Alternative Polyadenylation Modifies Target Sites of MicroRNAs with Clinical Potential for Breast Cancer Progression.

Soyeon Kim[1,2], Yulong Bai[3], Brenda Diergaarde[3,4], George C. Tseng[5], Hyun Jung Park[3*†]

[1]Department of Pediatrics, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

[2]Division of Pulmonary Medicine, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, Pennsylvania, USA

[3]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[4]Hillman Cancer Center, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

[5]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[†] Corresponding author: hyp15@pitt.edu (H.J.P.)

Abstract

Background: Alternative polyadenylation (APA) is a post-transcriptional mechanism that contributes to transcriptomic diversity. APA causes shortening or lengthening of the 3′-untranslated region (3′-UTR) associated with prognosis and drug sensitivity in cancer. We

recently identified a *trans* mechanism of 3′-UTR shortening that disrupts microRNA (miRNA) target sites for tumorigenesis. However, due to confounding interactions between APA events and miRNA target site modifications, it is challenging to identify which miRNAs contribute to tumorigenesis through APA events.

Methods: To identify miRNAs affected by APA events, we developed Probabilistic Inference of MicroRNA Target Site Modification through APA (PRIMATA-APA), a mathematical model that globally estimates significance of the APA-mediated target site modification for each miRNA.

Results: PRIMATA-APA identified that global APA events in TCGA breast tumor samples significantly modify target sites of particular miRNAs (tamoMiRNAs). TamoMiRNAs are enriched for miRNA biomarkers known for breast cancer etiology and treatments and their target genes are enriched in cancer-associated pathways such as "growth factor" and "signaling" pathways. As tamoMiRNAs are evolutionary more conserved and bind more genes than other miRNAs, APA events should effectively regulate the cancer-associated pathways. Moreover, the significant correlation between 3′-UTR usage and gene expression through miRNAs that are associated with the 3′-UTR usage suggests that APA events heavily contribute to miRNA-derived interpatient tumor heterogeneity.

Conclusions: Due to the high impact of APA *trans* effect on miRNA target site modification in cancer, we integrated miRNA target site modification and miRNA expression level profile, resulting in a systematic understanding of miRNA function for breast cancer.

Introduction

The dynamic usage of the messenger RNA 3′-untranslated region (3′-UTR) through alternative polyadenylation (APA) plays an important role in transcription[1], [2]. APA events use alternative sites of polyadenylation (polyA) resulting in transcription of distinct isoforms with 3′-UTR shortening and lengthening. 3′-UTR lengthening (3′UL) was recently reported to regulate cell senescence[3] with implication for tumor suppressive pathways such as cell cycle inhibitors, DNA damage markers, and tumor suppressors[4]–[7]. Widespread 3′-UTR shortening (3′US) was reported to be the strong prognostic marker[8] and was also found associated with drug sensitivity[9].

We recently discovered a tumorigenic mechanism of 3′US[10] that removes microRNA (miRNA) target sites in the distal region of the 3′-UTRs. Then, the miRNAs released from 3′US would be redirected to bind to the genes that would be competing for miRNA targeting (competing-endogenous RNA, ceRNA[11]) with the 3′US genes. Since the genes in the ceRNA relationship with the 3′US genes are likely to be tumor suppressors, the miRNA-mediated repression would then effectively promote tumorigenesis. These results imply that APA events globally disrupt the miRNA target sites (APA-derived miRNA target site modification) for tumor initiation/progression.

However, it is challenging to identify for which miRNAs APA events collectively modify the target sites in tumor samples in the following reasons. First, while 3′US removes miRNA target sites in the 3′-UTRs, 3′UL plays a confounding role by adding the target sites back. Second, APA events and their associated miRNAs are on many-to-many relationships, making it difficult to pinpoint miRNAs whose target sites are modified by particular APA events. To address the challenges, we developed a mathematical model that estimates the statistical significance of target site modifications due to APA for each miRNA, Probabilistic Inference of

MicroRNA Target Site Modification through APA (PRIMATA-APA). Further, we applied

PRIMATA-APA to analyze TCGA breast cancer data[12] and identified the functional

consequence of APA-derived miRNA target site modification.

Methods

TCGA breast tumor RNA-seq and miRNA-Seq data

Quantified gene expression files (RNASeqV1) for primary breast tumors and their matching

solid normal samples were downloaded from TCGA Data Portal[13]. We used 106 breast tumor

samples that have matched normal tissues. 10,868 expressed RefSeq genes (FPKM ≥ 1 in > 80%

of all samples) were selected for downstream analyses. To better quantify gene expression in the

presence of 3′-UTR shortening, we only used coding regions (CDS). Exon and CDS annotation

for TCGA data and miRNA expressions (syn1445790) were downloaded from Sage

Bionetworks' Synapse database.

Selection of miRNAs and genes

Predicted miRNA target sites were obtained from TargetScanHuman version 6.2[14]. Only those

with a preferentially conserved targeting score (Pct) more than 0 were used[8]. Experimentally

validated miRNA-target sites were obtained from TarBase version 5.0[15], miRecords version

4[16] and miRTarBase version 4.5[17]. The target sites found in indirect studies such as

microarray experiments and high-throughput proteomics measurements were filtered out [18].

Another source is the microRNA target atlas composed of public AGO-CLIP data[19] with

significant target sites (q-value < 0.05). The predicted and validated target site information was then combined to use in this study.

Probabilistic <u>I</u>nference of <u>M</u>icroRN<u>A</u> <u>T</u>arget Site Modification through <u>APA</u> (PRIMATA-APA)

For transcript x with a constitutive proximal 3′-UTR (pUTR) and a distal 3′-UTR (dUTR), we previously defined the amount of target sites for miRNA miR_j in all copies of transcript x as follows[10].

$$MiRs\_PDUI_t(x, miR_j) \qquad \text{(Eq. 1)}$$
$$= \left(pUTR(x, miR_j) + dUTR(x, miR_j) \times PDUI_t(x) \right)$$
$$\times CDS_t(x),$$

where $pUTR(x, miR_j)$ and $dUTR(x, miR_j)$ are the numbers of miR_j target sites in pUTR and dUTR of x. $PDUI_t(x)$ is the Percentage of dUTR Usage Index[8] of x and $CDS_t(x)$ is the expression level of x in a tumor sample. Note that $MiRs\_PDUI_n(x, miR_j)$ can be calculated for a normal sample with $PDUI_n(x)$ and $CDS_n(x)$. If APA-derived miRNA target site modification is not considered, the amount of target sites for miR_j in all copies of transcript x would be calculated as follows:

$$MiRs_t(x, miR_j) = \left(pUTR(x, miR_j) + dUTR(x, miR_j) \right) \times FPKM_t(x) \qquad \text{(Eq. 2)}$$

Based on **Eq.1** and **Eq.2**, PRIMATA-APA calculates $MiRs\_PDUI_t(miR_j)$ and $MiRs_t(miR_j)$ defined as below.

$$MiRs\_PDUI_t\big(miR_j\big) = \sum_x MiRs\_PDUI_t\big(x, miR_j\big)$$

$$MiRs_t\big(miR_j\big) = \sum_x MiRs_t\big(x, miR_j\big)$$

With $MiRs\_PDUI_t\big(miR_j\big)$, $MiRs_t\big(miR_j\big)$, $MiRs\_PDUI_n\big(miR_j\big)$, and $MiRs_n\big(miR_j\big)$ in a contingency table, PRIMATA-APA estimates significance of target site modifications for miR_j by testing nonrandom association in tumor and normal states (using χ^2 test), followed by FDR control using FowardStop[20] (FDR < 0.01).

