

Race Disparities in the Contribution of miRNA Isoforms and tRNA-Derived Fragments to Triple-Negative Breast Cancer

Aristeidis G. Telonis and Isidore Rigoutsos



Abstract

Triple-negative breast cancer (TNBC) is a breast cancer subtype characterized by marked differences between White and Black/African-American women. We performed a systems-level analysis on datasets from The Cancer Genome Atlas to elucidate how the expression patterns of mRNAs are shaped by regulatory noncoding RNAs (ncRNA). Specifically, we studied isomiRs, that is, isoforms of miRNAs, and tRNA-derived fragments (tRF). In normal breast tissue, we observed a marked cohesiveness in both the ncRNA and mRNA layers and the associations between them. This cohesiveness was widely disrupted in TNBC. Many mRNAs become either differentially expressed or differentially wired between normal breast and TNBC in tandem with isomiR or tRF dysregulation. The affected pathways included energy metabolism, cell signaling, and immune responses. Within TNBC, the wiring of the affected pathways with isomiRs and tRFs differed in each race. Multiple isomiRs and tRFs arising from specific miRNA loci (e.g., miR-200c,

miR-21, the miR-17/92 cluster, the miR-183/96/182 cluster) and from specific tRNA loci (e.g., the nuclear tRNA^{Gly} and tRNA^{Leu}, the mitochondrial tRNA^{Val} and tRNA^{Pro}) were strongly associated with the observed race disparities in TNBC. We highlight the race-specific aspects of transcriptome wiring by discussing in detail the metastasis-related *MAPK* and the *Wnt/β-catenin* signaling pathways, two of the many key pathways that were found differentially wired. In conclusion, by employing a data- and knowledge-driven approach, we comprehensively analyzed the normal and cancer transcriptomes to uncover novel key contributors to the race-based disparities of TNBC.

Significance: This big data-driven study comparing normal and cancer transcriptomes uncovers RNA expression differences between Caucasian and African-American patients with triple-negative breast cancer that might help explain disparities in incidence and aggressive character. *Cancer Res*; 78(5); 1140–54. ©2017 AACR.

Introduction

Breast cancer is the most common type of cancer among women in the United States with nearly 250,000 new cases in 2016 (1). Breast cancer's clinical classification relies on the expression levels of three receptors on which the tumor relies for its growth (2): these are the receptors for the hormones estrogen and progesterone (ER and PR, respectively), and the human epidermal growth factor receptor 2 (HER2). In the breast cancer subtype known as "triple negative" (TNBC), none of the three receptors is expressed.

TNBC accounts for 10% to 20% of all breast cancer cases and is the most aggressive subtype. In the absence of an expressed receptor, chemotherapy and surgery remain the only options for TNBC patients (2). Epidemiologically, TNBC is more frequent among Black/African-American (B/Aa) premenopausal women than is among White (Wh) women (3, 4). In B/Aa patients, TNBC exhibits higher grade, is more aggressive, and has higher metastatic potential (3, 5). Even though socioeconomic factors are

known contributors to the worse clinical picture of B/Aa TNBC patients, differences between the two races continue to remain after adjusting for these factors (5).

At the genetic level, the mutational profile of tumors for established cancer-linked genes is indistinguishable in the two races, and, thus, cannot explain the observed differences in outcome (5, 6). This observation together with the persistent disparities suggests other possibilities. For example, there may be consequential mutations that lie beyond the exons of the cancer-linked genes, or dynamic processes may be at work that are currently not understood and do not relate to mutations. Other possibilities could include expression differences in signaling cascades that are found upregulated in B/Aa patients, such as insulin-like growth factor 1 receptor (IGF1R) (7) or *Wnt/β-catenin* (5).

Noncoding RNAs (ncRNA) represent a recently emerged component of TNBC biology. Work by others (8, 9) and us (10, 11) has shown important differences in the transcriptomic profiles of ncRNAs among individuals who differ by sex, population origin, or race. Such transcripts include long ncRNAs (8), the shorter miRNAs and their isoforms (isomiRs), as well as the tRNA-derived fragments (tRF; refs. 9–11). Importantly for this presentation, we showed in recent work that isomiRs and tRFs are expressed in a tissue state-specific- and race-specific manner (10, 11). Given that isomiRs and tRFs can directly regulate mRNA abundance in multifaceted ways, for example, through RNAi or by displacing RNAs from RNA-binding proteins (10–13), it is conceivable that they are also associated with the race-based health disparities of TNBC.

In this presentation, we build on our previous work (10, 11) by approaching the transcriptomic differences between B/Aa and Wh TNBC patients from a systems-biology and network-

Computational Medicine Center, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania.

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

Corresponding Authors: Isidore Rigoutsos, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107. Phone: 215-503-4219; Fax: 215-503-0466; E-mail: Isidore.Rigoutsos@jefferson.edu; and Aristeidis G. Telonis, E-mail: Aristeidis.Telonis@jefferson.edu

doi: 10.1158/0008-5472.CAN-17-1947

©2017 American Association for Cancer Research.

biology perspective (14, 15). As before, we base our studies on the TNBC datasets of The Cancer Genome Atlas (TCGA) repository (16). Our analysis takes a holistic approach to differentially expressed (DE) transcripts and their potential interactions. We use gene and isomiR/tRF coexpression networks to study the coordination of tissue-specific biological processes with short ncRNAs in normal breast and in TNBC. We integrate data from other modalities and mine the mRNA networks to identify actors and effectors. We do so separately for each tissue state (=normal or TNBC) and for each race (=B/Aa or Wh). We place particular emphasis on how mRNAs are wired with the two categories of regulatory ncRNAs of interest, the isomiRs and the tRFs, separately for each state and race. We also compare how these relationships are rewired across tissue states and across races. Our analyses have two goals: identify novel molecules and novel interactions in the context of racial disparities and TNBC and elucidate whether isomiRs and tRFs associate with molecular pathways that have known links to disparities.

Materials and Methods

Acquisition of the molecular profiles

As in our previous work (10, 11, 17, 18), we utilized the molecular profiles of the TCGA breast cancer samples (16) to identify which samples to use in our analysis. We excluded all breast cancer samples from male patients. We employed all normal breast female samples that are available in TCGA (94 from Wh and 6 from B/Aa subjects), irrespective of the adjacent breast cancer subtype. Of the TNBC samples, we excluded those that TCGA annotates as potentially problematic (18), which left us with 66 samples from Wh and 32 from B/Aa women donors. Each small RNA-seq dataset was analyzed with our recently published method for defining the lowest abundance threshold (19), and subsequently mined using our tRF (17) and isomiR (11, 18) pipelines. We only considered those tRFs and isomiRs with a median abundance of ≥ 2 RPM. For the mRNA profiles, we used TCGA's *rsem_genes.normalized_results* files and filtered out those genes whose average abundance was less than the median of the means of abundances of all mRNAs. In the case of mRNA isoforms, we used TCGA's *rsem.isoforms.normalized_results* files and filtered out isoforms that had a mean expression lower than the 25th percentile. Protein abundance data (Level 4) were downloaded from The Cancer Proteome Atlas (20).

Statistical and pathway enrichment analyses and data visualization

Analyses were carried out in R and Python and data were \log_2 -transformed before further processing. Variability analysis was done by computing the coefficient of variations for each variable (mRNA, isomiR, or tRF) and by plotting their distributions in \log_{10} scale. For univariate analyses, we used the Mann-Whitney *U* test and for multivariate the significance analysis of microarrays (21) at 5,000 permutations and a FDR threshold of 5%. The mRNA isoforms between two states were expressed as a percentage (%) of total gene expression by dividing the calculated expression of each isoform by the sum of expression of all isoforms from the current gene. Pathway enrichment was estimated with DAVID (22) using an FDR threshold of 5%. The terms that were queried were Gene Ontology (GO) terms for biological process, molecular function, and cellular compartment, as well as KEGG pathways. We used Cytoscape for network visualization (23). All other visualizations were carried out in R as described previously (10, 11, 18).

Target prediction and sequence similarity

To identify potential binding sites of miRNAs in mRNAs, we used the rna22 algorithm (24). We allowed up to one unpaired base (=mismatch or bulge) within the seed region and enforced a minimum of 12 matched bases in the putative heteroduplex. As the mechanisms of interactions for tRFs remain elusive, we opted for a more general approach to minimize false negative results and, therefore, identified potential mRNA targets by sequence similarity with the *gsearch36* algorithm (25) with default settings. For all searches involving mRNAs, we used all cDNA variants as reported in ENSEMBL75 (26). We used both the tRF sequence (in search of putative "decoyed" targets) and its reverse complement (in search of putative "direct" targets) against the mRNA sequences. We queried tRFs for instances of RNA-binding motifs. RNA-binding motifs were downloaded from RBPDB (27). In these analyses, each isomiR and tRF was considered separately from the rest.

Gene coexpression networks and identification of commonly and differentially wired features

We constructed gene coexpression networks using the methodology of previous studies (28–30) and performed correlation analyses by computing all pair-wise Spearman ρ correlation coefficients among all mRNAs, tRFs, and isomiRs (six sets of correlations in all). For each case (Wh-Normal, Wh-TNBC, B/Aa-TNBC), the correlations were run independently and *P* values were corrected to FDR values. Our threshold choices here were $|\rho| = 0.333$ and FDR of 1%. In each case, we kept not more than the top 5,000 positive and the top 5,000 negative ρ coefficients and used them to form the set of significant correlations in each case. This process allowed us to create coexpression tables and convert the significant correlations to network connected nodes. In each coexpression network, the nodes represent features (mRNAs, isomiRs, or tRFs) and the edges denote the existence of a significant correlation between the connecting nodes. In the rendered networks, to improve clarity, we collapsed isomiRs and tRFs to the miRNA arm or tRNA of origin, for example, nodes with names like "miR-200c-3p" and "mtLysTTT" stand for isomiRs originating from the miR-200c-3p arm, and for tRFs overlapping with the mature mitochondrial (MT) tRNA^{LysTTT}, respectively. Network edges are undirected as they represent correlation (positive or negative) between two features, and not direct molecular coupling between the features. IsomiR–mRNA edges and tRF–mRNA edges indicate a putative regulatory action on the mRNA by the isomiR or the tRF, respectively.

