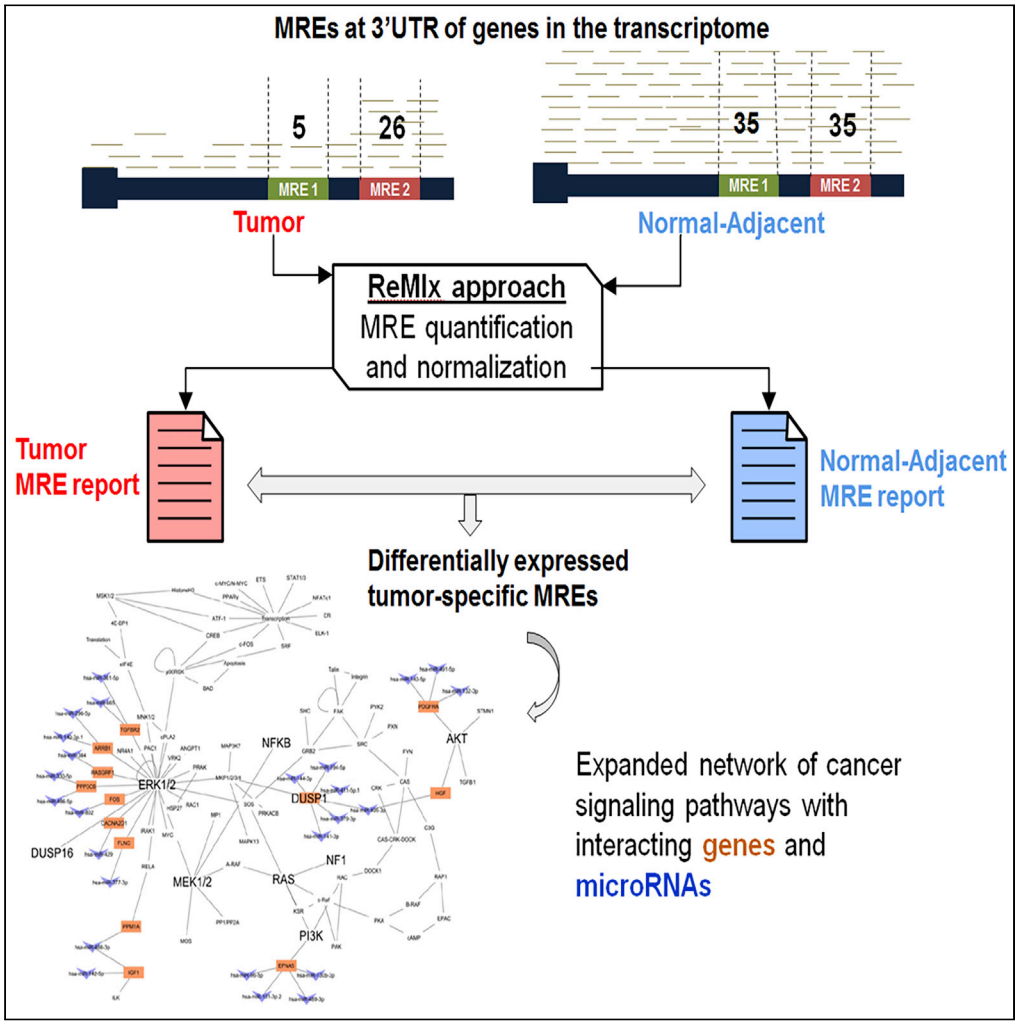
iScience

CellPress
OPEN ACCESS

Article

Frequency of MicroRNA Response Elements Identifies Pathologically Relevant Signaling Pathways in Triple-Negative Breast Cancer



Asha A. Nair, Xiaojia Tang, Kevin J. Thompson, Peter T. Vedell, Krishna R. Kalari, Subbaya Subramanian

kalari.krishna@mayo.edu (K.R.K.)
subree@umn.edu (S.S.)

HIGHLIGHTS

Bioinformatics method ReMIx identify differential microRNA response rlements (MRE)

Tumor-specific MREs frequency observed in triple-negative breast cancer (TNBC)

MRE analysis identify MAPK signaling genes as therapeutic target for TNBC

MREs frequency can be used to identify pathologically relevant pathways

Nair et al., iScience *23*, 101249
June 26, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.101249

iScience

CellPress
OPEN ACCESS

Article

Frequency of MicroRNA Response Elements Identifies Pathologically Relevant Signaling Pathways in Triple-Negative Breast Cancer

Asha A. Nair,[1] Xiaojia Tang,[1] Kevin J. Thompson,[1] Peter T. Vedell,[1] Krishna R. Kalari,[1,*] and Subbaya Subramanian[2,3,4,5,*]

SUMMARY

Complex interactions between mRNAs and microRNAs influence cellular functions. The mRNA-microRNA interactions also determine the post-transcriptional availability of mRNAs and unbound microRNAs. MicroRNAs binds to one or more microRNA response elements (MREs) located on the 3′UTR of mRNAs. In this study, we leveraged MREs and their frequencies in cancer and matched normal tissues to obtain insights into disease-specific interactions between mRNAs and microRNAs. We developed a bioinformatics method "ReMIx" that utilizes RNA sequencing (RNA-Seq) data to quantify MRE frequencies across the transcriptome. We applied ReMIx to triple-negative (TN) breast cancer tumor-normal adjacent pairs and identified MREs specific to TN tumors. ReMIx identified candidate mRNAs and microRNAs in the MAPK signaling cascade. Further analysis of MAPK gene regulatory networks revealed microRNA partners that influence and modulate MAPK signaling. In conclusion, we demonstrate a novel method of using MREs in the identification of functionally relevant mRNA-microRNA interactions in TN breast cancer.

INTRODUCTION

Regulatory interactions between coding and non-coding RNAs in cells determine the post-transcriptional availability of protein-coding mRNA transcripts (Chiang et al., 2010; Eichhorn et al., 2014; Garcia et al., 2011; Guo et al., 2010, 2014; Lee and Jiang, 2017; Rissland et al., 2017; Shin et al., 2010; Volinia and Croce, 2013; Wu and Bartel, 2017). MicroRNAs use seed sequences (6–8 bases long) to bind to microRNA response elements (MREs) predominantly located on the 3′UTRs of mRNAs. mRNAs can have one or more distinct MRE sites, thus being targets to multiple microRNAs. Similarly, microRNAs also bind to MRE sites of several different target genes (Krek et al., 2005; Lim et al., 2005). Thus, alterations in target gene expression via microRNA binding can affect several cellular processes such as cell proliferation and apoptosis during cancer development, progression, and metastasis. Thus, elucidating critical players among the mRNA-microRNA interacting networks can yield novel therapeutic targets and biomarkers in cancers, especially for cancer subtypes that are least responsive to current modalities of treatment.

Expression profiles of microRNAs and mRNAs (Illumina TruSeq libraries enriched for poly(A) RNAs) across many cancer types in The Cancer Genome Atlas (TCGA) were used to infer active and functional microRNA-target interactions in different cancer types (Jacobsen et al., 2013). Alternative polyadenylation of 3′UTRs in bladder cancer can lead to shortened 3′UTR affecting mRNA stability and attenuated protein translation (Han et al., 2018). Studies have also shown that the presence of single nucleotide polymorphisms (SNPs) in the 3′UTR of transcripts can affect microRNA binding and are associated with multiple cancer subtypes (Pelletier and Weidhaas, 2010). Here we extrapolate TCGA RNA sequencing (RNA-Seq) data to analyze MRE sites to obtain insights into unique interactions between mRNAs and microRNAs at the 3′UTRs of the tumor and normal-adjacent datasets.

We developed a new bioinformatics approach called ReMIx (pronounced "remix")—m**R**NA-**M**icroRNA **I**ntegration—which leverages RNA-Seq data to quantify MRE sites at the 3′UTR sequence across the transcriptome. ReMIx profiles MRE sites in tumor and matched normal samples separately, which enables the identification of differential frequency of MREs that are statistically significant in tumor samples. Because

[1]Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

[2]Department of Surgery, University of Minnesota, 420 Delaware St SE, Minneapolis, MN 55455, USA

[3]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

[4]Center for Immunology, University of Minnesota, Minneapolis, MN 55455, USA

[5]Lead Contact

*Correspondence: kalari.krishna@mayo.edu (K.R.K.), subree@umn.edu (S.S.)

https://doi.org/10.1016/j.isci.2020.101249

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

MRE is the interacting link between mRNAs and microRNAs, ReMIx brings together mRNAs with tumor-specific MREs and microRNAs that have the potential to bind to these MRE sites. ReMIx also reports potential mRNA-microRNA candidates that have unique tumor-specific interactions and potential disease-driving functions. This method can be applied to study any cancer type or complex diseases along with their normal tissue sets. To demonstrate the utility of ReMIx, we applied it to the largest RNA-Seq dataset of breast cancer cases and normal-adjacent tissues from TCGA (Cancer Genome Atlas Network, 2012; Ciriello et al., 2015). Using this method, we specifically identified MREs in estrogen receptor positive (ER+), ErbB2 overexpressed–HER2 positive (HER2+), triple-negative tumors, and normal-adjacent tissues. Triple-negative breast cancers (TNBC) are highly heterogeneous and one of the most severe forms of breast cancer subtypes with no targeted treatments currently available. In this study, we applied ReMIx and identified mRNA-microRNA candidates unique to the TNBC and not present in ER + or HER2+ subtypes. Analysis of TNBC data identified MAPK signaling pathway targets as a potential disease driver and target.

