Cancer Research

Identification of New MicroRNAs in Paired Normal and Tumor Breast Tissue Suggests a Dual Role for the *ERBB2/Her2* Gene

Helena Persson[1], Anders Kvist[1], Natalia Rego[2], Johan Staaf[1,3,4], Johan Vallon-Christersson[1,3,4], Lena Luts[5], Niklas Loman[1], Goran Jonsson[1,3,4], Hugo Naya[2], Mattias Hoglund[1,3], Ake Borg[1,3,4], and Carlos Rovira[1,3,4]

Abstract

To comprehensively characterize microRNA (miRNA) expression in breast cancer, we performed the first extensive next-generation sequencing expression analysis of this disease. We sequenced small RNA from tumors with paired samples of normal and tumor-adjacent breast tissue. Our results indicate that tumor identity is achieved mainly by variation in the expression levels of a common set of miRNAs rather than by tissue-specific expression. We also report 361 new, well-supported miRNA precursors. Nearly two-thirds of these new genes were detected in other human tissues and 49% of the miRNAs were found associated with Ago2 in MCF7 cells. Ten percent of the new miRNAs are located in regions with high-level genomic amplifications in breast cancer. A new miRNA is encoded within the *ERBB2/Her2* gene and amplification of this gene leads to overexpression of the new miRNA, indicating that this potent oncogene and important clinical marker may have two different biological functions. In summary, our work substantially expands the number of known miRNAs and highlights the complexity of small RNA expression in breast cancer. *Cancer Res; 71(1); 78–86. ©2011 AACR.*

Introduction

The identification of regulatory ~22 nucleotide (nt) noncoding RNAs (1, 2) has been one of the major findings in biology in recent years. Classical microRNAs (miRNAs) are produced from transcripts that form a fold-back hairpin structure, which is cropped by the RNase III enzyme Drosha to a ~70 nt long pre-miRNA. This is exported to the cytoplasm and further processed by a second RNase III enzyme, Dicer, into the mature ~22 nt miRNA and its complementary star sequence. Mature miRNAs regulate gene expression by guiding effector complexes formed by Argonaute proteins to cognate targets leading to translational repression or enhanced mRNA degradation. Regulation depends on base pairing between the miRNA and target sites on the mRNA. The seed sequence, nucleotides 2 to 7 at the 5′ end of the miRNA, is especially important for this interaction.

Authors' Affiliations: [1]Department of Oncology, Clinical Sciences, Lund University, Lund, Sweden; [2]Bioinformatics Unit, Institute Pasteur Montevideo, Montevideo, Uruguay; [3]SCIBLU Genomics, Centre for Integrative Biology at Lund University, [4]CREATE Health, Strategic Centre for Translational Cancer Research, and [5]Department of Clinical Pathology, Clinical Sciences, Lund University, Lund, Sweden

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

Corresponding Author: Carlos Rovira, Department of Oncology, Clinical Sciences, Lund University, BMC C13, Klinikgatan 28, 221 84 Lund, Sweden. Phone: +46462224773; Fax: +46462224773.
E-mail: carlos.rovira@med.lu.se

doi: 10.1158/0008-5472.CAN-10-1869

Dysregulation of miRNA function has frequently been implicated in cancer. Calin *et al.* (3) first observed that many miRNAs were located in genomic regions that are deleted or amplified in different cancer types. Many studies have analyzed the global expression of miRNAs in breast cancer using large-scale hybridization-based methods (4, 5). These studies showed a roughly 50% overlap between tumors and normal tissues for the analyzed miRNA genes, leading to the hypothesis that well-determined miRNA subsets may be expressed in a tissue-specific pattern (6, 7).

In total, 904 different mature miRNAs have been identified and reported to miRBase (release 14), the central repository of miRNA sequences. Genome-wide predictions of hairpin structures, the most prominent feature of miRNA genes, have estimated that several thousand may be encoded in the human genome (8). The search for new miRNAs has been biased toward highly expressed and evolutionarily conserved genes. Lately, application of highly sensitive approaches has revealed the presence of weakly expressed miRNAs that are not maintained by natural selection (9, 10). Next-generation sequencing small RNA data available in public databases includes little data for breast tissue. To perform a comprehensive study of breast cancer, we analyzed paired samples of normal, tumor-adjacent and tumor tissue from 5 patients and generated more than 6 million aligned small RNA reads per sample.

We observed a considerable overlap between miRNAs expressed in normal and tumor samples and also between different tumors, implying that miRNA expression levels, rather than miRNA identity, characterize tumor and normal tissues. Furthermore, we identified 535 new mature miRNAs.

Nearly two-thirds of the new genes were also detected in other tissues. Mature miRNAs from 49% of all new precursors were found in immunoprecipitates of Argonaute proteins. Common genomic changes in breast cancer could significantly increase the expression of several of these new genes. This is exemplified by a miRNA encoded within *ERBB2/Her2*, which suggests a new function as a miRNA host transcript for this clinically important oncogene. In summary, our results considerably increase the size of the human miRNome and add a new level of complexity to the role of miRNAs in breast carcinogenesis.

Materials and Methods

Sample collection

For each breast cancer patient 3 samples corresponding to normal (N), tumor-adjacent (TN), and tumor (T) tissue were prepared by a pathologist from surgically removed material. Tumor-adjacent samples were carefully dissected next to the tumor to include as little contamination of cancerous tissue as possible, while leaving the tumor samples free from normal tissue. The tissue removed during surgery varies from small resections to complete ablation of the breast. Therefore, the distances between tumor and normal samples differ between patients. Normal breast tissue from healthy donors was isolated from samples obtained from women undergoing reduction mammoplasty. The tissues were kept on dry ice at all times during tissue sampling and handling. Tissue samples were fresh-frozen in −70°C. The study was approved by the local medical ethics committee. Total RNA was extracted with Trizol following the manufacturer's instructions.

