



Published in final edited form as:

*Cancer Res.* 2019 June 15; 79(12): 3034–3049. doi:10.1158/0008-5472.CAN-19-0789.

## tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities

Aristeidis G. Telonis<sup>1</sup>, Phillipe Loher<sup>1</sup>, Rogan Magee<sup>1</sup>, Venetia Pliatsika<sup>1</sup>, Eric Londin<sup>1</sup>, Yohei Kirino<sup>1</sup>, and Isidore Rigoutsos<sup>1</sup>

<sup>1</sup>Computational Medicine Center, Thomas Jefferson University, Philadelphia, PA 19107, USA

### Abstract

tRNA-derived fragments (tRFs) are a class of potent regulatory RNAs. We mined the datasets from The Cancer Genome Atlas representing 32 cancer types with a deterministic and exhaustive pipeline for tRNA fragments. We found that mitochondrial tRNAs contribute disproportionately more tRFs than the nuclear ones. Through integrative analyses, we uncovered a multitude of statistically significant and context-dependent associations between the identified tRFs and mRNAs. In many of the 32 cancer types, these associations involve mRNAs from developmental processes, receptor tyrosine kinase signaling, the proteasome, and metabolic pathways that include glycolysis, oxidative phosphorylation, and ATP synthesis. Even though the pathways are common to multiple cancers, the association of specific mRNAs with tRFs depend on and differ from cancer to cancer. The associations between tRFs and mRNAs extend to genomic properties as well: specifically, tRFs are positively correlated with shorter genes that have a higher density in repeats, such as ALUs, MIRs, and ERVLS. Conversely, tRFs are negatively correlated with longer genes that have a lower repeat density, suggesting a possible dichotomy between cell proliferation and differentiation. Analyses of bladder, lung, and kidney cancer data indicate that the tRF-mRNA wiring can also depend on a patient's sex; sex-dependent associations involve cyclin-dependent kinases in bladder cancer, the MAPK signaling pathway in lung cancer, and purine metabolism in kidney cancer. Taken together, these findings suggest diverse and wide-ranging roles for tRFs and highlight the extensive interconnections of tRFs with key cellular processes and human genomic architecture.

---

Corresponding author: isidore.rigoutsos@jefferson.edu Phone: +1 215 503 4219, FAX: +1 215 503 0466 Computational Medicine Center, Jefferson Alumni Hall #M81, 1020 Locust Street, Philadelphia, PA 19107, USA.

#### Author contributions

IR and AGT designed the study with contributions from all co-authors. PL downloaded the data and generated the tRF profiles. RM contributed the sequencing thresholds. PL and VP designed and implemented the underlying infrastructure. AGT mined the tRF profiles and their correlations with mRNAs. AGT, PL, VP and IR analyzed the cleavage patterns. RM and IR analyzed the His(-1U) tRFs. AGT, IR and RM wrote the manuscript with contributions from PL, EL, YK, and VP. All authors have read and approved the final manuscript.

**Conflicts of interest:** The authors declare no conflicts of interest.

#### Data access

All tRFs can be bulk-downloaded from <https://cm.jefferson.edu/tcga-mintmap-profiles> or examined interactively through MINTbase (4,23) at <https://cm.jefferson.edu/MINTbase>.

## Keywords

transfer RNA; tRNA; tRNA-derived fragments; tRFs; 5'-tRFs; i-tRFs; 3'-tRFs; 5'-tRNA halves; 5'-tRHs; 3'-tRNA halves; 3'-tRHs; cancer; sex-dependence of tRFs; race-dependence of tRFs; tRNA<sup>HisGTG</sup>; repeat elements; The Cancer Genome Atlas; TCGA

---

## Introduction

tRNA-derived fragments (tRFs) are a new molecular category of short non-coding RNAs (ncRNAs) that are produced from specific cleavage of precursor and mature transfer RNA (tRNA) molecules (1,2). For tRFs that overlap the mature tRNA, four structural categories were reported initially: 5'-tRFs, 3'-tRFs, 5'-tRNA halves (5'-tRHs), and 3'-tRNA halves (3'-tRHs) (3). Our group recently identified and reported a fifth category, the internal tRFs or i-tRFs, with numerous and abundant members (4,5).

The production of tRFs changes in response to factors like diet variations (6) and trauma (7). It has also been shown that tRFs are produced in a tissue-dependent manner (5), exhibit differential abundances in cancer compared to normal tissue (8,9), and are involved in trans-generational inheritance (10).

Mechanistically, tRFs can regulate translation (11) as well as interact with the ribosome and aminoacyl tRNA synthetases (11,12). The specific category of tRHs can be produced under stress conditions (11,13) as well as constitutively (5) and its members can have endpoint modifications that render them invisible to standard RNA-seq (14). Some of the shorter tRFs can interact with Argonaute proteins (3,15,16) in a cell-type specific manner (5). tRFs that enter the RNA interference (RNAi) pathway follow base-pairing rules that match those of microRNAs (miRNAs) (3,17). Interestingly, tRF loading on Argonaute can be both Dicer-dependent and Dicer-independent (3,17). Moreover, the decoying of an RNA-binding protein (RBP) by tRFs (18) and the potentially extensive binding of tRFs by RBPs (19) can affect cancer molecular biology and metastasis.