RESULTS

ReMIx: An Automated Bioinformatics Approach for MRE Quantification

We developed an innovative bioinformatics approach called ReMIx to quantify the expression of MRE sites at the 3′UTRs of mRNAs using RNA-Seq data. ReMIx uses reads aligned to 3′UTRs of genes in a given transcriptome and scans them for evidence of MRE sequences (see Transparent Methods). All known MREs for genes in the reference genome, as reported by TargetScan—human version 7.0 (Agarwal et al., 2015), are quantified for their level of expression at the 3′UTR of all genes. After quantification, ReMIx normalizes the raw counts of MREs to account for sample library size, 3′UTR length, and 3′UTR GC content per gene. Finally, for every gene and for every conserved microRNA that targets the gene, the normalized MRE counts are reported in a tab-delimited format for each gene-microRNA pair in the transcriptome analyzed. The ReMIx workflow is fully automated and designed to run in a multithreaded cluster environment to analyze paired-end transcriptome samples. A flowchart of the ReMIx approach is shown in Figure 1 (see Transparent Methods).

ReMIx Identified 221 Triple-Negative Breast-Cancer-Specific MRE Sites

The 3′UTR sequences of individual genes (n = 12,455, TargetScan v7.0 (Agarwal et al., 2015)) were obtained using the reference human genome hg19 build. Reads aligned to these 3′UTR sequences were obtained using the TCGA Breast Cancer transcriptome dataset for 13 pairs (Tumor and Normal-Adjacent) from the TNBC subtype, 56 pairs of ER+, and 20 pairs of HER2+ subtypes and were provided as input to the ReMIx workflow (see Transparent Methods). The pre-computed MRE sequences (n = 329, TargetScan 7.0) were also provided as input to ReMIx to count reads mapped to individual MREs located on each gene. The raw MRE counts were then normalized by factoring library size, 3′UTR lengths, and 3′UTR GC content of individual genes. MRE quantification process identified normalized counts of 111,521 MRE sites in tumor and normal adjacent sample sets for each subtype (Data S1, S2, and S3 for TN, ER+, and HER2+, respectively).

Next, ReMIx results were used to identify MRE sites that had unique and significant levels of expression (high or low) in TNBC tumors in comparison to ER + tumors, HER2+ tumors as well as TN, ER+, and HER2+ normal-adjacent cases. The Dunnett-Tukey-Kramer (DTK) pairwise multiple comparison statistical test was applied to the tumor and normal-adjacent cases across all subtypes (six groups in total) to highlight MREs that were unique only to TNBC (p value < 0.05) when compared with other two subtypes and all normal-adjacent cases. This resulted in identifying 614 MREs unique to TNBC (Data S4). In addition, the edgeR bioinformatics package (Robinson et al., 2010) was applied to identify differentially expressed MREs by comparing 13 TN tumor and the respective normal-adjacent cases (FDR <5% and log2FC |2|) and reported 3,053 significant and differentially expressed MREs (Data S5). By adopting the approach of taking the intersection of MREs reported to be statistically significant and differentially expressed by the two complementary approaches, i.e., DTK (n = 614 MREs) and edgeR (n = 3,053 MREs), we identified a common set of 221 TN tumor-specific MRE sites (Figure S1). The 221 TNBC MREs are provided in Data S6. The distinct expression profile of these MRE sites in TNBC with respect to other subtypes and normal-adjacent cases are shown in the heatmap (Figure 2).

TNBC-Relevant MREs Are Associated with 88 mRNAs and 125 microRNAs

The unique feature of MRE is that it is the interactive site between mRNA and microRNA. Hence, for MRE sites of interest, we can decode and obtain information about the mRNA and its interacting microRNA by identifying the relevant MREs and decoupling them into their respective mRNA and microRNA pairs. Thus, for the TNBC tumor-specific MRE sites, we deciphered such information for the 221 MREs and obtained a total of 88 mRNAs and 125

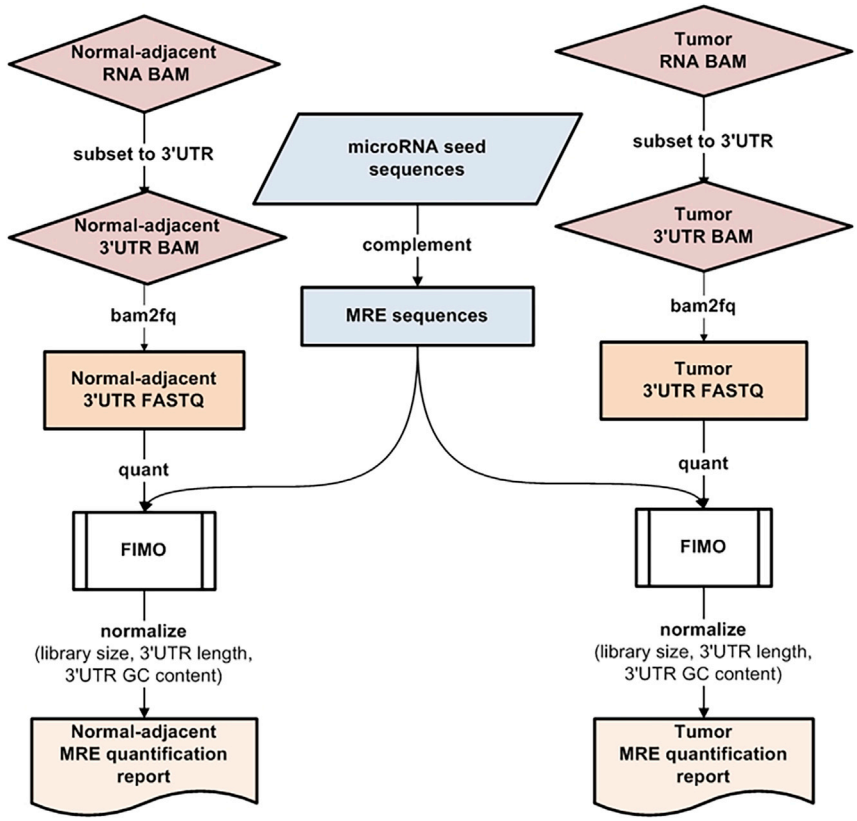


Figure 1. Flowchart of MRE Frequency Quantification from RNA-Seq BAM

The RNA-Seq BAM is subset to 3′UTRs of all genes, converted to FASTQ and processed through FIMO to obtain raw MRE counts per microRNA for every target gene. The raw MRE counts were normalized to account for library size, 3′UTR length, and 3′UTR GC content, and individual tumor and normal-adjacent quantification reports are generated. See also Data S1, S2, and S3.

microRNAs. Tables listing 88 mRNAs and 125 microRNAs along with their expression levels in TNBC are provided in Data S7 and S8, respectively. With these ReMIx analyses, we deduce that over half of the mRNAs (48 out of 88) were used repeatedly and these mRNAs had multiple MREs that were used as interactive sites by different microRNAs.

Unsupervised hierarchical clustering of 221 MREs based on expression profiles of these MREs showed that the TN cases clustered within their tumor and normal-adjacent groups. The differential expression pattern for 221 MREs are shown in Figure 3. Notably, unsupervised clustering of the corresponding 88 mRNAs and 125 microRNAs also showed a separation of TNBC into tumor and normal-adjacent groups (Figure 3).

Next, we evaluated the mRNAs and microRNAs identified by ReMIx using a standalone approach to analyze their predominance in terms of differential expression within the respective RNA and microRNA expression datasets of 13 TNBC tumor and normal-adjacent pairs and identified canonical pathways that were associated with 88 mRNAs and 125 microRNAs.

The differential expression analysis using RNA-Seq data for 13 TNBC tumor and normal-adjacent pairs showed that a total of 2,250 genes were differentially expressed (edgeR package (Robinson et al., 2010); statistical significance threshold at FDR <5% and log2FC |2|). Notably, out of the 88 mRNAs identified by the ReMIx analysis, we found that 68 (77%) were also differentially expressed at the gene level between TNBC cases. This indicated a high likelihood of microRNA-mediated gene expression regulation resulting in their differential expression in TNBC tumors compared with their normal-adjacent counterparts. Notably, these 68 mRNAs found in the 13 paired TNBC cases were also consistently differential expressed in a larger cohort of 120 TCGA-TNBC and 13 normal-adjacent samples (Figure S2). This suggests that mRNA expression observed in a smaller sample size is potentially reflective of mRNA expression in a larger

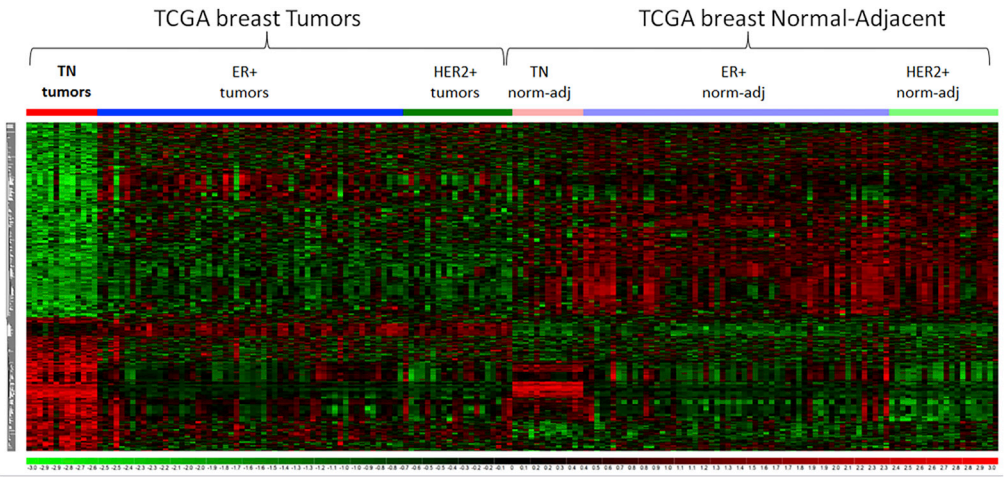


Figure 2. Heatmap of 221 TN Tumor-Specific MREs

The normalized conditional quantile normalization (CQN) values of 221 MREs were obtained for TN, ER+, and HER2+ tumors and normal-adjacent (norm-adj) cases. As shown in the heatmap, these MREs have a distinct expression in TN tumors in comparison to the other subtypes as well as TN normal-adjacent cases. See also Figure S1; Data S1, S2, S3, S4, S5, S6, and S7.

cohort. Out of 68 mRNAs, 41 had multiple MREs targeted by different microRNAs. A table listing these 68 mRNAs that are both differentially expressed and have interacting MRE sites are given in Data S9.