Small RNA sequencing

Sequencing of small RNAs was performed on the Illumina Genome Analyzer. Libraries of 18 to 50 nt small RNAs were prepared with the Small RNA Sample Prep Kit v1 (Illumina). Sequences were aligned to the human genome (UCSC hg18 assembly) and identified by comparison against the positions of mature miRNAs and miRNA precursors from miRBase release 14, and also protein-coding and noncoding genes and repeats using annotation from the UCSC Genome Browser. Significance analysis of microarrays (SAM) was done in MeV version 4.3.02 for comparison of expression in tumors versus normal tissue and clustering of the samples. For pairwise analysis of differential miRNA expression between samples, the *P* values were calculated using the *Z* test for proportions with Bonferroni correction for multiple testing.

Identification of new miRNAs

Sequences that did not match miRNAs or repetitive DNA were combined for all samples and submitted to the miRanalyzer web server (11) that identified a large number of candidate miRNA precursors. These were used as scaffolds for alignment of all sequences and data for 24 human cell lines and tissue samples available in the Gene Expression Omnibus database (GEO). Candidates were then selected for further study only if they passed criteria for (i) well-defined mature product and few other sequences from precursor region, (ii) hairpin structure and stability, (iii) sequence complexity, (iv) favorable genomic location, and (v) a requirement for sequences from both precursor arms consistent with Drosha and Dicer processing or sequences from several samples. Entry names are given for novel miRNA sequences deposited in miRBase. Provisional names are used for the rest.

For evolutionary conservation analysis, the basis of criteria was sequence and structure similarity, sufficient secondary structure stability, and conservation of the mature miRNA seed sequence (nucleotides 2–7).

Validation of candidate miRNAs

Expression of miRNAs was analyzed by real-time PCR using iTaq SYBR green supermix (Bio-Rad) with denaturation at 95°C 3 minutes, followed by 40 cycles of 95°C 15 seconds, 60°C 45 seconds, and a final melting curve analysis from 60°C to 95°C to verify that a single product was amplified. The cDNA was synthesized using stem-loop primers as described in reference (12). For Ago2 immunoprecipitation, MCF7 cell pellets were resuspended in lysis buffer: 20 mmol/L of Tris-HCl (pH = 7.5), 200 mmol/L of NaCl, 2.5 mmol/L of $MgCl_2$, 0.5% Triton X-100, 0.1 U/μL RNase inhibitor and 1× EDTA-free complete protease inhibitor (Roche Applied Science). Cells were sonicated and the extract was centrifuged. The supernatant was incubated with Ago2 antibodies with gentle rotation at 4°C. Blocked Dynabeads Protein G (Invitrogen) were added to the lysate and incubated at 4°C. Bound complexes were washed in lysis buffer followed by LiCl buffer (20 mmol/L of Tris-HCl (pH 8.0), 1 mmol/L of EDTA, 250 mmol/L of LiCl, 1% NP-40, and 1% sodium deoxycholate) and finally treated with DNase and proteinase K before phenol–chloroform extraction and RNA precipitation.

Array comparative genome hybridization

BAC (bacterial artificial chromosome) aCGH and gene expression data for 359 breast tumors were obtained from Jonsson et al. (13)

Detailed information is provided in the Supplementary Materials and Methods.

Results

A comprehensive analysis of miRNA expression in normal and tumor breast tissue

We performed next-generation sequencing of small RNAs to create a catalog of their expression in breast cancer. We extracted total RNA, fractionated 18 to 50 nt small RNAs and produced libraries for sequencing from paired normal (N), tumor-adjacent (TN), and tumor (T) breast tissue from 5 patients. The tissue removed during surgery was macrodissected by a pathologist to ensure that tumor-adjacent samples were devoid of malignant cells and that tumor samples were not contaminated with noncancerous tissue. The normal samples were taken as remotely as possible from the affected area. Two of the patients (TAX577740 and TAX577579) had estrogen and progesterone receptor-negative (ER^-/PR^-) invasive ductal carcinoma (IDC) with axillary lymph node involvement (LN^+). TAX577453 came from a BRCA2 mutation carrier

Table 1. Clinical data for sequenced breast tumors

Patient ID	Age, y	Histologic type	Grade	LN	ER	PR	ERBB2	Comment
TAX577740	57	IDC	2	+ (14/15)	−	−	−	
TAX577453	48	IDC	3	+ (10/17)	+	+	−	BRCA2 4486delG
TAX577745	46	IDC	2	− (0/16)	+	+	−	
TAX577579	90	IDC	2	+ (1/14)	−	−	−	
TAX577580	52	Fibroadenoma	NA	NA	NA	NA	−	

Abbreviations: IDC, invasive ductal carcinoma; LN, lymph node status (affected/tested lymph nodes within parentheses); ER, estrogen receptor alpha expression; PR, progesterone receptor expression; ERBB2, ERBB2 status.

(4486delG) with an $ER^+/PR^+/LN^+$ IDC. Patient 4 (TAX577745) had an $ER^+/PR^+/LN^-$ IDC and patient 5 had a fibroadenoma with no previous history of malignancy (TAX577580). None of the patients had *ERBB2/Her2* amplifications. To note, TAX577740 is a triple-negative tumor ($ER^-/PR^-/ERBB2^-$) with a BRCA1-like phenotype in a patient from a non-BRCA mutant family with a history of breast cancer (Table 1).

