


RESEARCH ARTICLE

Open Access



The subcellular distribution of miRNA isoforms, tRNA-derived fragments, and rRNA-derived fragments depends on nucleotide sequence and cell type

Tess Cherlin^{1,3}, Yi Jing¹, Siddhartha Shah¹, Anne Kennedy^{1,4}, Aristeidis G. Telonis^{1,5}, Venetia Pliatsika^{1,6}, Haley Wilson^{1,4}, Lily Thompson^{1,7}, Panagiotis I. Vlantis^{1,8}, Phillipe Loher¹, Benjamin Leiby² and Isidore Rigoutsos^{1*} 

Abstract

Background MicroRNA isoforms (isomiRs), tRNA-derived fragments (tRFs), and rRNA-derived fragments (rRFs) represent most of the small non-coding RNAs (sncRNAs) found in cells. Members of these three classes modulate messenger RNA (mRNA) and protein abundance and are dysregulated in diseases. Experimental studies to date have assumed that the subcellular distribution of these molecules is well-understood, independent of cell type, and the same for all isoforms of a sncRNA.

Results We tested these assumptions by investigating the subcellular distribution of isomiRs, tRFs, and rRFs in biological replicates from three cell lines from the same tissue and same-sex donors that model the same cancer subtype. In each cell line, we profiled the isomiRs, tRFs, and rRFs in the nucleus, cytoplasm, whole mitochondrion (MT), mitoplast (MP), and whole cell. Using a rigorous mathematical model we developed, we accounted for cross-fraction contamination and technical errors and adjusted the measured abundances accordingly. Analyses of the adjusted abundances show that isomiRs, tRFs, and rRFs exhibit complex patterns of subcellular distributions. These patterns depend on each sncRNA's exact sequence and the cell type. Even in the same cell line, isoforms of the same sncRNA whose sequences differ by a few nucleotides (nts) can have different subcellular distributions.

Conclusions SncRNAs with similar sequences have different subcellular distributions within and across cell lines, suggesting that each isoform could have a different function. Future computational and experimental studies of isomiRs, tRFs, and rRFs will need to distinguish among each molecule's various isoforms and account for differences in each isoform's subcellular distribution in the cell line at hand. While the findings add to a growing body of evidence that isomiRs, tRFs, rRFs, tRNAs, and rRNAs follow complex intracellular trafficking rules, further investigation is needed to exclude alternative explanations for the observed subcellular distribution of sncRNAs.

Keywords Small non-coding RNAs, sncRNAs, microRNAs, miRNAs, miRNA isoforms, isomiRs, tRNA-derived fragments, tRFs, rRNA-derived fragments, rRFs, subcellular distribution

Co-authors Kennedy, Wilson and Thompson were with the Computational Medicine Center when the work was carried out.

*Correspondence:

Isidore Rigoutsos
isidore.rigoutsos@jefferson.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Key points and significance

- We studied the subcellular distribution of microRNA isoforms (isomiRs), tRNA-derived fragments (tRFs), and rRNA-derived fragments (rRFs) in deep sequencing data from biological replicates of three cell lines.
- We focused on four fractions:
 - Nucleus
 - Cytoplasm
 - Whole mitochondrion (MT)
 - Mitoplast (MP)
- To minimize the number of independent variables, we chose cell lines from a single tissue (breast) and same-sex donors (women) that model the same cancer subtype (triple negative breast cancer).
- To account for cross-fraction contamination and technical errors, we developed a rigorous mathematical model to adjust the abundances of isomiRs, tRFs, and rRFs in each fraction.
- We find that the subcellular distribution of small RNAs is more complex than has been assumed to date:
 - IsomiRs, tRFs, and rRFs from the same parental RNA whose sequences differ by as few as 1–2 nucleotides can have different subcellular distributions.
 - The subcellular distribution of an isomiR, tRF, or rRF depends on its exact nucleotide sequence and the cell line at hand.
- The findings suggest that an isomiR, tRF, or rRF could function differently in different cell lines because of differences in their subcellular distributions.
- The findings add to a growing body of evidence supporting a complex subcellular trafficking program for small RNAs. Further investigation is required to exclude alternative explanations.

Background

We previously showed that personal attributes (e.g., sex, ancestry, age) and context (e.g., cell type, tissue type, disease) modulate the sequences and abundances of isomiRs, tRFs, and rRFs [1–11]. The observed differences include abundance changes and sequence changes that can alter the identity of the targeted mRNAs [5]. Because of the sheer number of isomiRs [6, 12], tRFs [9, 13], and rRFs [1, 14], we know little about their function. Knowing *which* isomiRs, tRFs, and rRFs are produced in each cell and *where* within a cell they localize can provide valuable

initial clues about their function. Despite that, systematic studies of the subcellular distribution of sncRNAs have been limited [15–27] by comparison to analogous studies of long non-coding RNAs (lncRNAs) [25, 28–36].

IsomiRs, tRFs, and rRFs arise from templates in the genomes of the nucleus and MT. Others and we have provided evidence that the nuclear genome harbors many thousands of miRNA precursors [37–42]. However, these findings were recently questioned [43]. The nuclear and MT genomes also harbor multiple tRNA genes [44] and multiple copies of four nuclear rRNA genes (5S, 5.8S, 18S, 28S) and the two MT rRNAs (12S, 16S) [45–47]. Numerous miRNA precursors have been shown to produce templated and non-templated [3, 6, 12] isomiRs simultaneously, but the biogenesis mechanisms are not well understood [48, 49]. Nuclear and MT tRNAs and rRNAs carry out their canonical functions during protein translation [50, 51] while also producing tRFs [52, 53] and rRFs [54–56], respectively. Notwithstanding a few exceptions [52, 57–59], the mechanism that produces tRFs and rRFs is poorly understood. Regarding function, some isomiRs, tRFs, and rRFs regulate mRNA and protein abundance by entering the RNA interference pathway [52, 55, 60]. For tRFs and rRFs, additional modes of action are known [53, 55].

While miRNAs are typically expected to localize to the cytoplasm, previous studies showed they can also localize and function in the nucleus [17, 21, 23, 24, 61–65] and *at* or *in* the MT [25–27]. However, these studies did not investigate potential differences in the subcellular distribution of individual isomiRs or a possible dependence on cell type. Regarding tRFs, the findings of a previous large-scale study [9] suggest that tRFs localize to the nucleus, cytoplasm, and MT in various combinations [53]. Even less is known about the subcellular distribution of rRFs, a recently emerged class of sncRNAs [1, 14, 55].

In what follows, we systematically explore the subcellular distribution of isomiRs, tRFs, and rRFs in three cell lines (BT-20, MDA-MB-231, MDA-MB-468) that model triple-negative breast cancer (TNBC). We chose these cell lines to minimize the number of independent variables (e.g., sex, tissue, disease) on which the expression of isomiRs, tRFs, and rRFs is known to depend. For each cell line, we generated and analyzed data from several biological replicates, four subcellular compartments (nucleus, cytoplasm, MT, and MP), and the whole cell. Using a mixed-effects model, we accounted for cross-fraction contamination and technical errors, reconstructed the true abundance of each molecule in each fraction, and analyzed the resulting distributions.

We note that the presented work focuses only on those sncRNAs visible to the “standard RNA-seq protocol,” which has been the overwhelming method of choice for the last 15 years and used to generate most publicly available datasets. Using this protocol, one can enumerate only molecules whose termini contain a 5′-P and a 3′-OH [56]. We emphasize this point because others and we have shown that many typical sncRNAs have co-expressed *atypical* variants that are very abundant [56, 66]. These atypical variants are “invisible” to standard deep sequencing but “visible” to northern blots, albeit subject to the latter method’s sensitivity.

Results

We used three cell lines that model TNBC: BT-20, MDA-MB-231, and MDA-MB-468. From each cell line, and separately for each biological replicate, we isolated the nuclear, cytoplasmic, MT (whole MT), and MP fractions (Additional file 1: Fig. S1). We assayed their purity using

fraction-specific protein markers (see the “Methods” section and Additional file 2: Fig. S2). For three biological replicates for each of BT-20 and MDA-MB-231 cells and four replicates for MDA-MB-468, we deep-sequenced the corresponding nuclear, cytoplasmic, MT, and MP fractions and total RNA (a total of 50 preparations). We profiled the isomiRs, tRFs, and rRFs separately in each preparation. Figure 1 captures the workflow. Throughout this presentation, we will refer to the physical material that we isolated as cell “fractions” and the source organelles—nucleus, cytoplasm, (whole) MT, and MP—as “compartments.”

Each TNBC model cell line has a unique sncRNA profile

We reasoned that if these cell lines, which are from the same tissue (breast), have meaningful differences, then these differences should be evident if we considered *only* their abundant molecules. Therefore, we initially focused on only those sncRNAs whose abundance in total RNA

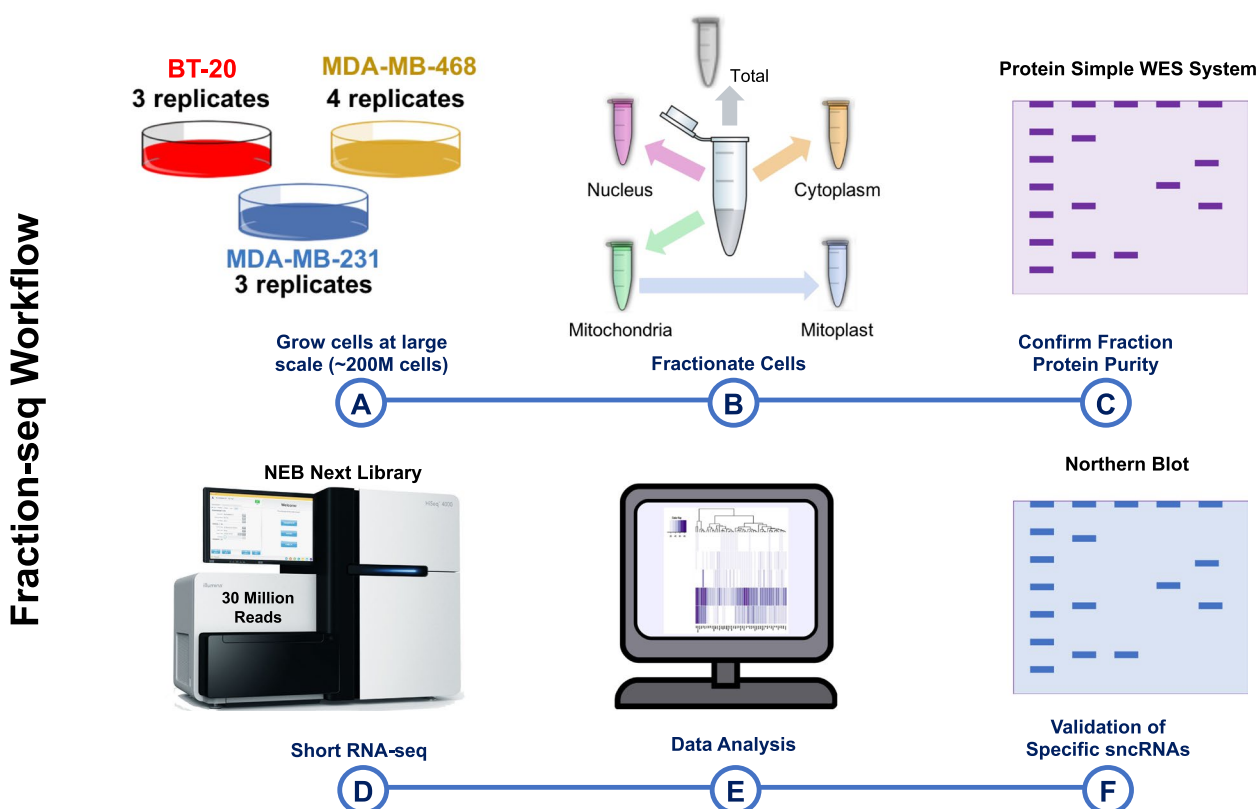


Fig. 1 Overview of the Fraction-seq workflow. **A** Three TNBC cell lines BT-20 (three replicates), MDA-MB-231 (three replicates), and MDA-MB-468 (four replicates) were grown to ~200 M cells and harvested. **B** From the same starting material, we separated cells into total, nuclear, cytoplasmic, MT, and MP subcellular compartments. **C** We used the WES protein detection assay to analyze cell compartment markers in each cell fraction. **D** We prepared libraries for short RNA-seq using NEBNext with 100 ng RNA from each sample, followed by RNA sequencing using the Illumina NextSeq 500 platform at 75 cycles and an average depth of 30 million reads per biosample. **E** After quality trimming and adapter removal, we mapped reads to isomiRs, tRFs, and rRFs, considering only sncRNAs with lengths between 18 and 50 nts and normalizing their abundance by the respective biospecimen’s sequencing depth. Reads with normalized abundances ≤ 10 RPM were not considered. **F** We used northern blotting to validate the subcellular enrichments of select sncRNAs. Parts of this image were created with BioRender.com

averaged across each cell line's replicates was ≥ 10 RPM (Additional file 3: Table S1). Figure 2A shows a Principal Component Analysis (PCA) of the three cell lines using the 7165 isomiRs, tRFs, and rRFs that satisfy the 10 RPM threshold. The replicates of each cell line—BT-20 (orange), MDA-MB-231 (green), MDA-MB-468 (cyan)—cluster with one another. Figure 2B shows a hierarchical clustering and heatmap of the cell line replicates after computing a Pearson correlation on the same 7165 sncRNAs. Figure 2C–E show that the clustering of each cell line's replicates persists when we apply the PCA method to each RNA class separately.