Next, using the microRNA expression data for 13 TNBC pairs, we found that out of a total of 2,245 microRNAs that were quantified for expression in the tumor and normal-adjacent cases, 778 microRNAs were differentially expressed in tumors (limma package (Ritchie et al., 2015); adjusted p value < 0.05). Examining the number of microRNAs identified by ReMIx that were also differentially expressed between TNBC tumor and normal-adjacent, we found that 64 out of 125 microRNAs (51%) were statistically different in expression (FDR <5%). A table and heatmap listing 64 microRNAs that are both differentially expressed and participate in MRE-mediated gene expression regulation can be found in the Data S10 and Figure S3.

Finally, using RNA-Seq and microRNA differential expression results, the magnitude and direction of change for 221 MREs and their associated genes and microRNAs in 13 TNBC tumor and normal-adjacent pairs were combined (Data S11). We observed that the majority of MREs follow the direction as their parent genes, with very few exceptions, likely due to the nature of TNBC sequencing libraries (Illumina TruSeq).

MRE-Associated 125 microRNAs Are Implicated in TN Breast Carcinoma

Further analysis of 125 microRNAs using the TAM 2.0 tool for microRNA set enrichment analysis revealed that these microRNAs were associated with cancer pathways as shown in Table S1. Specifically, 14 out of 125 microRNAs are also reported in other TNBC studies and are upregulated with an FDR $<2.87 \times 10^{-5}$. Similarly, 55/125 microRNAs are reported in breast carcinoma studies (FDR $<8.18 \times 10^{-18}$) and 34/125 in breast neoplasms (FDR $<6.12 \times 10^{-13}$). Information of these microRNAs are provided in Table S1.

Pathway Analysis of 88 Genes Identified MAPK Signaling Pathway

Eighty-eight genes obtained from ReMIx were analyzed to identify their associated signaling pathways. Using gene set enrichment analysis (GSEA) (Subramanian et al., 2005) on KEGG and REACTOME databases, the mitogen-activated protein kinase (MAPK) signaling cascade was identified among the top significant pathways. In addition, application of the signaling pathway impact analysis (SPIA) package also confirmed that the MAPK signaling pathway was activated in TN tumors. The GSEA and SPIA pathway results are provided in Data S12 and S13, respectively.

Further examination of genes in the MAPK pathway was conducted by juxtaposing the expression of these genes, obtained from RNA-Seq data of TNBC with the KEGG-based network of the MAPK pathway. Our analysis revealed that oncogenes *KRAS, NRAS, AKT*, and *NFKB* were notably activated and tumor

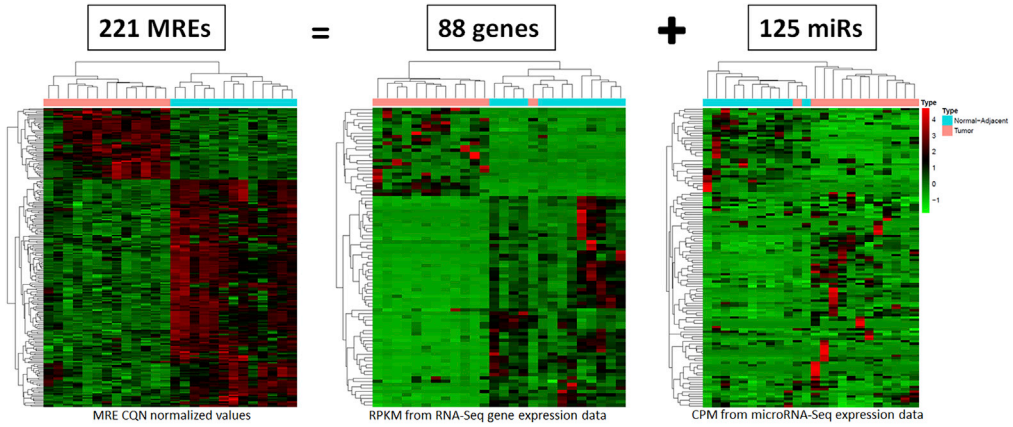


Figure 3. Unsupervised Clustering and Heatmap Representation of 221 TN Tumor-Specific MREs and Their Associated mRNAs and microRNAs

The 221 MREs were associated with 88 mRNAs and 125 microRNAs. Conditional quantile normalization values were obtained for the MRE sites. Reads per kilobase per million mapped reads (RPKM) normalized values from RNA-Seq and counts per million (CPM) normalized values from microRNA-Seq were obtained for the 13 pairs of TN tumor and normal-adjacent cases. Unsupervised clustering of the cases indicates that tumor (pink) and normal-adjacent (blue) were clustered well within the corresponding groups. See also Figure S1; Data S7, S8, S9, S10, and S11.

suppressor *PTEN* was repressed. Figure 4 illustrates the KEGG pathview for the MAPK signaling cascade. MAPK signaling pathway is an extensive cascade with connections to several biological pathways downstream such as proliferation, cell cycle, glycolysis, apoptosis, and protein synthesis.

Expanded MAPK Endogenous RNA Network including TNBC-Specific mRNA-microRNA Candidates

Based on MRE results from the ReMIx, we investigated relevant genes that were associated with MAPK signaling in TN tumors. We found 12 out of 294 genes (~4%) in MAPK pathway (https://www.genome.jp/dbget-bin/get_linkdb?-t+genes+path:hsa04010) that have MRE sites with the potential of differential binding of microRNAs. These 12 mRNAs with tumor-specific MRE sites and microRNAs with the potential to bind to these sites are provided in Table 1. Next, we expanded the MAPK gene network in TN tumors by including interacting microRNAs that are essential members of the pathway. Figure 5 shows the MAPK endogenous RNA network that represents the genes identified by ReMIx, interacting microRNAs, and other mRNAs that are likely to interact with each other and regulate expressions of key genes, such as *PI3K, AKT, RAS, NFKB,* and *PTEN*. Furthermore, *ERK1/2*, critical genes in the MAPK/ERK signaling cascade, were also directly associated with 7 of the 12 genes (Figure 5). Taken together, we present an expanded network of MAPK signaling cascade and provide a list of potential mRNA-microRNA candidates that interact with each other and could potentially be therapeutic targets for TN tumors.

DISCUSSION

The regulatory interactions between non-coding and protein-coding RNAs have been well recognized, where the mRNA-microRNA interactions are widely studied. Although there are several microRNA target prediction tools such as TargetScan (Agarwal et al., 2015), miRBase (Kozomara et al., 2019), DIANA (Vlachos et al., 2012), PicTar (Krek et al., 2005), miRwayDB (Das et al., 2018), miRanda (Betel et al., 2008), PITA (Kertesz et al., 2007), RNA22 (Loher and Rigoutsos, 2012), and miRTar (Hsu et al., 2011), not many computational tools have been developed that enable the integration of mRNA and microRNA expression datasets. MAGIA is a web-based tool for microRNA and gene integrated analysis that brings together target predictions and gene expression profiles using different functional measures for both matched and unmatched samples (Sales et al., 2010). The tool miRmapper uses mRNA-microRNA predictions and a list of differentially expressed mRNAs to identify top microRNAs and recognizes similarities between microRNAs based on commonly regulated mRNAs (da Silveira et al., 2018). HisCoM-mimi is a hierarchically structured component analysis method that models biological relationships as structured components to efficiently yield integrated mRNA-microRNA markers (Kim et al., 2018). These tools use prior knowledge of microRNA target predictions and are developed using unique methodologies to derive mRNA-microRNA interactions. Furthermore, tools such as miRmapper have the ability to