Close to 170 million sequences were produced for the 15 samples with a final number of quality filtered, aligned sequences of 95 million, giving an average of 6.4 million per sample (Supplementary Material S1). Mature miRNAs make up the vast majority of sequences in the 18 to 25 nt size range with a clear peak around 22 nt in all samples (Fig. 1A and B). Sequences that are annotated as intronic or intergenic, the most common locations for miRNA genes, also accumulate at 22 nt, suggesting that a number of miRNAs may still be unidentified. All libraries had a similar proportion of miRNAs independently of sample origin, ranging from 78% in the fibroadenoma (TAX577580T) to 91% for the tumor from the *BRCA2* mutation carrier, TAX577453T (Supplementary Material S2). Only 0.2% of the sequences were miRNA precursor fragments different from the functional mature miRNA or its complementary star sequence.

Differential expression of a large and common miRNA set characterizes tumor versus normal tissue identity

We first analyzed the expression of all human miRNAs included in miRBase release 14 with the addition of 155 new star sequences identified from our sequence data (Supplementary Material S3). All samples expressed a large and highly similar set of miRNA genes, irrespective of patient and sample type. We used a limit of expression of 0.5 counts per million reads (cpm) to minimize spurious detection due to sequencing errors. This corresponds to approximately 3 reads for an average library. The total number of mature miRNAs expressed in at least 1 sample was 703 with a sample average of 535 miRNAs. The intersection of expressed miRNAs between samples is very large, typically 70% to 80% both for samples from the same patient and for samples of the same tissue type. Most of the non-overlapping miRNA genes have low expression. Exceptions include miR-767-5p and miR-1269, which are expressed almost exclusively in TAX577740T and TN. The majority of the detected miRNAs have lower expression in tumors than in normal tissue, in agreement with the reported association between transformation and global downregulation of miRNA expression (14). However, there is an increase in the total number of miRNA reads for several tumor and tumor-adjacent samples compared with normal breast, but this is linked to a marked increase of a limited number of known oncogenic miRNAs such as miR-21 and some members of the let-7 family.

Significance analysis of differential miRNA expression and hierarchical clustering analysis clearly separated tumors from normal tissue samples (Fig. 1C). The samples of tumor-adjacent tissue were divided between the normal and tumor clusters, independently of their clinicopathologic characteristics. Although this may reflect the degree of tumor or inflammatory cell infiltration, differential expression suggests that the region may be more than a transition. A set of miRNAs was upregulated in tumor-adjacent samples compared with the surrounding normal and tumor tissue in 4 of 5 patients. This set includes both the pro-metastatic miR-10b and miRNAs with proposed metastasis-inhibiting or tumor suppressor activity such as miR-100, miR-101, miR-146b, mir-148a, mir-205, miR-99a, miR-196, miR-199, miR-30, miR-99, miR-26, miR-27, and miR-29.

Two-thirds of the differentially expressed genes are downregulated in tumors compared with normal samples for 3 patients (TAX577453, TAX577745, and TAX577580). Interestingly, the 2 tumors with the lowest percentage of downregulated miRNAs (50% and 37% for TAX577740 and TAX577579, respectively) are both ER-negative. The fibroadenoma had the smallest number of differentially expressed genes compared with normal tissue. A group of 4 miRNAs were upregulated in all tumor samples, including miR-107, miR-191, and the oncogenic miR-21 and miR-200a.

Pair-wise comparison between the 4 malignant tumors and their corresponding normal samples identified a common set of 9 upregulated and 21 downregulated miRNAs. The list of upregulated genes consists of miRNAs previously associated with cancer such as miR-21, miR-155/BIC, and miR-196a. Four of the 21 miRNAs that are downregulated in all tumors have been reported to be upregulated posttranscriptionally by TP53 upon DNA damage (15); miR-143 that targets *KRAS*, miR-145/145* (targets *CDK6*), and miR-26a (targets *CCND2* and *CCNE2*). These miRNAs have been shown to be underexpressed in several cancers. The miRNA miR-140 is downregulated in all tumor samples. A number of these downregulated miRNAs have, to our