Many abundant isomiRs, tRFs, and rRFs are unique to a cell line

We focused on total RNA and compared the 10% most abundant sncRNAs across the three cell lines (Additional file 4: Table S2). Figure 3A shows the resulting overlap. Note how the most abundant sncRNAs differ among the three cell lines, with only 17 molecules common to all three. Even though these 17 molecules are common to all

three cell lines, their abundance profiles in each cell line differ characteristically (Additional file 5: Fig. S3).

We also extended the analyses using DESeq2 to identify all sncRNAs in total RNA that are differentially abundant (DA) between pairs of cell lines. We only considered sncRNAs with length ≥ 18 nts with abundance ≥ 10 RPM. At a false discovery rate (FDR) threshold of 0.05 and requiring $|\log_2(\text{fold change})| \geq 1$, we found 2983 isomiRs, 532 tRFs, and 4765 rRFs that are DA in various pairwise cell-line combinations (Additional file 6: Table S3).

Some isomiRs, tRFs, and rRFs show preferential enrichments in specific compartments

Case 1: same compartment in different cell lines

In analogy to comparing total RNA from different cell lines, we can compare the contents of the same fraction in different cell lines. For those molecules whose average abundance in total RNA is ≥ 10 RPM (Additional file 7: Fig. S4), we found many molecules in each of the four compartments that showed statistically significant (FDR ≤ 0.05) abundance differences (Additional file 8: Table S4) across cell lines. Most of the DA molecules are rRFs.

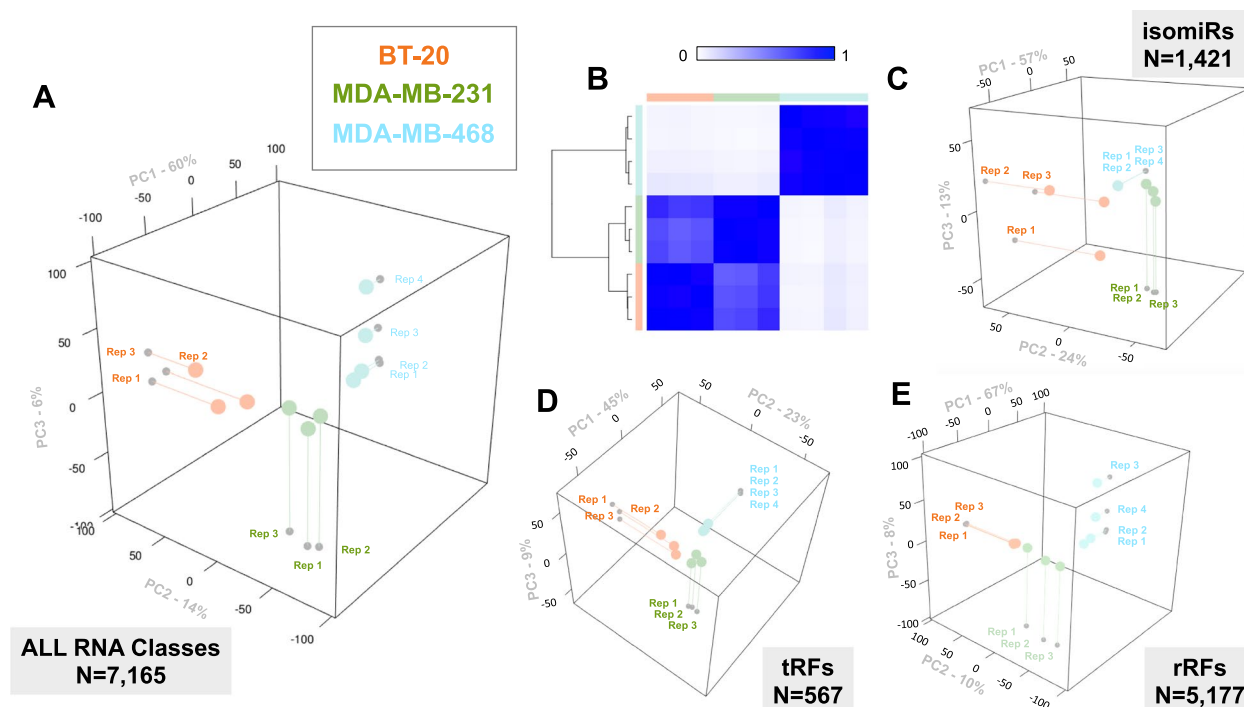


Fig. 2 The profiles of sncRNAs among the replicates are reproducible. **A** Principal Component Analysis (PCA) visualization of the sncRNAs with an average abundance ≥ 10 RPM across replicates ($n = 7165$). **B** Hierarchical clustering of the Pearson correlations for replicates from each cell line with sncRNA abundances ≥ 10 RPM. The color bar represents the Pearson correlation values, which range from 0 (white, uncorrelated) to 1 (blue, perfectly correlated). **C** PCA visualization of isomiRs in each sample with an average abundance ≥ 10 RPM across replicates ($n = 1421$). **D** PCA visualization of the tRFs in each sample with an average abundance ≥ 10 RPM across replicates ($n = 567$). **E** PCA of the rRFs in each sample with an average abundance ≥ 10 RPM across replicates ($n = 5177$). All panels: BT-20 samples are represented by orange, MDA-MB-231 samples are represented by green, and MDA-MB-468 samples are represented by cyan

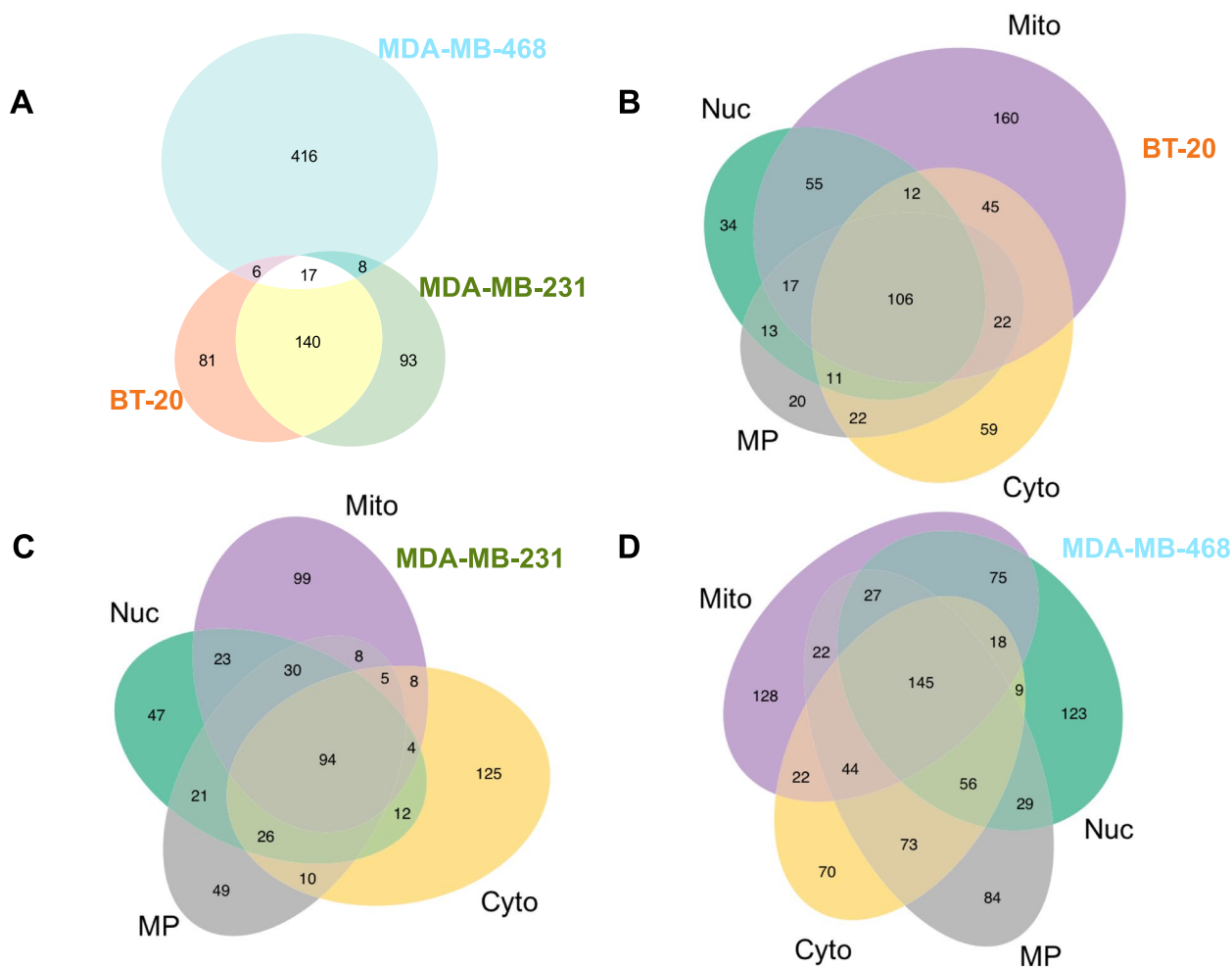


Fig. 3 The most abundant sncRNAs are cell-line- and subcellular-compartment-specific. **A** Venn diagram of the 10% most abundant sncRNAs in the three TNBC cell lines (BT-20, MDA-MB-231, and MDA-MB-468) based on average abundance. **B–D** Venn diagrams of the 10% most abundant sncRNAs in each compartment of BT-20 (**B**), MDA-MB-231 (**C**), and MDA-MB-468 (**D**) cells based on average abundance across replicates. Colors for panels **B–D**: nucleus (Nuc)—green, cytoplasm (Cyto)—yellow, MT—purple, and MP—gray

Case 2: different compartments in the same cell line

Unlike the comparisons of total RNA from different cell lines, we cannot compare *quantitatively* two fractions from the *same* cell line [67] without the rigorous mathematical modeling described in the next section. Nonetheless, we can compare two fractions *qualitatively* and still reach important conclusions. Focusing on the 10% most abundant sncRNAs in a fraction (Additional file 9: Table S5) and comparing them pairwise for each cell line in turn results in the Venn diagrams of Fig. 3B–D. We can see that there are abundant RNAs that are highly enriched in specific compartments.

We can also use a molecule's abundance in a fraction to determine its rank in that fraction, then convert the rank to a percentile value (100th percentile = most abundant sncRNA, 0th percentile = least abundant sncRNA). Then, we can use the percentile values of sncRNAs in the

different fractions to provide initial evidence of compartment-specific preferences. Figure 4A–B and Additional file 10: Fig. S5 show several examples for sncRNAs from MDA-MB-231, MDA-MB-468, and BT-20. The figures show heatmaps of color-coded percentiles averaged over the corresponding replicates. Each heatmap lists three groups of sncRNAs with dotted red lines separating the groups. The top, middle, and bottom groups comprise sncRNAs that rank highest in the nucleus, cytoplasm, and MT/MP, respectively. In all three heatmaps, we can see multiple pairs of sncRNAs whose percentile values in total RNA are very similar; this indicates that they have the same abundance in total RNA. However, the same pairs have different relative abundances (i.e., percentile) in one or more of the shown fractions. Or it may be that one member of the pair is absent from a fraction. These observations suggest differential subcellular distributions

for the sncRNAs forming the pair, as contamination cannot adequately explain the changes in the sncRNAs' relative abundances.

Many pairs for which the relative abundance in total RNA is not mirrored in fractions of the same cell line include sncRNAs with similar sequences and sncRNAs that arise from the same parental RNA. An example of the former includes isomiRs from the miRNAs miR-30a, miR-30e, and miR-30c in MDA-MB-468 cells. An example of the latter includes rRFs from 28S rRNA between locations 11600 and 13000 of the 45S cassette in MDA-MB-468.

We can adjust measured abundances by determining cross-fraction contamination and errors

While effective for very abundant sncRNAs, the percentile approach described in the previous section cannot be used with less abundant sncRNAs or in cases where the preferential enrichment in a subcellular compartment is less prominent. To identify such sncRNAs, we developed a mixed-effects model (see [Methods](#)) that accounts for technical errors and estimates and removes cross-fraction contamination from the various replicates. Essentially, the method reconstructs the true value of each sncRNA in each fraction by appropriately adjusting the values obtained from deep sequencing. Lastly, and separately for each cell line, we rescaled each sncRNA's reconstructed values in the nuclear, cytoplasmic, and MT fractions by constraining them by their values in total RNA.

With the recalculated abundances for the various sncRNAs, we can now *quantitatively* compare fractions from the same cell line and determine each sncRNA's subcellular distribution. Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8 list the reconstructed average values for each sncRNA and subcellular compartment for BT-20, MDA-MB-231, and MDA-MB-468. Only molecules whose recalculated abundance exceeds threshold are kept. In BT-20, there are 2185 molecules of which 972 are isomiRs, 275 tRFs, and 938 rRFs. In MDA-MB-231, there are 2347 molecules of which 847 are isomiRs, 296 tRFs, and 1204 rRFs. Lastly, in MDA-MB-468, there are 3732 molecules of which 142 isomiRs, 174 tRFs, and 3416 rRFs.