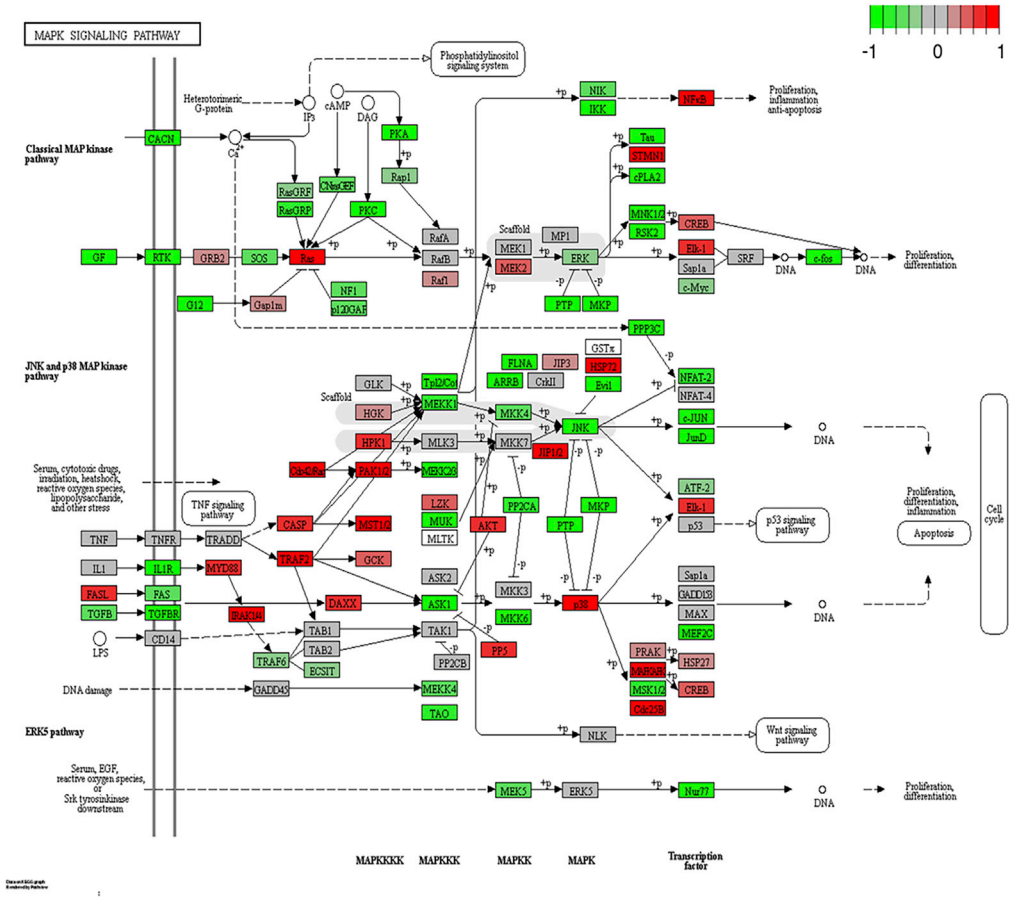


Figure 4. MAPK Signaling Pathway

Genes in the pathway are colored based on their expression in TN tumors. Oncogenes NFKB and AKT are activated in this pathway. See also Data S12 and S13.

highlight key microRNAs based on the number of connections it possesses in a given network. However, the underlying methodologies of all these tools are to use the expression of either mRNAs alone or both mRNAs and microRNAs to model their correlation and derive mRNA-microRNA relationships.

With the advent of RNA-Seq technology, profiling of the transcriptome is now possible at the base-prevision level. It is a known fact that microRNAs predominantly bind to the 3′UTRs of mRNAs to induce their regulatory effects and thereby impact mRNA expression and protein translation. Studies have shown that shortening of 3′UTR is a frequent phenomenon in cancer to evade oncogenes from microRNA suppression (Xue et al., 2018), repress tumor suppressor genes (Park et al., 2018), and enhance metastatic burden (Andres et al., 2019). Therefore, it is important not only to know which mRNAs are differentially expressed between a tumor and normal pair but also to determine which integration sites or microRNA response elements (MREs) are available along the 3′UTRs of the tumor mRNAs. Identification of MREs that are either present/absent/highly expressed/low expressed in the tumor can provide mechanistic insights of tumor progression. Although mRNA-microRNA integration tools exist, and may be applied to the tumor and normal datasets, no tools, to our knowledge, have the ability to precisely report mRNA-microRNA interactions that are solely based on the availability of MREs at the 3′UTRs. MREs are short 6–8 base segments and without appropriate bioinformatics methods, screening RNA-Seq data for MRE sites can yield highly non-specific and erroneous results. This could be a possible reason why this simple, but highly relevant, concept has not been explored to date.

In this study, we developed an innovative bioinformatics method "ReMIx" that uses RNA-Seq data to identify and quantify microRNA-binding sites (known as microRNA response elements [MREs]) at 3′UTRs. A hypothetical example of this approach is illustrated in Figure 6. We applied ReMIx to TCGA-paired tumors and normal-adjacent

mRNAs	microRNAs
MAPK Signaling Pathway	
CACNA2D1	hsa-miR-429
PPP3CB	hsa-miR-330-5p; hsa-miR-486-5p
RASGRF1	hsa-miR-384
IGF1	hsa-miR-142-5p; hsa-miR-488-3p
HGF	hsa-miR-495-3p
EFNA5	hsa-miR-101-3p.2; hsa-miR-130b-3p; hsa-miR-489-3p; hsa-miR-96-5p
PDGFRA	hsa-miR-132-3p; hsa-miR-140-5p; hsa-miR-491-5p
FOS	hsa-miR-802
TGFBR2	hsa-miR-361-5p; hsa-miR-665
FLNC	hsa-miR-377-3p
ARRB1	hsa-miR-140-3p.1; hsa-miR-296-5p
PPM1A	hsa-miR-488-3p

Table 1. Gene-microRNA Pairs with Distinct TN-Specific MRE Sites that Are Part of the MAPK Pathway
Table lists genes that are a subset of the 88 genes obtained from ReMIx and that are members of the MAPK signaling pathway. The microRNAs that bind to the MRE sites that were found to have distinct counts in TN tumors are also provided. Related to Figure 5.

breast cancer cases for TN, ER+, and HER2+ subtypes. Using two complementary statistical approaches, we identified 221 MRE sites that have a distinct expression in TN tumor-normal adjacent pairs. Upon decoupling, we found that the 221 MRE sites corresponded to 88 mRNAs and 125 microRNAs. By reviewing fold-changes of these MREs, mRNAs, and microRNAs, we observed that most of the MREs followed the same direction as their parent gene transcript. We postulate that this was likely driven by the sequencing library preparation kit (Illumina TruSeq). However, we also found MREs with the opposite trend, suggesting an alternative 3′UTR mechanism. Furthermore, we found mRNAs and MREs with positive expression in TNBC tumors but repressed microRNAs, likely denoting the effect of competing endogenous RNAs (ceRNAs) on the microRNAs. Canonical pathway analysis of both mRNAs and microRNAs revealed cancer-related pathways specific to breast cancer. Significantly, miR-27a and miR-143 were associated with breast cancer and TNBC, respectively (Jiang et al., 2018) (Deng et al., 2018). Upregulation of miR-27a induced epithelial-to-mesenchymal transition and increased cell migration in breast cancer (Jiang et al., 2018). Also, miR-143-3p was implicated in drug resistance; overexpression of miR-143-3p inhibits cytokine-induced apoptosis inhibitor 1 (CIAPIN1), enhancing the sensitivity of drug-resistant TNBC cells (Deng et al., 2018). Specifically, mRNAs revealed by ReMIx signified the MAPK signaling cascade in TNBC. The mRNAs determined by ReMIx represented about 4% of gene members in the MAPK signaling pathway. Based on TNBC-specific results reported by ReMIx, we expanded the MAPK endogenous RNA network by including mRNAs with TNBC-specific MRE sites. Further we also included the corresponding microRNAs that have the potential to bind to these MRE sites and other protein-coding RNAs in the network that have the ability to interact with each other and regulate expression of primary oncogenes and tumor suppressors in the MAPK signaling pathway. Based on the results, we provide a list of potential mRNA-microRNA candidates that interact with each other at the network level of the MAPK signaling cascade and could be possible therapeutic targets for TN tumors.

Because the TCGA breast cancer cases, including TNBC, have whole-exome sequencing data available, we sought to check for any copy number variant (CNV) events at the 3′UTRs. We also examined whether