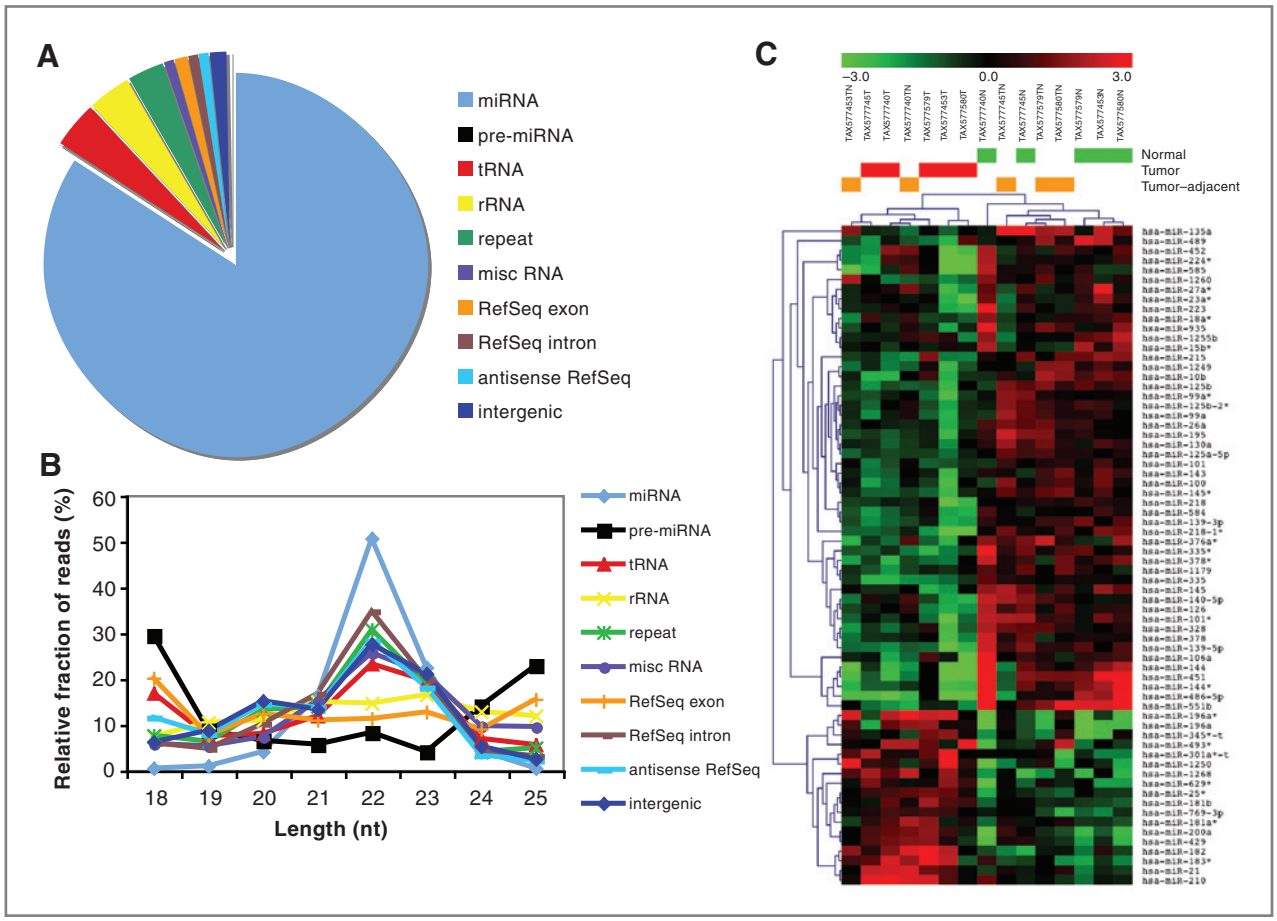

Figure 1. A, aligned 18 to 25 nt reads divided into the main categories. B, the relative distribution of reads within this size range has a clear peak at 22 nt for mature miRNAs and for other categories. C, hierarchical clustering of the sequenced samples based on a SAM analysis of differentially expressed miRNAs clearly separates normal and tumor samples.

knowledge, not previously been associated with breast cancer, such as miR-139-5p, miR-144/144*, miR-223, and miR-378.

Several studies have linked downregulation of the let-7 miRNAs to tumorigenesis and poor cancer prognosis. By sequencing, we see a complex pattern with both up- and downregulation of individual family members. For example, for the tumor TAX577579, the family members let-7a, -7b, -7c, and -7e are downregulated; whereas, let-7d and -7f are upregulated. The net result is a slight downregulation of the common seed sequence GAGGTA.

In summary, our sequence analysis accurately classifies normal and tumor tissue and shows that tumors and normal breast express a large, highly similar set of miRNAs.

Identification of a Large Number of New miRNAs

We analyzed the data for the presence of candidate new miRNAs using the miRanalyzer web server (11). This analysis reported more than 700 candidate precursor loci, which we used to produce alignments for the complete sequence data, and data for breast cancer cell lines produced in our laboratory and a number of small RNA sequence data sets available in the GEO database (GEO, NCBI). Candidate miRNAs with good sequence support were selected and only candidates meeting strict requirements were retained (see Material and Methods for further description). The GEO data sets (including ovary, cervix, colon, lung, kidney, human embryonic stem cells, nasopharynx, and bone) provided independent support for 63% of our new miRNAs and, convincingly, mature sequences from both 5′ and 3′ precursor arms were identified for 174 of the precursors. The most common sequences were selected as representative mature miRNAs. The final set contained 361 new miRNA precursors with 535 mature miRNAs (Supplementary Material S4), corresponding to 50% and 59%, respectively, of the total number of human miRNAs in miRBase release 14.

A majority of the precursors, 191 genes (53%), are encoded in introns and 37 (10%) are located in intergenic regions. Nineteen (5%) mapped within coding exons and another 19 to the 5′ or 3′ untranslated regions (UTR) of annotated transcripts. Forty-two novel miRNAs mapped to the antisense strand of mRNAs. Notably, 12 of our new miRNAs are transcribed antisense to known miRNAs. The expression of most