Results

Collective impact of APA genes for the *trans* effect

To identify APA events in large-scale data, several computational tools have been developed that use RNA-Seq data[8], [21]–[24]. For example, statistically significant APA genes can be defined using the difference in Percentage of Distal polyA site Usage Index (ΔPDUI)[8]. Current analyses have focused mostly on a subset of the significant APA genes that strongly changed 3′-UTR usage in tumor by employing an additional cutoff. Specifically, recent large-scale APA studies focused on the strong APA genes (ΔPDUI < -0.2 for 3′US and ΔPDUI > 0.2 for 3′UL) selected from significant APA genes (FDR < 0.05) in TCGA human cancer[8] and cell line data[9]. However, strong APA genes account for only a small portion of all significant APA genes. For example, in TCGA tumor-normal sample pair BH-A1FJ with the greatest number of significant APA genes, the strong APA genes account only for 50.5% (1,523) of 3,015 significant APA genes (**Fig. 1A**). Across 106 breast tumor-matched normal sample pairs in

TCGA, this trend is more pronounced in that only 40.5% of significant APA genes are strong APA genes on average (**Fig. 1B**).

We found that significant APA genes, strong or not, together elucidate the common *trans* mechanism of APA. In 106 normal/tumor sample pairs, we identified totally 6,825 significant 3′US genes, of which 82.4% (5,626) are strong in some sample pairs, demonstrating that significant 3′US genes in a sample pair are likely to undergo a strong 3′-UTR shortening in other sample pairs. Further, while considering all significant 3′US genes increases only 15.9% (6,825 vs. 5,626) more 3′US genes in total, it enables us to identify much more 3′US genes in common. For 20% (21) of the 106 normal/tumor sample pairs, 5.5-fold more 3′US genes are shared if all significant 3′US genes are considered (613 significant vs. 110 strong only, **Fig. 1C**). Especially, to characterize the common *trans* mechanism, all significant 3′US genes need to be considered. Since strong 3′US genes are likely strong *cis* targets of 3′US, they would be reasonable candidates to study the 3′US *cis* effect. However, as 3′US events exert the *trans* mechanism through modifying miRNA target sites[10], significant but not strong 3′US genes would also contribute to the *trans* mechanism by modifying miRNA target sites on their 3′UTRs[11], especially if they are highly expressed.

In the same sense, 3′UL genes also need to be considered, since they would increase miRNA target sites to compensate for those decreased by 3′US genes. Previous studies reported smaller numbers of 3′UL genes than 3′US from TCGA cancer patients[8] or cancer cell lines[9] of diverse types. This was partly because they focused only on strong APA genes common to the samples. Considering all significant APA genes in our TCGA breast cancer analysis, we found that 3′UL is as widespread as 3′US. For example, in TCGA tumor-normal sample pair AC-A2FB, which has the smallest ratio of strong APA genes to significant APA genes, 90.1% of the

significant APA genes are 3′UL genes (**S. Fig 1B**). In 106 breast tumor-matched normal sample pairs in TCGA, 56.4% of average significant APA genes are 3′UL genes (**S. Fig. 1C**). As in the case of 3′US, considering significant 3′UL genes enables to identify much more common 3′UL genes than considering only strong 3′UL genes. In our TCGA breast cancer data, 83.7% of the total significant 3′UL genes (6,081 of 7,265) are strong and 3.5-fold more 3′UL genes are shared for 20% of the sample pairs when all significant 3′UL genes are considered (993 significant vs. 288 strong only, **S. Fig. 1A**). Based on the results, investigating the common *trans* mechanisms of APA requires to consider all significant APA genes of both 3′UL and 3′US.

Probabilistic Inference of MicroRNA Target Site Modification through APA (PRIMATA-APA)

To quantify miRNA target site modifications due to all significant APA events, either 3′-UTR shortening and lengthening, we developed a mathematical model, Probabilistic Inference of MicroRNA Target Site Modification through APA (PRIMATA-APA). Previously, we successfully predicted gene expression changes based on the estimated number of miRNA target sites in the presence of 3′US ($MiRs\_PDUI_t(x, miR_j)$, **Eq. 2**)[10]. By extending this estimation, PRIMATA-APA estimates the total number of target sites for each miRNA with and without consideration of APA events. Based on the difference of the estimations, PRIMATA-APA quantifies how much of the target sites are modified, either increased or decreased, by APA events for each miRNA (see Methods).

Global miRNA target site modification due to alternative polyadenylation

To study APA-derived miRNA target site modification in cancer, we ran PRIMATA-APA on the 70 breast tumor-matched normal samples out of 106 pairs for which miRNA expression information was available. In the data, we considered 588 moderately expressed miRNAs (> 1 and < 100 FPM on average) and 3,318 expressed genes likely affected by miRNAs (≥ 5 miRNA target sites) to focus on active miRNA targets[10], [25]. In 39 (55.7%) of the 70 breast tumor samples, PRIMATA-APA identifies significant target site modifications (FDR < 0.01) for more than 100 miRNAs (**Fig. 2A, B**). Further, in each sample pair, miRNA target sites are either mostly increased or decreased, which makes a negative correlation between the number of target sites increased and decreased across tumor-normal sample pairs (P=0.006, **Fig. 2C**).

Additionally, we found that APA events modify miRNA target sites in a subtype-specific manner. The five subtypes of breast cancer by PAM50 are known to involve distinct molecular pathways with different clinical outcomes[26]. In our TCGA breast cancer data, none of basal and Her2 subtype samples increases target sites for > 100 miRNAs, while 57.1% (4/7) of the samples decrease target sites for > 100 miRNAs, indicating a significant (P=0.009) bias toward miRNA target site decrease. However, other PAM50 subtypes (Luminal A, B, and Normal-like) do not show such a bias (**S. Fig. 3**). Since both basal and Her2 subtypes are close in terms of molecular pathways and worse prognosis (reviewed in [27], [28]), their common pattern in miRNA target site modification suggests a similar APA landscapes between them. Altogether, the results show that APA globally modifies miRNA target sites for breast cancer in a non-random and a subtype-specific manner.