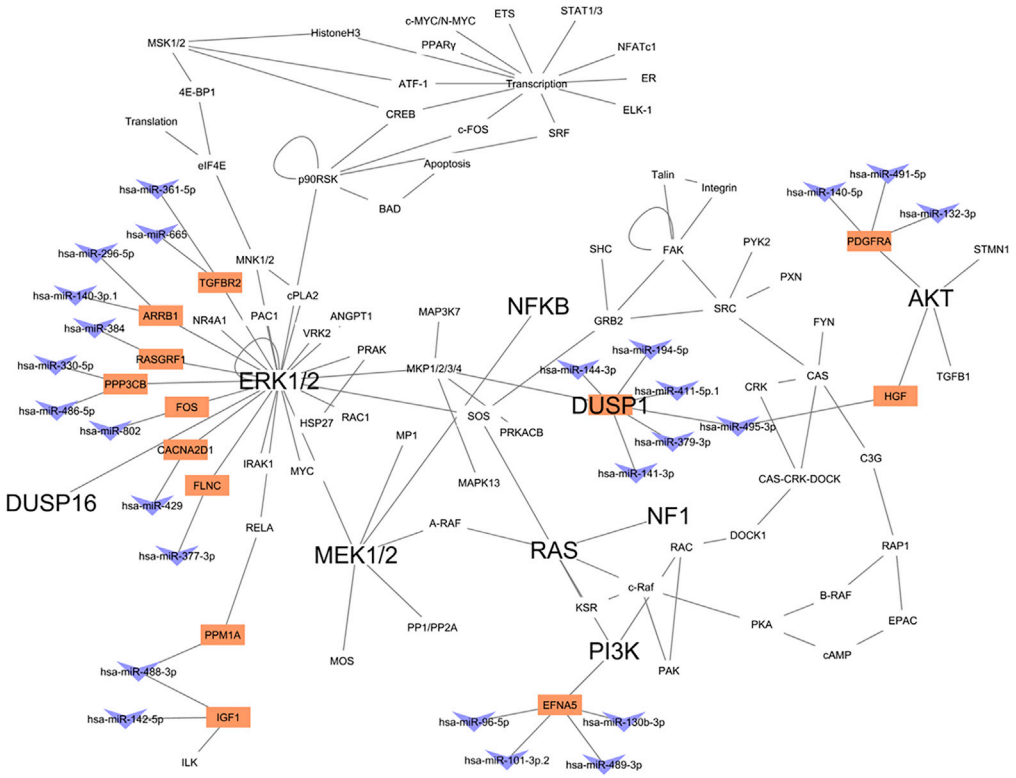


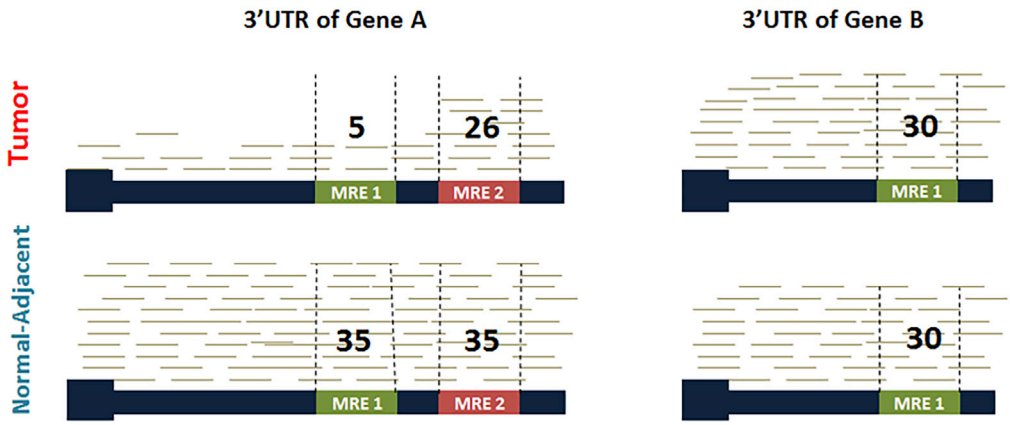
Figure 5. MAPK Endogenous RNA Network

This figure shows the network of interacting protein-coding mRNAs and non-coding microRNAs in the MAPK singling pathway. The mRNAs and microRNAs reported by ReMIx are represented in colors orange and blue, respectively. Oncogenes AKT, RAS, NFKB, PI3K, ERK, and MEK are shown to interact either directly or indirectly with the mRNA-microRNA candidates. See also Table 1.

differentially expressed mRNAs, specifically those found by ReMIx, were related to CNV events. The metadata for exome capture kit for the 26 TNBC tumor and normal-adjacent cases was accessed using the NCI-GDC API (NCI Genomic Data Commons Application Programming Interface). For the majority of cases (18 out of 26), the "hg18nimblegenexomeversion2" capture kit was used. For the remaining 8 out of 26, "NimblegenEZExomev3.0" and "SureSelectHumanAllExon38Mbv2" capture kits were utilized for four cases each. Upon checking the overlap of these exome capture kits to 3′UTRs, we found that these kits do not cover the 3′UTR. We only observed a 0.4% overlap of 3′UTRs with "hg18nimblegenexomeversion2," 13.8% with "NimblegenEZExomev3.0," and 1.8% with "SureSelectHumanAllExon38Mbv2." As a result, we could not verify the role of CNV on the TNBC cases for this study.

Recently, there has been an increasing focus to explore and identify therapeutic strategies to better treat and improve survival of TNBC patients. Activation of the MAPK pathway has been implicated in the proliferation and survival of cancer cells (Saini et al., 2013). Previous studies have shown this pathway to be highly prevalent in TN breast cancer as opposed to other breast cancer subtypes (Balko et al., 2012; Hashimoto et al., 2014; Hoeflich et al., 2009), thus supporting our findings. Studies have also shown that activation of MAPK pathway significantly correlates with disease progression in TN tumors (Eralp et al., 2008; Gholami et al., 2014; Giltnane and Balko, 2014; Hashimoto et al., 2014; Loi et al., 2016; Qi et al., 2015). MAPK pathway is a sequentially activated cascade consisting of key genes such as Ras, Raf, MEK, and ERK. Activation of Ras leads to the phosphorylation of Raf, thereby promoting the activation of MEK and ERK downstream and finally results in tumor proliferation and cell survival.

In conclusion, we demonstrate a novel method of using MREs in the identification of functionally relevant mRNA-microRNA interactions that can be potential targets in TNBC. Further, experimental validations of these interactions are warranted in developing novel therapeutic targets.



Gene	MRE sites	Tumor MRE counts	Normal-adjacent MRE counts	Differential MRE counts?
Gene A	MRE 1	5	35	Yes
Gene A	MRE 2	26	35	No
Gene B	MRE 1	30	30	No

Figure 6. Hypothetical Representation of MRE Frequency Counting Using RNA-Seq Data

The example shows a tumor and normal-adjacent sample with reads mapped to the 3′UTRs of two genes, Gene A and Gene B that consist of 2 and 1 MRE sites, respectively, with a common site (MRE 1). Reads that align with individual MRE sites (vertical dotted lines) are quantified. For every MRE site that belongs to a gene, the counts are then statistically evaluated between tumor and normal-adjacent for evidence of differential frequency (as shown in the inset table).

Limitations of the Study

One of the limitations of this study is that although ReMIx enables identification of candidate mRNA and microRNA players via MRE analysis using RNA-Seq data, this does not establish the fact that the identified microRNAs are indeed present and expressed in the particular disease, in this case, TN tumors. ReMIx results only confirm that the sites on 3′UTR of mRNAs show distinct expression profiles in tumor and thus have the potential to be regulated by microRNAs in a disease-specific manner. To complement these results, microRNA expression profiles can be used to validate the existence of microRNAs and to check for expression correlation with the corresponding mRNA target(s) identified by ReMIx. This study is limited to microRNA-mediated interactions; however, several other mechanisms modulate gene expression. At the post-transcriptional level, the interplay of other noncoding RNAs, such as long non-coding RNAs, circular RNAs, and pseudogenes can collectively form the ceRNA network and compete with protein-coding genes for microRNA binding, thereby influencing their ultimate impact on gene expression. Also, during transcription, structural and chemical changes such as histone acetylation to determine the accessibility of chromatin domains, and DNA methylation to silence genes, are well-established modes of regulation, especially in cancer.

Resource Availability

Lead Contact

Further information and requests for resources, code, and scripts should be directed to and will be fulfilled by the Lead Contact, Subbaya Subramanian (subree@umn.edu) or Asha Nair (Nair.Asha@mayo.edu).

Materials Availability

This study did not generate new unique reagents.

The 26 samples (13 tumor and normal-adjacent pairs) of the TCGA TNBC cohort for which REMIx results were obtained in this study are: TCGA-BH-A0B3-01A-11R-A056-07,TCGA-BH-A0B3-11B-21R-A089-

07,TCGA-BH-A0BW-01A-11R-A115-07,TCGA-BH-A0BW-11A-12R-A115-07,TCGA-BH-A0E0-01A-11R-A056-07,TCGA-BH-A0E0-11A-13R-A089-07,TCGA-BH-A18Q-01A-12R-A12D-07,TCGA-BH-A18Q-11A-34R-A12D-07,TCGA-BH-A18V-01A-11R-A12D-07,TCGA-BH-A18V-11A-52R-A12D-07,TCGA-BH-A1EW-01A-11R-A137-07,TCGA-BH-A1EW-11B-33R-A137-07,TCGA-BH-A1F6-01A-11R-A13Q-07,TCGA-BH-A1F6-11B-94R-A13Q-07,TCGA-BH-A1FC-01A-11R-A13Q-07,TCGA-BH-A1FC-11A-32R-A13Q-07,TCGA-E2-A158-01A-11R-A12D-07,TCGA-E2-A158-11A-22R-A12D-07,TCGA-E2-A1L7-01A-11R-A144-07,TCGA-E2-A1L7-11A-33R-A144-07,TCGA-E2-A1LH-01A-11R-A14D-07,TCGA-E2-A1LH-11A-22R-A14D-07,TCGA-E2-A1LS-01A-12R-A157-07,TCGA-E2-A1LS-11A-32R-A157-07,TCGA-GI-A2C9-01A-11R-A21T-07,TCGA-GI-A2C9-11A-22R-A21T-07.

Data and Code Availability

The code for the ReMIx workflow is available through GitHub at https://github.com/nairasha/ReMIx.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101249.

ACKNOWLEDGMENTS

This work is supported by the Center for Individualized Medicine, Mayo Clinic (CIM) and by the Mayo Clinic Breast Specialized Program of Research Excellence (SPORE) (P50CA116201) at the Mayo Clinic in Rochester, MN, and funds from the University of Minnesota, Medical School Innovation award and the Department of Surgery, Research funds. We also thank Ce Yuan and the anonymous reviewers for their input and insightful comments that helped improve the quality of our manuscript. We thank Dr. Matthew Robertson and Ms. Sruthi Subramanian for assistance with manuscript editing.

AUTHOR CONTRIBUTIONS

SS conceived the idea. SS and KK designed the study. AN performed the analyses. KK and SS supervised and helped AN interpret the data. XT, PV, and KT helped with subtyping the TCGA breast cancer RNA-Seq data and provided useful suggestions regarding this study. AN, KK, and SS wrote the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 15, 2019
Revised: March 24, 2020
Accepted: June 3, 2020
Published: June 26, 2020

REFERENCES

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*, e05005.

Andres, S.F., Williams, K.N., Plesset, J.B., Headd, J.J., Mizuno, R., Chatterji, P., Lento, A.A., Klein-Szanto, A.J., Mick, R., Hamilton, K.E., et al. (2019). IMP1 3′ UTR shortening enhances metastatic burden in colorectal cancer. Carcinogenesis *40*, 569–579.