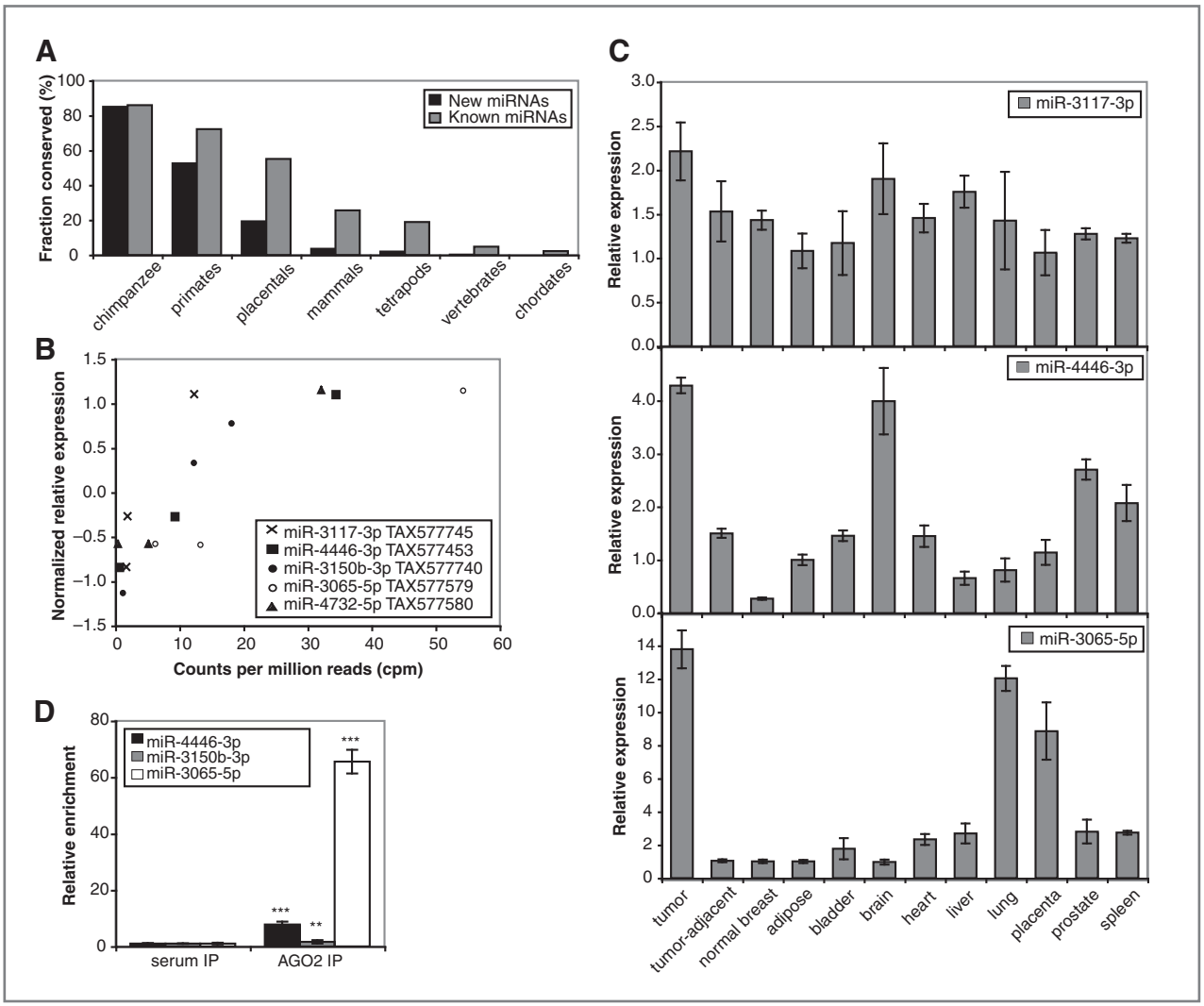

Figure 2. A, most new miRNAs have a recent evolutionary origin compared with previously known miRNAs, although a small number are deeply conserved among vertebrates. B, a representative group of five new miRNAs were detected by qPCR in the sequenced tissues and correlated well with the cpm values from sequencing, $r = 0.96$ to 1.00 within patients. C, three miRNAs were also detected in a panel of nine human tissues, whereas the remaining two appear to be breast tissue specific. D, three of the 5 miRNAs were expressed in MCF7 cells and showed significant enrichment in RNA immunoprecipitated with AGO2 ($P = 2.2 \times 10^{-7}$, 0.023 and 3.0×10^{-7} for miR-4446-3p, miR-3150b-3p, and miR-3065-5p, respectively). Error bars indicate SD, $^{**}P < 0.05$, $^{***}P < 0.01$, $n = 4$ in (C) and (D), P values were determined using Student's t test.

of the new miRNAs lies within the lower range of expression of currently known miRNAs (Supplementary Material S5a).

The new miRNAs identified here have remained undetected in spite of a number of studies in different organisms. The phylogenetic distribution clearly shows a lower degree of conservation beyond primates for the new miRNAs than for known miRNAs (Fig. 2A). Seven percent of our miRNAs may be specific to the human lineage, but a majority is also found in closely related species such as chimpanzee. Seventy-eight (21%) are detected in at least one of the nonprimate placental mammals and a handful appear to be deeply conserved.

Experimental validation of new miRNAs

We selected a number of new miRNAs for experimental validation to confirm their expression. Precursor structures for these miRNAs (miR-3117-3p, miR-4446-3p, miR-3150b-3p, miR-3065-5p, and miR-4732-5p) are shown in Supplementary Material S5b. They were chosen randomly among candidates that had high expression in at least 1 sequenced sample. We first measured the expression of these miRNAs by real-time PCR in the tissue samples analyzed by sequencing, in which their expression ranged from 0 to 54 cpm. All 5 miRNAs were detected above background by PCR, and their relative expression between samples showed excellent correlation with the cpm values from sequencing ($r = 0.96$–1.00 within patients; Fig. 2B). In addition, to test their presence in independent samples we measured their expression in a panel of 9 human tissues (Fig. 2C). Two of the 5, miR-3150b-3p and miR-4732-5p, were not detected in any other tissue, suggesting that they may be specific to breast.

Argonaute proteins are the key components of the miRNA-induced silencing complex (miRISC). To generate functional evidence for a large number of new miRNAs, we immunoprecipitated Ago2 from MCF7 cells and analyzed the associated small RNAs. To verify enrichment, we first tested the new miRNAs that had been experimentally validated by real-time PCR. Three miRNAs (miR-4446-3p, miR-3150b-3p and miR-3065-5p) were expressed in MCF7 cells and showed significant enrichment in the Ago immunoprecipitate compared with the control (Fig. 2D). We then sequenced the immunoprecipitated RNA and detected mature miRNAs from 176 (49%) of our candidate precursors. These experimental results provide strong support for the expression, processing and biological function through association with miRISC for our set of candidate new miRNAs.