Previously, we showed that tRFs have links to Precision Medicine and hold promise for furthering our understanding of homeostasis and disease. Specifically, we demonstrated that the identity and abundance of tRFs depend on a person's sex, population origin, and race/ethnicity, as well as tissue, tissue state, and disease type (5,8,19). We also found that the associations ('wiring') of tRFs with mRNAs and molecular pathways is race/ethnicity-specific in prostate cancer (PRAD) (8) and triple negative breast cancer (19).

The current literature highlights the heterogeneity of the tRFs' roles, mechanisms of action, and functional impact. Given this heterogeneity, we pursued our studies in an integrated manner. We holistically investigated unexplored areas of tRF expression and patterns of inter-transcript (tRFs-mRNAs) associations in 32 cancer types of The Cancer Genome Atlas (TCGA) cohort. Our analysis distinguished tRFs based on whether they can be traced back to the nuclear or the mitochondrial (MT) genome. We also examined the attributes of the mRNAs with which the tRFs are correlated and sought potential dependencies of the tRF-

mRNA associations on cancer type. Lastly, we examined possible links between tRFs and the repeat-element content of genes, and between tRFs and sex disparities.

## Materials and Methods

### Data acquisition and tRNA fragment profile generation

We downloaded 11,198 short RNA-seq datasets (sequenced samples) from TCGA's Cancer Genomic Hub (CGHub) and the respective clinical metadata from TCGA's data portal. We converted TCGA's small RNA-seq datasets to FASTQ format using *bam2FastQ* (<http://genome.sph.umich.edu/wiki/BamUtil-v1.0.10>). To ensure consistency with previously reported TCGA analyses, we worked with the GRCh37/hg19 genome assembly. We used MINTmap (20,21) to exhaustively and deterministically mine tRFs in all 11,198 datasets. These tRFs were recently made available through an interactive database (4). We computed an *adaptive minimum-support threshold* for each dataset using the Threshold-seq algorithm (22) retaining only those tRFs that exceeded threshold and also had mean normalized abundance  $\geq 1$  reads-per-million (RPM) in at least one of the 32 cancer types. We refer to tRFs using the "license plate" naming scheme that we introduced previously (5,23). We tag tRF sequences as "exclusive" if they exist only within the span of mature tRNAs that contain a CCA (the "tRNA space") and appear nowhere else on the genome; otherwise we tag them as *ambiguous*. In the presented analyses we included the 10,274 "white-listed" datasets that contain no special annotations in the associated clinical metadata.

### Race/ethnicity

We adhere to the NIH/TCGA designations. White (Wh) refers to person with origins in any of the original peoples of the far Europe, the Middle East, or North Africa. Black or African American (B/Aa) refers to persons with origins in any of the black racial groups of Africa.

### tRF abundances/networks

We define the 'normalized abundance' of an isoacceptor (in RPM) for a dataset as the sum of the (normalized) abundances of all the tRFs that the isoacceptor produces. We define the 'normalized abundance' of an isoacceptor for a cancer type as the average of the isoacceptor's abundances across all the datasets of this cancer type. For the 5'-tRFs from tRNA<sup>HisGTG</sup>, and separately for each sample/dataset, we computed the ratio of abundance of 5'-tRF ending at consecutive positions, then  $\log_2$ -transformed it, filtered out infinite values in the ratios (divisions by 0), and computed mean and standard deviation separately for tumor and normal samples. We distinguished 5'-tRFs starting at position -1 from 5'-tRFs starting at position +1 of tRNA<sup>HisGTG</sup>. Univariate statistical comparisons in abundance were carried out with the non-parametric Mann-Whitney U-test. For network visualizations, we collapsed all tRFs to the isoacceptor level: e.g., a node labeled "nAspGTC" represents *all* expressed tRFs that overlap the mature *nuclear* tRNA<sup>AspGTC</sup>.

### tRNA base modifications and mapping with mismatches

We leveraged the known modifications of human tRNAs and the respective sequence alignments contained in MODOMICS (24). Per MODOMICS, a base's "frequency of modification" is defined as the ratio of tRNAs with a modification at that base over the

number of considered tRNAs. Intuitively, had a modified base at position  $N$  of the mature tRNA stopped reverse transcription during library preparation, then tRFs from this tRNA would have appeared to possess a “pseudo-5’” terminus at position  $N+1$ . We examined this possibility at each mature tRNA position by counting the number of tRFs starting at that position and comparing it to the modification frequency that MODOMICS lists for the position immediately upstream.

Additionally, for 3’-tRFs, we also evaluated whether there is benefit in mapping tRFs after allowing a single nucleotide mismatch. Such a mismatch would, in principle, alleviate the potential impact on the sequenced reads of the known  $m^1A_{58}$  modification (methylated adenine at position 58 of the mature tRNA). To this end, for all 3’-tRFs, we enumerated all possible sequence variations that result from changing exactly one nucleotide anywhere along the 3’-tRF at hand. We then examined which of these derivative “3’-tRFs” are: (1) supported by the available RNA-seq data; and, (2) can now be found in the genome outside of the sequence space of tRNAs identically (without when allowing no mismatch).