The mixed-effects model also allows us to estimate the relative contribution to total RNA by each subcellular compartment. The estimated “nucleus:cytoplasm:MT” ratios for the three cell lines are: 23.0%:61.5%:15.5% for BT-20; 36.5%:31.0%:32.5% for MDA-MB-231; and 32.2%:50.8%:17.0% for MDA-MB-4682.

Many sncRNAs, including multiple isoforms of the same sncRNA, exhibit complex subcellular distributions

Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8 show multiple sncRNAs with different subcellular distributions in different cell lines. Rather notable are the cases of isoforms of the same sncRNA, which typically differ by a few nts, and of sncRNAs produced from different regions of the same parental RNA yet exhibit different subcellular distributions within the same cell line.

For example, we highlight the case of the polycistronic miR-17/92 cluster [68, 69]. The cluster, also known as “oncomiR-1,” arises as a *single transcript* that spans several hundred nucleotides and harbors six miRNA precursors (mir-17, mir-18, mir-19a, mir-19b, mir-20a, and mir-92a). Figure 5 shows recalculated abundances (color-coded RPM values) for all isomiRs that pass threshold in either the nucleus, cytoplasm, or whole MT of BT-20 and MDA-MB-231 cells (Additional file 11: Table S6 and Additional file 12: Table S7). The isomiRs highlighted in green are the reference isoforms one finds in miRBase [41]. We can see that the abundance and subcellular distribution of these isomiRs, including the six reference miRNAs, differ widely even though all arise from the same polycistronic transcript [68] and many have similar sequences. Note also how many isomiRs with complex subcellular distributions are non-canonical or non-templated (the non-templated nts are marked in red, bold-face) and have subcellular distributions that differ by cell line. For some of the precursors, the reference mature miRNA is absent: e.g., miR-19a-3p|0|0| is absent from both cell lines.

Another notable finding is the prevalence of *non-templated* isomiRs containing a single guanine addition to their 3' terminus (“+1G”) in the nuclear fraction of MDA-MB-231 cells. On the other hand, *templated* isomiRs that end with guanine do not exhibit this localization preference. Our finding is concordant with a previous report

(See figure on next page.)

Fig. 4 Examples of subcellular enrichments using abundance-based ranking within a fraction. **A** Percentile-colored heatmap for select sncRNAs in total RNA and the nucleus, cytoplasm, MT, and MP for MDA-MB-231. **B** An analogous heatmap for MDA-MB-468 cells and a different collection of sncRNAs. For each sncRNA, we list its nucleotide sequence, the parental RNA from which the sncRNA arises, and the location of the sncRNA within the parental RNA. Red dotted lines separate each list into three groups: sncRNAs that are enriched primarily in the nucleus, the cytoplasm, or the MT, respectively

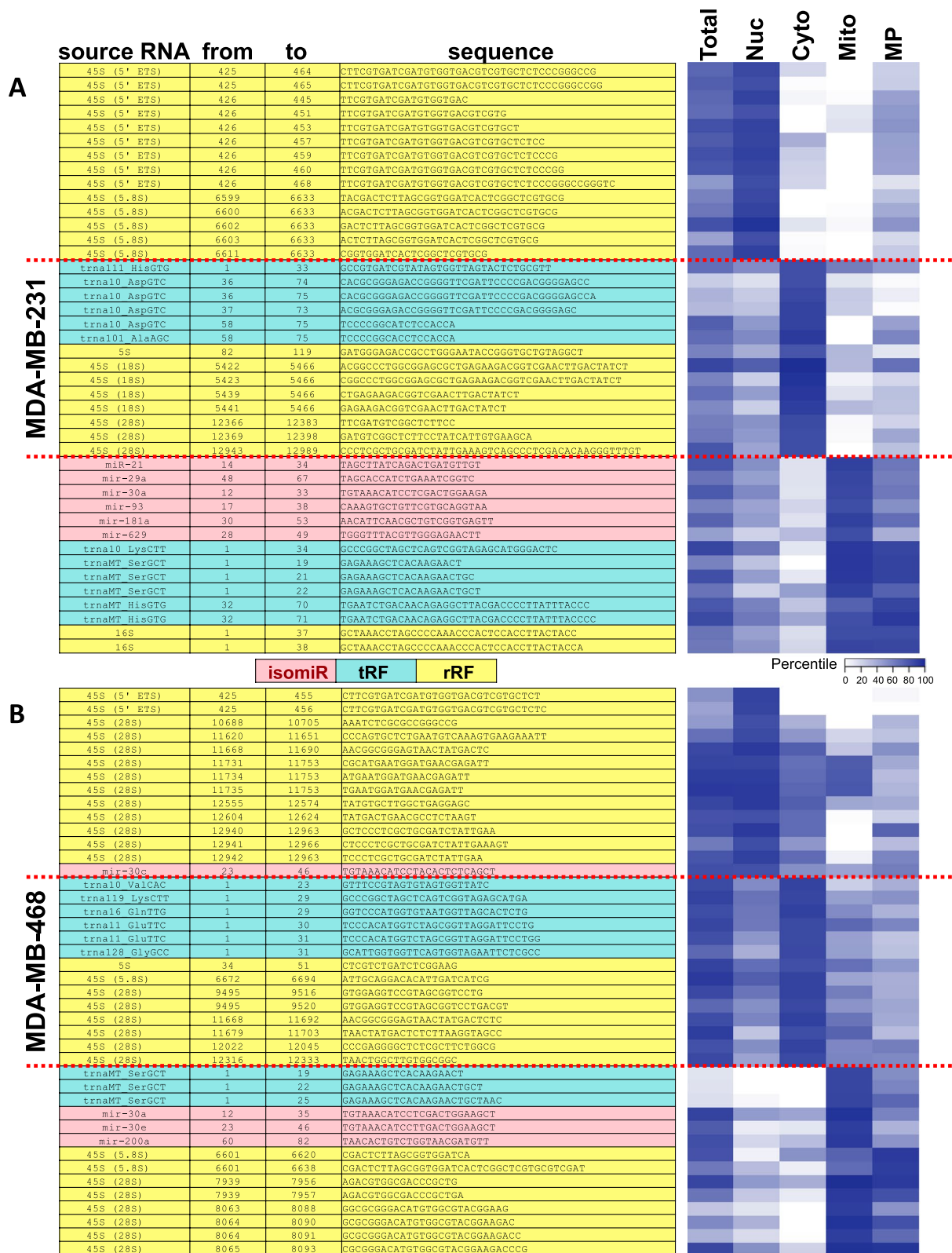


Fig. 4 (See legend on previous page.)

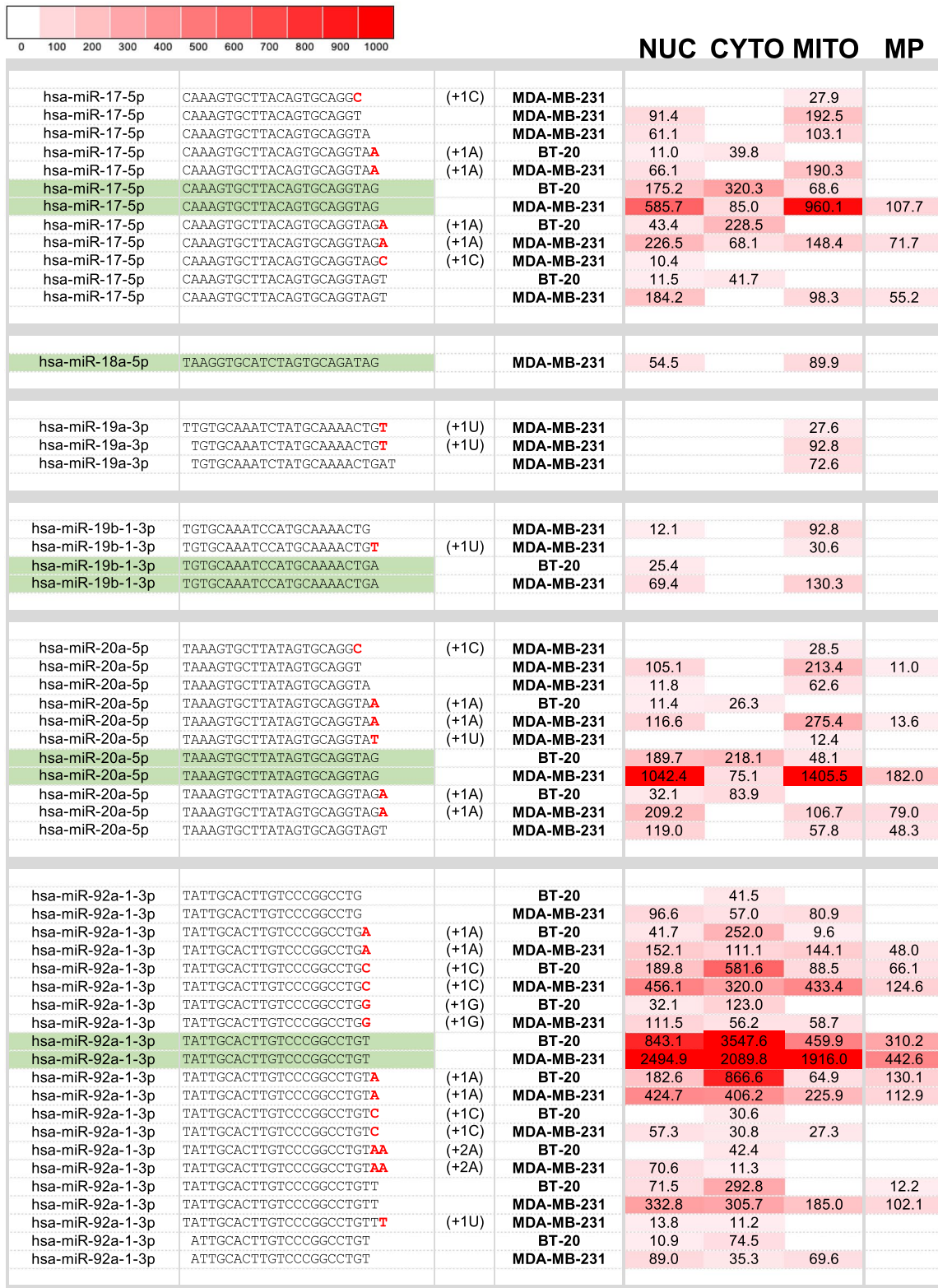


Fig. 5 Model-adjusted abundances of isomiRs from the miR-19/92 cluster in different subcellular compartments. Heatmap of average reconstructed abundances of isomiRs from the miR-17/92 cluster ("oncomiR-1") in BT-20 and MDA-MB-231 cell fractions. The shown abundances are model-adjusted RPM values. The isomiRs highlighted in green are the reference isoforms found in miRBase

that non-templated “+1G” isomiRs localize in the nuclei of post-mitotic neurons [19] and additionally suggests that this event is cell-type-specific.

The co-existence of sncRNAs with typical and atypical terminal phosphate states leads to differences between deep sequencing profiles and northern blots

We recently showed that tRNAs belonging to the same isodecoder produce different tRFs in different tissues and different cell lines [9, 70]. Figure 6A shows two tRNAs from the CTT isodecoder of tRNA^{Lys}. The first tRNA, trna10-LysCTT.chr16, is located between positions 3241501 and 3241573 on the forward strand of chromosome 16 (human genome assembly GRCh37). The second tRNA, trna119-LysCTT.chr1, is located between positions 145395522 and 145395594 on the reverse strand of chromosome 1 (human genome assembly GRCh37). The two tRNAs differ by a single nucleotide at position 29. We used the recalculated abundances of sncRNAs (Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8) to determine the relative abundance of tRFs from these two tRNAs in total RNA and the cytoplasmic fraction.

Figure 6B shows a heatmap of the abundance of 20 tRFs from these two tRNAs. For 18 of the 20 tRFs, the parental tRNA can be determined unambiguously, but not for the last two (red box). Note how a 35-mer from trna10-LysCTT.chr16 is most abundant in BT-20 cells, less so in MDA-MB-231, and absent from MDA-MB-468 cells. On the other hand, a 30-mer from trna119-LysCTT.chr1 is most abundant in MDA-MB-468 cells and absent from BT-20 and MDA-MB-231 cells. A notable observation is the absence of any tRFs with lengths 33–37 nts from MDA-MB-468 cells.

We designed a northern blot probe against the region of trna119-LysCTT.chr1 highlighted in yellow in Fig. 6A and profiled total and cytoplasmic RNA from all three cell lines. This probe will hybridize to tRFs from trna10-LysCTT.chr16 and trna119-LysCTT.chr1 that contain the highlighted region. Figure 6C shows the sums of the recalculated abundances of same-length tRFs from these two tRNAs, whereas Fig. 6D shows the northern blots

for total RNA and the cytoplasmic fraction. Additional file 14: Fig. S6 shows the original uncropped blots.