Coexpression networks can have many structures. However, biological networks exhibit a scale-free topology with a few, central "hubs," that is, nodes of high degree, and many "peripheral" nodes of low degree (31). To evaluate the scale-free topology of the gene coexpression network, we plotted the distribution of node degrees (number of connections with other nodes in the network) and performed linear regression on the log scale of the distributions.

In each mRNA–mRNA network, we assessed how the 50 mRNA hubs with the highest degree are connected to other mRNA nodes by computing pair-wise "mRNA–Jaccard" indices. We define the "mRNA–Jaccard" index between two mRNA hubs as the ratio of the number of correlated mRNAs that are common to these two mRNA hubs over the number of all the distinct mRNAs with which these two mRNA hubs are correlated. We then performed hierarchical clustering (metric: Euclidean distance) in each of the three (square) matrices that we obtained.

In each mRNA–mRNA network, we also assessed how the 50 mRNA hubs with the highest degree are connected to isomiRs and tRFs by computing pairwise "isomiR/tRF–Jaccard" indices. We define the "isomiR/tRF–Jaccard" index between two mRNA hubs as the ratio of the number of correlated isomiR/tRF nodes that are common to these two mRNA hubs over the number of all the distinct isomiRs/tRFs with which these two hubs are correlated. This allowed us to obtain a global perspective of which isomiRs/tRFs are connected with which mRNA hubs in the coexpression networks.

We compared commonly wired and differentially wired pairs of transcripts in the following two cases: (i) Wh women in the normal and cancer states; and (ii) Wh women and B/Aa women in the cancer state. A specified pair of transcripts is "commonly wired" (CW) in two different states, if and only if the two transcripts are statistically significantly coexpressed in both states and with the same correlation sign (either positive or negative) in both states. Analogously, a specified pair of transcripts is "differentially wired" (DW) in two states, if and only if the two transcripts are statistically significantly coexpressed in exactly one state, or statistically significantly coexpressed in both states but with opposite correlation signs. To further limit potential false positive DW pairs, in those cases where there was no change of sign, we required that the absolute ρ coefficient of the nonsignificant correlation be <0.2 . Thus, our differentially wired pairs were defined qualitatively, and not quantitatively, due to the relatively low statistical power (30).

Attributing mRNA correlation pairs to upstream or downstream factors or to isomiRs and tRFs

We leveraged information from several public databases and projected it on the mRNA correlation pairs. For the cellular compartment of the mRNAs' protein products, we used the manually reviewed human proteome of UniProt (accessed on November 27, 2016; ref. 32). Protein–protein interactions, normalized at the gene level with default filtering, were drawn from the v2.1 release of the PICKLE database (accessed on September 13, 2017; ref. 33). Collections of genes that share the same transcription-binding site were downloaded from MSigDB (C3. TFT set accessed on March 23, 2017; ref. 34). To find genes that are exclusively (or predominantly) expressed in immune cells as compared with normal breast tissue, we used the data from the Expression Atlas of EMBL–EBI (accessed on February 14, 2017; ref. 35). We first formed a collection of "immune tissues" comprised of the entries "EBV-transformed lymphocyte," "blood," "small intestine Peyer patch," and "spleen." To identify the relevant genes, we ranked each gene's expression across the 53 tissues and kept those genes whose four highest expressing tissues included at least three of the four immune tissues that also showed low expression (rank ≥ 20) in breast. We obtained transcription factor (TF) information from ENCODE's chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) experiments (36, 37) that we downloaded from the UCSC Genome Browser *wgEncodeRegTfbsClusteredV3* track (accessed on April 6, 2017). We linked genes with TF-binding sites by searching for binding sites not more than 5,000 base pairs upstream or 5,000 bases or downstream from the gene's transcriptional start site. In the case of multiple start sites, we considered the union of the respective 10 Kbp genomic regions around the multiple start sites. Before using a binding event, we required that it be observed in four or more independent experiments and that the maximum signal

strength, as defined by the UCSC track, to be ≥ 100 . Finally, we considered two mRNAs to be part of the same pathway, if they were listed under the same biological process by DAVID.

If both mRNAs in a correlation pair encode for proteins that directly interact, are in the same cellular compartment, are part of the same biological process as identified by DAVID, or are encoded by genes that are under the control of the same transcription factor per the ENCODE ChIP-seq data, then we can conclude that this mRNA pair can be "explained" by the respective common feature(s). In the case of immune system, we considered correlations to be "immune-system explained" if at least one mRNA in the mRNA pairs is predominantly expressed in the immune system tissues as outlined above. To link mRNA pairs to isomiRs and tRFs, we examined whether both mRNAs of the pair were significantly correlated with the same isomiR or tRF.

Monte Carlo simulations

We examined whether we could attribute the mRNA correlation pairs to factors (e.g., encoded proteins localize to the same cell compartment) that can potentially cause the observed correlations and, thus, could account for (i.e., explain) a certain portion of the correlations. To examine whether the accounted-for-portion was due to chance, we ran Monte Carlo simulations. For each of the three cases (Wh-Normal, Wh-TNBC, B/Aa-TNBC), we randomly selected the same number of unique mRNA pairs and calculated the same percentage. Repeating this random selection 20,000 times yielded the distribution of percentage values for each factor that are expected by chance. Finally, using this distribution, we calculated the respective z -score for the percentage value that we observed. z -scores with absolute values ≥ 2.0 were deemed statistically significant.

Data availability

The original datasets that we analyzed are available in the TCGA repository (<https://gdc.cancer.gov>). The tRF profiles that we generated in each case are in MINTbase (<https://cm.jefferson.edu/MINTbase/>). The results from our correlation analysis and their attributes to other levels of cellular function are available in the Supplementary Tables that are included with this presentation.

Results

TNBC progression in B/Aa and Wh patients follows parallel yet distinct trajectories

In our previous work, we showed that TNBC samples can be distinguished from normal breast tissue using isomiR or tRF profiles (10, 11). Also, we showed that isomiRs and tRFs can also distinguish TNBC samples by race (10, 11). Having reported recently the isomiR and tRF profiles for all TCGA datasets (17, 18), we sought to examine how the expression patterns of mRNAs, isomiRs, and tRFs differ between normal breast and TNBC, and also between the two races in TNBC.

Looking at the variability of expression for mRNAs, isomiRs, and tRFs in normal and TNBC for each race, we found that TNBC exhibits a significant increase in average variability compared with normal breast, in all three classes of molecules (Supplementary Fig. S1A and S1B). This result reflects the considerable heterogeneity of TNBC.

We also performed a significance analysis to find which features are DE between tissue states, or, between races. These analyses

(Supplementary Table S1). We also found that in both races, tRNA^{HisGTC}, tRNA^{GlnTTG}, and tRNA^{GlnCTG} give rise to tRFs whose abundance decreases significantly in TNBC compared with normal breast.

In a handful of cases, we observed an intriguing dual behavior for miRNAs and tRNAs (Fig. 1A and B). For example, isomiRs from the 5p arm of mir-146b are upregulated in TNBC, whereas isomiRs from its 3p arm are downregulated in TNBC (Supplementary Table S1). The miR-30 family provides another intriguing example of dual behavior (see Fig. 1A and B; Supplementary Table S1, and Discussion). This phenomenon of opposing changes in abundance is also present in the case of tRFs, for example, see tRNA^{HisGTC} in Supplementary Fig. S1D. This behavior suggests multiple biogenesis mechanisms for tRFs.

IsomiRs and tRFs are known to load on Argonaute (AGO) and to directly target mRNAs through RNAi (10, 11). Thus, we sought possible links between DE isomiRs and tRFs, on one hand, and mRNAs that are DE in the opposite direction and contain a predicted binding site (in the case of isomiRs) and a predicted binding site or sufficient sequence similarity (in the case of tRFs) on the other. For both races, we could link 26% of the downregulated mRNAs to upregulated short ncRNAs. However, of the upregulated mRNAs, only 7% could be linked to a downregulated short ncRNA. Our DE analyses provide some first insight into the evidently complex regulatory processes that underlie TNBC.

To further demonstrate this complexity, we selected important pathways that we found enriched in these analyses and generated a network of mRNAs, miRNAs, and tRNA isoacceptors that are DE independent of race. The miRNA arms and tRNA isoacceptors of Fig. 1C produce isomiRs and tRFs, respectively, that are DE in a direction opposite to that of the connected mRNAs. The green links in Fig. 1C indicate that the archetype miRNA is among the DE isoforms produced from the shown mRNA locus. Recall here that by archetype we refer to the miRNA isoform that is listed in the public databases. It is important to stress here that the archetype neither has to be the corresponding locus' most abundant isoform nor is the only functionally important isoform, as we showed in other settings (11). Figure 1 makes evident that the vast majority of the miRNA:mRNA links, which we uncovered, involve isoforms other than the archetype: These isoforms have not been studied before, and thus, their functional roles have not been elucidated, especially in the context of these molecular processes in TNBC.

Among the dysregulated RNAs are transcripts and pathways that are DE between TNBC and normal breast in only one of the two races. Specifically, there are pathways that are DE between normal breast and TNBC exclusively in Wh patients (Fig. 1D and E, top). Analogously, other pathways are DE exclusively in B/Aa patients (Fig. 1D and E, bottom). We note the multiple manifestations of immune system responses ("primary immunodeficiency" and "antigen processing and presentation") in Wh TNBC patients (Fig. 1D). Similarly, there are several entries for cell signaling deregulation that are exclusive to B/Aa TNBC patients (oxytocin, Ras, MAPK, and Jak-STAT signaling; Fig. 1D and E). It is interesting to note that although "amino acid metabolism" is common to both groups, it appears as a decrease in the degradation of branched-chain amino acids in Wh TNBC patients and as an increase in the biosynthesis of amino acids in B/Aa TNBC patients.