APA modifies target sites of miRNAs associated with cancer

For further analyses based on miRNAs, we identified target miRNAs whose target sites are modified by APA events. Based on the number of tumor samples where the target sites are increased or decreased for each miRNA, we selected top half (289) of 588 moderately expressed miRNAs, whose target sites are more often modified than the other (299) miRNAs, which will be termed target site modified miRNA (tamoMiRNA). APA events modify target sites of tamoMiRNAs significantly more than the other miRNAs (P-value=$1.12e^{-28}$, **Fig. 3A**). The 289 tamoMiRNAs are significantly (P=$5.8e^{-5}$) enriched for cancer etiology and treatments compared to the other miRNAs. Specifically, tamoMiRNAs are enriched in the miRNAs that are dysregulated in breast cancer with clinical and biological implications[29], regulating diverse mechanisms for breast cancer[30], regulatory elements in either adaptive or innate immune system[31], or potential prognostic and predictive biomarkers identified for breast cancer[32](**Fig. 3B**, S. Table 1). Among 43 tamoMiRNAs found in the categories, 31 (72.1%) occur only in one of the categories (**S. Fig. 2**), confirming that the high enrichment of tamoMiRNAs to the multiple categories reflects their important roles in tumor, not redundancy in data curation. Also, we estimated conservation score (PhyloP[33], 46 way Placental) of 202 tamoMiRNAs and 191 other miRNAs for which miRBase[34] curated the genomic locations. TamoMiRNAs have significantly (P= $7.19e^{-5}$) larger conservation scores than the other miRNAs (**Fig. 3C**). Altogether, the results indicate that, by selection or design, APA modifies target sites of miRNAs that are evolutionary conserved and functionally important for cancer etiology and treatments.

APA modifies target sites of miRNAs to effectively regulate biological processes.

To investigate the function of the APA-derived miRNA target site modification in cancer, we focused on the target genes of tamoMiRNAs (see Methods). First, we identified GO terms enriched for tamoMiRs using MiEAA[35]. Probably due to the many-to-many relationships between miRNAs and target genes[36]–[38], inputting all tamoMiRNAs and all the other miRNAs to MiEAA web server returns mostly under-represented terms. So, we focus MiEAA analysis on 99 tamoMiRNAs (with the greatest number of samples in which modified) and 105 other miRNAs (with the least number of samples in which modified). 125 and 1 biological terms are significantly (FDR < 0.01) enriched for tamoMiRNAs and for the other miRNAs respectively (S. Table 2). The significant bias (P-value=5.0×10^{-5}) of the number of enriched biological terms to tamoMiRNAs suggests that APA events effectively regulate biological functions. Additionally, compared to the other miRNAs, tamoMiRNAs are exclusively enriched for pathways with keyword "growth factor", "signaling", and "circadian", (**Fig. 4A**, S. Table 2), which are essential for tumor initiation and progression[39].

To understand their impact in regulating gene expression, we evaluated the number of genes targeted by tamoMiRNAs. Among 3,318 expressed genes in the breast tumor data that are likely controlled by miRNAs (> 5 miRNA target sites), 3,177 genes (95.7%) have more target sites for tamoMiRNAs than for the other miRNAs (**Fig. 4B**). Further, 911 of 3,177 (27.4%) genes have target sites only for tamoMiRNAs in their 3′-UTRs. While expression fold change (tumor vs. normal) does not differ between tamoMiRNAs and other miRNAs (P=0.1, **S. Fig. 2**), 911 genes targeted only by tamoMiRNAs are significantly more down-regulated in tumor (P=$3.9e^{-23}$) than the same number of genes affected by other miRNAs (**Fig. 4C**), indicating that tamoMiRNAs effectively regulate gene expressions of target genes in tumor. Altogether, the

results indicate that APA modifies target sites of miRNAs that effectively regulate genes in tumor-associated pathways.

APA regulates tumor-specific progression in association to interpatient heterogeneity.

We further studied the role of APA-derived miRNA target site modification on the observed role of miRNAs for tumor interpatient heterogeneity[40]. In particular, since 3′-UTR shortening regulates gene expression by modifying miRNA target sites[10], we hypothesized that variation in the degree of APA events across tumor samples diversifies the effect of miRNAs on the target genes for tumor interpatient heterogeneity. To test the hypothesis, we compared the expression variation of tamoMiRNAs, the other miRNAs and 911 of their target genes defined above. While the expression variation across the sample pairs changes equally in tumor for both tamoMiRNAs and the other miRNAs (P=0.4, **Fig. 5A**), the expression variation of genes that are targeted by tamoMiRNA is significantly higher than that of the other miRNAs (P=4.9e^{-15}, **Fig. 5B**) in tumors. Since the degree of APA events varies significantly more (P=0.004) in tumor (**S. Fig. 5B**), the high variation in tamoMiRNA target gene expression in tumor is attributable to the APA events modifying tamoMiRNA target sites in the genes.

An example is myocin heavy chain 11 (MYH11) that promotes tumorigenesis of various cancer types[41]–[45]. In our miRNA-target site information, MYH11 is predicted to have a target site for a single miRNA (miRNA-124/124ab/506). Although the expression of MYH11 is not correlated with the miRNA expression (**S. Fig. 5A**), its expression is significantly correlated with ΔPDUI values of its ceRNA partner gene, platelet-activating factor acetylhydrolase IB subunit alpha (PARAH1B1) (Pearson's r=0.4, P-value=0.00014, **Fig. 5C**). The functional role of

PARAH1B1 in tumor has not been widely studied, especially in association with MYH11, partly because their expression levels are not correlated (e.g. Pearson's r=0.1, P-value=0.25 in TCGA breast cancer). Our analysis found that 3′-UTR of PARAH1B1 undergoes a significant (T-test statistic=2.65, P=0.004) shortening in tumors. Since negative ΔPDUI values represent 3′-UTR shortening in tumor[8] which is expected to release miRNAs to repress MYH11 (**Fig. 5D**), the positive correlation between MYH11 expression and ΔPDUI of PARAH1B1 (**Fig. 5C**) supports the role of 3′US *trans* effect differentiating MYH11 expression.

Further enrichment analyses support that APA-associated, miRNA-mediated[46], [47] transcriptomic diversity contributes to interpatient difference in tumor progression. First, 911 genes targeted only by tamoMiRNAs include significantly more oncogenes than the same number of genes targeted by other miRNAs (43 vs. 28, P-value=0.03), suggesting that varying degree of APA diversifies the effect of miRNA target activity on oncogenic processes. Second, our MiEAA analysis showed that tamoMiRNAs regulate tumor progression pathways such as "growth factor" pathways (**Fig. 4A**). Third, Ingenuity Pathway Analysis showed that cancer progression and migration pathways are more enriched in tamoMiRNA target genes than the target genes of the other miRNAs (P-value $< 10^{-3}$, **Fig. 5E**, S. Table 3). Specifically, they are implicated for breast cancer often through miRNAs, e.g. with miR-494 suppressing chemokine (C-X-C motif) receptor 4 (CXCR4) for breast cancer progression[48], miR-200c regulating Protein kinase A subunits for cancer cell migration[49], and miR-520b targeting Interleukin-8 for breast cancer cell migration[50] (see other examples in **Fig. 5E**). Altogether, APA regulates tumor-specific progression by diversifying miRNA target site landscape.