The most notable difference between the deep sequencing and northern blot profiles is the presence in the blots of tRFs with lengths 33–37 nts (red rectangles, Fig. 6C–D). This seeming discordance helps highlight two points:

- First, it shows that distinct tRFs with near-identical sequences and differing abundances are co-expressed and produced by different tRNAs. Deep sequencing can easily distinguish among them and determine which tRNA produces a given size fragment, whereas northern blotting cannot. These differences are not unique to the tRNA^{Lys} isodecoder: indeed, we recently reported multiple similar cases for many different tRNAs in various cell lines, tissues, and diseases [70].
- Second, it shows that northern blotting can reveal atypical sncRNAs (with terminal phosphate states other than 5'-P/3'-OH) that cannot be profiled using standard deep sequencing [56], an observation that others and we documented previously [56, 66].

To demonstrate the second point, namely that the difference results from atypical terminal phosphate states, we isolated total RNA from MDA-MB-468 and treated it with T4 polynucleotide kinase (T4 PNK) before library preparation; doing so enables the ligation of standard adapters. Our subsequent analysis confirmed the presence of abundant tRFs from tRNA^{LysCTT} with lengths and relative abundances that match the northern blots of Fig. 6D (Additional file 15: Table S9). The seemingly “missing” tRFs became “visible” after modification of the standard sequencing protocol.

Northern blots show complex patterns of subcellular distributions for sncRNAs

Figures 5 and 6 highlight the difficulty of using northern blots to reproduce findings obtained by analyzing *standard* deep-sequencing datasets. Nonetheless, we can still use northern blots to demonstrate differential subcellular distributions while keeping in mind that some of the sncRNAs that these blots capture may have modified

(See figure on next page.)

Fig. 6 Abundance and subcellular compartment enrichments of sncRNA from tRNAs belonging to the same isodecoder. **A** Nucleotide sequences and an alignment of two tRNAs belonging to the same isodecoder (LysCTT). The tRNAs differ by a single nucleotide at position 29. The highlighted text shows the region that our northern blot probe targets. **B** A heatmap showing abundances for 20 distinct tRFs from the same two tRNAs, in total RNA and the cytoplasmic fraction averaged across the replicates of BT-20, MDA-MB-231, and MDA-MB-468 cells. These abundances correspond to measurements after they were adjusted by our mixed-effects model to remove cross-fraction contamination and errors. **C** Summed adjusted abundances of tRFs bound by the probe shown in panel **A** above. **D** A northern blot showing the short RNAs (range: 20–40 nts) that we detected using 5 µg of total and cytoplasmic RNA. Full-length tRNA^{LysCTT} was used as a loading control

A **trna10-LysCTT. chr16**
GCCCCGGCTAG CTCAGTCGGT AGAGCATGGG ACTCTTAATC TCAGGGTCGT GGGTTCGAGC CCCACGTTGG GCG
|||||
GCCCCGGCTAG CTC **AGTCGGT AGAGCATGAG ACTCTTAATC** TCAGGGTCGT GGGTTCGAGC CCCACGTTGG GCG
trna119-LysCTT. chr1

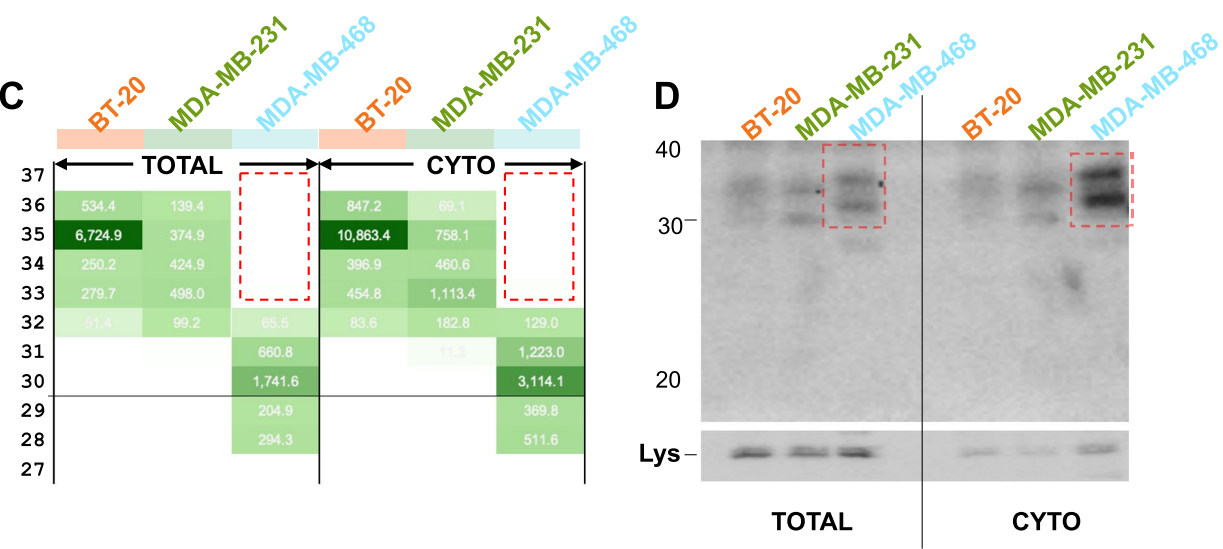
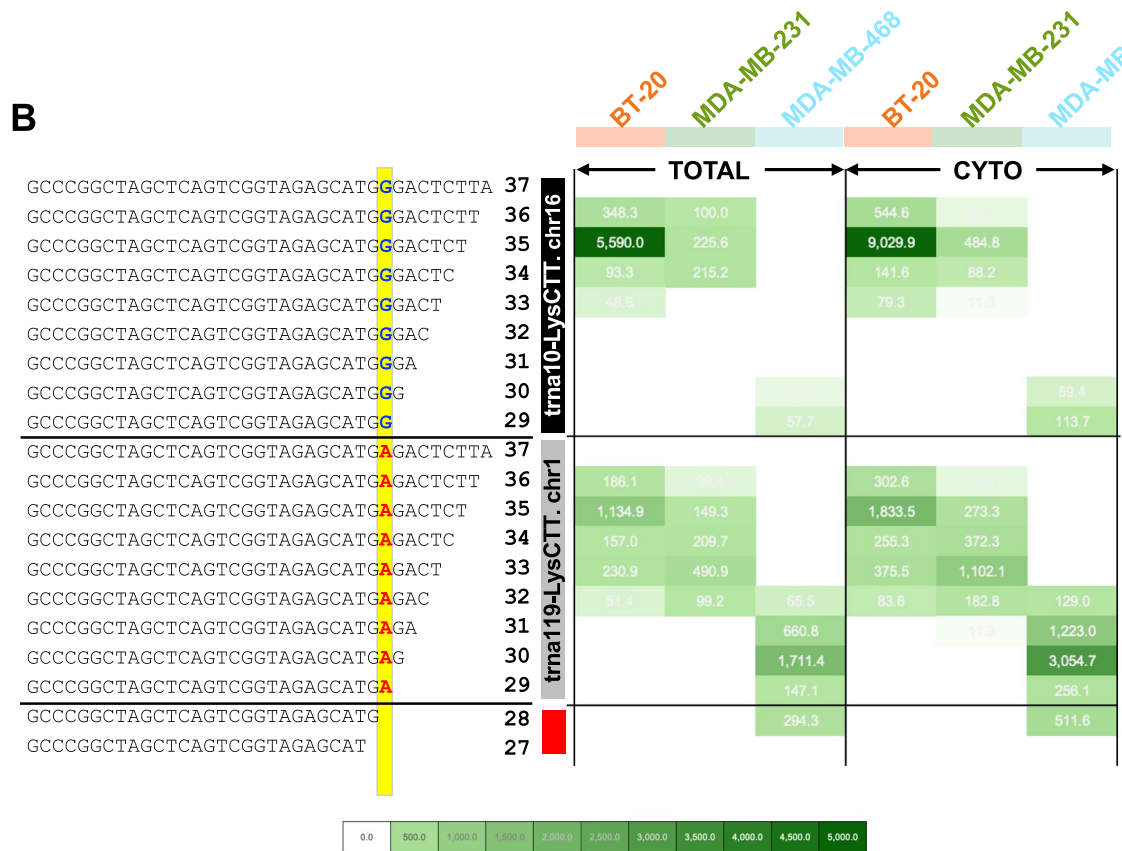


Fig. 6 (See legend on previous page.)

termini and be absent from Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8, in analogy to what is shown in Fig. 6B–D.

Previously, others and we showed that the 5' and 3' regions of the 28S rRNA are hotspots for producing highly abundant rRFs in multiple cell types and tissues [1, 55, 71–73]. Figure 7A shows northern blots for total RNA and the cytoplasmic, nuclear, and MT fractions of MDA-MB-231 cells. The probe targets the sequence TCAGATCAGACGTGGCGA from the 5' region of 28S rRNA. Figure 7B shows analogous blots using a probe

that targets the sequence CTCGCTGCGATCTAT TGAAAG from the 3' region of 28S rRNA. Additional file 16: Fig. S7 shows the original uncropped blots. Additional file 17: Fig. S8 shows the full WES membrane for these fractions. Figure 7C shows analogous blots using a probe that targets the sequence CGACTCTTAGCG GTGGATCA of 5.8S rRNA, which is part of the 45S cassette together with the 28S and 18S rRNAs. The panels reveal the existence of abundant rRFs with lengths between 50 and 100 nts, with subcellular distribution preferences. Note that the low presence of U1 RNA in

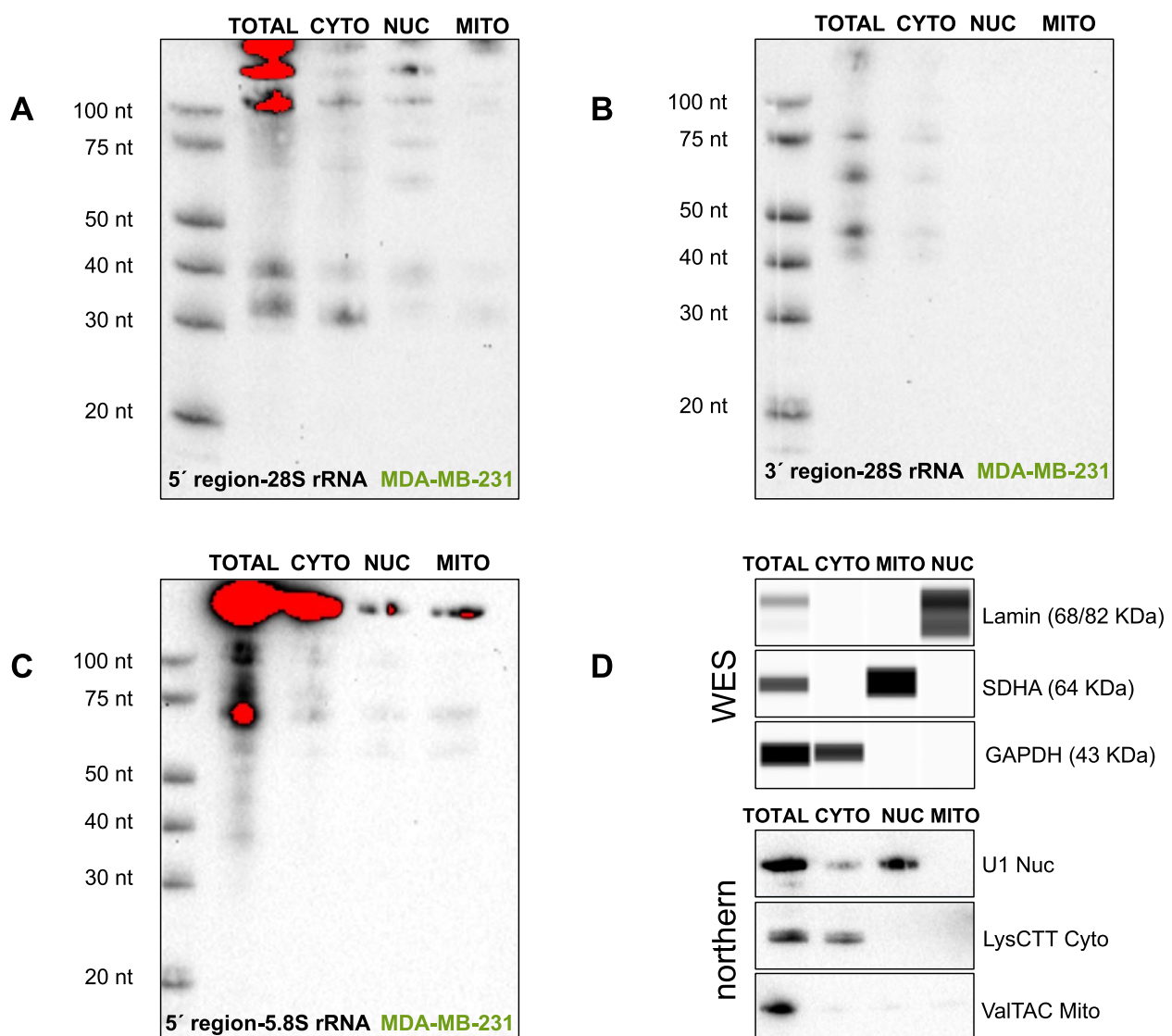


Fig. 7 Subcellular compartment enrichments for 45S-derived rRFs. **A** Northern blots showing rRFs that arise from the 5' region of 28S rRNA in total RNA and the cytoplasmic, nuclear, and MT fractions. **B** Analogous blots for rRFs that arise from the 3' region of 28S rRNA. **C** Analogous blots for an rRF from the 5.8S rRNA (see also text). All lanes were loaded with equivalent RNA. All experiments were carried out with MDA-MB-231 cells. **D** WES markers to assay the protein purity of the fractions and northern blot RNA markers to assay the RNA composition of the fractions. Note the different order of the lanes in the WES and northern panels

the cytoplasmic fraction (Fig. 7D) is supported by previous work [74–76], including ENCODE’s profiling of RNAs in the cytoplasm and nucleus of multiple cell lines [77]. Another observation is the *potential* presence in the MT of the full-length (157 nts) and shorter rRFs (~60 and ~75 nts) from 5.8S rRNA. However, we cannot confidently conclude this from this experiment given the low abundance of MT RNA that is estimated by the MT tRNA^{ValTAC} probe. We note that the MT localization of the full-length 5.8S rRNA in the MT has been reported before [78] as has the MT localization of 5S rRNA [79–81].