In light of the race-specific differences across tissue states (i.e., normal breast and TNBC), we hypothesized that there

may also be race-specific differences within a tissue state. We compared the expression profiles of B/Aa and Wh patients in (i) normal breast and (ii) TNBC, respectively. In the normal breast state, compared with Wh donors, B/Aa donors exhibit increased expression of key genes (e.g., *CD2* and *CXCL9*) that are linked to high tissue inflammation (Supplementary Tables S1 and S2). Analogously, in the TNBC state, key genes from cell signaling pathways (e.g., *MAP2K7*, *PIM1*, and *PIM2*) are dysregulated in B/Aa patients (in agreement with the results of Fig. 1D and E).

Collectively, the above results show extensive and very concrete molecular differences between B/Aa and Wh patients at the two physiologic endpoints, that is, normal breast and TNBC. During the transition from normal tissue physiology to TNBC (38), the patients in the two race groups follow parallel yet distinct trajectories whose specifics are influenced sharply by the patients' race.

The transcriptomic networks in TNBC exhibit pronounced heterogeneity

To investigate the nature of correlations among mRNAs and short ncRNAs from a network perspective (28, 39, 40), we constructed gene coexpression networks (see Materials and Methods; Supplementary Table S3). We focused on the three groups, or "patient states," for which there are adequate samples in the TCGA repository for the purposes of correlation analysis: normal breast samples from Wh donors, TNBC samples from Wh donors, and TNBC samples from B/Aa donors. The available normal breast samples from B/Aa donors are not adequate in this regard.

The three resulting networks are found to be scale-free networks with the number of edges per number of nodes following a power-law distribution (Supplementary Fig. S2A). We note that the correlations we observe in normal samples are not affected by the adjacent tumor subtype (Supplementary Fig. S2B). Despite the similarity in their topology, the three networks exhibit extensive differences in terms of (i) the genes they comprise (Supplementary Fig. S2C) and (ii) the correlations in which these genes participate (Supplementary Fig. S2C and S2D).

In the normal state, the hubs of the mRNA-mRNA coexpression networks correspond to "cell growth and survival" regulators (e.g., *RAB25* and *EPS15*) as well as "lipid metabolism and storage" genes like *PLIN1*. In TNBC, for both Wh and B/Aa patients, the hubs correspond to "transcription factors/regulators" (like *REST* and *NCOA2*), "cell signaling" (e.g., *APC* and *TAOK1*) and "immune system" (e.g., *CD4* and *CD48*) genes, as well as enzymes that participate in "oxidative phosphorylation" (Supplementary Table S4). The KEGG pathway "ribosome" is the only one that is enriched in all three groups (Supplementary Table S4). "Immune system responses" and "oxidative phosphorylation" were enriched only in the TNBC states, which supports the view of higher tissue- and cellular heterogeneity in the cancer state. We examine the similarities and differences of the three mRNA-mRNA networks below.

There are two distinct subnetworks of mRNAs that are coexpressed with isomiRs and tRFs in TNBC and correspond to nonoverlapping modules

We focused on the genes that are hubs in the mRNA-mRNA coexpression networks and examined the immediate mRNA neighbors of the 50 nodes with the highest degree. Specifically, for each pair of hubs, we evaluated their "mRNA overlap" by

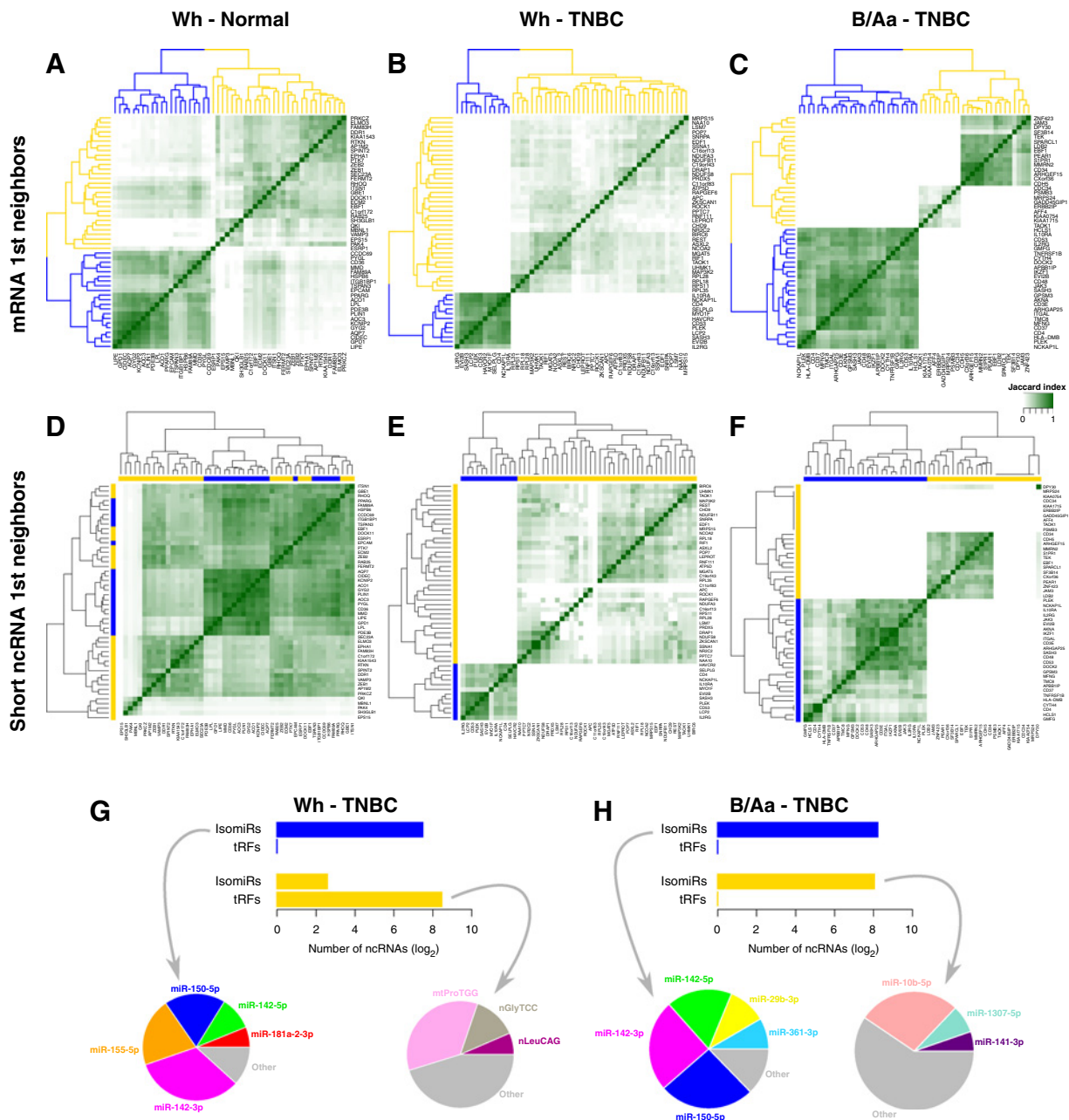


Figure 2.

mRNA hubs and their neighbors. The heatmaps of **A-F** visualize the matrix of Jaccard indices of the shared connections for the top 50 mRNA hubs. **A-C**, The hubs are compared by examining their immediate mRNA neighbors (mRNA-Jaccard index). **D-F**, The hubs are compared by examining their immediate short ncRNA neighbors (isomiR/tRF-Jaccard index). The higher the Jaccard index between two hubs, the higher the number of shared first neighbors (see Materials and Methods). **A** and **D**, Normal breast, Wh donors. **B** and **E**, Wh TNBC patients. **C** and **F**, B/Aa TNBC patients. **G-H**, Number of isomiRs/tRFs that are first neighbors of the mRNA hubs in Wh-TNBC (**G**) and B/Aa-TNBC (**H**). IsomiRs and tRFs are broken down by cluster (orange or blue, as identified in **B** and **C** for each race). The pie charts show the composition of the respective bar of the histogram.

computing their mRNA-Jaccard index (see Materials and Methods). This index gauges the number of immediate mRNA neighbors that the two hubs have in common.

In the normal state, the top 50 hubs form two clusters that overlap considerably, that is, hubs not belonging to the same cluster share many first neighbors (Fig. 2A). In TNBC, we observe

again two clusters, for both the Wh and B/Aa patients. However, the clusters are now very cleanly separated, and hubs belonging to different clusters share no first mRNA neighbors (Fig. 2B and C). Strikingly, one cluster comprises multiple genes involved in immune system responses and immune system pathways, whereas the other comprises multiple genes from oxidative

phosphorylation and ribosomal components. These findings strongly suggest that the stroma significantly influences gene expression (in this case of the immune system): This stromal influence is present in TNBC but absent from normal breast. We note that the "orange" clusters in Wh-TNBC and B/Aa-TNBC patients, respectively, share only one gene. The "blue" cluster contains several common genes in the two races.

Next, for the 50 highest degree hubs of the mRNA-mRNA coexpression networks, we examined their immediate *isomiR/tRF* neighbors. We hypothesized that the distinction in two clusters in the TNBC state might be reflected on the *isomiRs* and *tRFs* with which the mRNAs are correlated. Specifically, for each pair of hubs, we evaluated their "isomiR/tRF overlap" by computing their *isomiR/tRF*-Jaccard index (see Materials and Methods). This index gauges the number of immediate *isomiR/tRF* neighbors that the two hubs have in common.