Discussion

Here, we studied the effect of miRNA target site modification together with miRNA expression. As miRNAs bind and repress their target genes, a comprehensive understanding of miRNAs' function requires investigating not only their expression level but also their target site landscape. Unlike previous works focusing on the changes in miRNA expressions, we investigated aberration in miRNA target site landscape by taking widespread 3′-UTR shortening and lengthening into a mathematical model, PRIMATA-APA.

Further, our work will shed novel insights into the development of therapeutic miRNAs. The high enrichment of tamoMiRNAs to miRNAs validated with clinical implication (**Fig. 3B**) suggest that APA events employ the *trans* effect as patient-specific APA events[8] are associated with tumor progression[10], prognosis[8] and treatment outcomes[51]. Based on this observation, further inspection of tamoMiRNAs would effectively narrow down search space to identify therapeutic miRNAs. Further, identifying APA events associated with the interpatient heterogeneity will help regulate interpatient tumor heterogeneity, which is essential for the success of early cancer detection and the development of new effective therapies[52], [53]. For example, when MYH11's oncogenic effect is associated with a varying degree of APA events of PARAH1B1 through miRNA-124/124ab/506 in breast cancer, molecular agents for miRNA-124/124ab/506 may help normalize different MYH11 effect on cancer patients.

References

[1]   R. Elkon *et al.*, "E2F mediates enhanced alternative polyadenylation in proliferation.," *Genome Biol.*, vol. 13, no. 7, p. R59, Jul. 2012.

[2]   R. Sandberg, J. R. Neilson, A. Sarma, P. A. Sharp, and C. B. Burge, "Proliferating Cells Express mRNAs with Shortened 3′ Untranslated Regions and Fewer MicroRNA Target Sites," vol. 320, no. June, pp. 1643–1647, 2008.

[3]   M. Chen *et al.*, "3 ′ UTR lengthening as a novel mechanism in regulating cellular senescence," pp. 285–294, 2018.

[4]   G. P. Dimri *et al.*, "A biomarker that identifies senescent human cells in culture and in aging skin in vivo.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 20, pp. 9363–7, 1995.

[5]   R. A. Busuttil, M. Rubio, M. E. T. Dollé, J. Campisi, and J. Vijg, "Oxygen accelerates the accumulation of mutations during the senescence and immortalization of murine cells in culture.," *Aging Cell*, vol. 2, no. 6, pp. 287–294, 2003.

[6]   C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "The hallmarks of aging," *Cell*, vol. 153, no. 6, 2013.

[7]   D. Muñoz-Espín and M. Serrano, "Cellular senescence: From physiology to pathology," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 7, pp. 482–496, 2014.

[8]   Z. Xia *et al.*, "Dynamic Analyses of Alternative Polyadenylation from RNA- Seq Reveal Landscape of 3 ' UTR Usage Across 7 Tumor Types," *Nat. Commun.*, pp. 1–38, 2014.

[9]   Y. Xiang *et al.*, "Comprehensive Characterization of Alternative Polyadenylation in Human Cancer," vol. 110, no. November 2017, pp. 1–11, 2018.

[10]   H. J. Park *et al.*, "3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk," *Nat. Genet.*, 2018.

[11]   L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi, "A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?," *Cell*, vol. 146, no. 3, pp. 353–8, Aug. 2011.

[12]   T. C. G. A. Network, "Comprehensive molecular portraits of human breast tumours.," *Nature*, vol. 490, no. 7418, pp. 61–70, Oct. 2012.

[13]   M. Goldman *et al.*, "The UCSC Cancer Genomics Browser: update 2013.," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D949-54, Jan. 2013.

[14]   B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.," *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005.

[15]   G. L. Papadopoulos, M. Reczko, V. a Simossis, P. Sethupathy, and A. G. Hatzigeorgiou, "The database of experimentally supported targets: a functional update of TarBase.," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D155-8, Jan. 2009.

[16]   F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li, "miRecords: An integrated resource for microRNA-target interactions," *Nucleic Acids Res.*, vol. 37, no. November 2008, pp. 105–110, 2009.

[17]   S.-D. Hsu *et al.*, "miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions.," *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D78-85, Jan. 2014.

[18]   H. Dvinge *et al.*, "The shaping and functional consequences of the microRNA landscape

in breast cancer.," *Nature*, vol. 497, no. 7449, pp. 378–82, May 2013.

[19] M. P. Hamilton *et al.*, "Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif.," *Nat. Commun.*, vol. 4, p. 2730, Jan. 2013.

[20] M. G. G'Sell, S. Wager, A. Chouldechova, and R. Tibshirani, "Sequential Selection Procedures and False Discovery Rate Control," *arXiv Prepr.*, p. 27, 2013.

[21] K. C. H. Ha, B. J. Blencowe, and Q. Morris, "QAPA : a new method for the systematic analysis of alternative polyadenylation from RNA-seq data," pp. 1–18, 2018.

[22] C. Ye, Y. Long, G. Ji, Q. Q. Li, and X. Wu, "Genome analysis APAtrap : identification and quantification of alternative polyadenylation sites from RNA-seq data," vol. 34, no. January, pp. 1841–1849, 2018.

[23] Z. Huang and E. C. Teeling, "ExUTR: A novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data," *BMC Genomics*, vol. 18, no. 1, pp. 1–11, 2017.

[24] W. Wang, Z. Wei, and H. Li, "A change-point model for identifying 3'UTR switching by next-generation RNA sequencing," *Bioinformatics*, vol. 30, no. 15, pp. 2162–2170, 2014.

[25] U. Ala *et al.*, "Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 18, pp. 7154–9, Apr. 2013.

[26] T. Cancer Genome Atlas Network, "Comprehensive molecular portraits of human breast tumours," 2012.

[27] X. Dai *et al.*, "Breast cancer intrinsic subtype classification, clinical use and future trends," *Am J Cancer Res*, vol. 5, no. 10, pp. 2929–2943, 2015.

[28] X. Dai, H. Cheng, Z. Bai, and J. Li, "Breast cancer cell line classification and Its relevance with breast tumor subtyping," *J. Cancer*, vol. 8, no. 16, pp. 3131–3141, 2017.

[29] Y. Peng and C. M. Croce, "The role of MicroRNAs in human cancer," *Signal Transduct. Target. Ther.*, vol. 1, no. November 2015, p. 15004, 2016.

[30] S. Kurozumi, Y. Yamaguchi, M. Kurosumi, M. Ohira, and H. Matsumoto, "Recent trends in microRNA research into breast cancer with particular focus on the associations between microRNAs and intrinsic subtypes," vol. 62, no. 1, pp. 15–24, 2016.

[31] A. Mehta and D. Baltimore, "MicroRNAs as regulatory elements in immune system logic," *Nat. Rev. Immunol.*, vol. 16, no. 5, pp. 279–294, 2016.

[32] E. van Schooneveld, H. Wildiers, I. Vergote, P. B. Vermeulen, L. Y. Dirix, and S. J. Van Laere, "Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management," *Breast Cancer Res.*, vol. 17, no. 1, p. 21, 2015.

[33] K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, and A. Siepel, "Detection of nonneutral substitution rates on mammalian phylogenies," *Genome Res.*, vol. 20, no. 1, pp. 110–121, 2010.