Discussion

In this presentation, we reported our findings on the subcellular distribution preferences of isomiRs, tRFs, and rRFs in three cell lines that model TNBC. Knowledge of these subcellular distributions can provide clues as to the sncRNAs’ potential function and help choose among them when designing experiments. We arrived at our results by developing a new integrated approach, Fraction-seq (see the “Methods” section). Our findings from using Fraction-seq show complex sub-cellular distributions for sncRNAs that depend on sncRNA class, sequence, and cell type.

While several recent studies examined the subcellular distribution of long non-coding RNAs [67, 82–86] and sncRNAs [26, 27, 72, 84, 87], they did not account for cross-fraction contamination or technical errors. Moreover, we are unaware of any systematic treatise of replicates from multiple cell lines from the same tissue, same-sex donors, and modeling the same disease subtype while additionally correcting for cross-fraction contamination and errors.

Fraction-seq is an integrated approach that combines a novel cell fractionation method we developed for this project, deep sequencing, and a mixed-effects model that reconstructs each molecule’s true abundance by accounting for cross-fraction contamination and experimental errors (Fig. 1). We used Fraction-seq to study multiple biological replicates from three cell lines that model TNBC. For each replicate, we profiled the sncRNAs in total RNA and the nuclear, cytoplasmic, MT, and MP fractions, deriving these five preparations from the same starting cell material (see the “Methods” section and Additional file 1: Fig. S1). We generated and analyzed a total of 50 datasets.

Our analysis of total RNA confirmed the congruence of the biological replicates (Fig. 2). The analysis also established the unique character of each cell line’s sncRNA profile (Fig. 3A). Indeed, the 10% most abundant sncRNAs in total RNA in the three cell lines, 761 unique molecules, have only 17 sncRNAs in common. Even the

most similar cell lines, BT-20 and MDA-MB-231, share not more than 45.5% of their most abundant sncRNAs (Fig. 3A). By computing differential abundance across cell lines, we found numerous statistically significant changes in total RNA and counterpart subcellular compartments (Additional file 6: Table S3 and Additional file 9: Table S5). This further confirms the uniqueness of each cell line’s sncRNA contents.

In regards to the difference between MDA-MB-468 and both BT-20 and MDA-MB-231 cells (Fig. 3A and Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8), we note that MDA-MB-468 was isolated from a Black TNBC patient, whereas MDA-MB-231 and BT-20 were isolated from White patients. The difference we observed in this study is concordant with differences by ancestry that we reported previously for TNBC patients [5, 10]. However, before a conclusion can be drawn, it will be necessary to apply our Fraction-seq approach to additional TNBC cell lines isolated from patients with different ancestries. This is the topic of ongoing research in our laboratory.

Percentile-based rankings reveal sncRNA preferences for specific compartments

Before embarking on the strategy that uses a mixed-effects model, we used a more straightforward approach to show that different sncRNAs have different subcellular distributions. The approach relies on comparing the abundance-based ranking of two molecules of interest in two preparations, e.g., total RNA and a fraction, or, two fractions. Any substantial changes in the molecules’ relative ranking in the two preparations would indicate a differential distribution of one of the molecules in the corresponding compartment(s). This is because cross-fraction contamination would act uniformly on both molecules and not affect their relative abundance. This simple approach helped identify differences in the subcellular distributions of various sncRNAs with highly overlapping sequences (Fig. 4 and Additional file 10: Fig. S5). However, this approach works only for very abundant sncRNAs.

Mixed-effect modeling can correct for cross-fraction contamination and technical errors

The sncRNA abundances in the various fractions that we can measure using deep sequencing can be influenced by possible cross-fraction contamination and experimental errors. While fraction-specific protein markers can help assay possible contamination (Additional file 2: Fig. S2), they are accurate within the sensitivity capabilities of these assays. To account for contamination that may still be present but not detectable by protein-based assays, we

used a “mixed-effects” approach to model the underlying process and correct the abundances we measured with deep sequencing.

For each combination of biological replicate and cell line, the model assumes that the measured abundance of a sncRNA in a fraction is the sum of its true abundance in that fraction, a contribution from a replicate-specific technical error, and a contribution from contamination by other fractions. The model further constrains its calculation by using the measured abundances of the sncRNAs in total RNA. We applied this model to each cell line separately, simultaneously using the available biological replicates for the cell line, and recovered the true abundance of sncRNAs in each fraction (Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8).

SncRNAs from the same parental RNA can have different subcellular distributions

The reconstructed abundances of Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8 suggest a complex subcellular distribution for isomiRs, tRFs, and rRFs. Figure 5 shows several examples for the miR-17/92 miRNA cluster and helps highlight three points. First, the subcellular distribution of the canonical miRNAs found in public databases is cell-type dependent. For example, the canonical isoform of miR-17 (highlighted in green) is most enriched in the nucleus and the MT in MDA-MB-231 cells, and the cytoplasm in BT-20 cells. Second, sncRNAs from the same parental RNA whose sequences differ by 1–2 nts generally have different subcellular distributions. Third, the typical miRNA arm produces multiple non-canonical and non-templated miRNA isoforms with widely varying subcellular distributions and cell-type-dependent abundances. For example, in the case of miR-92a-1-3p, the non-canonical isoform TATTGCACTTGTCCCGGCCTGTT and the non-templated isoform TATTGCACTTGTCCCGGCCTGTA are equally abundant in both the nucleus and the cytoplasm of MDA-MB-231 cells. But, in BT-20 cells, both are primarily cytoplasmic.

Nucleus-encoded sncRNAs are enriched in the MT fraction

Our study suggests that multiple sncRNAs encoded by the nuclear genome are enriched in the MT fraction in a sequence- and cell-type-dependent manner (Fig. 4 and Additional file 10: Fig. S5). Most of these sncRNAs are likely “attached” to the MT’s outer membrane. For example, in the case of miRNAs, any miRNAs/isomiRs that are part of Argonaute complexes attached to the MT membrane would be unaffected by our RNase A treatment and, thus, contribute to the contents of the MT fraction. Nonetheless, our analysis also generated

evidence suggesting that some nucleus-encoded sncRNAs, including tRFs and rRFs, enter the MT. Our findings are in concordance with previous reports that reported full-length tRNAs [88–91] and full-length rRNAs [78, 79, 81] in the MT. For example, we find multiple rRFs from the nuclear 45S rRNA that are strongly enriched in the MP fraction but less so in the MT fraction (Fig. 4 and Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8), indicating that these rRFs are inside the MT. Also, we find multiple tRFs from tRNA^{Lys}CTT that are strongly enriched in the MT or MP fractions. For the various nucleus-encoded sncRNAs that we find enriched in the MT fraction, it is unclear whether they are shuttled to this compartment or produced from the local processing of longer precursor RNAs.

Northern blots reveal even more sncRNAs with unique subcellular distributions

Our study identified complex subcellular distributions for multiple sncRNAs. Since we used the standard deep-sequencing protocol, these sncRNAs have the typical 5′-P and 3′-OH termini. Our northern blots, which can profile molecules without regard to their terminal phosphate state, revealed an even richer sncRNA-ome that includes atypical molecules “invisible” to standard sequencing. These atypical molecules exhibit complex subcellular distributions as well (Figs. 6 and 7), and their abundance depends on cell type and nucleotide sequence, in complete analogy to the standard sncRNAs.

The tRFs produced by two tRNAs from the same iso-decoder (CTT) of tRNA^{Lys} are good examples of standard sncRNAs that arise from the same parental RNA and whose lengths and subcellular enrichments depend on the cell type (Fig. 6). These standard sncRNAs co-exist with atypical variants with similar lengths and subcellular distributions of their own. The mechanisms and conditions that decide a sncRNA’s terminal phosphate state are poorly understood [56, 92]. These observations likely have significant ramifications in many contexts, including translational efficiency [93] or targeted therapeutics [5, 9, 10].

The 28S rRNA also revealed a richness in the rRFs that it produces from its 5′ and 3′ regions. These rRFs span a wide range of lengths and abundances and comprise standard (Additional file 11: Table S6, Additional file 12: Table S7, and Additional file 13: Table S8) and atypical variants (Fig. 7) with similar lengths and unique subcellular distribution preferences and cell type dependencies. A probe against an rRF from the 5.8S rRNA helped demonstrate that additional uncategorized sncRNAs with lengths outside the range considered here (18–50 nts) also exhibit preferences in their subcellular distribution (Fig. 7C).

Conclusions

The findings suggest that isomiRs, tRFs, and rRFs have unique subcellular distribution preferences. These preferences depend on each sncRNA's sequence and cell type. Even sncRNAs that have similar sequences or arise from the same precursor RNA can have different subcellular distributions. Our study's results add to the existing literature that sncRNAs follow complex intracellular trafficking rules. Nonetheless, further investigation is needed to exclude other alternatives.

While we derived our findings from the study of three cancer cell lines modeling TNBC, the findings presumably extend beyond cell lines and contexts beyond cancer. Indeed, the combined experimental and analytical approach we presented can be applied to any cell line and bulk tissue. It can also be used to study atypical sncRNAs with modified termini, provided a modified deep sequencing protocol is used throughout [94].

The data also indicate that different cell lines modeling the same disease cannot always be used interchangeably. To design experiments to determine a sncRNA's function, it is essential to know where in the cell line of interest the sncRNA localizes. We hope Fraction-seq will help create an "Atlas" of subcellular distributions for sncRNAs in multiple cells and tissues. Such an Atlas would help prioritize among isomiRs, tRFs, and rRFs, provide valuable constraints on their possible functions, and assist in designing improved diagnostics, prognostics, and therapeutics.

Methods

For this project, we developed a new multi-stage method, "Fraction-seq." Fraction-seq comprises an experimental and an analytical component. Next, we describe each component of Fraction-seq, followed by information on other methods we used.

Cell lines and cell culture

We used BT-20 (ATCC #HTB-19), MDA-MB231 (ATCC #HTB-26), and MDA-MB468 (ATCC #HTB-132) cells for our work. BT-20 and MDA-MB-231 were originally isolated from White patients. MDA-MB-468 was originally isolated from a Black patient. All three are basal-type breast cancer cell lines, contain a mutant (missense) p53 variant, and express c-MYC [95]. P21 is absent in BT-20 and barely detectable in MDA-MB-231 and MDA-MB-468 [95, 96]. Ki-67 is moderate in BT-20 and high in MDA-MB-231 and MDA-MB-468 [96]. We independently validated the origin of all three cell lines by STR. We grew these cell lines according to standard cell growth conditions using Complete DMEM (Gibco Dulbecco's modified Eagle's medium).

We have been testing all cell lines regularly to ensure they are mycoplasma-free.