A dual function for the *ERBB2/Her2* gene?

Most new miRNA genes are lowly expressed in our samples. However, expression levels could be influenced by genomic gains or losses of chromosomal regions, a common mechanism of gene dysregulation in cancer. We, therefore, searched for recurrent genomic amplifications affecting the expression of our new miRNAs in a set of 359 breast tumors (13) with samples representing breast cancer subtypes associated with different clinicopathologic characteristics (16). We specifically sought to identify genes encoded in regions with high-level genomic amplifications and found 39 new miRNAs distributed in 19 regions with significant gains in 4% or more of the samples in at least 1 tumor group based on subtype, ER-status, or hereditary mutations (Table 2). Amplification of 17q12 is a clinically important genomic aberration that causes overexpression of the ERBB2 receptor. Five of our new miRNAs (miR-4726, miR-4727, Candidate_296, miR-4728, and miR-4734) map to this region. Strikingly, miR-4728 is located in an intron of the *ERBB2* gene itself (Fig. 3A). Both mature miRNAs from this candidate, miR-4728-5p and miR-4728-3p, were found among the Ago2-associated RNAs in MCF7 cells. Closer inspection of this new miRNA shows that the 5′ end of miR-4728-5p is in the middle of the intron; whereas, miR-4728-3p ends just before the 5′ splice site of exon 24 (Fig. 3B). This suggests that the new miRNA is a half-mirtron (17, 18) and that its processing may depend on splicing of the *ERBB2* RNA itself.

We analyzed the expression of this miRNA in cell lines with *ERBB2* amplification (BT-474, JIMT-1, and SK-BR-3) and 2 control cell lines (MCF7 and MCF10A). All 3 ERBB2$^+$ cell lines overexpressed both *ERBB2* exon 24 and miR-4728-3p, whereas the expression of miR-4728-5p was lower and only showed a marginal increase (Fig. 3C). To confirm these results, we selected a group of breast tumors and 4 normal breast tissue samples from healthy women obtained from reduction mammoplasties. Six tumors were ERBB2$^+$ and the remaining 12 were ERBB2$^-$, equally divided between the luminal A, luminal B, basal, and normal-like subtypes. Four of the 6 tumors with ERBB2$^+$ overexpressed both miR-4728-3p and *ERBB2* exon 24 (Fig. 3D). These results uncover a new genetic component within the *ERBB2/Her2* amplicon, one of the major predictive markers of breast cancer, underscoring the value of our data.

Discussion

We aimed to characterize the entire miRNA complement and detected 703 known mature miRNAs expressed in at least 1 sample with a sample average of 535 miRNAs using a 0.5 cpm cut-off. If we include all miRNAs with at least 1 read, the total number of mature miRNAs increases to 876 and the average to 661, suggesting that the sequencing depth did not reach saturation. We also detected 155 star sequences for known miRNAs that had not been previously annotated. The miRNA sets that are expressed in normal, tumor, and tumor–adjacent samples were largely overlapping, and miRNAs with sample-specific expression always had low expression. This implies that breast tumors differ from normal breast by variation in the expression level of similar sets of miRNAs, rather than tissue-restricted expression of specific miRNAs. We also identified a large number of new miRNAs with strong support based on careful analysis of sequence data. In total, we found 535 mature miRNAs from 361 new precursors, numbers that represent 59% and 50%, respectively, of all human miRNAs in miRBase release 14. Release 15, which became public during the final preparation of this manuscript, contains 60 of 86 identified new mature miRNAs for 54 of our candidate precursors, providing further confirmation for our set of new miRNA genes.

The new miRNAs are generally less evolutionarily conserved than known miRNAs, with a large fraction unique to humans or primates. The importance of nonconserved, weakly expressed miRNAs has probably not been fully appreciated. Although conservation generally indicates functional importance, the contrary is not necessarily true for miRNAs, as shown by the lack of conservation of miR-296 target sites in Nanog and Oct4 (19). Furthermore, expression of functionally important miRNAs can be restricted to only a few specialized cells (20) giving constantly low counts in sequencing screens. We recently reported that the endogenous levels of lowly expressed small vault RNAs efficiently downregulate the expression of highly expressed reporters in a miRNA-like manner (21). Weakly expressed sequences are sometimes dismissed as noise or random degradation products. In our case, processing by the miRNA pathway is confirmed by the almost exclusive presence of mature miRNA sequences, often derived from both precursor arms and detected in several independent samples, together with a negligible number of sequences from other parts of the predicted precursors. We could also show association with Argonaute proteins for a large fraction of our candidate new miRNAs, which strongly indicates that they are new functional miRNA genes.

More than 10% of the new miRNAs map to regions that display high-level amplifications in breast tumors. Among these, miR-4728 is encoded within an intron of *ERBB2*, a gene that is frequently amplified and associated with more aggressive disease and shorter patient survival. This new miRNA is also significantly overexpressed in tumors and cell lines with *ERBB2* amplification. The discovery of a new gene that has escaped previous identification in a well-studied region, known to be potently transforming upon overexpression, was highly unexpected. It is tempting to speculate that expression of miR-4728 might explain why overexpression of ERBB2, rather than

Table 2. New miRNAs located in regions amplified in breast cancer as detected by BAC array CGH