[34] A. Kozomara and S. Griffiths-Jones, "miRBase: annotating high confidence microRNAs using deep sequencing data.," *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D68-73, Jan. 2014.

[35] C. Backes, Q. T. Khaleeq, E. Meese, and A. Keller, "MiEAA: MicroRNA enrichment analysis and annotation," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W110–W116, 2016.

[36] C. H. Ooi *et al.*, "A densely interconnected genome-wide network of micrornas and oncogenic pathways revealed using gene expression signatures," *PLoS Genet.*, vol. 7, no. 12, 2011.

[37] K. J. Mavrakis *et al.*, "A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (T-ALL)," *Nat. Genet.*, vol. 43, no. 7, pp. 673–678, 2011.

[38] Y. Hashimoto, Y. Akiyama, and Y. Yuasa, "Multiple-to-Multiple Relationships between MicroRNAs and Target Genes in Gastric Cancer," *PLoS One*, vol. 8, no. 5, 2013.

[39] F. Sanchez-Vega *et al.*, "Oncogenic Signaling Pathways in The Cancer Genome Atlas," *Cell*, vol. 173, no. 2, p. 321–337.e10, 2018.

[40] S. P. Nana-Sinkam and C. M. Croce, "MicroRNA regulation of tumorigenesis, cancer progression and interpatient heterogeneity: Towards clinical use," *Genome Biol.*, vol. 15, no. 9, pp. 1–9, 2014.

[41] N. Vickaryous *et al.*, "Smooth-muscle myosin mutations in hereditary non-polyposis colorectal cancer syndrome," *Br. J. Cancer*, vol. 99, no. 10, pp. 1726–1728, 2008.

[42] P. Liu *et al.*, "Fusion Between Transcription Factor CBFβ / PEBP2β and a Myosin Heavy Chain in Acute Myeloid Leukemia Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/2881859 JSTOR is a not-for-profit service tha," vol. 261, no. 5124, pp. 1041–1044, 2019.

[43] S. Huang, Z. G. Gulzar, K. Salari, J. Lapointe, J. D. Brooks, and J. R. Pollack, "Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness," *Oncogene*, vol. 31, no. 37, pp. 4164–4170, 2012.

[44] L. Burkhardt *et al.*, "CHD1 Is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer," *Cancer Res.*, vol. 73, no. 9, pp. 2795–2805, 2013.

[45] P. Alhopuro *et al.*, "Unregulated smooth-muscle myosin in human intestinal neoplasia," *Proc. Natl. Acad. Sci.*, vol. 105, no. 14, pp. 5513–5518, 2008.

[46] I. Gupta *et al.*, "Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA–protein interactions," *Mol. Syst. Biol.*, vol. 10, no. 719, p. 176, 2014.

[47] W. Chen, Q. Jia, Y. Song, H. Fu, G. Wei, and T. Ni, "Alternative Polyadenylation: Methods, Findings, and Impacts," *Genomics. Proteomics Bioinformatics*, vol. 15, no. 5, pp. 287–300, Oct. 2017.

[48] Li. Song *et al.*, "miR-494 suppresses the progression of breast cancer in vitro by targeting CXCR4 through the Wnt/β-catenin signaling pathway," *Oncol. Rep.*, vol. 34, no. 1, pp. 525–531, 2015.

[49] F. C. Sigloch, U. C. Burk, M. L. Biniossek, T. Brabletz, and O. Schilling, "miR-200c dampens cancer cell migration via regulation of protein kinase A subunits," *Oncotarget*, vol. 6, no. 27, 2015.

[50] N. Hu *et al.*, "MiR-520b regulates migration of breast cancer cells by targeting hepatitis B X-interacting protein and interleukin-8," *J. Biol. Chem.*, vol. 286, no. 15, pp. 13714–13722, 2011.

[51] Y. Xiang *et al.*, "Comprehensive Characterization of Alternative Polyadenylation in Human Cancer," *JNCI J. Natl. Cancer Inst.*, vol. 110, no. March, pp. 1–11, 2017.

[52] D. Endesfelder *et al.*, "Intratumor Heterogeneity and Branched Evolution Revealed by

Multiregion Sequencing," *N. Engl. J. Med.*, vol. 366, no. 10, 2012.

[53]   P. L. Bedard, A. R. Hansen, M. J. Ratain, and L. L. Siu, "Tumour heterogeneity in the clinic," *Nature*, vol. 501, pp. 355–364, 2013.

[54]   M. Mancini and A. Toker, "NFAT proteins: Emerging roles in cancer progression," *Nat. Rev. Cancer*, vol. 9, no. 11, pp. 810–820, 2009.

[55]   G. C. Harburg and L. Hinck, "Navigating breast cancer: Axon guidance molecules as breast cancer tumor suppressors and oncogenes," *J. Mammary Gland Biol. Neoplasia*, vol. 16, no. 3, pp. 257–270, 2011.

[56]   C. Corzo *et al.*, "The MYC oncogene in breast cancer progression: From benign epithelium to invasive carcinoma," *Cancer Genet. Cytogenet.*, vol. 165, no. 2, pp. 151–156, 2006.

[57]   J. Cai, "Decreased Pigment Epithelium-Derived Factor Expression in Human Breast Cancer Progression," *Clin. Cancer Res.*, vol. 12, no. 11, pp. 3510–3517, 2006.

[58]   M. L. Nierodzik and S. Karpatkin, "Thrombin induces tumor growth, metastasis, and angiogenesis: Evidence for a thrombin-regulated dormant tumor phenotype," *Cancer Cell*, vol. 10, no. 5, pp. 355–362, 2006.

[59]   K. K. Chan *et al.*, "Erythropoietin drives breast cancer progression by activation of its receptor EPOR," *Oncotarget*, vol. 8, no. 24, pp. 38251–38263, 2015.

[60]   Z. Li *et al.*, "miR-29a regulated ER-positive breast cancer cell growth and invasion and is involved in the insulin signaling pathway," *Oncotarget*, vol. 8, no. 20, pp. 32566–32575, 2017.

[61] F. J. Sulzmaier, C. Jean, and D. D. Schlaepfer, "FAK in cancer: Mechanistic findings and clinical applications," *Nat. Rev. Cancer*, vol. 14, no. 9, pp. 598–610, 2014.

[62] K. Banerjee and H. Resat, "Constitutive activation of STAT3 in breast cancer cells: A review," *Int. J. Cancer*, vol. 138, no. 11, pp. 2570–2578, 2016.