Balko, J.M., Miller, T.W., Morrison, M.M., Hutchinson, K., Young, C., Rinehart, C., Sanchez, V., Jee, D., Polyak, K., Prat, A., et al. (2012). The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. Proc. Natl. Acad. Sci. U S A *109*, 221–226.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. (2008). The microRNA.org resource: targets and expression. Nucleic Acids Res. *36*, D149–D153.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. *24*, 992–1009.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al. (2015). Comprehensive molecular portraits of invasive lobular breast cancer. Cell *163*, 506–519.

da Silveira, W.A., Renaud, L., Simpson, J., Glen, W.B., Jr., Hazard, E.S., Chung, D., and Hardiman, G. (2018). miRmapper: a tool for interpretation of miRNA(-)mRNA interaction networks. Genes (Basel) *9*, 458.

Das, S.S., Saha, P., and Chakravorty, N. (2018). miRwayDB: a database for experimentally validated microRNA-pathway associations in pathophysiological conditions. Database (Oxford) *2018*, bay023.

Deng, Y.W., Hao, W.J., Li, Y.W., Li, Y.X., Zhao, B.C., and Lu, D. (2018). Hsa-miRNA-143-3p reverses multidrug resistance of triple-negative breast cancer by inhibiting the expression of its target protein cytokine-induced apoptosis inhibitor 1 in vivo. J. Breast Cancer *21*, 251–258.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol. Cell *56*, 104–115.

Eralp, Y., Derin, D., Ozluk, Y., Yavuz, E., Guney, N., Saip, P., Muslumanoglu, M., Igci, A., Kucucuk, S., Dincer, M., et al. (2008). MAPK overexpression is associated with anthracycline resistance and increased risk for recurrence in patients with triple-negative breast cancer. Ann. Oncol. *19*, 669–674.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146.

Gholami, S., Chen, C.H., Gao, S., Lou, E., Fujisawa, S., Carson, J., Nnoli, J.E., Chou, T.C., Bromberg, J., and Fong, Y. (2014). Role of MAPK in oncolytic herpes viral therapy in triple-negative breast cancer. Cancer Gene Ther. *21*, 283–289.

Giltnane, J.M., and Balko, J.M. (2014). Rationale for targeting the Ras/MAPK pathway in triple-negative breast cancer. Discov. Med. *17*, 275–283.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, L., Zhao, Y., Yang, S., Zhang, H., and Chen, F. (2014). Integrative analysis of miRNA-mRNA and miRNA-miRNA interactions. Biomed. Res. Int. *2014*, 907420.

Han, S., Kim, D., Shivakumar, M., Lee, Y.J., Garg, T., Miller, J.E., Kim, J.H., Kim, D., and Lee, Y. (2018). The effects of alternative splicing on miRNA binding sites in bladder cancer. PLoS One *13*, e0190708.

Hashimoto, K., Tsuda, H., Koizumi, F., Shimizu, C., Yonemori, K., Ando, M., Kodaira, M., Yunokawa, M., Fujiwara, Y., and Tamura, K. (2014). Activated PI3K/AKT and MAPK pathways are potential good prognostic markers in node-positive, triple-negative breast cancer. Ann. Oncol. *25*, 1973–1979.

Hoeflich, K.P., O'Brien, C., Boyd, Z., Cavet, G., Guerrero, S., Jung, K., Januario, T., Savage, H., Punnoose, E., Truong, T., et al. (2009). In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. Clin. Cancer Res. *15*, 4649–4664.

Hsu, J.B., Chiu, C.M., Hsu, S.D., Huang, W.Y., Chien, C.H., Lee, T.Y., and Huang, H.D. (2011). miRTar: an integrated system for identifying miRNA-target interactions in human. BMC Bioinformatics *12*, 300.

Jacobsen, A., Silber, J., Harinath, G., Huse, J.T., Schultz, N., and Sander, C. (2013). Analysis of microRNA-target interactions across diverse cancer types. Nat. Struct. Mol. Biol. *20*, 1325–1332.

Jiang, G., Shi, W., Fang, H., and Zhang, X. (2018). miR27a promotes human breast cancer cell migration by inducing EMT in a FBXW7dependent manner. Mol. Med. Rep. *18*, 5417–5426.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. *39*, 1278–1284.

Kim, Y., Lee, S., Choi, S., Jang, J.Y., and Park, T. (2018). Hierarchical structural component modeling of microRNA-mRNA integration analysis. BMC Bioinformatics *19*, 75.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. Nat. Genet. *37*, 495–500.

Lee, S., and Jiang, X. (2017). Modeling miRNA-mRNA interactions that cause phenotypic abnormality in breast cancer patients. PLoS One *12*, e0182666.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769–773.

Loher, P., and Rigoutsos, I. (2012). Interactive exploration of RNA22 microRNA target predictions. Bioinformatics *28*, 3322–3323.

Loi, S., Dushyanthen, S., Beavis, P.A., Salgado, R., Denkert, C., Savas, P., Combs, S., Rimm, D.L., Giltnane, J.M., Estrada, M.V., et al. (2016). RAS/MAPK activation is associated with reduced tumor-infiltrating lymphocytes in triple-negative breast cancer: therapeutic cooperation between MEK and PD-1/PD-L1 immune checkpoint inhibitors. Clin. Cancer Res. *22*, 1499–1509.

Park, H.J., Ji, P., Kim, S., Xia, Z., Rodriguez, B., Li, L., Su, J., Chen, K., Masamha, C.P., Baillat, D., et al. (2018). 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. Nat. Genet. *50*, 783–789.

Pelletier, C., and Weidhaas, J.B. (2010). MicroRNA binding site polymorphisms as biomarkers of cancer risk. Expert Rev. Mol. Diagn. *10*, 817–829.

Qi, X., Yin, N., Ma, S., Lepp, A., Tang, J., Jing, W., Johnson, B., Dwinell, M.B., Chitambar, C.R., and Chen, G. (2015). p38gamma MAPK is a therapeutic target for triple-negative breast cancer by stimulation of cancer stem-like cell expansion. Stem Cells *33*, 2738–2747.

Rissland, O.S., Subtelny, A.O., Wang, M., Lugowski, A., Nicholson, B., Laver, J.D., Sidhu, S.S., Smibert, C.A., Lipshitz, H.D., and Bartel, D.P. (2017). The influence of microRNAs and poly(A) tail length on endogenous mRNA-protein complexes. Genome Biol. *18*, 211.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Saini, K.S., Loi, S., Azambuja, E., Saini, M.L., Ignatiadis, M., Dancey, J.E., and Piccart-Gebhart, M.J. (2013). Targeting the PI3K/AKT/mTOR and Raf/MEK/ERK pathways in the treatment of breast cancer. Cancer Treat Rev *39*, 935–946.

Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. (2010). MAGIA, a web-based tool for miRNA and genes integrated analysis. Nucleic Acids Res. *38*, W352–W359.

Shin, C., Nam, J.W., Farh, K.K., Chiang, H.R., Shkumatava, A., and Bartel, D.P. (2010). Expanding the microRNA targeting code: functional sites with centered pairing. Mol. Cell *38*, 789–802.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Vlachos, I.S., Kostoulas, N., Vergoulis, T., Georgakilas, G., Reczko, M., Maragkakis, M., Paraskevopoulou, M.D., Prionidis, K., Dalamagas, T., and Hatzigeorgiou, A.G. (2012). DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. Nucleic Acids Res. *40*, W498–W504.

Volinia, S., and Croce, C.M. (2013). Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. Proc. Natl. Acad. Sci. U S A *110*, 7413–7417.

Wu, X., and Bartel, D.P. (2017). Widespread influence of 3'-end structures on mammalian mRNA processing and stability. Cell *169*, 905–917.e911.

Xue, Z., Warren, R.L., Gibb, E.A., MacMillan, D., Wong, J., Chiu, R., Hammond, S.A., Yang, C., Nip, K.M., Ennis, C.A., et al. (2018). Recurrent tumor-specific regulation of alternative polyadenylation of cancer-related genes. BMC Genomics *19*, 536.

iScience, Volume *23*

Supplemental Information

Frequency of MicroRNA Response Elements Identifies Pathologically Relevant Signaling Pathways in Triple-Negative Breast Cancer

Asha A. Nair, Xiaojia Tang, Kevin J. Thompson, Peter T. Vedell, Krishna R. Kalari, and Subbaya Subramanian