For normal breast, we did not observe any notable clustering of the mRNA hubs based on their *isomiR/tRF*-Jaccard index (Fig. 2D). However, in TNBC, we found that the mRNA hubs are split again into two nonoverlapping clusters (Fig. 2E and F). When we examined the formed mRNA clusters and their corresponding first *isomiR/tRF* neighbors, we made three striking observations:

- (a) The gene hubs are split again into an "immune response" cluster (blue) and an "oxidative phosphorylation and ribosomal components" cluster (orange). These are exactly the same two groups we obtained when we clustered the hubs based on their mRNA neighbors (Fig. 2B and C).
- (b) The ncRNAs that are immediate neighbors of the blue cluster in the TNBC state belong exclusively to the *isomiR* category in both races (Fig. 2E-H). *IsomiRs* originate from the same miRNA arms largely irrespective of race, for example, from miR-150-4p and both arms of mir-142 (Fig. 2G and H).
- (c) The ncRNAs that are immediate neighbors of the orange cluster belong overwhelmingly to the *tRF* category in Wh patients and exclusively to the *isomiR* category in B/Aa patients (Fig. 2G and H). Moreover, the *isomiR*-producing miRNA loci that are linked to the orange cluster in B/Aa patients, for example, miR-10b-5p and miR-141-3p (Fig. 2H; right pie chart), have no overlap with the miRNA loci that are linked to the blue cluster (Fig. 2H; left pie chart).

These findings argue strongly that the biological processes driven by the genes of each cluster are associated with a distinct type and identity of short ncRNAs. They also show that the intercluster coordination of *isomiRs* and *tRFs* in normal breast (Fig. 2A and D) has given its place to an intracluster coordination in the TNBC state (Fig. 2B, C, E, and F). Moreover, the results reveal an important race disparity in TNBC that pertains to how immune system genes could be regulated by *isomiRs* and *tRFs*.

The short ncRNAs are likely the most important modulators of the mRNA transcriptome in normal breast and in TNBC

We should not lose sight of the fact that gene coexpression networks capture correlation patterns. Thus, attempts to infer causality from these undirected graphs should be based on additional evidence. With that in mind, we focused on the mRNA coexpression networks and sought to identify factors ("drivers") that could potentially explain the correlated pairs. Specifically, we first examined five categories of drivers:

- (a) protein products that belong to the same pathway, as per DAVID (22);
- (b) protein products that localize to the same cellular compartment, as per UniProt (32);
- (c) protein products that directly interact with each other, as per PICKLE (33);
- (d) gene promoter regions that share the same regulatory motif, as per MSigDB (34); and
- (e) genes expressed predominantly in immune cells, as per the Expression Atlas (35).

Colocalization in the same cell compartment and a shared DNA motif accounted evenly for a combined approximately 44% of the correlation pairs in each of the three states (Fig. 3A). Of note, protein-protein interactions accounted for a much lower, yet statistically significant, fraction of the mRNA-mRNA correlation pairs (Fig. 3A). We also observed that in the case of TNBC, many of the correlated mRNAs belong to the same pathway (42% for Wh and 34% for B/Aa patients). Still, in all cases, between 36% and 51% of the correlations cannot be explained by any of the above five potential "drivers."

An important limitation of the above analyses is that they rely on static data. With the exception of the Expression Atlas, this is unavoidable given the nature of the databases. Dynamic dimensions, such as tissue-specific expression and posttranslational modifications, would help select among the universe of possible interactions but are not available.

To identify potential dynamic regulators that could be behind some of the observed correlations, we did the following two things:

- (a) we used the ENCODE data to check whether the corresponding gene promoters are bound by the same TF (36, 37); and
- (b) we examined which of the mRNA-mRNA correlation pairs can be projected to *isomiRs* and *tRFs*. There are two known mechanisms of actions that justify the inclusion of *tRFs*: (i) the previous demonstration that *tRFs* can act as miRNAs by entering the RNAi pathway through AGO-loading (10, 13); and (ii) the previous demonstration that *tRFs* can regulate mRNA levels *in trans* by decoying RNA-binding proteins (12).

Our results show that the portion of the mRNA-mRNA correlations that can be explained by TF-binding remains essentially unchanged at approximately 82% across normal breast and TNBC. However, we note that this could be a significant overestimation (see Discussion). Unlike TF binding, *isomiRs* or *tRFs* exhibit a much higher dependence on the tissue and state under consideration.

More specifically, in normal breast, *isomiRs* can explain 68% of the mRNA-mRNA pairs whereas *tRFs* can explain 86% (Fig. 3B), that is, there is at least one *isomiR* or *tRF* that is correlated with both mRNAs of most pairs. In TNBC, *isomiRs* can link comparatively fewer mRNA pairs: 21% in Wh and 35% in B/Aa patients. Analogously, *tRFs* can still account for 42% of the mRNA-mRNA pairs in Wh TNBC patients. However, there is a very drastic drop in the case of B/Aa TNBC patients: a mere 1% of the mRNA-mRNA pairs can be linked to *tRFs*. These data strongly indicate that the disruption of the regulatory layer affected by *isomiRs* and *tRFs* is system wide and has a race-dependent component.

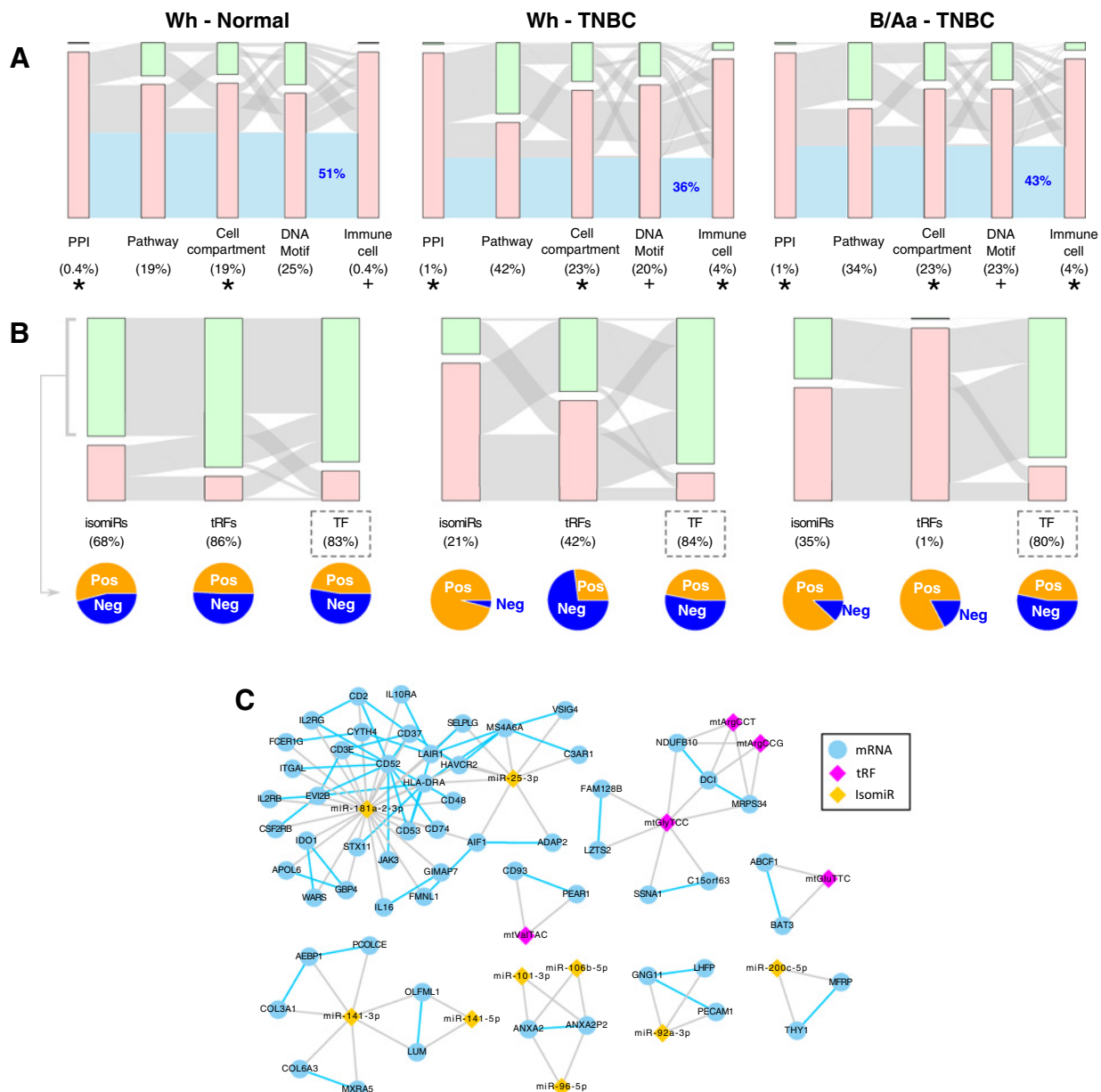


Figure 3.

The mRNA associations within a state can be largely explained by short ncRNAs and common transcription factors. **A**, Alluvial diagrams representing the percentage of mRNA couplings that are associated with each factor. The percentage in blue shows the fraction of mRNA couplings that cannot be explained by any of the considered factors. An asterisk (*) indicates that this percentage is significantly higher than expected by chance, whereas a cross (+) indicates that it is lower. Statistical significance was assessed with Monte Carlo simulations with a threshold of an absolute z-score of at least 2.0 as compared with the constructed expected distribution. **B**, Alluvial diagrams showing the percentage (also shown below each label) of mRNA coexpression pairs that can be projected to the same isomiR, tRF, or TF in each state. The gray fields are proportional to the overlap among the clusters. The pie charts show the sign (negative, blue; positive, orange) for those of the coexpressed pairs that are also associated with the corresponding regulator. The TF values are boxed to indicate the possibility that the percentage is an overestimate (see also text). **C**, Example subnetworks comprising mRNAs (circles) that are coexpressed with isomiRs or tRFs from the shown miRNA (orange diamonds) or tRNA (magenta diamonds) loci in Wh TNBC patients. Blue edges indicate coexpressed and positively correlated mRNA-mRNA pairs. Gray edges represent miRNA or tRNA loci whose isomiRs or tRFs are coexpressed with the corresponding mRNAs. All shown miRNA-mRNA pairs capture negative correlations. The shown tRF-mRNA pairs capture either positive or negative correlations.