### Correlations

We computed positive and negative tRF-mRNA Spearman *rho* correlation coefficients using only the tumor datasets, and separately for each cancer type. For bladder urothelial carcinoma (BLCA), lung adenocarcinoma (LUAD) or kidney renal clear cell carcinoma (KIRC), we split the tumor datasets by sex and carried out the correlations separately for each sex. For KIRC, we only considered samples that belonged in the ccA cluster in the TCGA analysis (25). For increased stringency, we further required that the median normalized abundance of each tRF be  $\geq 2$  RPM within the group of considered samples. For the mRNA profiles, we used TCGA’s files of normalized results (“*rsem\_genes.normalized\_results*”) filtering out any genes whose average abundance was less than the median of the means of abundances of all mRNAs across the primary tumor samples of the group under consideration. We determined which tRFs and mRNAs enter the correlation analyses separately for each cancer type (or, sex in the case of BLCA, LUAD, and KIRC-ccA). We used Python’s *numpy* (version 1.11.1) and *scipy* (version 0.18.1) packages. We kept tRF-mRNA pairs with Spearman correlation  $\leq -0.333$  or  $\geq 0.333$  and an associated False Discovery Rate (FDR)  $\leq 5\%$ . For most cancer types, we found tens of thousands of tRF-mRNA correlations satisfying these constraints. To focus on the strongest of the correlations and to balance stringency and specificity, we analyzed only tRF-mRNA pairs corresponding to the 5,000 highest (positive) and 5,000 lowest (negative) correlation values. All retained correlation pairs, correlation values, and FDRs are listed in the Supplement.

We computed commonly and differentially correlated tRF-mRNA pairs as we have previously done (19). Specifically, a tRF-mRNA pair is *commonly correlated* in both cancer types (e.g. LUAD and LUSC) or sexes (e.g. BLCA Male and BLCA Female) if it is listed among the significant pairs with the same sign in the two categories being compared. A pair is *differentially correlated* if it is listed among the significant pairs in exactly one of the two categories being considered or listed among the significant pairs in both categories but with opposite signs.

## Computing enrichments

For hierarchical clustering as well as visualizations, we used R and Cytoscape, as described previously (5,19). Protein-protein interactions were drawn from PICKLE (26). We carried out pathway enrichment analysis with DAVID (27) using as “background” all the genes that passed the expression filtering for the respective cancer type. We examined GO terms for biological processes (GOTERM\_BP\_FAT), molecular function (GOTERM\_MF\_FAT), cell compartment (GOTERM\_CC\_FAT), and KEGG pathways (KEGG\_PATHWAY) for enrichment filtering at an FDR threshold of 0.05. BP terms in more than 10 cancer types were grouped into clusters based on the pairwise Jaccard index matrix measuring how common two BP GO terms are in terms of the genes they contain and are correlated with tRFs. We reduced and summarized the grouped GO terms with REVIGO (28) (allowed similarity=medium, similarity measure=normalized Resnik).

To construct the network of glycolysis/gluconeogenesis we connected the genes whose encoded proteins interact with the same metabolite: we downloaded the hsa00010 KEGG pathway structure, and identified the genes in the KEGG modules hsa\_M00001, hsa\_M00002, and hsa\_M00003. We treated two genes as connected if and only if the reactions catalyzed by the encoded proteins include at least one common metabolite as substrate or product, collapsing duplicate edges to a single node. We constructed the network of purine metabolism in a similar manner by connecting all genes the enzymes in the biosynthesis of IMP from Ribose 1-Phosphate (KEGG pathway hsa00230). The gene collections for ribosomal and proteasomal proteins are from HGNC (<https://www.genenames.org/cgi-bin/genefamilies/>).

Protein localization data are from UniProt (29) (October 10, 2017). For each cancer type and the mRNAs participating in correlations with tRFs, we identified the cellular compartments and destinations of the corresponding encoded proteins. We used a  $\chi^2$  test of homogeneity of proportions to determine enrichment and/or depletion of compartments within this mRNA set compared to the mRNAs not participating in the correlations. We only used enriched or depleted compartments with an FDR of <5% and residual scores  $+3$  (i.e. enriched and colored gold) or  $-3$  (i.e. depleted and colored purple), respectively.