ID	Cytoband	Host gene	Luminal A	Luminal B	ERBB2	Basal	Normal-like	Unclassified	ER^+	ER^-	BRCA1	BRCA2	Total
Candidate_11	1p12	*PHGDH*	0	0	0	3	0	0	0	2	6	3	1
hsa-mir-3121	1q25.3	*ACBD6*	1	0	6	1	0	0	0	3	6	0	1
hsa-mir-3682	2p16.2	*LOC100302652*	0	0	0	3	0	0	0	2	6	0	1
hsa-mir-4778	2p14		0	0	0	1	0	0	0	1	6	0	0
hsa-mir-4635	5p15.33	*SLC12A7*	0	1	0	1	0	0	0	1	6	0	1
hsa-mir-4469	8p11.21	*RNF170*	1	6	0	0	0	0	2	1	0	3	1
hsa-mir-4661	8q21.3	*LRRC69*	0	4	0	3	0	2	1	3	0	6	2
hsa-mir-3150b	8q22.1		0	9	3	3	0	2	2	4	0	10	3
hsa-mir-3150a	8q22.1		0	9	3	3	0	2	2	4	0	10	3
hsa-mir-4471	8q22.2		0	7	6	5	0	2	3	4	0	10	3
Candidate_163	8q22.3	*PABPC1*	0	9	6	5	0	2	3	5	0	10	4
Candidate_164	8q24.13	*DERL1*	1	13	0	9	0	4	6	4	0	16	5
hsa-mir-4663	8q24.13		1	13	0	9	0	4	6	4	0	16	5
hsa-mir-4662b	8q24.13		1	16	3	8	0	9	7	5	0	16	6
hsa-mir-4662a	8q24.13		1	16	3	8	0	9	7	5	0	16	6
hsa-mir-4665	9p24.1	*KIAA2026*	0	0	0	3	0	0	0	1	6	0	1
hsa-mir-4683	10p11.21	*FZD8*	0	0	0	3	0	0	0	2	6	0	1
hsa-mir-4691	11q13.2	*NDUFS8*	0	6	0	0	0	0	2	0	0	0	1
Candidate_213	11q13.3	*IGHMBP2*	1	11	0	1	6	0	4	3	0	0	3
hsa-mir-3664	11q13.4	*SHANK2*	1	10	0	1	3	2	4	2	0	0	3
hsa-mir-3913-1	12q15		0	7	0	0	3	0	3	0	0	3	2
hsa-mir-4723	17q11.2	*TMEM199*	0	0	12	3	0	2	1	3	0	0	2
hsa-mir-451b	17q11.2		0	0	24	4	0	2	2	6	0	0	3
hsa-mir-4732	17q11.2		0	0	24	4	0	2	2	6	0	0	3
hsa-mir-4734	17q12		0	4	32	1	3	4	3	9	0	3	5
hsa-mir-4726	17q12	*MLLT6*	1	4	38	3	3	4	4	11	0	3	6
hsa-mir-4727	17q12		1	4	38	3	3	4	4	11	0	3	6
Candidate_296	17q12	*LASP1*	1	4	44	4	3	4	4	13	0	3	7
hsa-mir-4728	17q12	*ERBB2*	2	7	85	9	15	6	8	27	0	10	14
hsa-mir-3614	17q23.2	*TRIM25*	1	3	15	1	0	4	2	5	0	3	3
hsa-mir-4736	17q23.2		0	3	12	1	0	2	1	4	0	3	2
hsa-mir-4729	17q23.2	*YPEL2*	0	4	12	3	0	0	2	4	0	3	3
hsa-mir-4737	17q23.2		0	4	15	1	0	4	2	6	0	3	3
hsa-mir-3064	17q24.1	*DDX5*	0	3	9	0	0	2	1	3	0	0	2
hsa-mir-4524	17q24.3	*ABCA6*	0	0	9	0	0	0	0	2	0	0	1
Candidate_299	17q25.1	*KIAA0195*	0	1	9	1	0	2	1	3	0	0	2
hsa-mir-4738	17q25.1		0	1	6	1	0	2	1	3	0	0	1
hsa-mir-3065	17q25.3		0	4	3	0	0	0	1	1	0	3	1
hsa-mir-3194	20q13.2	*NFATC2*	0	4	0	0	0	0	1	0	0	0	1
Total sample number			95	70	34	79	34	47	227	119	17	31	359

NOTE: High-level gains were identified in a set of 359 tumors classified according to subtype (luminal A, luminal B, ERBB2-positive, basal, and normal-like), ER expression and BRCA1 and BRCA2 mutation status. Shown are the percentages of tumors with gains within each subgroup.