Another important finding pertains to the concordance between the ability of isomiRs and tRFs to account for mRNA-mRNA correlations. In normal breast, the concordance

is exceptionally high. Indeed, most of the mRNA correlation pairs that can be linked with isomiRs can also be linked with tRFs, and vice versa (Fig. 3B, left). This is true for both

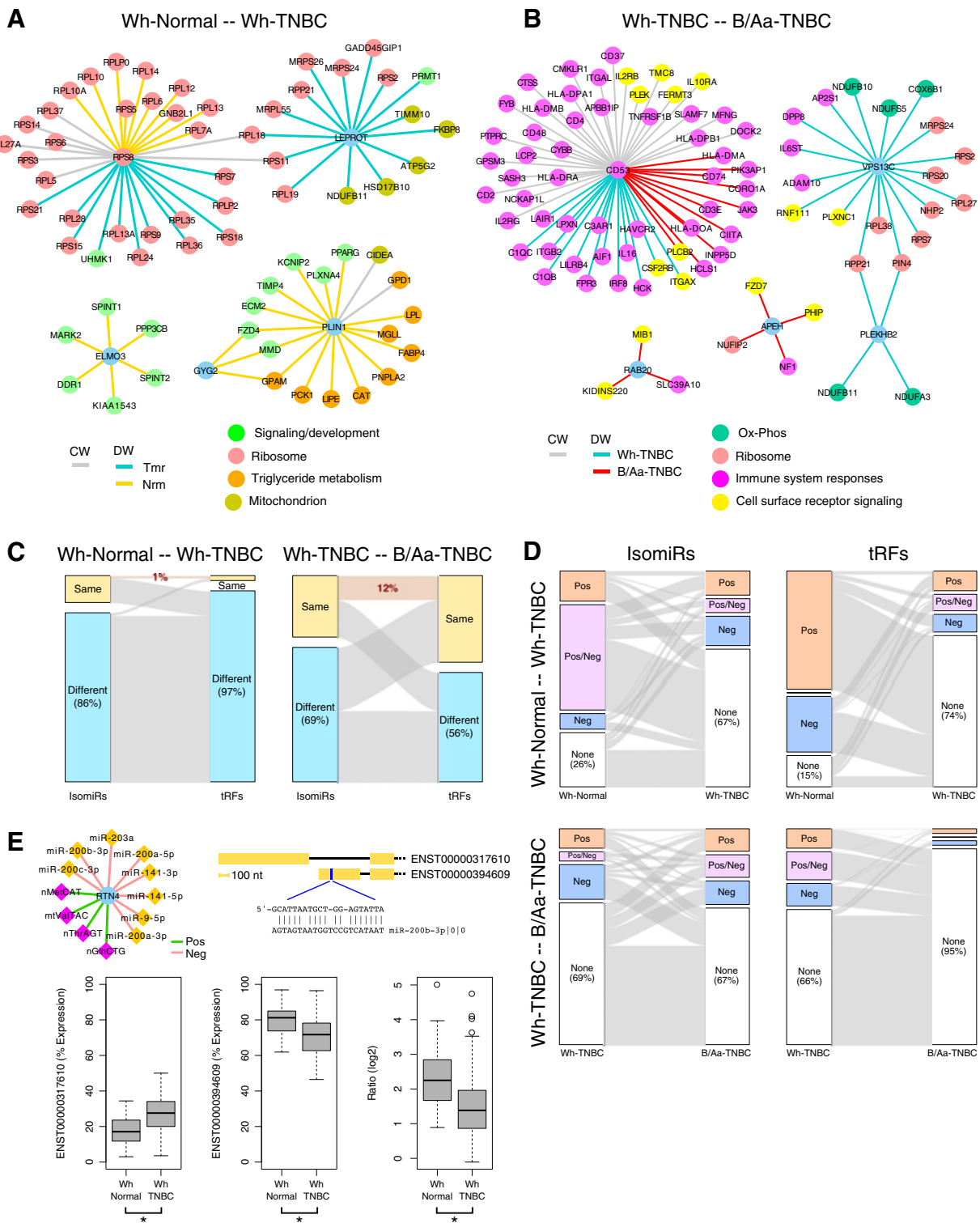


Figure 4. IsomiRs and tRFs as potential drivers of differentially wired mRNAs. **A** and **B**, The panels show CW and DW mRNAs when comparing normal breast and TNBC in Wh donors (**A**), or, the TNBC state between Wh and B/Aa patients (**B**). The mRNA nodes are colored by pathway. **C**, Alluvial diagram showing the percentage of the DW associations that can be attributed to differential wiring ("Different") with an isomiR (left column of each panel) or a tRF (right column of each panel). The box labeled "Same" captures the percentage of DW mRNA-mRNA pairs that are either not correlated with isomiRs/tRFs or the correlations with isomiRs/tRFs are CW between the two tissue states or races. (Continued on the following page.)

Downloaded from <http://aacrjournals.org/cancerres/article-pdf/78/5/1140/2604213/1140.pdf> by guest on 16 January 2025

positively- and negatively correlated pairs (see accompanying pie charts) and suggests that isomiRs and tRFs mediate highly coordinated regulatory programs in normal breast tissue.

However, this concordance/coordination is reduced radically in the cancer state wherein isomiRs and tRFs are linked to different mRNA pairs (Fig. 3B, middle/right). Moreover, there is an apparent dichotomy in the case of Wh TNBC patients: Whereas the isomiRs are linked primarily to positively correlated mRNA–mRNA pairs, the tRFs are linked primarily to negatively correlated pairs in these patients (see pie charts). We visualize examples of how miRNAs and tRFs can potentially explain positively correlated mRNAs in Fig. 3C.

The outcome of this analysis further supports the previous findings (Figs. 1 and 2) that the TNBC states differ considerably from the normal state. The extensive differences between any two of the three patient states (Wh-normal, Wh-TNBC, and B/Aa-TNBC) suggest broad disruptions of the wiring at the molecular level in TNBC, and also between races. At the same time, the relatively large percentage of mRNA–mRNA pairs in both Wh TNBC and B/Aa TNBC patients that cannot be explained by the categories we considered here suggests either the existence of an additional layer of regulation whose details elude us, or a profound alteration of the posttranscriptional control implemented by isomiRs and tRFs.

The wiring of coexpressed mRNAs is disrupted extensively between normal breast and TNBC

In this section, we focus on mRNAs that either remain CW or become DW (29, 41) between states (see Materials and Methods for definitions). Intuitively, two specified mRNAs that remain CW across multiple states can be assumed to belong to "housekeeping" processes. Conversely, two mRNAs that become DW in a state-dependent manner are likely related to each state's underlying biology (41) and can be informative just like DE mRNAs (14, 15, 39–41).

First, we compared normal breast and TNBC in Wh patients and found that a mere 2% of the mRNAs remain CW. Perhaps not surprisingly, these few CW pairs comprise positively correlated mRNAs that code for ribosomal proteins (for the full list of results, see Supplementary Table S5). Of the mRNA pairs that are coexpressed in normal breast, 45% are lost in TNBC (DW mRNAs). Of the mRNA pairs that are coexpressed in TNBC, 9% are not present in normal breast, and are DW as well.

These results speak to profound differences in coexpression between normal breast and TNBC. We highlight this observation with a few example subnetworks in Fig. 4A. In all these instances, the coexpression is absent in one of the states. Note that many

associations of the 40S ribosomal protein S8 (*RPS8*) mRNA with other mRNAs encoding ribosomal proteins remain unchanged (CW) between normal breast and TNBC. Other associations, however, are tissue-state specific. This indicates an extensive rewiring of the ribosomal protein network in Wh-TNBC through losses and gains of correlation.

Note also how the correlation of the signaling processes with triglyceride metabolism by means of the *PLIN1* hub mRNA is lost in TNBC. At the same time, the leptin receptor overlapping transcript (*LEPROT*) mRNA emerges as a hub in TNBC that links many mRNAs encoding ribosomal and mitochondrial proteins (Fig. 4A). These results indicate a wide-ranging deregulation of normal cell morphology and tissue physiology in the TNBC state.

The wiring of coexpressed mRNAs differs considerably between Wh TNBC patients and B/Aa TNBC patients

Next, we analyzed coexpressed mRNAs in Wh TNBC patients and B/Aa TNBC patients and found significant rewiring between the two races. This is consistent with the view that these two groups of patients are characterized by extensive molecular differences. Among the coexpressed mRNA pairs that are present in either patient group, 17% are CW in Wh and B/Aa TNBC patients, and include important components of tumor biology such as the ribosome, oxidative phosphorylation, and immune system responses. This is exemplified by the subnetwork shown in Fig. 4B (see also Supplementary Table S5). Here too, many correlations are gained and lost between the two races in TNBC.

A modest, yet considerable, fraction (7%) of coexpressed mRNA pairs are DW between the two states. These mRNAs belong to pathways that include proteasome, cell signaling, oxidative phosphorylation, transcription regulation, and other (Supplementary Table S5). They are captured by the "nonalcoholic fatty liver disease" label due to the latter's extensive molecular similarities with those pathways. As evidenced from Supplementary Table S5, for a few pathways (e.g., immune system, oxidative phosphorylation, Wnt signaling), a subset of their genes is CW between Wh and B/Aa patients, whereas a different subset is DW. This indicates that the corresponding mRNAs make race-specific contributions to the biology of TNBC.

Finally, we note that thioredoxin (*TXN*) is among the genes that are DW between TNBC patients of different race. *TXN* participates in multiple negative correlations in the B/Aa TNBC patients that are absent from the Wh TNBC patients. These negative correlations include Aquaporin 1 (*AQP1*) and leptin (*LEP*), suggesting a race-specific coordination of genes involved in redox balance and energy signaling/homeostasis.