We used RepeatMasker (<http://www.repeatmasker.org>; hg19 version 4.0.5) to find the overlap of human genes with repeat elements. The coordinates of all introns and exons are from ENSEMBL 75. We formed a gene’s genomic span by taking the union of all its unspliced variants. We define the *exonic* portion of a gene as the union of all its exons. The gene’s *intronic* portion is what remains of the gene’s span after removing its exonic portion. We ran all Monte-Carlo simulations for 10,000 iterations, choosing genes randomly from the pool of background genes in each iteration, i.e. the genes that participated in the correlation analyses. For the randomly-chosen genes, we computed: the average span of the chosen genes, and the average density of each repeat element family, and did so using: (i) only the genes’ exonic portion to which we refer as the “*mRNA space*” of a gene; and, (ii) only the genes’ intronic portion. We defined a repeat family’s density (“repeat content”) in a genomic region as the fraction of the region that is annotated as belonging to the family. Upon completion of all the iterations, we built the ‘expected’ distribution for each parameter and used Z-scores to evaluate the enrichment/depletion of the ‘observed’ parameter. We used the

mean across the gene set that is correlated (either positively or negatively) with tRFs as the 'observed' value. In total, we carried out 128 rounds of Monte-Carlo simulations (each with 10,000 iterations). Shades of gold color represent enrichment (Z-score  $+2$ ). Shades of purple color represent depletion (Z-score  $-2$ ). To account for multiple testing, we also carried out Kolmogorov-Smirnov tests checking whether the observed distribution is statistically significantly different than the background one. The resulting P values were corrected to FDR values and are included in the Supplemental Tables.

For the case of BLCA, we carried out two different Monte-Carlo simulations with 10,000 iterations. In the first simulation, we examined whether the deviations from the diagonal (where the number of correlations of an isoacceptor in males equals the ones in females) are random. For each iteration, separately for males and females, we randomly selected (with replacement) the same number of tRFs as counted in the correlations. We then collapsed the tRFs to the isoacceptor level and for each one we calculated the distance from the diagonal, i.e. its deviation from exhibiting the same number of correlations in both sexes and computed the median distance across them. Thus, we built an expected distribution. We also computed the median distance of the data from the diagonal and used the expected distribution to calculate the Z-Score. In the second simulation, and separately for each sex, we reassigned to each tRF the mRNAs with which it formed correlated pairs: we chose only among mRNAs that already participated in tRF-mRNA correlations while maintaining the number of correlations for each tRF unchanged: this allows us to estimate the significance of the differentially correlated observed tRF-mRNA pairs. At the end of the simulation (10,000 iterations), we had an expected distribution of the differences in correlation coefficients that we plotted against the observed coefficient differences. Additionally, we point out that out of the 10,000 iterations, we found only 12 simulated tRF-mRNA pairs that would be considered as differentially correlated based on their coefficient and FDR values – this is vastly smaller than the 5,520 tRF-mRNA pairs that emerge from our analysis of the BLCA datasets, which emphasizes the statistical significance of the findings.

### Repeat element density and its correlation with biological processes

In the context of the recent literature around repeats and tRFs, a central finding in our work was the link of tRFs with mRNAs with specific distribution in repeats. It is important to note that we came to these results by analyzing the *relative abundances* of transcripts. We did not take into account the magnitude of transcript abundance, or any genomic properties. In the context of genome organization, several publications previously highlighted the non-randomness of repeat elements and their correlations with other gene properties, like length and GC content (see Results and Discussion). Therefore, it seems reasonable to hypothesize that the tRF-mRNA correlations at the *transcriptomic* level are the dynamic manifestation of the (static) genomic landscape. We emphasize the dynamic nature of these transcriptomic correlations because we observed that the same genomic properties significantly emerge in multiple contexts (cancer types), even though the tRF-mRNA pairs exhibit a strong context-specific signature. From this perspective, we examined how repeat element distribution is associated with biological processes, or, in other words, what is the overarching *genomic* architecture that is intertwined with the *transcriptomic* results.

We focused on the repeat elements that are mostly correlated with tRFs, namely SINE (ALU and MIR), LINE (L1 and L2), LTR (ERVL-MaLR) and DNA transposons (TcMar-Tigger and hAT-Charlie), and their instances that are sense to the whole genomic span of protein-coding genes. These analyses do not aim at providing a thorough investigation but rather at demonstrating the links between biological processes and these repeats. For each repeat family, we computed the repeat content as described above and ranked the genes. Then, for each gene, we calculated the average rank and sorted the genes based on this metric. We considered the top 3,000 and the bottom 3,000 genes corresponding to the genes with the highest and lowest, respectively, density in repeats. We note that for the majority of the genes with low repeat density, the mean density value is zero, or close to zero. On these two lists, we carried out examined enrichments using DAVID with the same parameters as described above.

## Results

We mined all the datasets (see Methods) of the TCGA repository and identified 23,413 tRFs that overlap mature tRNAs. The tRFs can be bulk-downloaded from <https://cm.jefferson.edu/tcga-mintmap-profiles>, or, examined interactively through MINTbase at <https://cm.jefferson.edu/MINTbase/> (4,23). We focus the below analyses on 10,274 white-listed TCGA samples and the corresponding 20,722 tRFs with significant expression in at least one of these datasets (Supplemental Table S1). 16,133 (78%) of the discovered fragments belong to the new category of i-tRFs, in complete analogy to what we reported previously (5,8,19). We also identified 1,717 5'-tRFs (8% of all the identified tRFs), 2,840 3'-tRFs (14%) and 32 5'-tRHs. Fragments with lengths  $\geq 28$  nt are likely truncated versions of longer fragments that have been (artificially) shortened as a result of the 30-cycle limitation of the TCGA sequencing protocol (30). This limitation results in an under-representation of halves among the identified tRFs. Of the 20,722 tRFs, 13,904 (67%) have sequences that can be found only inside the tRNA space ("exclusive" tRFs). The sequences of the remaining 6,818 (33%) tRFs are of ambiguous genomic origin, i.e., one third of the identified tRFs may not arise from tRNA genes.

