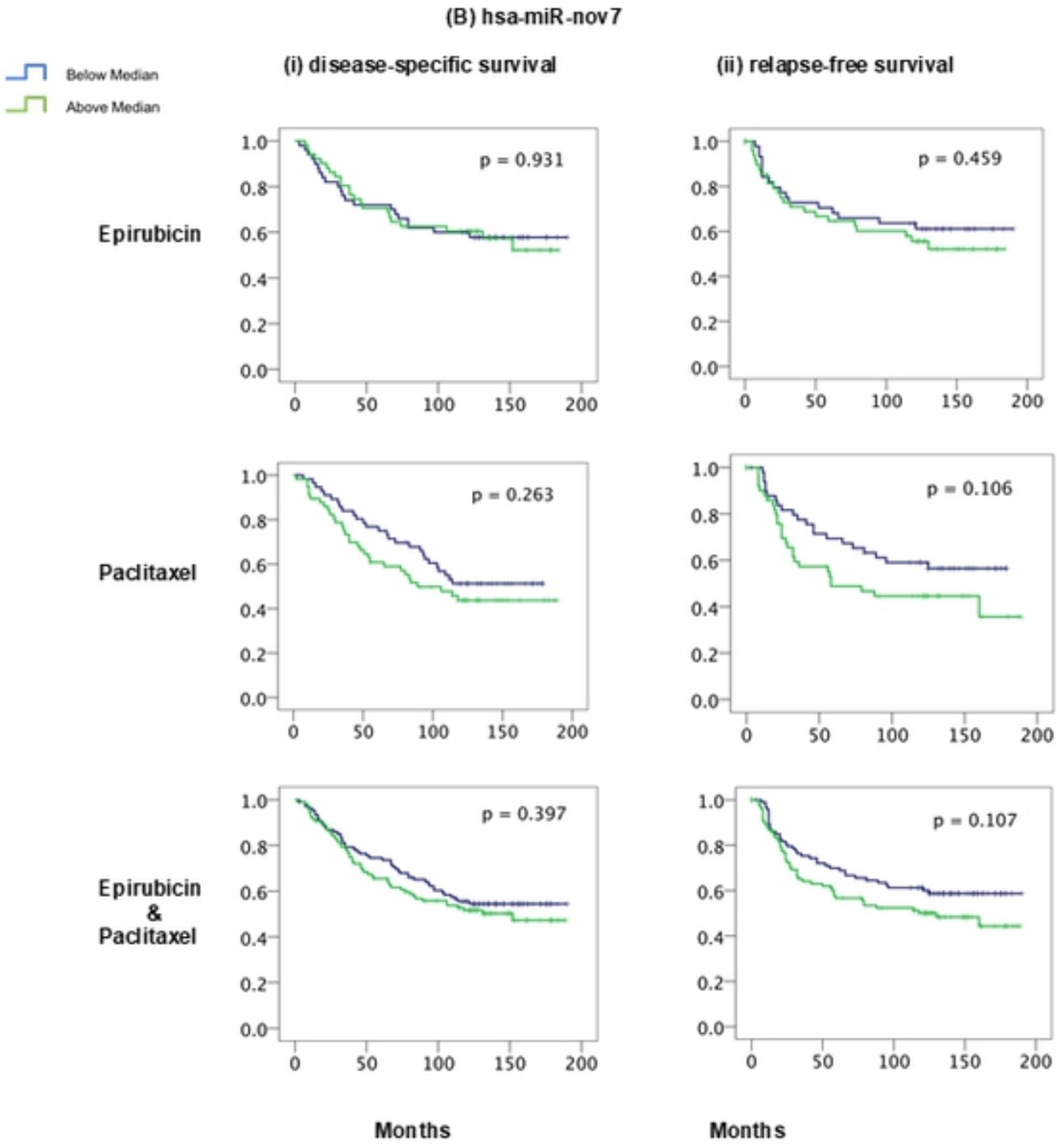


Figure

Figure 8



Figure


(B) hsa-miR-nov7
Below Median
Above Median
(i) disease-specific survival
(ii) relapse-free survival
Epirubicin
Paclitaxel
Epirubicin & Paclitaxel
p = 0.931
p = 0.459
p = 0.263
p = 0.106
p = 0.397
p = 0.107
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
Months
Months

Figure