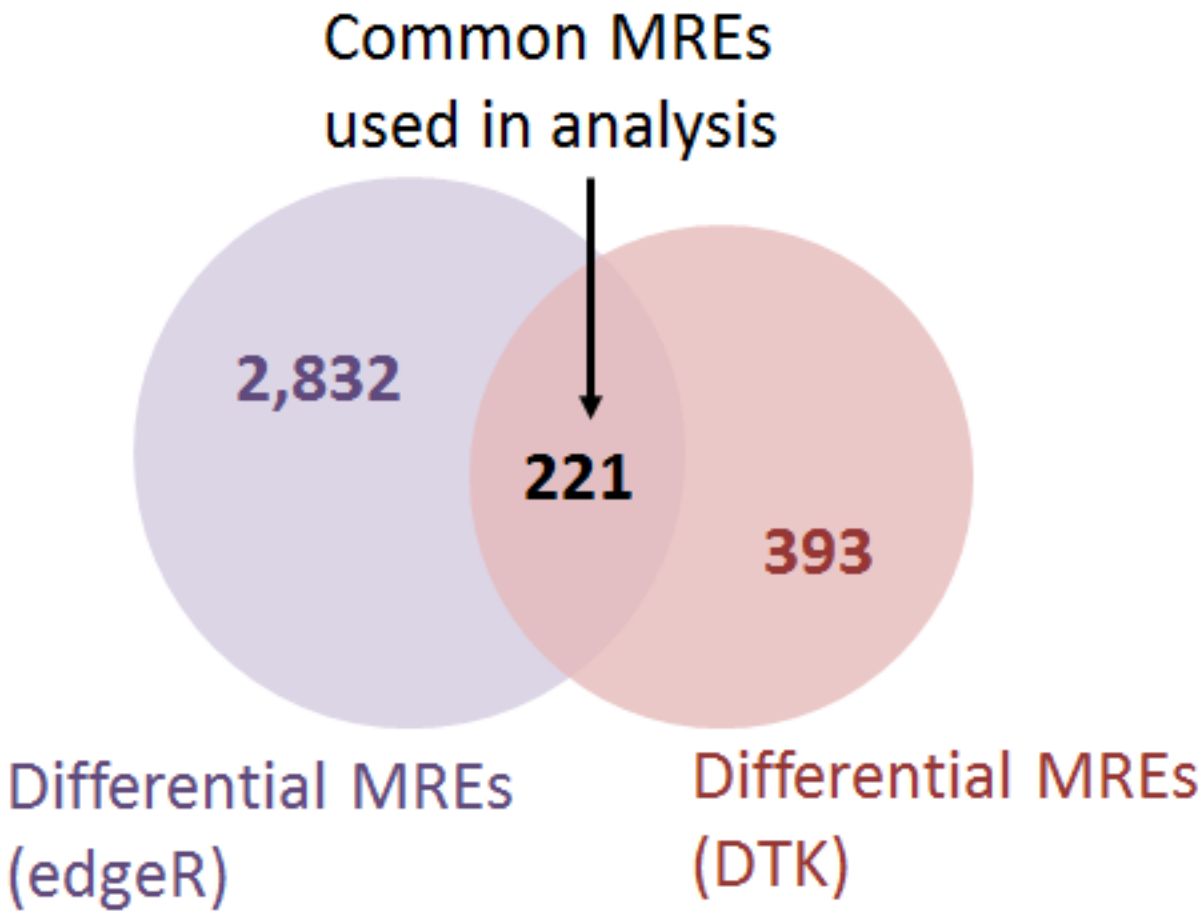


Figure S1: Common set of 221 TN tumor-specific MRE sites. Related to Figures 2 and 3

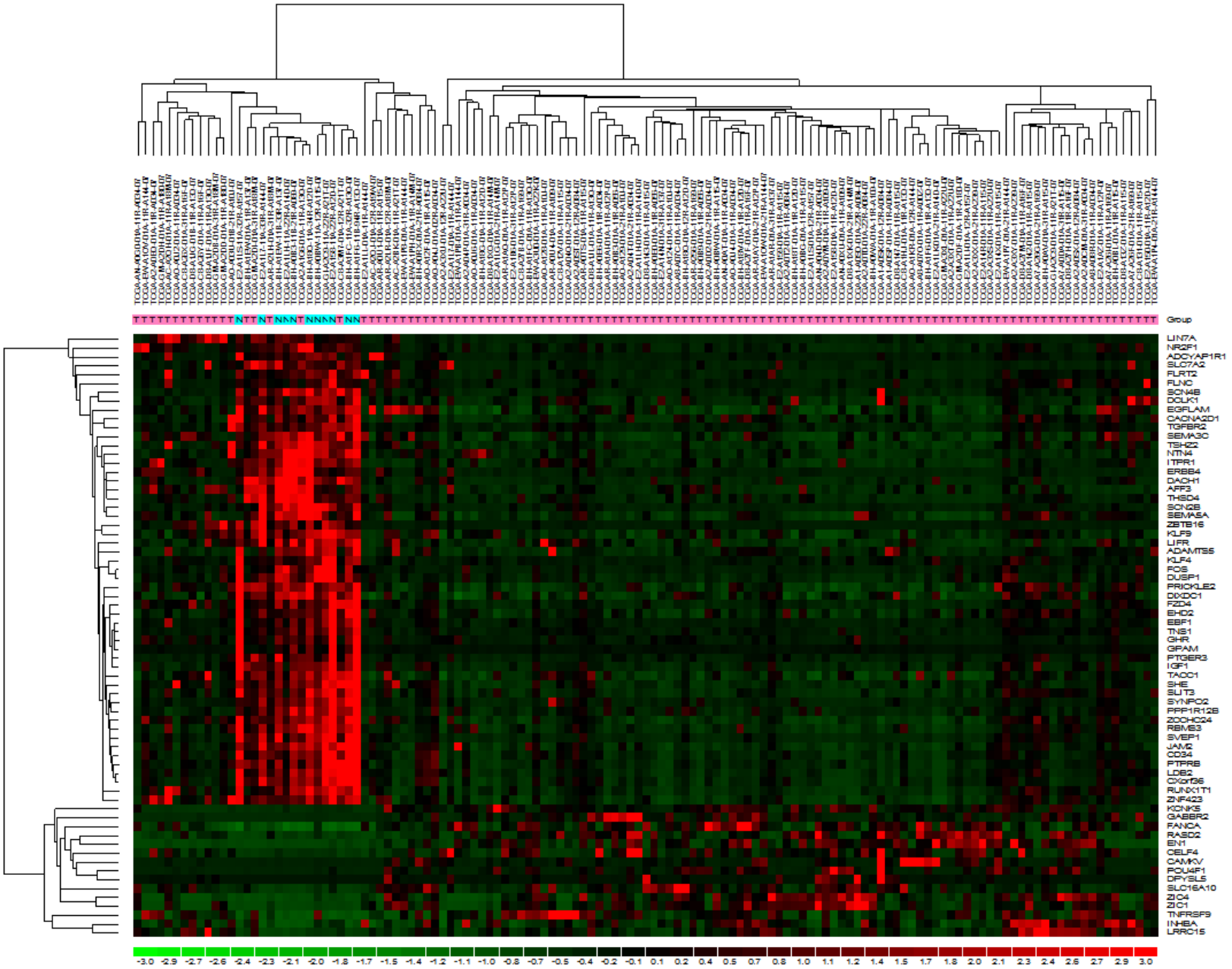


Figure S2: Differential expression of 68 mRNAs in a larger cohort of 120 TCGA-TNBC and 13 normal-adjacent samples. Related to Figure 3.

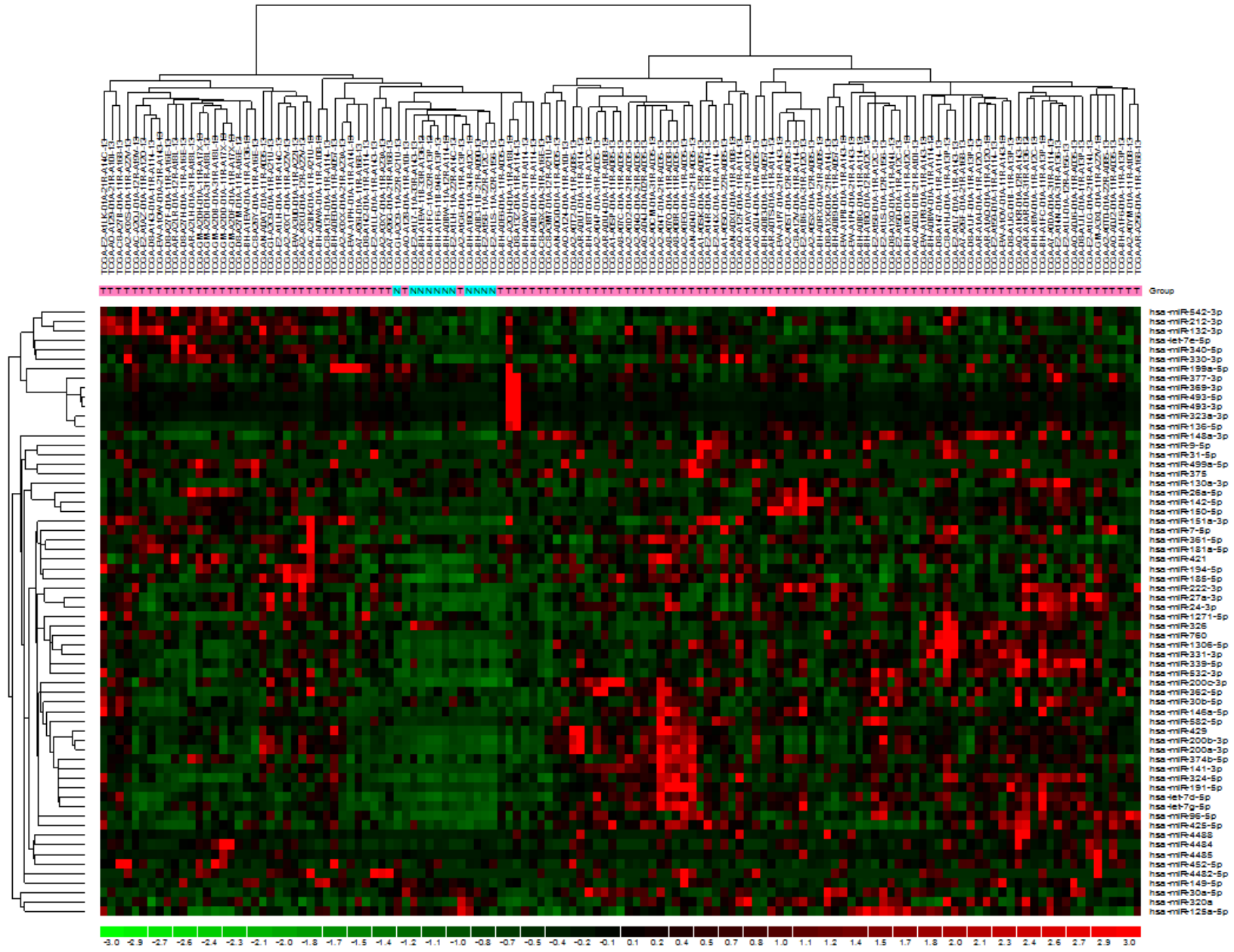


Figure S3: Heatmap listing 64 microRNAs that are both differentially expressed and participate in MRE mediated gene expression regulation. Related to Figure 3.