Because the tRFs we analyze are believed to originate from mature tRNAs, they are expected to inherit the parental molecule's base modifications. In principle, such modifications could hinder the reverse-transcription step during sequencing leading to an artificial 5' endpoint for some tRFs, or, to misreading the nucleotide at the modified location. Our analysis did not find evidence that the presence of base modifications affects the identity of tRFs derived from TCGA (Supplemental Fig. S1A-B). Our analysis also confirmed that permitting nucleotide mismatches when mapping reads to the genome greatly hinders one's ability to distinguish among tRFs and non-tRFs (Supplemental Fig. S1C). Thus, we enforced exact matching during read-mapping, which is a key property of MINTmap (20,21) – see Methods.

### The case of 5'-tRFs from the nuclear tRNA<sup>His</sup>GTG

Among the many diverse tRFs, the nuclear tRNA<sup>His</sup>GTG stands apart as a notable exception (Supplemental Figure S1D-G). We previously reported in a model human cell line (BT-474)

that 5'-tRNA halves from tRNA<sup>HisGTG</sup> have the *expected* guanosine added to their 5' termini ("position -1") as well as an *unexpected* uracil (31). We refer to these molecules using the "His(-1G)" and "His(-1U)" qualifier, respectively. Examination of the -1 positions of all 5'-tRFs from tRNA<sup>HisGTG</sup> across all TCGA samples revealed unexpectedly that His(-1U) is the most abundant modification (Fig. 1A). A smaller portion of the 5'-tRFs from tRNA<sup>HisGTG</sup> contain an adenine at the -1 position, or no modification. Even fewer 5'-tRFs from tRNA<sup>HisGTG</sup> contain a guanosine or a cysteine (Fig. 1A). We also found that the His(-1U) 5'-tRFs exhibit a notable property: as nucleotides are progressively added to their 3' terminus, their abundance levels oscillate through position 23: while the *absolute abundances* of these 5'-tRFs change among cancer types (Supplemental Fig. S1D), the *abundance ratios* of His(-1U) 5'-tRFs that differ by 1 base at their 3' terminus is conserved (see Fig. 1B for an example; full data matrix included in (Supplemental Table S2). This 'see-saw' pattern persists across all 32 TCGA cancer types and extends to the normal tissue samples as well (Supplemental Table S2; Supplemental Fig. S2A). Note that the His(+1G), i.e. the unmodified 5'-tRFs beginning at position +1 of tRNA<sup>HisGTG</sup> or other isoacceptors do not exhibit this exact pattern or can exhibit other patterns (Supplemental Fig. S2B). We note that analogous patterns have been reported for tRNA-derived piRNAs in *Bombyx mori* (32).

Interestingly, the MT tRNA<sup>HisGTG</sup> does not generate any 5'-tRFs even though it produces a similar number of i-tRFs and 3'-tRFs as the nuclear tRNA<sup>HisGTG</sup>. Comparison of the relative abundances of nuclear and MT tRNA<sup>HisGTG</sup> fragments showed that they are not correlated (Fig. 1C). Moreover, across all TCGA datasets and cancer types, tRFs from the MT tRNA<sup>HisGTG</sup> are considerably less abundant than their nuclear counterparts (Fig. 1C, diagonal of Fig. 1D, and Supplemental Fig. S1E-F).

### tRF lengths and tRNA cleavage patterns depend on the genome of origin

Motivated by the differences between the nuclear and MT tRNA<sup>HisGTG</sup>, and analogous differences we reported previously (5) in healthy and diseased samples, we extended our analyses and comparisons to the rest of the nuclear and MT isoacceptors.

In terms of unique tRF sequences, the contribution by the 22 MT tRNAs comparatively eclipses that by the 610 nuclear tRNAs. We stress that this statement is about the diversity in the *identity* of the produced molecules and not about their relative abundance levels. The 22 MT tRNAs (3.5% of all tRNAs) are responsible for 6,031 (29%) of all distinct tRFs we find in TCGA. We note here that the human nuclear chromosomes are riddled by MT-like sequences (NUMTs) as well as by hundreds of tRNA-lookalikes (33). Thus, it is conceivable that MT tRFs are not the product of the MT genome exclusively.

In terms of length distributions, there are concrete differences between tRFs produced from nuclearly-encoded tRNAs and their MT-encoded counterparts. These differences persist in all analyzed cancer types (Supplemental Fig. S1G and Supplemental Table S3), mirror our previous findings in lymphoblastoid cells from healthy individuals (5), and suggest potentially distinct roles for nuclear and MT tRFs in cancer.