Fraction-seq: cellular fractionation step

Additional file 1: Fig. S1 provides a pictorial summary of this step. Our method leverages deep sequencing to enumerate sncRNAs in the nuclear, cytoplasmic, MT, and MP fractions of a cell starting from the same material. Specifically, we grow cells in 20–40 10-cm dishes depending on cell type in Complete DMEM Media. When cells reach 90% confluence, we wash the plates twice with ice-cold 1X PBS. We add 2 ml of 1X ice-cold PBS to each plate and then mechanically scrape cells using a cell scraper. We pool cells into two 50 ml ice-cold conical tubes that we spin at 500 g for 10 min. We remove the supernatant and resuspend the cells in 2×packed cell volume of 0.9% NaCl followed by incubation on ice for 10 min. We then aliquot cells into 10 ice-cold 1.5 ml microcentrifuge tubes and spin them at 500 g for 10 min. We remove the supernatant by aspirating. Using the Qiagen Mitochondria Isolation Kit (Qiagen #37612), we resuspend cell pellets in ice-cold Lysis Buffer + Protease Inhibitor and 1 mM EGTA and incubate for 10 min rotating (end-over-end shaker) at 4 °C. We save 4–10% of lysate from each tube for the total cell sample. We centrifuge the lysate at 1000 g for 10 min at 4 °C, then carefully remove the supernatant and transfer it to fresh 1.5 ml microcentrifuge tubes—this is the *crude* cytoplasmic fraction. We spin the cytoplasmic fraction at 16,000 g for 10 min at 4 °C to remove additional cell debris and transfer the supernatant to ultracentrifuge tubes that we centrifuge at 100,000 g for 30 min at 4 °C. The supernatant now contains the *pure* cytoplasmic fraction. We resuspend the remaining cell lysate pellets in 1.5 ml ice-cold Disruption Buffer (+protease inhibitor). We pass the pellets slowly through a blunt-ended 26-gauge needle syringe, 20× on ice (making sure to avoid bubbles). We centrifuge the lysate at 1000 g for 10 min at 4 °C, then transfer the supernatant to two clean 1.5 ml microcentrifuge tubes. We save the pellets as a *crude* nuclear fraction and centrifuge the supernatant at 6000 g for 10 min at 4 °C. The pellets that remain after we remove and discard the supernatant comprise the *crude* MT fraction. We combine the two *crude* MT pellets into one and carefully resuspend it in 750 µl Mitochondria Purification Buffer (+protease inhibitor) using a 1 ml pipette. We add 750 µl Mitochondria Purification Buffer to a 2-ml microcentrifuge tube. We also carefully add 500 µl Disruption Buffer underneath (see image on page 13 of the QProteome Mitochondria Isolation Handbook). Finally, we carefully add the MT suspension to the top. We centrifuge this multilayer solution at 14,000 g for 15 min at 4 °C, which produces a delicate, soft, white band at the

bottom of the tube. After carefully removing and discarding 1.5 ml of the supernatant, we resuspend the remaining 500 μ l (MT pellet and buffer) and transfer it to clean 1.5 ml tubes. We resuspend 1 ml of Mitochondria Storage Buffer into the solution and centrifuge at 8000 g for 10 min at 4 °C. We carefully remove and discard 1.5 ml of the supernatant and resuspend it in 1.5 ml Mitochondria Storage Buffer, then centrifuge it at 8000 g for 10 min at 4 °C. We repeat this process until a pellet forms at the bottom of the tube. We combine these *pure* MT fraction pellets in 100 μ l Mitochondrial Storage Buffer (+ protease inhibitor). We measure the two crude nuclear fraction pellets and resuspend them in the appropriate volume of CER I buffer (+ protease inhibitors) from the NE-PER Nuclear Cytoplasmic Extraction Kit (Thermo Fisher Scientific #78833). We vortex the tubes at the highest setting for 15 s and incubate on ice for 10 min. We add ice-cold CER II buffer to the tubes. We vortex the pellet for 5 s at the highest setting and then incubate on ice for one additional minute. Again, we vortex the pellet for 5 s on the highest setting, then centrifuge it at maximum speed (~16,000 g). We remove and discard the supernatant. We then resuspend the *insoluble* pellet in NER buffer from the NE-PER kit (+ protease inhibitors), vortex the pellet at the highest setting for 15 s, incubate on ice for 10 min (one time only), and centrifuge it at max speed (~16,000 g) for 10 min at 4 °C, and transfer the supernatant, which contains the *pure* nuclear fraction, to a clean 1.5 ml tube on ice. Following the assessment of the MT fraction's purity using protein-based methods (described below), we divide the pure MT fraction into thirds: 1/3rd represents the pure MT fraction; the remaining 2/3rds are transferred to a new tube to be purified into the MP fraction. We centrifuge the pre-MP pellet at 8000 g for 10 min, then discard the supernatant. We resuspend the pellet in 10 \times pellet volume 100 mM Swelling Buffer (NaPO₄, pH 7.4) (+ protease inhibitor) and incubate on ice for 20 min. The hypotonic reaction is stopped by adding 3.75 \times pellet volume of 60% sucrose to the solution. We centrifuge the solution at 12,000 g for 15 min, then remove the supernatant. We wash the pellet twice in 1 ml Mitochondrial Storage Buffer (+ protease inhibitor). Lastly, we resuspend the *pure* MP fraction in 50 μ l Mitochondrial Storage Buffer (+ protease inhibitor). We store all fractions in -20 °C.

Fraction-seq: WES "blotting" step

We subject all fraction lysates (total, nuclear, cytoplasmic, MT, and MP) to the Bradford Protein Assay (BioRad #5000006) to calculate protein concentration. We use the Western size-based Assay from Protein Simple (WES) to evaluate the purity of the cell fractions based on protein profiling. We carry out WES validation according to the

company's specifications (<https://proteinsimple.com>) by loading 0.2 mg/ml of each sample to a 12–230 kDa Jess/Wes Separation Module, 8 \times 25 Capillary Cartridge (Protein Simple #SM-W004). Primary antibodies, Cytochrome C—Mitoplast (BD Pharmagen #556433), GAPDH—Cytoplasm (NEB #2118), LAMIN A/C—Nucleus (NEB #4777), SDHA—Mitochondria/Mitoplast (NEB #11998), and TFAM—Mitochondria/Mitoplast (NEB #8076) are used as cell fraction markers. Secondary Anti-Rabbit (Protein Simple #DM-001) and Anti-Mouse (Protein Simple #DM-002) antibodies are used. We follow the company's protocol to analyze the WES results and calculate the area under the curve for each antibody.

Fraction-seq: RNA isolation step

After we assess the purity of the cell fractions, we subject the MT fraction to incubation with 2 mg/ml RNase A for 15 min on ice. Immediately, we add 900 μ l of TRIzol (Invitrogen) to the Total, Nuclear, Cytoplasmic, MT, and MP fractions. We use a 1 ml pipette to resuspend the samples in Trizol and then incubate them at room temperature for 5 min to equilibrate. We add 180 μ l chloroform to the samples and shake the Trizol/chloroform solution vigorously for 30 s. We incubate the samples at room temperature for 3 min and then centrifuge them at 12,000 g for 15 min. We transfer the clear supernatant from each sample to clean tubes and add 500 μ l 100% isopropanol to each tube. We incubate the samples at 20 °C for 2 h, allowing the RNA to precipitate out of solution. We centrifuge the samples at 12,000 g for 10 min at 4 °C and remove the supernatant. We wash the samples twice with 1 ml 70% ethanol and centrifuge them at 7600 g for 5 min. We remove all ethanol and allow the pellets to dry for 15 min. We then resuspend the pellets in molecular-grade DNase-, RNase- and Protease-Free water (Fisher #BP2819-1). We detect the RNA concentrations using a spectrophotometer.

Fraction-seq: deep sequencing step

We assess the integrity and purity of the RNA of the validated preparations (above) using the Agilent 2100 Bioanalyzer. We prepare the validated RNAs for sequencing using the NEBNext Library preparation method (#E7330) with the kit's standard protocol. The NEBNext 3'-adapter is AGATCGGAAGAGCACACGTCT. All samples are sequenced at an average depth of 30 million reads (single-end, 75 bases).

Fraction-seq: profiling isomiRs, tRFs, and rRFs step

We profile isomiRs using isoMiRmap [12], tRFs using MINTmap [97], and rRFs with the approach that we described previously [1]. Before mapping, we use *cutadapt* to quality trim and remove the 3' adapter [98]. We threshold the profiled sncRNAs using the adaptive,

sample-specific Threshold-seq tool [99] (default parameter settings). Since the MP fraction is not treated with RNase A because of its limiting yield, we apply additional computational filters: specifically, for a given biological replicate, we remove any sncRNA that is present in the MP fraction above threshold but absent from the MT fraction. We only consider sncRNAs (isomiRs, tRFs, and rRFs) whose lengths are ≥ 18 nts. To account for different sequencing depths, we express sncRNA abundances in reads-per-million (RPM).

Differential abundance of sncRNAs

We use the DESeq2 package [100] in R to determine sncRNAs that are DA in total RNA in two cell lines or the same fraction in two cell lines. For the values input into the DESeq2 package, we use a mean raw read threshold of 300 for either of the two groups being compared. For normalization, we used DESeq2's default parameters which calculates size factors based on the median of ratios. DA is measured in $|\log_2| \geq 1$ change. We determine statistical significance using a false discovery rate (FDR) threshold of 0.05.

Northern blotting

We run 5 or 10 μg of RNA from each cell line on a 15% acrylamide/8 M urea gel at 300 V until the lower dye front reaches the bottom of the gel. We transfer the gel to HybondTM-N⁺ membrane (Amersham Biosciences, catalog number: RPN303B) at 400 mA for 30 min. We dry the membrane and then cross-link twice at 120,000 $\mu\text{J}/\text{cm}^2$. We pre-hybridize the membranes in a hybridization buffer (PerfectHybTM Plus Hybridization Buffer: H7033-1L) for 30 min, rotating at 37 °C. We prepare northern probes using the DIG labeling kit (DIG Oligonucleotide 3'-End Labeling Kit, 2nd Gen: 3353575910). Detection of membranes is done using the DIG detection kit (DIG Wash and Block Buffer Set: 11585762001, Anti-Digoxigenin-AP, Fab fragments: 11093274910, CDP-Star Chemiluminescent Substrate: C0712-100ML)

derived from nuclear tRNAs; (iii) tRFs derived from MT tRNAs; (iv) ambiguous tRFs whose sequences are present in both nuclear and MT tRNAs; (v) rRFs derived from the nuclear 5S rRNA; (vi) rRFs derived from the nuclear 45S rRNA cassette; and (vii) rRFs derived from two MT rRNAs, 12S and 16S. The model treats each sncRNA as an independent molecule, including those that are adjacent to or even overlap one another on the parental RNA from which they arise. Furthermore, the model incorporates loss of material during the fractionation process, technical errors, and cross-fraction contamination separately for each replicate.

For the estimation process, and separately for each cell line:

- We assume that the cytoplasmic fraction (“*cyto*”) is uncontaminated (Additional file 19: Text 1) and the measured abundance of the i th molecule in the j th replicate is linked to the i th molecule's true value as follows:

$$\log(1 + \text{cyto}_{i,j}^{\text{measured}}) = \alpha_{1,j} + \log(1 + \text{cyto}_i^{\text{True}}) + e_{1,i,j}$$

$$\text{where } \sum_{j=1}^J \alpha_{1,j} = 0, \log(1 + \text{cyto}_i^{\text{True}}) \sim N(\mu_{1,i}, \tau_1^2),$$

$e_{1,i,j} \sim N(0, \sigma_1^2)$, and J is the number of replicates. This is a mixed effects model where $\log(1 + \text{cyto}_i^{\text{True}})$ is the random effect. The mean of the log-transformed true value ($\mu_{1,i}$) is a function of group membership. We fit this model using restricted maximum likelihood estimation in SAS Proc Mixed. From the results of fitting this model, we can estimate the expected true value for the i th molecule from $\log(1 + \widehat{\text{cyto}}_i^{\text{True}})$, and the j th replicate of the i th molecule from $(\widehat{\alpha}_{1,j} + \log(1 + \widehat{\text{cyto}}_i^{\text{True}}))$.

- We assume that the nuclear fraction (“*nuc*”), MT fraction (“*mito*”), and mitoplast fraction (“*MP*”) are contaminating one another and the measured abundance of the i th molecule in the j th replicate is linked to the i th molecule's true value as follows:

$$\log(1 + \text{mito}_{i,j}^{\text{measured}}) = \alpha_{2,j} + \log(1 + \text{mito}_i^{\text{True}}) + \gamma_{1,j} \log(1 + \text{cyto}_{i,j}^{\text{measured}}) + \gamma_{2,j} \log(1 + \text{nuc}_{i,j}^{\text{measured}}) + e_{2,i,j}$$

$$\log(1 + \text{MP}_{i,j}^{\text{measured}}) = \alpha_{3,j} + \log(1 + \text{MP}_i^{\text{True}}) + \gamma_{3,j} \log(1 + \text{cyto}_{i,j}^{\text{measured}}) + \gamma_{4,j} \log(1 + \text{nuc}_{i,j}^{\text{measured}}) + \gamma_{5,j} \log(1 + \text{mito}_{i,j}^{\text{measured}}) + e_{3,i,j}$$

$$\log(1 + \text{nuc}_{i,j}^{\text{measured}}) = \alpha_{4,j} + \log(1 + \text{nuc}_i^{\text{True}}) + \gamma_{6,j} \log(1 + \text{cyto}_{i,j}^{\text{measured}}) + e_{4,i,j}$$

following the manufacturer's instructions. Additional file 18: Table S10 lists the probes we used in this study.

Mixed-effects modeling

We use a “mixed-effects” approach to model the fractionation process. Our model distinguishes among the following seven groups of sncRNAs: (i) isomiRs; (ii) tRFs

$$\text{where } \sum_{j=1}^J \alpha_{x,j} = 0, \log(1 + \text{mito}_i^{\text{True}}) \sim N(\mu_{2,i}, \tau_2^2),$$

$$\log(1 + \text{MP}_i^{\text{True}}) \sim N(\mu_{3,i}, \tau_3^2),$$

$\log(1 + \text{nuc}_i^{\text{True}}) \sim N(\mu_{4,i}, \tau_4^2)$, $e_{x,i,j} \sim N(0, \sigma_x^2)$, J is the number of replicates, and the γ 's represent the contamination.

Fitting these models directly would result in high estimates for the various γ 's as they would capture contamination and true correlation. We tackle this complication using a three-step process:

1. We fit a model with estimated *True* values as predictors. For example, for the nuclear fraction

$$\log(1 + nuc_{i,j}^{measured}) = \alpha_{4,j} + \log(1 + nuc_i^{True}) + e_{4,i,j}$$

where $\log(1 + nuc_i^{True}) \sim N(\mu_{4,i}, \tau_4^2)$ and $\mu_{4,i} = \sum_{g=1}^7$

$(\lambda_{0,g} + \lambda_{1,g} \log(1 + cyto_i^{True})) I(Group_i = g)$. That is, we allow the mean of the *i*th molecule's true value in the nuclear fraction to depend on the group type of the *i*th molecule and the molecule's estimated true cytoplasmic value, which captures the correlation.