(Continued.) The percentage numbers noted in red show the fraction of DW mRNA associations that cannot be attributed to DW with isomiRs or tRFs. For example, 86% and 97% of the DW associations between Wh-normal and Wh-TNBC can be attributed to DW with an isomiR and a tRF, respectively. The gray area connecting the columns represents the overlap. In the left plot, almost all of the DW mRNA pairs that are DW with isomiRs are also DW with tRFs. **D**, Alluvial diagrams showing how the isomiRs and tRFs are differentially wired with the DW mRNA associations between the normal breast and TNBC in Wh donors (top) and between Wh TNBC and B/Aa TNBC patients (bottom). Each "Pos" ("Neg" respectively) block indicates the portion of DW mRNA pairs that are positively (negatively, respectively) correlated with at least one isomiR/tRF in the corresponding state. The "Pos/Neg" block captures the portion of DW mRNA pairs that are both positively and negatively correlated with isomiRs/tRFs. "None" captures the portion of DW mRNA pairs that cannot be correlated with any isomiRs/tRF. **E**, DW of *RTN4* as a result of differential targeting by isomiRs. *RTN4* is DW with isomiRs and tRFs from several different loci when comparing the Wh-normal and Wh-TNBC state (top left). The archetype miRNA of miR-200b-3p has a putative target site only in one of the two mRNA isoforms of *RTN4* (top right). The *RTN4* mRNA isoforms are differentially expressed between the normal tissue and tumor in Wh patients (bottom boxplots). *, statistically significant differences ($P < 0.05$; Mann-Whitney U test). **A**, **B**, and **E**, The shown DW transcripts capture the absence of coexpression in one of the states being compared. **A**, **B**, and **E**, The edges connecting DW transcripts are colored by the state in which the transcripts are coexpressed.

The disruption of the wiring of coexpressed mRNAs in TNBC is race dependent and involves isomiRs and tRFs

Considering the above findings, we hypothesized that many of the DW mRNAs are associated with isomiRs or tRFs. Our analysis shows that for nearly all (99%) of the mRNA pairs that are DW between normal breast and TNBC in Wh patients, their wiring with isomiRs or tRFs is also DW between these states (shown in cyan in Fig. 4C). It is interesting to note that, despite being numerically fewer, tRFs can be linked with 10% more DW mRNAs than isomiRs. When comparing Wh TNBC with B/Aa TNBC patients, the results remain unchanged qualitatively, but now the two states exhibit more similarities (Fig. 4C). Characteristic rewiring examples are shown in Supplementary Fig. S3A and S3B.

Looking deeper into the links between DW mRNAs and isomiRs/tRFs in Wh patients, we find that the majority of DW mRNAs lose their links to isomiRs or tRFs between normal and TNBC (top half of Fig. 4D). Supplementary Figure S3C and S3D shows that isomiRs from the mir-200 family, which has been linked to metastasis in many settings (42, 43), and tRFs from the nuclear tRNA^{GlnCTG} are associated with several of the mRNAs that are DW between the Wh-normal and the Wh-TNBC states. When comparing Wh TNBC and B/Aa TNBC patients, we find that the DW mRNAs are linked with isomiRs or tRFs predominantly in one of the two races but not in the other (bottom half of Fig. 4D).

Taken together, these findings suggest that the associations between isomiRs/tRFs and mRNAs appear or disappear as a function of the patient's race and of the state of the tissue. On the basis of the fact that both isomiRs and tRFs enter the RNAi pathway, it is reasonable to assume that some of these associations are indicative of direct regulatory interactions that are active only in the specific biological context.

Potential contributors to the differential wiring of mRNAs could include splicing and decoying

One possible mechanism behind the DW between mRNAs and isomiRs could be the existence of splicing variants between normal breast and TNBC. For example, isomiRs could be targeting specifically one of the mRNA isoforms but not the other, possibly doing so in only one of the tissue states (44). We highlight this possibility through an example involving the Reticulon 4 (*RTN4*) gene. In Wh donors, *RTN4*'s mRNA is correlated with the mRNAs of other genes in normal breast but not in TNBC. The available data indicate that *RTN4* is correlated with several isomiRs and tRFs from multiple miRNA loci and tRNA isoacceptors (top left of Fig. 4E). *RTN4* has two mRNA isoforms. Using the rna22 miRNA target prediction tool (24), we find that only one of the two mRNA isoforms has a putative target site for the archetype miRNA of miR-200b-3p (top right of Fig. 4E). Interestingly, this putative target site is present in the mRNA of the transcript that is downregulated in Wh TNBC patients compared with normal breast (boxplots of Fig. 4E). The relative expression of these two mRNA isoforms changes between normal breast and tumor in Wh patients (Fig. 4E). Thus, it is conceivable that an alternative splicing event is responsible for the DW between miR-200b-3p and *RTN4* and the increase of the nontargeted isoform in TNBC.

Recently, it was shown that tRFs can also act by decoying RNA-binding proteins (12). In addition to expected alignments of tRFs with YBX1 and SSB RNA-binding motifs, our *in silico* studies show that HuR's binding motif matches tRF sequences from several

isoacceptors including the nuclear tRNA^{HisGTG} and tRNA^{AlaTGC} and the mitochondrial tRNA^{GluTTC} (Supplementary Table S6). HuR has long been studied for its ability to stabilize the mRNA transcripts of oncogenes, particularly under stress conditions (45). Thus, it is conceivable that tRFs interfere with HuR function, in a manner analogous to what was shown for YBX1 (12).

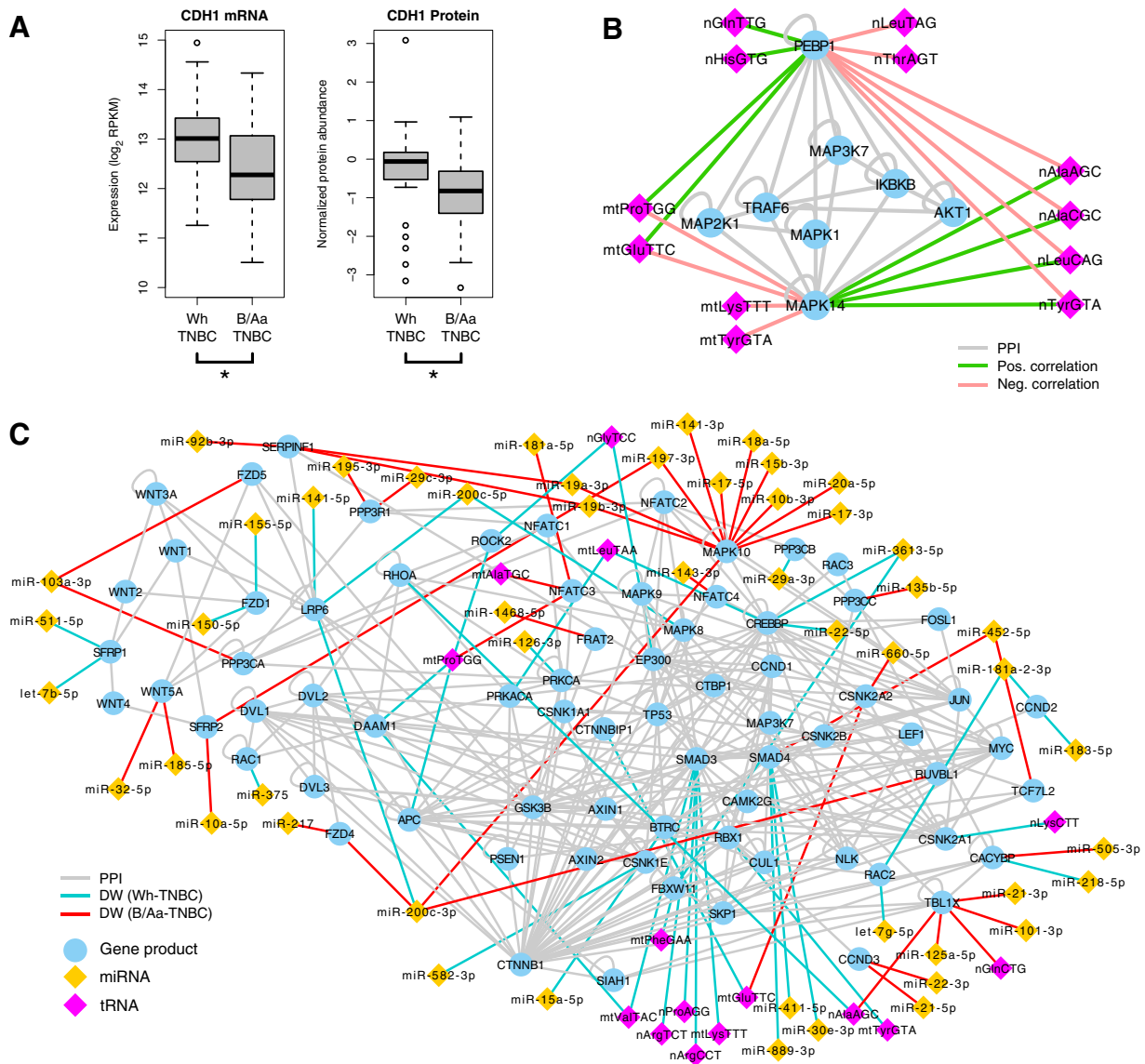
Genes from key metastasis-linked pathways show race-specific associations with isomiRs and tRFs

TNBC tumors in Wh and B/Aa patients exhibit striking differences in their metastatic potential (3, 5). Although not apparent from the genetic or clinical TCGA data (6), there is indirect evidence of an increased potential for metastases among the TCGA B/Aa patients. For example, tumors from these patients have lower expression of E-cadherin (*CDH1*) at both the mRNA and protein levels (Fig. 5A). Thus, we sought to understand the potential contribution of tRFs and isomiRs on the mRNA abundance of two known metastasis-associated pathways.