Figures and Tables
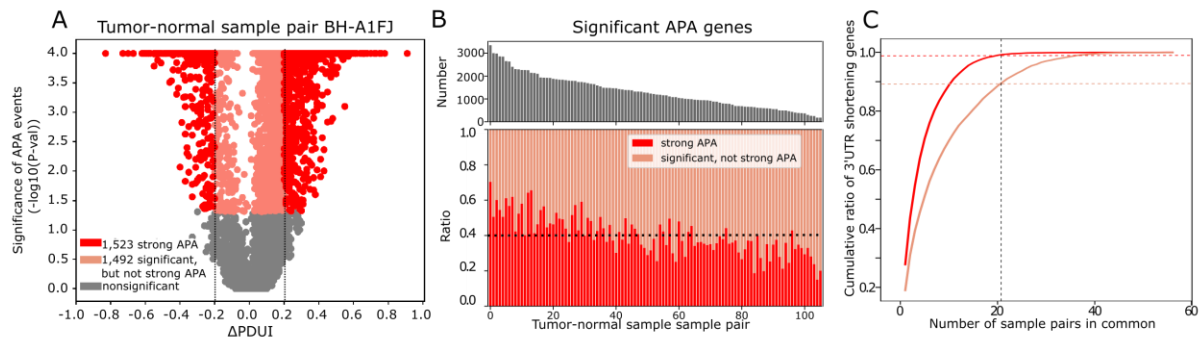
Figure 1. Collective impact of strong and significant APA events. A. Statistical significance of APA target genes in a breast tumor-normal sample pair (TCGA-BH-A1FJ) with their ΔPDUI (Percentage of Distal polyA site Usage Index) values (tumor-normal). Since PDUI represents the ratio of isoforms with distal 3′-UTR, negative ΔPDUI value represent 3′-UTR shortening target genes and positive ΔPDUI value 3′UL genes. Strong APA target genes are in red, significant but not strong ones in pink and not significant ones in gray. B. For 106 breast tumor-normal sample pairs sorted by the number of significant APA target sites, upper panel shows the total number of significant APA genes and the lower panel shows the ratio of the APA genes by whether it is significant but not strong (orange) or strong (red). Black dotted line represent the average ratio of strong APA genes. C. Cumulative ratio of 3′US genes shared by sample pairs. Red and orange dotted lines represent the ratio of strong and significant 3′US genes shared by < 21 sample pairs, respectively.
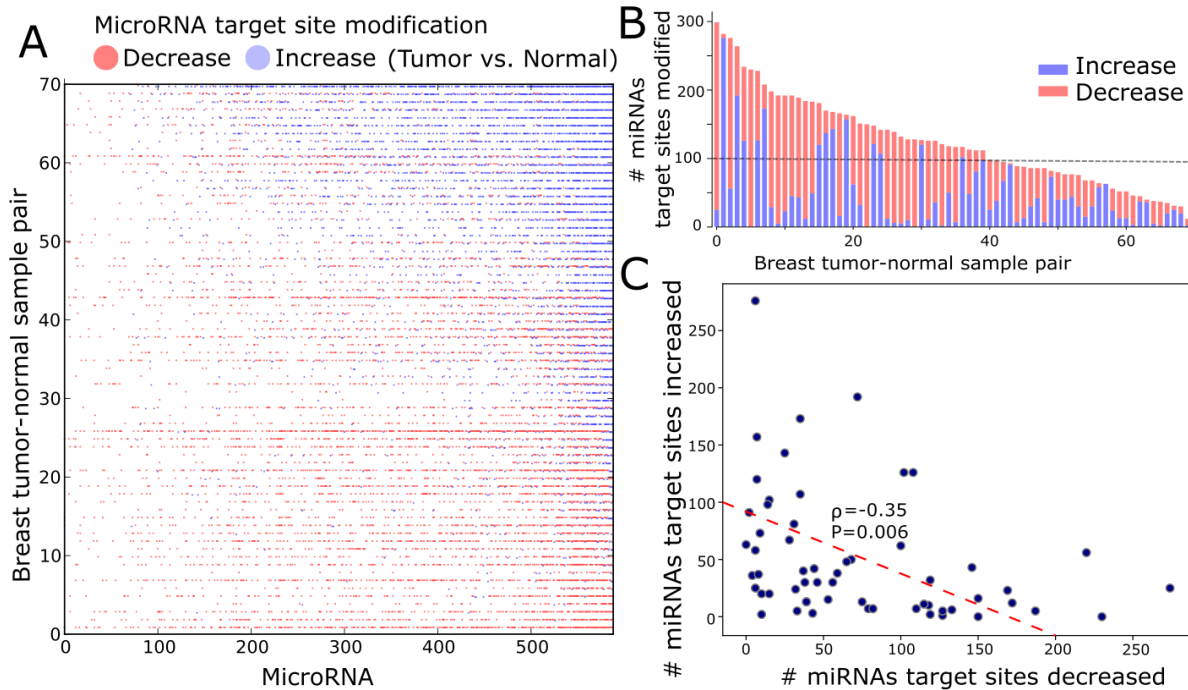
Figure 2. Tumor-specific APA-mediated microRNA binding site changes. A. The heatmap shows tumor-normal samples (row) where the total number of binding sites for each microRNA (column) is increased (blue) or decreased (red) due to APA. Not significant changes or no changes are not colored. Samples are sorted by the number of increased microRNA target site modicifcation. B. The total number of miRNA binding site changes, either increased (blue) or decreased (red) due to APA, in breast tumor-normal samples pair sorted by the total number of modification per sample pair. C. Number of miRNAs of which binding sites are increased (y-axis) or decreased (x-axis) for each tumor-normal sample. The red dotted line represents linear least-squares regression.
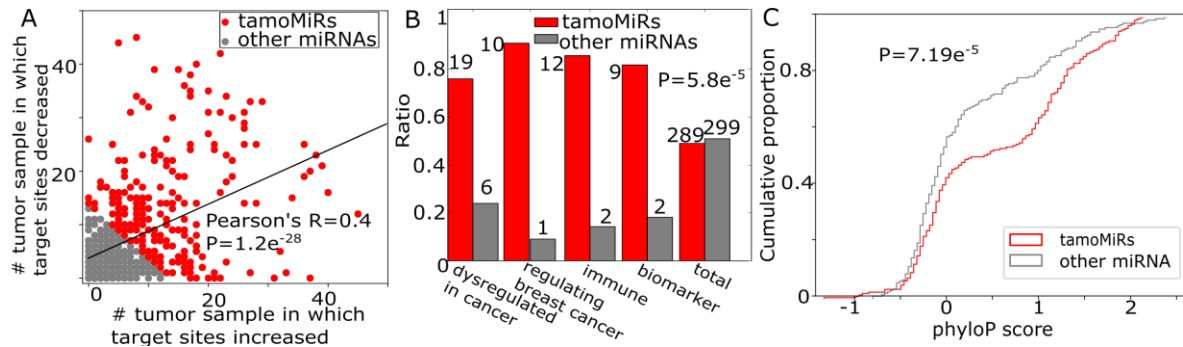