2. Using the estimated residuals from step 1, we fit the model

$$\hat{e}_{4,i,j} = \delta_{4,j} + \gamma_{6,j} \log(1 + cyto_{i,j}^{measured}) + \varepsilon_{4,i,j}$$

The idea here is that if all association between the *i*th molecule's measured value in the nuclear and cytoplasmic fractions is due to real correlation, then there should be no association between the residuals of step 1 and the measured cytoplasmic value. Any association that may be present indicates contamination.

3. We subtract the contamination from the measured value to estimate the uncontaminated value of the *i*th molecule in the *j*th replicate as:

$$\log(1 + nuc_{i,j}^{uncontam}) = \log(1 + nuc_{i,j}^{measured}) - \hat{\gamma}_{6,j} I(\hat{\gamma}_{6,j} > 0) \log(1 + cyto_{i,j}^{measured})$$

Any non-positive estimates for γ are ignored. We then refit the mixed effects model using the uncontaminated estimates as the dependent variable:

$$\log(1 + nuc_{i,j}^{uncontam}) = \alpha_{4,j} + \log(1 + nuc_i^{True}) + e_{4,i,j}$$

Using the results from the model, we estimate the true abundance of the *i*th molecule in the *j*th replicate in the nuclear fraction $\hat{nuc}_{i,j} = \exp(\hat{\alpha}_{4,j} + \log(1 + \hat{nuc}_i^{True})) - 1$.

We repeat the last three steps for the *mito* and *MP* fractions. In these two models, the estimated mean true abundance of the *i*th molecule in the nuclear and cytoplasmic fractions is included in step 1 (including the *mito* fraction in the *MP* results in non-convergence). To estimate the contamination, we include all factors in the contamination model.

- Finally, for each replicate, we fit the following equation for rescaling purposes:

$$total_{i,j}^{measured} = \beta_{0,j} + \beta_{1,j} \hat{cyto}_{i,j} + \beta_{2,j} \hat{nuc}_{i,j} + \beta_{3,j} \hat{mito}_{i,j} + e_{5,i,j}$$

where $\hat{cyto}_{i,j} = \exp(\hat{\alpha}_{1,j} + \log(1 + \hat{cyto}_i^{True})) - 1$, $\hat{nuc}_{i,j} = \exp(\hat{\alpha}_{4,j} + \log(1 + \hat{nuc}_i^{True})) - 1$, and $\hat{mito}_{i,j} = \exp(\hat{\alpha}_{4,j} + \log(1 + \hat{mito}_i^{True})) - 1$. The equation incorporates loss of material by scaling each fraction separately. This calculation excludes molecules that occupy the top 1% and bottom 1% by abundance (outliers).

We fit the mixed effects models using SAS Proc MIXED (SAS/STAT version 15.1, SAS Institute, Cary, NC) with restricted maximum likelihood estimation under default convergence criteria. We estimated scaling constants using SAS Proc GENMOD with Bayesian estimation (MCMC) with priors for the compartment coefficients having a normal distribution with mean equal to the coefficient from a linear regression model that assumes the same coefficient for each compartment and variance equal to 10 times the variance of the estimated coefficient. We ran three chains for each model and assessed convergence for these parameters using the Gelman-Rubin statistic [101].

Abbreviations

DA	Differentially abundant or differential abundance
FDR	False discovery rate
lncRNA	Long non-coding RNA
isomiR	miRNA isoform
miRNA	MicroRNA
MT	Mitochondrion
MP	Mitoplast
mRNA	Messenger RNA
nt	Nucleotide
PCA	Principal component analysis
rRF	rRNA-derived fragment
RPM	Reads-per-million
sncRNA	Small non-coding RNA
T4 PNK	T4 polynucleotide kinase
tRF	tRNA-derived fragment
TNBC	Triple-negative breast cancer

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01970-6>.

Additional file 1: Fig. S1. Our approach. Cells are washed 2x with 1X PBS on 20-40 Petri dishes when 90% confluent. Cells are then mechanically scraped and transferred to a 50 ml conical tube. Cells are centrifuged at 500 g for 10 minutes then the supernatant is removed. Cells are then resuspended in 2x packed cell volume 0.9% NaCl and left to incubate on ice for 10 minutes. A. Cells are transferred to 10 1.5 microcentrifuge tubes, where they are centrifuged at 500g for 10 minutes, and the supernatant is discarded. The pellets are resuspended in ice-cold Lysis Buffer + Protease Inhibitor and 1 mM EGTA and incubated for 10 minutes rotating at 4°C. B. 4-10% of lysate from each tube is saved for the total cell sample. C. Lysate is centrifuged at 1,000 g for 10 minutes at 4°C, and

supernatant is transferred to fresh 1.5 ml microcentrifuge tubes. D. Supernatant is centrifuged at 16,000 g for 10 minutes at 4°C to remove additional cell debris and new supernatant is transferred to ultracentrifuge tubes which are centrifuged at 100,000 g for 30 minutes at 4°C – the supernatant now contains the cytoplasmic fraction. E. The remaining cell lysate pellets are resuspended in 1.5 ml ice-cold Disruption Buffer. The pellets are passed slowly through a blunt-ended 26-gauge needle syringe, 20x on ice. F. The supernatant is transferred to clean tubes and processed according to the Qiagen Mitochondria Isolation Kit instructions. G. Purified MT pellets are combined in 100 µl Mitochondria Storage Buffer – this is the MT fraction. H. 1/3rd of the MT fraction is further processed to create the MP fraction. I–J. The pellet that formed from spinning the solution in E is further purified using the NE-PER isolation kit and the final supernatant is the nuclear fraction. Parts of this image were created with BioRender.com.

Additional file 2: Fig. S2. Protein marker validation of cell fractionation via WES. A–C. WES blots showing detection of cell fraction protein markers Lamin A/C, GAPDH, SDHA, Cytochrome C, TFAM. 2 mg/ml protein lysate from each sample was loaded into a given lane. The absence of Cytochrome C in the MP fraction indicates the removal of intermembrane material from the MT. A. BT-20 – three replicates. B. MDA-MB-231 – three replicates. C. MDA-MB-468 – four replicates

Additional file 3: Table S1. Data table of the average abundances for each short RNA found in RNA-seq at ≥ 10 RPM in a given cell line and cell fraction

Additional file 4: Table S2. Data table of the top 10% most abundant short RNA from each cell line based on average abundance of the cell line replicates

Additional file 5: Fig. S3. Abundances of 17 sncRNAs shared between TNBC cell lines. These are the 17 sncRNAs that are common to BT-20, MDA-MB-231, and MDA-MB-468 cell lines when comparing the top 10% most abundant RNAs from each cell line based on average abundance

Additional file 6: Table S3. Data table of the differentially abundant short RNAs when comparing the three cell lines pairwise

Additional file 7: Fig. S4. Comparisons of sncRNAs across cell lines reveal cell-line-specific features. A–D. Venn Diagrams of the 10% most abundant sncRNAs in BT-20, MDA-MB-231, and MDA-MB-468 cells based on average abundance in each cell fraction. A. Nucleus. B. Cytoplasm. C. MT. D. MP

Additional file 8: Table S4. Data table of the differentially abundant short RNAs

Additional file 9: Table S5. Data table of the top 10% most abundant short RNAs from each cell line for each cell fraction based on average abundance of the cell line replicates

Additional file 10: Fig. S5. Examples of subcellular enrichments using within-fraction abundance-based ranking. Percentile-colored heatmap for select sncRNAs in total RNA and the nucleus, cytoplasm, MT, and MP for BT-20. For each sncRNA, we list its nucleotide sequence, the parental RNA from which the sncRNA arises, and the location of the sncRNA within the parental RNA. Red dotted lines separate the list into three groups: sncRNAs that are enriched primarily in the nucleus, the cytoplasm, or the MT, respectively

Additional file 11: Table S6. The average reconstructed abundances in total RNA and each of the small RNAs in the cell fractions of BT-20 cells

Additional file 12: Table S7. The average reconstructed abundances in total RNA and each of the small RNAs in the cell fractions of MDA-MB-231 cells

Additional file 13: Table S8. The average reconstructed abundances in total RNA and each of the small RNAs in the cell fractions of MDA-MB-468 cells

Additional file 14: Fig. S6. Original uncropped blots used for Fig. 6.

Additional file 15: Table S9. Abundance of tRNA LysCTT-derived fragments in RNA-seq data generated by treating RNA with T4 PNK prior to library preparation

Additional file 16: Fig. S7. Original uncropped blots used for Fig. 7.

Additional file 17: Fig. S8. A full-length WES membrane for the MDA-MB-231 fractions that are shown in Fig. 7.

Additional file 18: Table S10. DNA oligo sequences ordered from Fisher/Invitrogen used for the DIG northern blot experiments in this study

Additional file 19: Text 1. Additional information about our modeling approach

Acknowledgements

The authors thank the TNBC patients who donated their samples for the three cell lines (BT-20, MDA-MB-231, and MDA-MB-468) used in this study. The authors also thank the members of the CMC for helpful discussions, assistance, and suggestions. The authors especially wish to thank Drs. Yohei Kirino, Megumi Hamasaki, Erin Seifert, Gyorgy Csordas, and Sergio De La Fuente Perez for their helpful conversations and guidance regarding the cell fractionation method.

Authors' contributions

IR conceived and supervised the study. TC and IR designed the experiments. TC, YJ, SS, AK, AGT, HW, and LT performed the experiments. VP and TC managed the submission and maintenance of original data deposited to NCBI. TC, IR, and BL designed the analysis methodology. TC, IR, BL, VP, PIV, and PL contributed to the analytical tools. TC and IR analyzed the data and designed the figures. TC and IR wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported partially by a William M. Keck Foundation grant (IR) and by Institutional Funds.

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The 50 RNA-seq datasets we generated using the standard deep-sequencing protocol can be found in the NCBI SRA repository <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA816866> [102]. The RNA-seq dataset we generated from MDA-MB-468 after T4 PNK pretreatment can be found in the NCBI SRA repository <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1096643> [103].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Computational Medicine Center, Thomas Jefferson University, Philadelphia, PA 19017, USA. ²Division of Biostatistics, Thomas Jefferson University, Philadelphia, PA 19017, USA. ³Present Address: University of Pennsylvania, Philadelphia, PA, USA. ⁴Present Address: Department of Pharmacology, Physiology and Cancer Biology, Thomas Jefferson University, Philadelphia, PA, USA. ⁵Present Address: University of Miami, Miami, FL, USA. ⁶Present Address: New York University, New York, NY, USA. ⁷Present Address: Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA, USA. ⁸Present Address: Independent Scholar, Athens, Greece.

Received: 7 December 2022 Accepted: 1 August 2024
Published online: 12 September 2024