The P38 pathway. We first examined metastasis suppressors that have been studied for their potential to represent rate-limiting steps in the metastatic cascade (46, 47). Among them, 13 are found correlated with isomiRs and/or tRFs in TNBC in exactly one of the two races (Supplementary Table S7). Two of the 13 genes are noteworthy because of their multiple associations with tRFs: mitogen-activated protein kinase 14 (*MAPK14* or P38) and Raf kinase inhibitor protein (*RKIP*). *RKIP* is also known as phosphatidylethanolamine binding protein 1 (*PEBP1*). The *MAPK14* and *RKIP* mRNAs are correlated with tRFs but only in Wh TNBC patients. Notably, the mRNAs' correlations with tRFs have opposite signs (Fig. 5B). Specifically, several tRFs from mitochondrially encoded tRNAs, many of them from the MT tRNA^{ProTGG}, are correlated negatively with *MAPK14* and positively with *RKIP*. On the other hand, tRFs from nuclearly encoded tRNAs exhibit the opposite pattern (Fig. 5B).

We stress that these are nonaccidental correlations. Indeed, these two genes belong to the same signaling pathway, and each of the respective protein products interacts directly with several protein partners (48). We stress that the MT-derived tRFs shown in Fig. 5B are downregulated in the B/Aa TNBC patients. The observed dynamics suggests that the tRFs compete with one another for dynamic control of the abundances of these two genes. The loss of the MT tRFs in B/Aa TNBC patients leads potentially to the rewiring of this module with a consequent adverse impact on *RKIP*'s ability to inhibit the metastatic cascade.

The Wnt/ β -catenin pathway. The Wnt pathway has been receiving a lot of attention as a potential contributor to metastasis (49) and, also, to the health disparities in TNBC (5). In Fig. 5C, we depict a portion of the Wnt signaling pathway (32, 48, 50) with the PPI among its components shown in gray. Information on whether a gene's mRNA is DW with isomiRs or tRFs between Wh TNBC and B/Aa TNBC patients is shown overlaid (the full list of DW isomiRs/tRFs and mRNAs appears in Supplementary Table S7). Surprisingly, we find isomiRs and tRFs to be DW with numerous components of the pathway, including cell surface Frizzled receptors, intracellular signaling proteins, and gene expression regulators (Fig. 5C). It can be seen that the mRNAs that are DW with isomiRs originate from loci with well-documented roles in metastasis (43, 51). These loci include the miR-200c-3p and miR-21-5p

**Figure 5.**

Race-specific metastasis-related genes and pathways. **A**, mRNA and protein levels of E-cadherin (CDH1), a marker of metastatic potential, show differential expression between the two races in TNBC. *, $P < 0.05$; Mann-Whitney U test. **B**, DW tRFs and mRNAs participating in MAPK signaling. The shown correlations between tRFs and mRNAs are present only in Wh TNBC patients. Shown are tRNA loci (diamonds) that produce tRFs that are correlated with the mRNAs of the shown genes (circles) as well as PPIs among the protein products of these genes. *PEBP1* is also known as *RKIP* and *MAPK14* is known as *P38*. **C**, PPI network and DW associations of the corresponding mRNAs with isomiRs/tRFs of the Wnt/β-catenin signaling pathway as integrated from KEGG, Pickle, and UniProt databases. The shown DW associations correspond to an absence of coexpression in one of the states. The edges connecting DW transcripts are colored by the state in which the transcripts are coexpressed. Gene nodes are sorted on the basis of the cellular compartment of the encoded protein, from extracellular (left) to nuclear (right). The full list of how the Wnt/β-catenin pathway genes are DW with isomiRs and tRFs is included in Supplementary Table S7.

arms, the miR-17/92 miRNA cluster, and the miR-183/96/182 miRNA cluster. As indicated by the color of the edges connecting the DW mRNA and miRNA transcripts, these transcript pairs are coexpressed almost exclusively in B/Aa TNBC patients. Also, the DW mRNA-tRF pairs of transcripts are coexpressed almost exclusively in Wh TNBC patients. Finally, we note that more than half of the shown DW mRNA-tRF pairs involve tRFs from MT tRNAs, particularly MT tRNA^{ProTGG} and MT tRNA^{ValTAC}.

The various associations between isomiRs/tRFs and mRNAs from the two pathways mentioned above are but a small fraction of numerous such associations. As we depict in Fig. 4, other important pathways with similar differentially wired associations include immune system responses, oxidative phosphorylation, and cell adhesion molecules. Nonetheless, even the two examples that we showed above for the P38 and Wnt pathways suggest widespread, race-specific regulatory events that involve isomiRs

and tRFs. Presumably, these events shift B/Aa TNBC patients to a molecular state that greatly favors metastasis.

Discussion

We analyzed the molecular profiles of TNBC tumors from B/Aa and Wh patients and compared them with the profile of normal breast. We identified mRNAs, isomiRs, and tRFs that, in various combinations, are either DE or DW between normal breast and TNBC, and between patients belonging to the two races. Our findings suggest pronounced disruption of the normal breast's molecular physiology in TNBC. By dissecting the underlying links, we flagged isomiRs and tRFs as potentially very important drivers of this disruption and of the racial health disparities in TNBC.

Components of the proteasome are significantly upregulated in TNBC compared with normal breast tissue as are ribosomal proteins and ribosome biogenesis factors, in agreement with reports on altered protein dynamics at multiple stages of tumor biology (52). In addition to the pathways of metabolism, signaling, and immune response that are DE between the normal breast and TNBC (Fig. 1A), we found that several circadian rhythm genes, like *CLOCK* and *PER1* (Fig. 1C; Supplementary Table S1), are also downregulated in both Wh and B/Aa TNBC patients, in concordance with previous reports (53).

Expanding on our previous work (10, 11), we identified isomiRs and tRFs that are DE between normal tissue and TNBC. IsomiRs from the miR-21-5p arm, the miR-200 and miR-10 families, and the miR-183/96/182 cluster, all of which are known to be dysregulated in breast cancer (54), are DE in TNBC compared with normal in both races. Similarly, tRFs from tRNA^{HisGTG}, tRNA^{GlnTTG}, and tRNA^{GlnCTG} are DE in TNBC compared with normal breast.

A new finding pertains to the discovery of miRNA and tRNA loci, each of which produces two groups of transcripts with a curious property: each of the two groups from the locus is DE between normal and TNBC, but the two groups exhibit opposite signs in their change of abundance. Examples include: miR-146b-5p, miR-146b-3p, miR-223-3p, and several members (mir-30b, mir-30c, mir-30d, and mir-30e) of the mir-30 family (Supplementary Table S1). This is concordant with our previous report that isomiRs from the 5p arm of mir-183 affect the abundance of mRNAs in TNBC in opposing ways (11).

DE transcripts provide invaluable insights into pathways that deviate from their normal function. However, expression changes do not always capture the changes in the relationships among transcripts (28, 39). To determine which network nodes are critical to such changes, we used gene coexpression analysis (39), an approach that has yielded valuable insights in multiple contexts (28, 30, 55).

We also integrated the observed mRNA correlations with multimodal information. This allowed us to link putatively mRNA correlations with events that are temporally upstream (e.g., regulation by a common TF, isomiR, or tRF) or downstream (e.g., participation of the protein product in common processes). We found that TFs, isomiRs, and tRFs often have a profound effect on mRNA wiring. The other modalities could not explain adequately the mRNA–mRNA correlations, something that was observed previously by others (55) and us (56).

With regard to TFs, in particular, we leveraged ENCODE's experimentally identified binding sites (36, 37) from multiple cellular settings. Because the ENCODE data reflect a TF's complete

ciptome and not events that are specific to a tissue or a tissue state (57), the portion of mRNA correlation pairs that TFs could account for remained unchanged across tissue states and patient races (Supplementary Table S3). This indicates a potential overestimation by our analyses of the TFs' influence (Fig. 3B).

Examination of the changes in the interaction network from normal breast to TNBC allowed us to trace the CW and DW of mRNAs to differences in the associations of mRNAs with isomiRs and tRFs. We found widespread loss of these associations in TNBC compared with normal breast. Presumably, some of these associations correspond to regulatory activity by the isomiRs/tRFs, which would suggest a profound loss of the isomiRs' and tRFs' regulatory capacity in TNBC. In fact, we find that only 2% of the mRNA–mRNA associations are CW in the normal breast and TNBC in Wh patients.

IsomiRs and tRFs appear to have a very broad impact on tumorigenesis and many other aspects of TNBC biology (43). We found analogous molecular differences between Wh TNBC and B/Aa TNBC patients, which adds support to previous reports by others (3, 5–9) and us (10, 11) that molecular causes underlie health disparities in TNBC. However, to resolve whether the loss of regulatory capacity drives or results from this widespread mRNA rewiring will require exhaustive time-series data. We also note that based on the TCGA data, there is a marked inability of tRFs to be connected with mRNAs in B/Aa TNBC patients.

With regard to disparities, we focused on metastasis, one aspect of TNBC where Wh and B/Aa patients exhibit well-documented differences. Starting with metastasis inhibitor genes and pathways (46, 47), we sought mRNAs that are DW with tRFs or isomiRs and provided evidence of race-based disparities in the context of two pathways:

- (a) P38: We found the axis between MAPK14 (P38) and RKIP to be DW with MT and nuclear tRFs. RKIP regulates Raf activity (46, 47), thereby inhibiting MAPK activation. The coordinated association of tRFs with these two mRNAs suggests that tRFs likely maintain a balance that becomes disrupted in B/Aa TNBC patients, thereby increasing the tumor's potential to metastasize.
- (b) Wnt/ β -catenin: We established associations of key Wnt members with several miRNA loci (e.g., miR-21, miR-29, miR-200, miR-19a, miR-150) and tRNA isoacceptors (e.g., MT tRNA^{ProTGG} and tRNA^{ValTAC}). Although the implications of these associations cannot be inferred directly from RNA expression studies, our results suggest extensive regulation of Wnt signaling by isomiRs and tRFs. The dynamics of this regulation appear to be affected by the biological context and to be altered in TNBC. We note that many of the identified miRNAs are linked to androgen receptor (AR) signaling (58), and AR is emerging as an important pathway in breast cancer and TNBC (59). Thus, our data may describe a race-specific interaction of AR and Wnt/ β -catenin signaling and miRNAs.