To investigate the relative contribution of the nuclear and MT genomes to the pool of present RNAs, we calculated the abundance of tRFs across cancer types at the isoacceptor level,



doing so separately for MT and nuclear isoacceptors. Use of unsupervised hierarchical clustering groups separates nearly all MT-encoded isoacceptors from their nuclear counterparts. However, Fig. 2A also shows that the abundance of MT tRFs depends on cancer type. We investigated this further by correlating the expression of tRFs per isoacceptor with the MT DNA copy number in 22 cancer types (34). With the exception of adrenocortical carcinoma (ACC) and kidney renal papillary cell carcinoma (KIRP), the average correlation coefficient of MT tRFs with mtDNA copy number was low (Spearman's  $\rho < 0.4$ ) (Supplemental Fig. S3).

We also examined the clustering of tRFs when we consider their structural category. Separately for MT and nuclear fragments, we computed the abundance levels of 5'-tRFs, i-tRFs, and 3'-tRFs. Because i-tRFs represent a more heterogeneous group of molecules, we divided them into six sub-categories based on the location along the mature tRNA of an i-tRF's 5' terminus. Hierarchical clustering reveals some groupings of the structural categories that persist across the 32 cancer types (Fig. 2B): e.g., nuclear and MT i-tRFs that begin in region A are correlated with those that begin in the D loop; nuclear i-tRFs that begin in region B are correlated with nuclear 3'-tRFs; etc. A detailed cleavage analysis of tRNAs, which we carried separately for each of the 32 cancer types and tRF structural categories, further supports these findings (Supplemental Fig. S4).

These findings suggest that tRF production strongly depends on their genomic origin, with nuclear and MT tRNAs producing characteristically different tRFs, in terms of identity and abundance. tRF production and the resulting tRF profiles are the outcome of currently-unknown mechanisms.

### The associations between tRFs and mRNAs depend on molecular context

To identify links between tRFs and biological processes, we studied each of the 32 cancers for patterns of statistically-significant correlations between tRFs and mRNAs. We filtered these tRF-mRNA correlations using stringent criteria (see Methods) while examining MT tRFs separately from nuclear tRFs. We find that the *identities* of the tRFs and mRNAs that are present in the analyzed samples remain largely unchanged across cancer types (Supplemental Fig. S5A-C, and Supplemental Tables S4 and S5). However, the identities of the tRFs and mRNAs that are statistically-significantly correlated with one another change dramatically from one cancer type to the next (Supplemental Fig. S5D-G). Intriguingly, we find that the correlations with MT tRFs primarily comprise 3'-tRFs whereas the ones with nuclear tRFs include a mixture of all structural categories with a preference for 5'-tRFs or i-tRFs in some cancer types (Fig. 2C).

To examine whether the observed correlation patterns reflect tissue-specific (and not cancer-type-specific) events, we considered lung adenocarcinoma (LUAD) and kidney renal cell clear carcinoma (KIRC). For these two cancer types, the TCGA has analyzed additional types from the same tissue: lung squamous cell carcinoma (LUSC), kidney chromophobe (KICH), and KIRP. Examining the number of commonly- and differentially-correlated tRF-mRNA pairs (19), we found that LUAD is closer to LUSC than to any other cancer type (Supplemental Fig. S5H): nonetheless, only a mere 14% of the significantly-correlated tRF-mRNA pairs in LUAD are shared with LUSC. For KIRC, we found that it is as far from

KIRP and KICH as it is from any other cancer type. These data indicate that in addition to cancer type representing a major contribution to our analyses (Supplemental Fig. S5I) there can also be a tissue-specific contribution for some cancers (Supplemental Fig. S5H). Since decoupling the contribution of each component is not comprehensively feasible with the data that is available in TCGA, we refer to these correlations as “context-specific.” This is a particularly notable observation that echoes our recent report on triple negative breast cancer (19).

### **tRFs are positively correlated with shorter mRNAs and negatively correlated with longer mRNAs**

With the tRF-mRNA correlation pairs in hand, we examined whether the involved mRNAs exhibit length biases. We used Monte-Carlo simulations and Kolmogorov-Smirnov tests corrected for multiple testing (see Methods, and Supplemental Table S6) to evaluate the length of the mRNA, as the length of the union of the respective exonic sequences. As Fig. 3A shows, the mRNAs that are positively correlated with tRFs are, on average, significantly shorter than the average length of the expressed mRNAs. Also, the mRNAs that are negatively correlated with tRFs are, on average, significantly longer. This holds true for both nuclear and MT tRFs, and most of the 32 cancer types. To highlight the differences in distributions we show in Supplemental Fig. S6 box-plots of the length distributions for each of the four combinations: two correlation signs x two genomes-of-origin. These length biases persist when instead of mRNAs we examine the length of the *intronic portion* for genes whose mRNAs are correlated with tRFs (Supplemental Fig. S7). These results suggest preferential interactions of tRFs with genes of specific genomic architecture.

### **The localization of the encoded proteins is dependent on the genomic origin of the correlated tRFs**

Next, we systematically examined the cellular localization of proteins encoded by mRNAs that participate in tRF-mRNA correlations. We considered seven destinations: nucleus, cytoplasm, endoplasmic reticulum (ER)-Golgi, mitochondrion, cell membrane, secreted, and “other” (e.g. vesicles, endosomes, etc.). Again, we separated positive from negative correlations, and nuclear tRFs from MT tRFs.