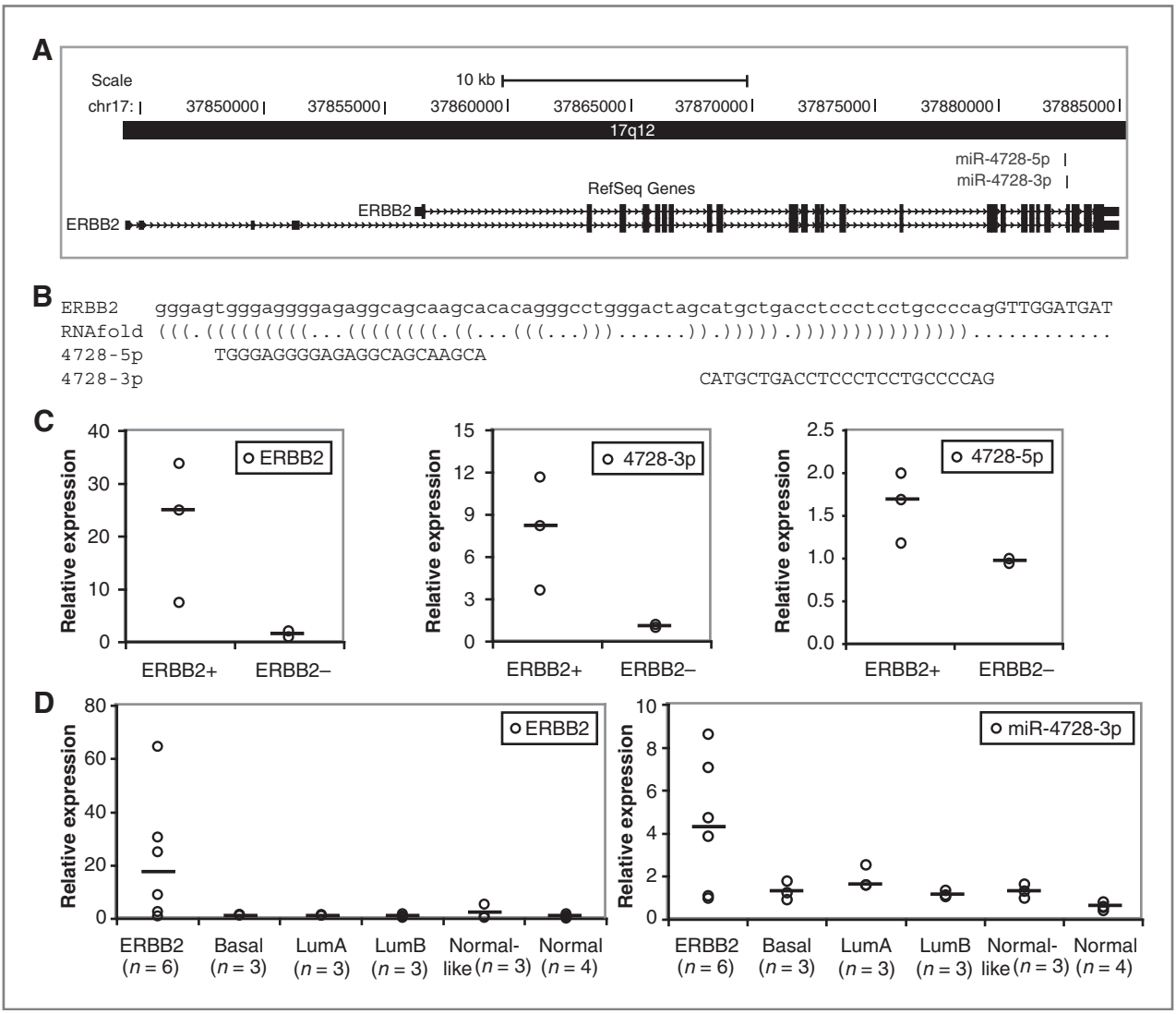

Figure 3. A, miR-4728 is encoded within an intron of the *ERBB2* gene. B, position of the 3′ terminus of miR-4728 suggests that processing depends on *ERBB2* splicing. Intronic sequence is shown in lower case letters. C, three ERBB2$^+$ breast cancer cell lines overexpressed both miR-4728 and *ERBB2*, although the expression of miR-4728-5p was lower and only showed a marginal increase. ERBB2$^+$ cells: BT-474, JIMT-1, SK-BR-3; ERBB2$^-$ cells: MCF7 and MCF10A. D, 4 out of 6 ERBB2$^-$ tumors overexpressed both miR-4728-3p and *ERBB2* compared with 12 ERBB2 tumors equally divided between the luminal A, luminal B, basal, and normal-like subtypes and 4 samples of normal breast tissue from healthy controls. Horizontal bars mark the median value for each group.

overactivity, is observed in human ERBB2$^+$ breast cancer. Gene ontology analysis of targets predicted by TargetScan (22) for miR-4728-3p highlights cellular processes such as small GTPase-mediated signal transduction, although enrichment for individual terms does not reach significance when correcting for multiple testing. However, MAPK1 and SOS1, downstream targets of ERBB2 activation and central components of the transduction of signals that control cell growth and differentiation, were among the top-ranked predicted targets (Supplementary Material S6). This could suggest that miR-4728-3p is part of a negative feedback loop that controls or fine-tunes the function of its host gene.

In summary, our results highlight the complexity of miRNA regulation in breast cancer and support the conclusion that stringent criteria for high expression and evolutionary conservation applied to miRNA discovery misses authentic miRNA genes, many of which could be of potential importance for human pathology.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgment

We thank Zissimos Mourelatos and Mikiko and Haruhiko Siomi for the kind gift of the Ago antibodies.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 25, 2010; revised September 8, 2010; accepted September 24, 2010; published online January 3, 2011.

References

1. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 1993;75:855–62.
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;75:843–54.
3. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A 2004;101:2999–3004.
4. 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;65:7065–70.
5. Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol 2007;8:R214.
6. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257–61.
7. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37: 495–500.
8. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 2006;126:1203–17.
9. Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, et al. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. Genome Res 2006;16: 1289–98.
10. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 2005;37:766–70.
11. Hackenberg M, Sturm M, Langenberger D, Falcon-Perez JM, Aransay AM. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 2009;37: W68–76.
12. Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 2007;3:12.
13. Jonsson G, Staaf J, Vallon-Christersson J, Ringnér M, Holm K, Hegardt C, et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res ;12:R42.
14. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834–8.
15. Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S, Miyazono K. Modulation of microRNA processing by p53. Nature 2009;460:529–33.
16. 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; 98:10869–74.
17. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. Mol Cell 2007;28:328–36.
18. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res 2008;18:957–64.
19. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature 2008;455 1124–8.
20. Chang S, Johnston RJ Jr., Frokjaer-Jensen C, Lockery S, Hobert O. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature 2004;430:785–9.
21. Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nat Cell Biol 2009;11:1268–71.
22. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007;27:91–105.

Identification of New MicroRNAs in Paired Normal and Tumor Breast Tissue Suggests a Dual Role for the *ERBB2/Her2* Gene

Helena Persson, Anders Kvist, Natalia Rego, et al.

Cancer Res 2011;71:78-86.