Interestingly, we could separate the normal breast samples by race (Fig. 1). B/Aa women are known to exhibit higher tissue inflammation that results in differences of fat tissue physiology compared with Wh women (5). This difference likely predisposes them to a more aggressive instance of TNBC. The existence of a different "starting" point (38) for B/Aa women is also supported by recent findings that stem cells in normal breast tissue exhibit ethnicity dependencies (60).

These race dependencies appear to extend to the trajectory's "ending" point as well, that is, to TNBC. Indeed, we found several immune system response genes to be DW between races in the TNBC state. In addition, we found several aspects of lipid metabolism to be DW between normal breast and TNBC in Wh women. These findings indicate that dissecting the complexity of racial disparities in TNBC will benefit from system biology approaches.

In summary, we provided a systems perspective on the transcriptomic differences between normal breast and TNBC, and, on the racial disparities in TNBC. We generated extensive evidence supporting a central role for miRNAs and tRFs in the regulation of mRNAs in TNBC (61). Our approach provides a framework in which to generate novel hypotheses that link specific mRNAs and cellular processes to specific isomiRs and tRFs. Our findings also underline the importance and value of integrative systems and network analyses in tackling the complexity of dynamic and context-dependent molecular interactions (14, 28, 62).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7–30.
2. Dawson SJ, Provenzano E, Caldas C. Triple negative breast cancers: clinical and prognostic implications. *Eur J Cancer* 2009;45Suppl 1:27–40.
3. Ademuyiwa FO, Edge SB, Erwin DO, Orom H, Ambrosone CB, Underwood W III. Breast cancer racial disparities: unanswered questions. *Cancer Res* 2011;71:640–4.
4. Williams DR, Mohammed SA, Shields AE. Understanding and effectively addressing breast cancer in African American women: Unpacking the social context. *Cancer* 2016;122:2138–49.
5. Dietze EC, Sistrunk C, Miranda-Carboni G, O'Regan R, Seewaldt VL. Triple-negative breast cancer in African-American women: disparities versus biology. *Nat Rev Cancer* 2015;15:248–54.
6. Ademuyiwa FO, Tao Y, Luo J, Weilbaeher K, Ma CX. Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians. *Breast Cancer Res Treat* 2017;161:491–9.
7. Lindner R, Sullivan C, Offor O, Lezon-Geyda K, Halligan K, Fischbach N, et al. Molecular phenotypes in triple negative breast cancer from African American patients suggest targets for therapy. *PLoS One* 2013;8:e71915.
8. Stewart PA, Luks J, Roycik MD, Sang QX, Zhang J. Differentially expressed transcripts and dysregulated signaling pathways and networks in African American breast cancer. *PLoS One* 2013;8:e82460.
9. Sugita B, Gill M, Mahajan A, Duttargi A, Kirolikar S, Almeida R, et al. Differentially expressed miRNAs in triple negative breast cancer from African-American and non-Hispanic white women. *Oncotarget* 2016;7:79274–91.
10. Telonis AG, Loher P, Honda S, Jing Y, Palazzo J, Kirino Y, et al. Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. *Oncotarget* 2015;6:24797–822.
11. Telonis AG, Loher P, Jing Y, Londin E, Rigoutsos I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res* 2015;43:9158–75.
12. Goodarzi H, Liu X, Nguyen HC, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* 2015;161:790–802.
13. Shigematsu M, Kirino Y. tRNA-derived short non-coding RNA as interacting partners of argonaute proteins. *Gene Regul Syst Bio* 2015;9:27–33.
14. Ahn AC, Tewari M, Poon CS, Phillips RS. The clinical applications of a systems approach. *PLoS Med* 2006;3:e209.
15. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12:56–68.
16. The Cancer Genome Atlas Research Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
17. Loher P, Telonis AG, Rigoutsos I. MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Sci Rep* 2017;7:41184.
18. Telonis AG, Magee R, Loher P, Chervoneva I, Londin E, Rigoutsos I. Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res* 2017;45:2973–85.
19. Magee R, Loher P, Londin E, Rigoutsos I. Threshold-seq: a tool for determining the threshold in short RNA-seq datasets. *Bioinformatics* 2017;33:2034–6.
20. Li J, Lu Y, Akbari R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. *Nat Methods* 2013;10:1046–7.
21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
22. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
24. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 2006;126:1203–17.
25. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 1988;85:2444–8.
26. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. *Database* 2016;2016:baw093.
27. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res* 2011;39:D301–8.
28. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol* 2007;3:78.
29. Hudson NJ, Reverter A, Dalrymple BP. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* 2009;5:e1000382.
30. Reznik E, Sander C. Extensive decoupling of metabolic genes in cancer. *PLoS Comput Biol* 2015;11:e1004176.

Authors' Contributions

Conception and design: A.G. Telonis, I. Rigoutsos
Development of methodology: A.G. Telonis, I. Rigoutsos
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.G. Telonis
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.G. Telonis, I. Rigoutsos
Writing, review, and/or revision of the manuscript: A.G. Telonis, I. Rigoutsos
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.G. Telonis, I. Rigoutsos
Study supervision: A.G. Telonis, I. Rigoutsos

Acknowledgments

This work was supported by a William M. Keck Foundation grant (I. Rigoutsos), NIH/NCIR21-CA195204 (I. Rigoutsos), and by Institutional Funds (I. Rigoutsos).

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

Received July 7, 2017; revised October 19, 2017; accepted November 30, 2017; published OnlineFirst December 11, 2017.

31. Albert R. Scale-free networks in cell biology. *J Cell Sci* 2005;118:4947–57.
32. Consortium TU. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;45:D158–D69.
33. Gioutlakis A, Klapa MI, Moschonas NK. PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS One* 2017;12:e0186039.
34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
35. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60.
36. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012;489:91–100.
37. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* 2012;22:1798–812.
38. Mar JC, Quackenbush J. Decomposition of gene expression state space trajectories. *PLoS Comput Biol* 2009;5:e1000626.
39. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav* 2014;13:13–24.
40. Padi M, Quackenbush J. Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst Biol* 2015;9:80.
41. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol* 2012;8:565.
42. Rigoutsos I, Lee SK, Nam SY, Anfossi S, Pasculli B, Pichler M, et al. N-BLR, a primate-specific non-coding transcript leads to colorectal cancer invasion and migration. *Genome Biol* 2017;18:98.
43. Shah MY, Ferrajoli A, Sood AK, Lopez-Berestein G, Calin GA. microRNA therapeutics in cancer - an emerging concept. *EBioMedicine* 2016;12:34–42.
44. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 2008;320:1643–7.
45. Abdelmohsen K, Gorospe M. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdiscip Rev RNA* 2010;1:214–29.
46. Bohl CR, Harihar S, Denning WL, Sharma R, Welch DR. Metastasis suppressors in breast cancers: mechanistic insights and clinical potential. *J Mol Med* 2014;92:13–30.
47. Stafford LJ, Vaidya KS, Welch DR. Metastasis suppressors genes in cancer. *Int J Biochem Cell Biol* 2008;40:874–91.
48. Klapa MI, Tsafou K, Theodoridis E, Tsakalidis A, Moschonas NK. Reconstruction of the experimentally supported human protein interactome: what can we learn? *BMC Syst Biol* 2013;7:96.
49. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* 2017;36:1461–73.
50. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45:D353–D61.
51. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* 2013;20:1603–14.
52. Demarchi F, Brancolini C. Altering protein turnover in tumor cells: new opportunities for anti-cancer therapies. *Drug Resist Updat* 2005;8:359–68.
53. Blakeman V, Williams JL, Meng QJ, Streuli CH. Circadian clocks and breast cancer. *Breast Cancer Res* 2016;18:89.
54. Kurozumi S, Yamaguchi Y, Kurosumi M, Ohira M, Matsumoto H, Horiguchi J. Recent trends in microRNA research into breast cancer with particular focus on the associations between microRNAs and intrinsic subtypes. *J Hum Genet* 2017;62:15–24.
55. Aure MR, Jernstrom S, Krohn M, Vollan HK, Due EU, Rodland E, et al. Integrated analysis reveals microRNA networks coordinately expressed with key proteins in breast cancer. *Genome Med* 2015;7:21.
56. Londin ER, Hatzimichael E, Loher P, Edelstein L, Shaw C, Delgrosso K, et al. The human platelet: strong transcriptome correlations among individuals associate weakly with the platelet proteome. *Biol Direct* 2014;9:3.
57. MacNeil LT, Pons C, Arda HE, Giese GE, Myers CL, Walhout AJ. Transcription factor activity mapping of a tissue-specific in vivo gene regulatory network. *Cell Syst* 2015;1:152–62.
58. ChunJiao S, Huan C, ChaoYang X, GuoMei R. Uncovering the roles of miRNAs and their relationship with androgen receptor in prostate cancer. *IUBMB Life* 2014;66:379–86.
59. Barton VN, D'Amato NC, Gordon MA, Christenson JL, Elias A, Richer JK. Androgen receptor biology in triple negative breast cancer: a case for classification as AR+ or quadruple negative disease. *Horm Cancer* 2015;6:206–13.
60. Nakshatri H, Anjanappa M, Bhat-Nakshatri P. Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. *Sci Rep* 2015;5:13526.
61. Bracken CP, Scott HS, Goodall GJ. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat Rev Genet* 2016;17:719–32.
62. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;144:986–98.