We find that tRFs are correlated with mRNAs whose protein products localize to various combinations of the seven destinations following localization patterns that are context-specific (Fig. 3B-C). As far as positive correlations are concerned, mRNAs encoding nuclear proteins are significantly *enriched* in some cancer types, e.g. breast cancer (BRCA) and testicular germ cell tumors (TGCT) (Fig. 3B). But in uveal melanoma (UVM) and thyroid cancer (THCA), mRNAs encoding nuclear proteins are significantly *depleted* (Fig. 3B). We note that mRNAs encoding proteins destined for the MT are consistently enriched among the positive correlations with MT tRFs in 30 of the 32 cancers.

Analogous observations can be made for the negative correlations (Fig. 3C). In colon adenocarcinoma (COAD), rectum adenocarcinoma (READ), esophageal carcinoma (ESCA), and TGCT the negative correlations of both MT and nuclear tRFs are *depleted* in mRNAs encoding proteins destined for the nucleus or the MT; yet, they are *enriched* in the mRNAs

of secreted or cell membrane proteins. In other cancer types like PRAD, lower grade glioma (LGG), KIRP, and LUAD, the negative tRF-mRNA correlations include *more* mRNAs whose proteins are destined for the nucleus than expected by chance (Fig. 3C). UVM is another interesting case: the negative correlations involving *nuclear* tRFs are *enriched* in mRNAs whose proteins are destined for the nucleus whereas those involving MT tRFs are depleted in this regard.

These findings suggest extensive and context-specific flow of information across cellular compartments.

### The mRNAs that are correlated with tRFs differ by cancer type but often belong to the same biological processes

Having established that the identity of mRNAs that are correlated with nuclear or MT tRFs depends on the molecular context, we examined whether this dependence extends to biological processes (Supplemental Table S7 and S8). We identified multiple examples of tRF-mRNA correlations and associated pathways that are enriched in only one cancer type (Supplemental Fig. S8A). For example, mRNAs from bile secretion, several amino acid metabolic pathways, and xenobiotic metabolism are exclusively prevalent among the tRF-mRNA correlations in liver hepatocellular carcinoma (LIHC). As another example, mRNAs from the steroid biosynthesis pathway are prevalent only among the tRF-mRNA correlations in ACC.

We also found pathways that are enriched among the tRF-mRNA correlations in multiple cancers. For example, the KEGG pathway “ribosome” (hsa03010) is significantly overrepresented in 21 cancer types (Supplemental Tables S7 and S8) and the corresponding mRNAs are correlated with both nuclear and MT tRFs (Fig. 4A). For clarity, each tRF node in this figure represents all tRFs from the corresponding isoacceptor. Note how four MT tRNAs (mt-tRNA<sup>ValTAC</sup>, mt-tRNA<sup>LeuTAA</sup>, mt-tRNA<sup>ProTGG</sup> and mt-tRNA<sup>GluTTC</sup>) have the highest out-degrees and are associated with all three groups of ribosomal proteins, including those forming the cytosolic LSU and SSU subunits. We also note that which tRFs are correlated with ribosomal proteins depends on the considered cancer type (Supplemental Fig. S8B).

We methodically pursued this further by seeking “Biological Process” (BP) Gene Ontology (GO) terms that are enriched in > 10 distinct cancer types. To account for the overlap in the included genes, we grouped them into clusters and identified the represented processes (Supplemental Fig. S8C and S9A-D). The BP GO terms formed four basic groups with a multitude of correlations involving nuclear and MT tRFs (Fig. 4B).

The largest group (“red”) of GO terms pertains to development, cell adhesion and signaling. Heart, blood vessel, and central nervous system development are included, as well as cell-matrix adhesion, and receptor tyrosine kinase (RTK) signaling. Genes that are exclusive to this cluster include *IGF2R*, *TGFBR2*, *ELK3*, *LRP1*, *ZEB2* and *EDF1*: all have positive and negative correlations with nuclear and MT tRFs in 28 of the 32 cancer types. The second largest group (“blue”) pertains to DNA and RNA metabolism, trafficking across compartments, and cell division. Notable genes in this category include *TOP3B*, *RECQL*,

*VDAC2*, *TRAM2*, *XPOT*, *IPO11*, *PQLC2*, *BOB1* and *RAD1*. The third group (“magenta”) pertains to “oxidative phosphorylation and ATP synthesis.” Lastly, the “green” group pertains to genes linked to proteasome degradation, protein ubiquitination, antigen presentation, and NF- $\kappa$ B signaling.

We stress that despite the presence of the same pathways in different cancer types, the tRF-mRNA correlations that fuel these findings involve *different* tRFs and *different* mRNAs in each cancer type. We demonstrate this with two examples, the proteasome and glycolysis pathways.