Table S1. TAM 2.0 results for microRNA set pathway analysis. Related to Figure 4.

Category: Disease	Count	Percent	Fold	P-value	Bonferroni	FDR
Carcinoma, Hepatocellular	60	17.7%	3.06	5.48e-28	5.73e-26	6.64e-23
Carcinoma, Gastric	55	20%	3.46	1.31e-24	1.37e-21	7.90e-22
Carcinoma, Colon	57	18%	3.13	1.48e-23	1.54e-20	5.96e-21
Carcinoma, Lung, Non-Small-Cell	47	21%	3.71	7.07e-21	7.40e-18	2.14e-18
Carcinoma, Breast	55	15.8%	2.73	3.38e-20	3.53e-17	8.18e-18
Carcinoma, Prostate	49	19%	3.30	1.34e-19	1.40e-16	2.70e-17
Neoplasms (unspecific)	44	21%	3.66	8.17e-19	8.55e-16	1.24e-17
Carcinoma, Hepatocellular (UP)	38	25%	4.44	7.82e-19	8.18e-16	1.35e-16
Glioblastoma	39	23%	3.99	1.38e-17	1.44e-14	1.86e-15
Leukemia, Myeloid, Acute	32	27%	4.65	4.34e-16	4.54e-13	4.87e-14
Glioma	41	20%	3.4	6.17e-16	6.45e-13	6.22e-14
Melanoma	33	25%	4.32	1.38e-15	1.44e-12	1.29e-13
Breast Neoplasms	34	23%	3.97	7.08e-15	7.40e-12	6.12e-13
Carcinoma, Breast, Triple Negative	14	23%	4.04	2.70e-06	2.83e-3	2.87e-5

TRANSPARENT METHODS

ReMIx – a novel methodology to compute MRE frequency from RNA-Seq data

We developed an innovative bioinformatics approach called ReMIx, which was used to quantify MRE sites at the 3'UTR regions of mRNAs from RNA-Seq data. ReMIx uses reads aligned to the 3'UTR of genes and scans them for evidence of any given MRE sequence. MRE sequences, which are complementary to the seed sequences of microRNAs, are searched in the 3'UTR of genes that are known to be associated with microRNAs (TargetScan – human version 7.0 (Agarwal et al., 2015)). A hypothetical example of this approach is illustrated in **Figure 6**. The reads aligned to 3'UTR regions of genes Gene A and Gene B are shown in Tumor and Normal-Adjacent samples. Gene A contains two MRE sites, and Gene B has one MRE site, with a common site (MRE1) in both genes. The number of reads mapped to these MRE sites are quantified for each gene and tabulated separately for Tumor and Normal-Adjacent samples. Later, MRE counts per gene are normalized and statically evaluated to identify differentially expressed MREs for downstream analysis.

MRE frequency analysis from RNA-Seq data

Seed sequences for all the conserved microRNA families (n=329) were downloaded from TargetScanHuman 7.1, and the corresponding complementary MRE sequences were derived using in-house bioinformatics scripts. **Figure 1** is a flowchart representation of the ReMIx methodology. In ReMIx, the RNA-Seq BAM files for both Tumor and Normal-Adjacent were subset to the 3'UTR regions for all genes using the SAMTools suite (Li et al., 2009). The newly obtained BAM files were converted into a FASTQ format for every gene, using the bam2fastx module from Tophat (Trapnell et al., 2009). Next, using MRE sequences of individual microRNAs and the FASTQ files for corresponding genes, the frequency of each MRE was quantified using FIMO (Grant et al., 2011). MRE sites with p-value < 0.05 were selected from the FIMO output for downstream processing. The raw MRE counts were then normalized to account for sample library size, 3'UTR length and 3'UTR GC content per gene. Finally, for every gene and for every conserved microRNA that targets the gene, the normalized MRE counts were reported in a tab-delimited format for each gene-microRNA pair in the Tumor and Normal-Adjacent cases separately.

3'UTR definitions obtained from TargetScan

Bartel's group developed an improved quantitative model to predict canonical targeting of microRNAs to 3'UTR regions of mRNA (Agarwal et al., 2015). A combination of 14 features in the model coupled with experimental approaches such as poly(A)-position profiling by sequencing called 3P-seq was used to define 3'UTR positions of genes in the transcriptome accurately. This data, available at the TargetScan Human 7.1 database, is what was used for 3'UTR definitions of genes in the MRE analysis study.

RNA-Seq and microRNA-Seq data from TCGA

The RNA-Seq and the microRNA Sequencing fastq files for the TCGA breast cancer samples were downloaded from the TCGA Research Network (http://cancergenome.nih.gov/) using the National Cancer Institute (NCI) Genomic Data Commons (GDC) resource (https://gdc.cancer.gov/). The RNA-Seq fastq files and aligned to the hg19/NCBI 37.1 human reference genome using the MAP-RSeq workflow (Kalari et al., 2014) and the microRNA fastq files were aligned using the CAP-miRSeq workflow (Sun et al., 2014). The normalized microRNA counts from CAP-miRSeq were used to obtain the microRNA expression values in the TNBC samples.

The differential expression analysis of the RNA-Seq data for the TNBC tumor and normal-adjacent pairs were obtained using the bioinformatics R package edgeR (Robinson et al., 2010). The statistical significance threshold used was FDR < 5% and log2FC |2|. For these 13 pairs of TNBC cases, differential expression analysis of the microRNA sequencing data was performed using the R bioinformatics package called limma (Ritchie et al., 2015). The statistical threshold used to identify significantly differential expressed microRNAs was adjusted p-value <0.05.

MRE site evaluation and activated pathway identification

Evaluation of MRE sites that represented distinct and TNBC-specific expression as opposed to ER+ and HER2+ subtypes and normal-adjacent cases were obtained using the R package Dunnett-Tukey-Kramer Pairwise Multiple Comparison Test Adjusted for Unequal Variances and Unequal Sample Sizes. Statistically significant MRE sites were selected using p-value cut-off < 0.05. The bioinformatics R

package edgeR (Robinson et al., 2010) was used to obtained differentially expressed MREs between TN tumors and matched normal-adjacent pairs at FDR <5% and log2FC |2|.

Pathway analysis for canonical pathways

The microRNA set analysis tool called TAM2.0 was used to identify cancer-related pathways for the 125 microRNAs. Likewise, enriched canonical pathway analysis for 88 genes was performed using KEGG and Reactome functional databases. Open source analysis toolkit WebGestalt (Wang et al., 2017) was also used for pathway identification by using the option to perform Gene set enrichment analysis (GSEA). Identification of the relevance and activation/inhibition status of pathways was evaluated using the R package called Signaling Pathway Impact Analysis (SPIA). The Bioconductor R package called Pathview (Luo et al., 2017) was used to map the gene expression data from TNBC and visualize the MAPK pathway using the KEGG-based network model of this pathway.

Statistical Methods

The various statistical tests performed in this study are summarized as follows –

1. MRE selection and normalization: (a) raw counts for MRE sites were quantified using the Find Individual Motif Occurrences (FIMO) tool and MREs with p-value < 0.05 were selected for downstream analysis, (b) raw counts were normalized using conditional quantile normalization (CQN) to account for sample library size, 3'UTR length and 3'UTR GC content per gene.
2. Identification of TNBC specific MREs: (a) first, Dunnett-Tukey-Kramer (DTK) pairwise multiple comparison statistical test was applied to tumor and normal-adjacent of all subtypes – TNBC, ER+ and Her2+ (6 groups in total) to obtain 614 TNBC MREs at p-value < 0.05, (b) second, edgeR bioinformatics package was applied to TNBC tumor and normal-adjacent samples to identify 3,053 MREs (FDR < 5% and log2FC |2|), (c) finally, union of the DTK and edgeR results were used to arrive at 221 MREs.
3. RNASeq data of TNBC tumor and normal-adjacent samples were compared for differential expression analysis using the edgeR package to obtain 2,250 genes at the statistical significance threshold of FDR (false discovery rate) < 5% and log2 fold change > 2 or < -2.

4. The microRNA sequencing data of TNBC tumor and normal-adjacent samples were compared for differential expression analysis using the limma package to obtain 778 microRNAs at the statistical significance threshold of adjusted p-value < 0.05.
5. The microRNA pathway analysis was performed using the TAM 2.0 tool for enrichment analysis of the 125 microRNAs found by ReMIx. Out of 125 microRNAs, 14 microRNAs were associated with the upregulation of disease at FDR < 2.87e-5, 55 microRNAs reported in breast carcinoma studies at FDR < 8.18e-18, and 34 microRNAs in breast neoplasms at (FDR < 6.12e-13).
6. The mRNA pathway analysis was performed using (a) gene set enrichment analysis (GSEA) – with the FDR values reported in Supplemental File 12, (b) SPIA - pGFdr and pGFWER are the False Discovery Rate and Bonferroni adjusted global p-values reported in Supplemental file 13.

Supplemental References

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Kalari, K.R., Nair, A.A., Bhavsar, J.D., O'Brien, D.R., Davila, J.I., Bockol, M.A., Nie, J., Tang, X., Baheti, S., Doughty, J.B.*, et al.* (2014). MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. BMC Bioinformatics *15*, 224.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Luo, W., Pant, G., Bhavnasi, Y.K., Blanchard, S.G., Jr., and Brouwer, C. (2017). Pathview Web: user friendly pathway visualization and data integration. Nucleic Acids Res *45*, W501-W508.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., and Kocher, J.P. (2014). CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. BMC Genomics *15*, 423.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res *45*, W130-W137.