Figure 3. APA modifies binding sites of miRNAs associated with cancer. A. The number of tumor-normal samples between which binding sites for each miRNA are increased (x-axis) or decreased (y-axis). For further analyses, we dichotomize miRNAs by the amount of binding site modification into tamo- (red) and the other (gray) miRNAs. B. Number of cancer-related miRNAs in tamo- (red) and the other (gray) miRNAs. C. Distribution of phyloP conservation score for 202 tamo and 191 the other miRNAs.
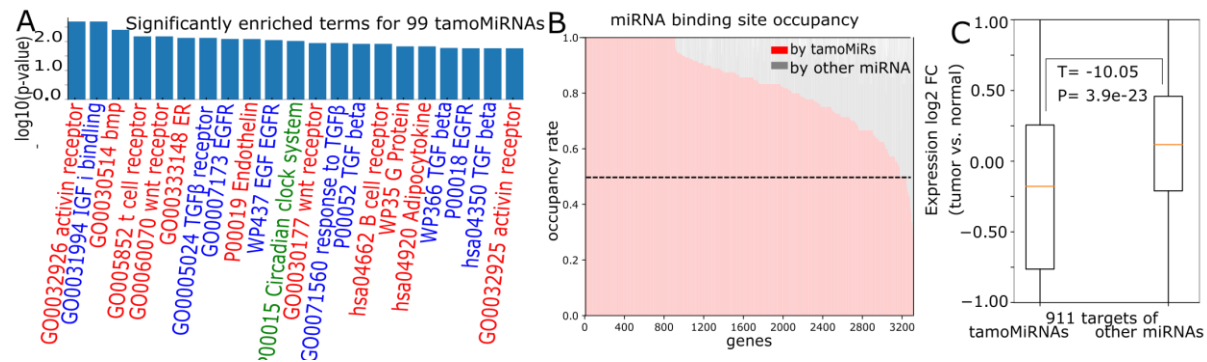
Figure 4. TamoMiRNAs effectively regulate biological processes. A. Cancer-associated pathways enriched for 99 tamoMiRNAs with their enrichment p-values (red for "signaling", blue for "GF" (growth factor), and green for "circadian"). B. Number of target sites for tamoMiRs and the other miRNAs in the genes with more than 5 target sites. C. Expression fold change (log2 tumor vs. normal) of 911 genes that are targets of tamoMiRs and other miRNAs.
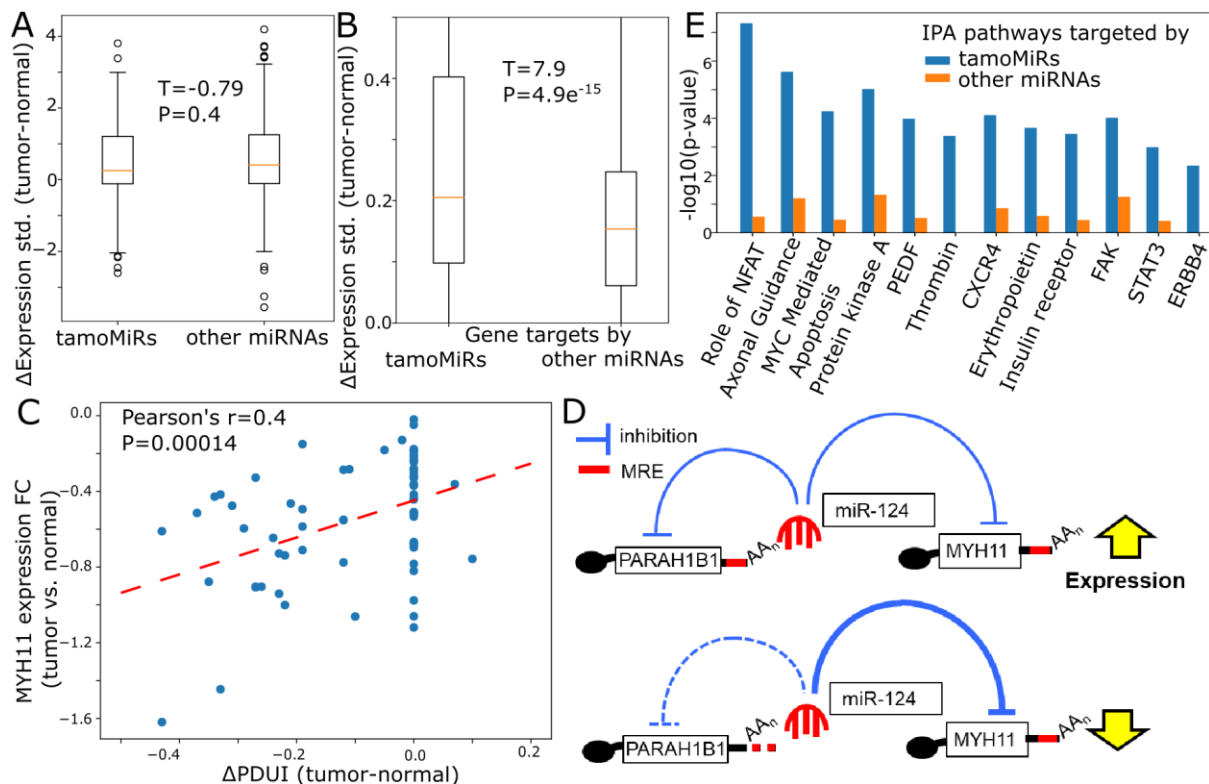
Figure 5. APA regulates tumor-specific progression in trans. Distribution of standard deviation values across sample pairs on the expression values of A. tamoMiRs and other miRNAs and B. their target genes. C. Scatterplot of MYH11 expression fold change values against its ΔPDUI values. The red dotted line represents linear least-squares regression. D. Illustration of the effect of PARAH1B1's 3′-UTR shortening on MYH11 expression mediated by miR-124/124ab/506. MRE stands for microRNA recognition element. E. IPA comparison analysis between gene targets by highly and lowly modified miRNAs for pathways implied for cancer progression and migration, NFAT[54], Axonal Guidance[55], MYC Mediated Apoptosis[56], Protein kinase A[49], Pigment epithelium-derived factor (PEDF)[57], Thrombin[58], CXCR4[48], Erythropoietin[59], Insulin receptor[60], FAK[61], STAT3[62].
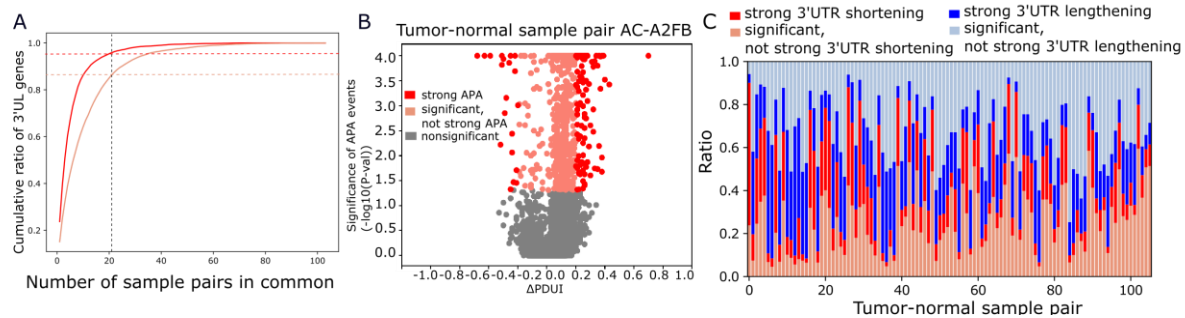