References

- Cherlin T, Magee R, Jing Y, Pliatsika V, Loher P, Rigoutsos I. Ribosomal RNA fragmentation into short RNAs (rRFs) is modulated in a sex- and population of origin-specific manner. *BMC Biol.* 2020;18(1):38.
- Giuliani A, Londin E, Ferracin M, Mensa E, Prattichizzo F, Raminini D, et al. Long-term exposure of human endothelial cells to metformin modulates miRNAs and isomiRs. *Sci Rep.* 2020;10(1):21782.
- Loher P, Londin ER, Rigoutsos I. IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget.* 2014;5(18):8790–802.
- Pliatsika V, Loher P, Telonis AG, Rigoutsos I. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics.* 2016;32(16):2481–9.
- 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(19):9158–75.
- 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(6):2973–85.
- 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(28):24797–822.
- Magee RG, Telonis AG, Loher P, Londin E, Rigoutsos I. Profiles of miRNA isoforms and tRNA fragments in prostate cancer. *Sci Rep.* 2018;8(1):5314.
- Telonis AG, Loher P, Magee R, Pliatsika V, Londin E, Kirino Y, et al. tRNA fragments show intertwining with mRNAs of specific repeat content and have links to disparities. *Cancer Res.* 2019;79(12):3034–49.
- Telonis AG, Rigoutsos I. Race disparities in the contribution of miRNA isoforms and tRNA-derived fragments to triple-negative breast cancer. *Cancer Res.* 2018;78(5):1140–54.
- Singh J, Boopathi E, Addya S, Phillips B, Rigoutsos I, Penn RB, et al. Aging-associated changes in microRNA expression profile of internal anal sphincter smooth muscle: role of microRNA-133a. *Am J Physiol Gastrointest Liver Physiol.* 2016;311(5):G964–73.
- Loher P, Karathanasis N, Londin E, Bray P, Pliatsika V, Telonis AG, et al. IsoMiRmap-fast, deterministic, and exhaustive mining of isomiRs from short RNA-seq datasets. *Bioinformatics.* 2021;37(13):1828–38.
- Pliatsika V, Loher P, Magee R, Telonis AG, Londin E, Shigematsu M, et al. MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic Acids Res.* 2018;46(D1):D152–9.
- Pliatsika V, Cherlin T, Loher P, Vliantis P, Nagarkar P, Nersisyan S, et al. MINRbase: a comprehensive database of nuclear- and mitochondrial-ribosomal-RNA-derived fragments (rRFs). *Nucleic Acids Res.* 2024;52(D1):D229–38.
- Bandiera S, Mateogot R, Girard M, Demongéot J, Henrion-Caude A. MitomiRs delineating the intracellular localization of microRNAs at mitochondria. *Free Radic Biol Med.* 2013;64:12–9.
- Catalanotto C, Cogoni C, Zardo G. MicroRNA in control of gene expression: an overview of nuclear functions. *Int J Mol Sci.* 2016;17(10):1712.
- Foldes-Papp Z, König K, Studier H, Buckle R, Breunig HG, Uchugonova A, et al. Trafficking of mature miRNA-122 into the nucleus of live liver cells. *Curr Pharm Biotechnol.* 2009;10(6):569–78.
- Hwang HW, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear import. *Science.* 2007;315(5808):97–100.
- Khudayberdiev SA, Zampa F, Rajman M, Schrott G. A comprehensive characterization of the nuclear microRNA repertoire of post-mitotic neurons. *Front Mol Neurosci.* 2013;6:43.
- Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, et al. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS One.* 2010;5(5):e10563.
- Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell.* 2004;15(2):185–97.
- Stavast C, Erkeland S. The non-canonical aspects of microRNAs: many roads to gene regulation. *Cells.* 2019;8(11):1465.
- Turunen T, Roberts T, Laitinen P, Väänänen M, Korhonen P, Malm T, et al. Changes in nuclear and cytoplasmic microRNA distribution in response to hypoxic stress. *Sci Rep.* 2019;9(1):10332.
- Zhang X, Shen B, Cui Y. Ago HITS-CLIP expands microRNA-mRNA interactions in nucleus and cytoplasm of gastric cancer cells. *BMC Cancer.* 2019;19(1):29.
- Vendramin R, Marine JC, Leucci E. Non-coding RNAs: the dark side of nuclear-mitochondrial communication. *EMBO J.* 2017;36(9):1123–33.
- Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AM, et al. The human mitochondrial transcriptome. *Cell.* 2011;146(4):645–58.
- Sripada L, Tomar D, Prajapati P, Singh R, Singh AK, Singh R. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS One.* 2012;7(9):e44873.
- Noh JH, Kim KM, Abdelmohsen K, Yoon JH, Panda AC, Munk R, et al. HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev.* 2016;30(10):1224–39.
- Xia M, Zhang Y, Jin K, Lu Z, Zeng Z, Xiong W. Communication between mitochondria and other organelles: a brand-new perspective on mitochondria in cancer. *Cell Biosci.* 2019;9:27.
- Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell.* 2009;33(6):717–26.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22(9):1775–89.
- Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA.* 2018;9(3):e1471.
- Zhou X, Yin C, Dang Y, Ye F, Zhang G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci Rep.* 2015;5:11516.
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 2010;464(7291):1071–6.
- Peng W, Koirala P, Mo Y. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene.* 2017;36(41):5661–7.
- Kugel JF, Goodrich JA. Non-coding RNAs: key regulators of mammalian transcription. *Trends Biochem Sci.* 2012;37(4):144–51.
- Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol.* 2003;13(10):807–18.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006;34(Database issue):D140–4.
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011;39(Database issue):D152–7.
- Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc Natl Acad Sci U S A.* 2015;112(10):E1106–15.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1):D155–62.
- Friedlander MR, Lizano E, Houben AJ, Bezdán D, Banez-Coronel M, Kudla G, et al. Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.* 2014;15(4):R57.
- Fromm B, Zhong X, Tarbier M, Friedländer MR, Hackenberg M. The limits of human microRNA annotation have been met. *RNA.* 2022;28(6):781–5.
- Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 2016;44(D1):D184–9.
- Tafforeau L, Zorbas C, Langhendries JL, Mullineux ST, Stamatopoulou V, Mullier R, et al. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol Cell.* 2013;51(4):539–51.
- Weinberg RA, Penman S. Small molecular weight monodisperse nuclear RNA. *J Mol Biol.* 1968;38(3):289–304.

47. Yu S, Lemos B. A portrait of ribosomal DNA contacts with Hi-C reveals 5S and 45S rDNA anchoring points in the folded human genome. *Genome Biol Evol.* 2016;8(11):3545–58.
48. Medhi R, Price J, Furlan G, Gorges B, Sapetschnig A, Miska EA. RNA uridylyl transferases TUT4/7 differentially regulate miRNA variants depending on the cancer cell type. *RNA.* 2022;28(3):353–70.
49. Yang A, Bofill-De Ros X, Stanton R, Shao TJ, Villanueva P, Gu S. TENT2, TUT4, and TUT7 selectively regulate miRNA sequence and abundance. *Nat Commun.* 2022;13(1):5260.
50. Brimacombe R, Stiege W. Structure and function of ribosomal RNA. *Biochem J.* 1985;229(1):1–17.
51. Wilson DN, Doudna Cate JH. The structure and function of the eukaryotic ribosome. *Cold Spring Harb Perspect Biol.* 2012;4(5):a011536.
52. Kumar P, Kuscic C, Dutta A. Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends Biochem Sci.* 2016;41(8):679–89.
53. Magee R, Rigoutsos I. On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Res.* 2020;48(17):9433–48.
54. Aubert M, O'Donohue MF, Lebaron S, Gleizes PE. Pre-ribosomal RNA processing in human cells: from mechanisms to congenital diseases. *Biomolecules.* 2018;8(4):123.
55. Lambert M, Benmoussa A, Provost P. Small non-coding RNAs derived from eukaryotic ribosomal RNA. *Noncoding RNA.* 2019;5(1):16.
56. Rigoutsos I, Lordin E, Kirino Y. Short RNA regulators: the past, the present, the future, and implications for precision medicine and health disparities. *Curr Opin Biotechnol.* 2019;58:202–10.
57. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* 2008;22(20):2773–85.
58. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA.* 2009;15(12):2147–60.
59. Li Z, Ender C, Meister G, Moore PS, Chang Y, John B. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res.* 2012;40(14):6787–99.
60. Pratt A, MacRae I. The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem.* 2009;284(27):17897–901.
61. Chu Y, Yokota S, Liu J, Kilikevicius A, Johnson KC, Corey DR. Argonaute binding within human nuclear RNA and its impact on alternative splicing. *RNA.* 2021;27(9):991–1003.
62. Clark PM, Loher P, Quann K, Brody J, Lordin ER, Rigoutsos I. Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci Rep.* 2014;4:5947.
63. Gagnon K, Li L, Chu Y, Janowski B, Corey D. RNAi factors are present and active in human cell nuclei. *Cell Rep.* 2014;6(1):211–21.
64. Jeffries CD, Fried HM, Perkins DO. Nuclear and cytoplasmic localization of neural stem cell microRNAs. *RNA.* 2011;17(4):675–86.
65. Sarshad AA, Juan AH, Muler AIC, Anastasakis DG, Wang X, Genzor P, et al. Argonaute-miRNA complexes silence target mRNAs in the nucleus of mammalian stem cells. *Mol Cell.* 2018;71(6):1040–50.e8.
66. Honda S, Kirino Y. SHOT-RNAs: a novel class of tRNA-derived functional RNAs expressed in hormone-dependent cancers. *Mol Cell Oncol.* 2016;3(2):e1079672.
67. Mas-Ponte D, Carlevaro-Fita J, Palumbo E, Pulido TH, Guigo R, Johnson R. LncAtlas database for subcellular localization of long noncoding RNAs. *RNA.* 2017;23(7):1080–7.
68. 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(12):1603–14.
69. Olive V, Jiang I, He L. mir-17-92, a cluster of miRNAs in the midst of the cancer network. *Int J Biochem Cell Biol.* 2010;42(8):1348–54.
70. Akins RB, Ostberg K, Cherlin T, Tsiouplis N, Loher P, Rigoutsos I. The typical tRNA co-expresses multiple 5' tRNA halves whose sequences and abundances depend on isodecoder and isoacceptor and change with tissue type, cell type, and disease. *Noncoding RNA.* 2023;9(6):69.
71. Guan L, Grigoriev A. Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. *Nucleic Acids Res.* 2021;49(7):4085–103.
72. Li S. Human 28s rRNA 5' terminal derived small RNA inhibits ribosomal protein mRNA levels. *bioRxiv.* 2019:618520. <https://doi.org/10.1101/618520>.
73. Chen Z, Sun Y, Yang X, Wu Z, Guo K, Niu X, et al. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS One.* 2017;12(4):e0176458.
74. Madore SJ, Wieben ED, Pederson T. Intracellular site of U1 small nuclear RNA processing and ribonucleoprotein assembly. *J Cell Biol.* 1984;98(1):188–92.
75. Terns MP, Dahlberg JE, Lund E. Multiple cis-acting signals for export of pre-U1 snRNA from the nucleus. *Genes Dev.* 1993;7(10):1898–908.
76. Vegvar HEND, Dahlberg JE. Nucleocytoplasmic transport and processing of small nuclear RNA precursors. *Mol Cell Biol.* 1990;10(7):3365–75.
77. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101–8.
78. Wong TW, Clayton DA. DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. *Cell.* 1986;45(6):817–25.
79. Entelis NS, Kolesnikova OA, Dogan S, Martin RP, Tarassov IA. 5S rRNA and tRNA import into human mitochondria. Comparison of in vitro requirements. *J Biol Chem.* 2001;276(49):45642–53.
80. Jeandard D, Smirnova A, Tarassov I, Barrey E, Smirnov A, Entelis N. Import of non-coding RNAs into human mitochondria: a critical review and emerging approaches. *Cells.* 2019;8(3):286.
81. Smirnov A, Tarassov I, Mager-Heckel AM, Letzelter M, Martin RP, Krashenninnikov IA, et al. Two distinct structural elements of 5S rRNA are needed for its import into human mitochondria. *RNA.* 2008;14(4):749–59.
82. Carlevaro-Fita J, Rahim A, Guigó R, Vardy L, Johnson R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA.* 2016;22(6):867–82.
83. Carlevaro-Fita J, Johnson R. Global positioning system: understanding long noncoding RNAs through subcellular localization. *Mol Cell.* 2019;73(5):869–83.
84. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
85. Guo C, Xu G, Chen L. Mechanisms of long noncoding RNA nuclear retention. *Trends Biochem Sci.* 2020;45(11):947–60.
86. Guo C, Ma X, Xing Y, Zheng C, Xu Y, Shan L, et al. Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell.* 2020;181(3):621–36.e22.
87. Van Bortle K, Marciano D, Liu Q, Chou T, Lipchik A, Gollapudi S, et al. A cancer-associated RNA polymerase III identity drives robust transcription and expression of snaR-A noncoding RNA. *Nat Commun.* 2022;13(1):3007.
88. Tarassov IA, Entelis NS. Mitochondrially-imported cytoplasmic tRNA(Lys) (CUU) of *Saccharomyces cerevisiae*: in vivo and in vitro targeting systems. *Nucleic Acids Res.* 1992;20(6):1277–81.
89. Rinehart J, Krett B, Rubio MA, Alfonso JD, Soll D. *Saccharomyces cerevisiae* imports the cytosolic pathway for Gln-tRNA synthesis into the mitochondrion. *Genes Dev.* 2005;19(5):583–92.
90. Laforest MJ, Delage L, Marechal-Drouard L. The T-domain of cytosolic tRNAVal, an essential determinant for mitochondrial import. *FEBS Lett.* 2005;579(5):1072–8.
91. Rubio MA, Hopper AK. Transfer RNA travels from the cytoplasm to organelles. *Wiley Interdiscip Rev RNA.* 2011;2(6):802–17.
92. Honda S, Loher P, Shigematsu M, Palazzo JP, Suzuki R, Imoto I, et al. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc Natl Acad Sci U S A.* 2015;112(29):E3816–25.
93. Geslain R, Pan T. Functional analysis of human tRNA isodecoders. *J Mol Biol.* 2010;396(3):821–31.
94. Honda S, Morichika K, Kirino Y. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat Protoc.* 2016;11(3):476–89.
95. Hollestelle A, Nagel JH, Smid M, Lam S, Elstrodt F, Wasielewski M, et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat.* 2010;121(1):53–64.
96. Subik K, Lee J-F, Baxter L, Strzepak T, Costello D, Crowley P, et al. The expression patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by immunohistochemical analysis in breast cancer cell lines. *Breast Cancer Basic Clin Res.* 2010;4:117822341000400000.
97. 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.

98. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–2.
99. Magee R, Loher P, Londin E, Rigoutsos I. Threshold-seq: a tool for determining the threshold in short RNA-seq datasets. *Bioinformatics.* 2017;33(13):2034–6.
100. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
101. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci.* 1992;7(4):457–72.
102. Cherlin T, Rigoutsos I. MiRNA isoforms, tRNA-derived fragments, and rRNA-derived fragments have enrichments that differ characteristically among sub-cellular compartments and among cell lines. National Institutes of Health, Sequence Read Archive (SRA). Accession #: PRJNA816866 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA816866>.
103. Telonis A, Rigoutsos I. MiRNA isoforms, tRNA-derived fragments, and rRNA-derived fragments have enrichments that differ characteristically among sub-cellular compartments and among cell lines (Supplemental). National Institutes of Health, Sequence Read Archive (SRA). Accession #: PRJNA1096643 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1096643>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.