S. Figure 1. A. Cumulative ratio of genes with lengthened 3′-UTRs shared by sample pairs. Cumulative ratio of 3′US genes shared by sample pairs. Red and orange dotted lines represent the ratio of strong and significant 3′US genes shared by < 21 sample pairs, respectively. B. Statistical significance of APA target genes in a breast tumor-normal sample pair (TCGA-BH-A1FJ) with their ΔPDUI (Percentage of Distal polyA site Usage Index) values (tumor-normal). C. In each of 106 breast tumor-normal sample pairs, the ratio of the APA target genes by whether it is significant and not strong or strong, also by whether it is 3'UTR shortening or lengthening. They are ordered in consistency with Fig. 1B.
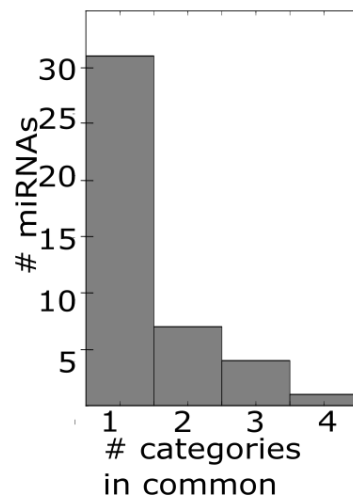
S. Figure 2. Number of miRNAs validated for breast tumor progression and treatment against how many validation category they are in.
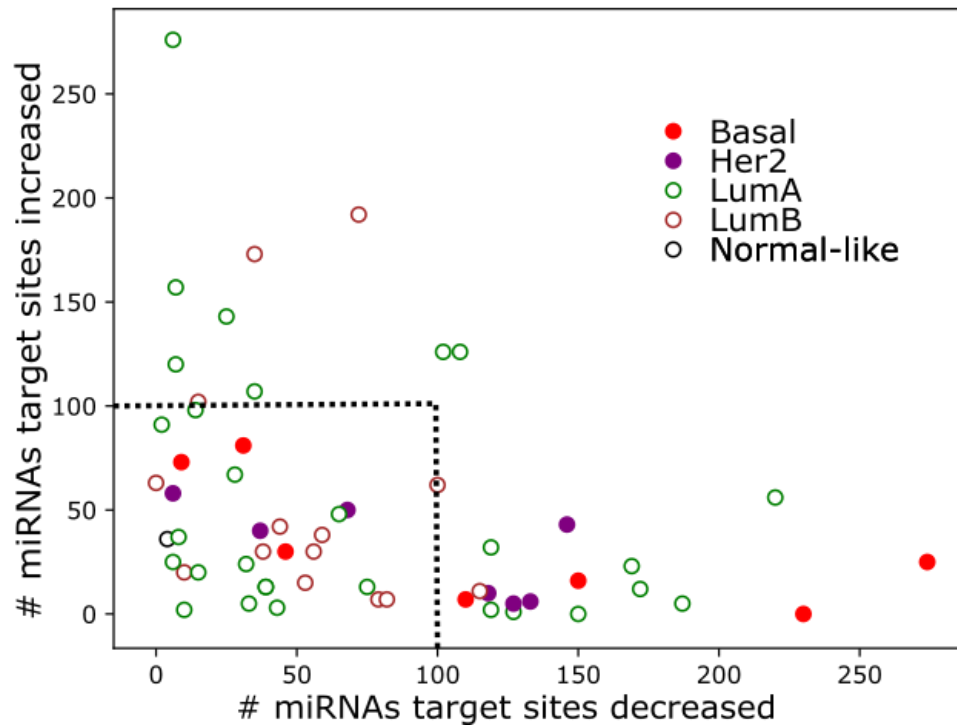
S. Figure 3. Number of miRNAs of which binding sites are increased (y-axis) or decreased (x-axis) for each tumor-normal sample colored by the breast tumor subtype. The black dotted rectangle represents not-high modifications.
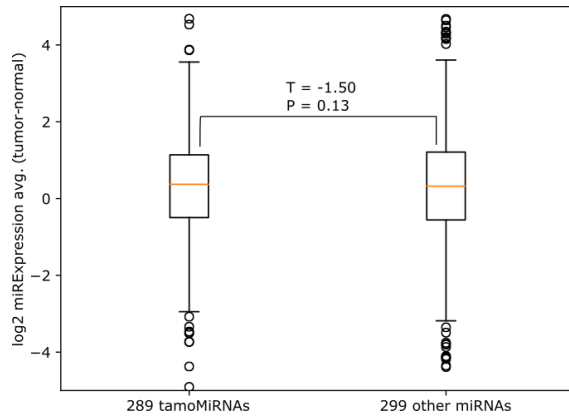
S. Figure 4. Expression difference (log 2, tumor vs. normal) of tamoMiRNAs and other miRNAs.
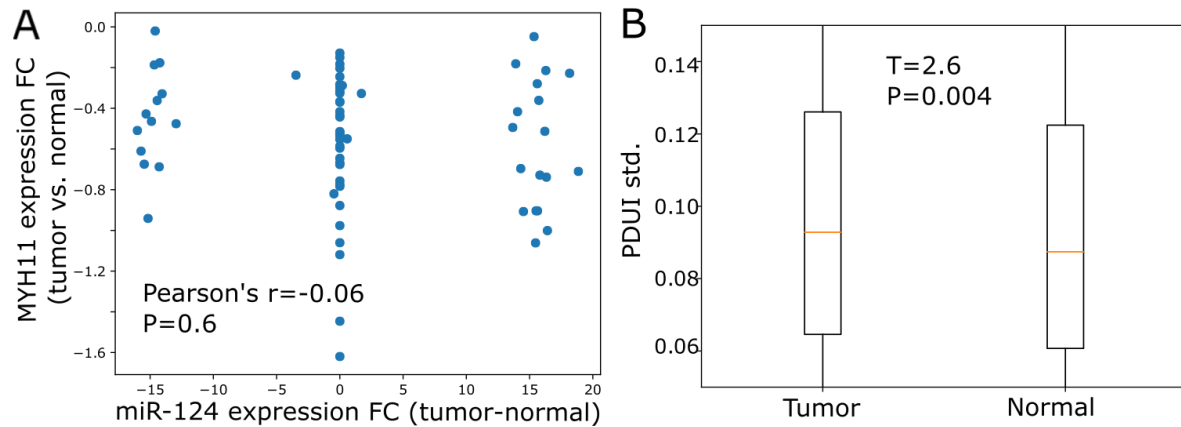
S. Figure 5 A. Scatterplot of MYH11 expression fold change values against expression fold change values of the mediating miRNA (miR-124/124ab/506). **B.** PDUI standard deviation of 2,862 genes with 3′-UTR usage status in tumor and normal whose PDUI values were available for all 70 tumor-normal pairs.

Supplementary Table 1. List of miRNAs with the number of patients in which APA modifies the target sites with the indication of whether the miRNA is found for each category of miRNAs with clinical potential.

Supplementary Table 2. MiEAA analysis result for 99 tamoMiRs and 109 other miRNAs.

Supplementary Table 3. Ingenuity Pathway Analysis report on tamoMiRNA target genes.