- For the proteasome, we analyzed the correlations from diffuse large B-cell lymphoma (DLBC) and KIRC. Fig. 4C shows the results in each case, with the proteasome genes in exactly the same placement to facilitate comparisons. The shown edges indicate correlations between tRFs and the respective mRNAs. Note how the identities of the correlated partners differ in the two cancers. For example, in DLBC, the expression of *PSMD5* and *PSMD9* exhibit the most correlations with tRFs whereas in KIRC it is *PSMB3* and *PSME2*. In DLBC, the nuclear tRNA<sup>AlaTGC</sup> and tRNA<sup>LeuCAG</sup> isoacceptors have the most links with mRNAs whereas in KIRC, it is MT tRNA<sup>ValTAC</sup>.
- For the glycolysis, we analyzed the correlations from ESCA, THCA, uterine corpus endometrial carcinoma (UCEC), and UVM. Specifically, we dissected the correlations of tRFs with metabolism-related genes. Fig. 4D presents a reconstructed network of the genes from this pathway. The genes are connected based on their known interactions with common metabolites as substrates/products of consecutive reactions in the pathway (see Methods). In ESCA, *TPII* and *GAPDH* attract the attention of positively-correlated tRFs from several nuclear and MT isoacceptors. In THCA, it is *GPI* and *PFKM* that have the most associations (positive and negative) with tRFs. UCEC and UVM exhibit yet different correlations patterns involving genes from this pathway.

These examples highlight the existence of strong associations between tRFs and core cellular processes (ribosome, proteasome, glycolysis, etc.). Also, the findings indicate that the exact details of these associations and underlying *putative* regulatory links depend strongly on cancer/tissue type.

### **The genomic spans of genes whose mRNAs are correlated with tRFs contain specific repeat elements**

In earlier work, we showed that the distribution of repeat elements in introns and exons is not random and that it captures functional conservation in the absence of sequence conservation (35,36). More recent efforts linked tRFs from tRNA<sup>GlyGCC</sup> to mRNAs in the mouse, through the MERVL family (6), and showed that tRFs can interfere with the reverse transcription or coding ability of two active mouse ERV families (37). We thus posited a link between tRFs and repeat elements at large in human cancers and investigated this possibility by mining the tRF-mRNA correlations at hand. Separately for each cancer type, we examined possible enrichment/depletion of transcripts in repeat elements distinguishing between sense and antisense orientations. We analyzed separately the intronic and exonic

portions of the genes (see Methods). We note here that, as our analyses showed, the abundance of mRNAs is not correlated with repeat element density in any of the cancer types (Supplemental Table S9).

We find multiple repeat families to be specifically enriched or depleted in genes whose mRNAs participate in correlations with tRFs (Supplemental Table S9). Fig. 5 shows the most frequently occurring ones. A striking link between the sign of the tRF-mRNA correlations and the repeat content of the corresponding mRNAs is evident: specifically, the introns and exons of genes whose mRNAs are positively correlated with tRFs are significantly enriched in several types of repeats. For mRNAs that are negatively correlated with tRFs, their genes are significantly depleted in these repeats. ALU and MIR retrotransposons are most significantly enriched or depleted across multiple cancer types. On average, exons show less pronounced enrichment/depletion scores compared to introns.

MT tRFs are correlated more strongly with repeats than nuclear tRFs. Generally, MT tRFs are *positively* (*negatively*, respectively) correlated with mRNAs whose introns have *high* (*low*, respectively) density in these repeats. This is true for exons as well. Unlike MT tRFs, nuclear tRFs show less pronounced associations with repeats through their correlated mRNAs (Fig. 5A-B). However, see the links between DLBC, KIRP, COAD, READ, head and neck squamous cell carcinoma (HNSC), pancreatic adenocarcinoma (PAAD), PRAD, and BRCA with ALUs and MIRs.

LINE elements warrant special mention. The introns of genes whose mRNAs are positively correlated with MT tRFs are frequently *depleted* in antisense L1s and consistently *enriched* in both sense and antisense L2s. For genes whose mRNAs are correlated (positively or negatively) with nuclear tRFs, their introns are only enriched in antisense L2s.

These results suggest that the links between tRFs and repeat elements are far-ranging and extend well beyond the recently-reported links with ERVs. Notably, the findings suggest links between MT tRFs and extra-mitochondrial transcripts and have direct implications for the roles of repeats in cancer biology.

### **The density of repeat element in human genes is indicative of the biological process**

In our results so far, we focused on elucidating the genomic properties of the mRNAs that are correlated with tRFs. In other words, the mRNAs that were linked with tRFs were singled out based on *abundance relationships* between mRNAs and tRFs, and not on *abundance magnitude* or any *genomic characteristics* selected *a priori*. As noted above, many of the identified genomic properties are not independent. It is known that gene length and repeat element density follow specific distributions based on the encoded proteins' involvement in biological processes (38), as well as the evolutionary trajectory of the genomic region (39). To link the underlying genome architecture with our results, we took an orthogonal approach and examined biological processes from a genomic perspective.

We considered all human genes, independent of expression, and ranked them based on the density of the repeat element categories of Figure 5. We then examined which pathways are over-represented in the most repeat-dense and the least repeat-dense gene sets (Supplemental

Table S10). We found a considerable number of enriched pathways with essentially no overlap. The gene set with low repeat density includes ribosomal proteins, homeobox genes, G-Protein coupled receptors, keratins, cytokines as well as FOX proteins. The corresponding enriched Gene Ontology (GO) terms include development, morphogenesis and differentiation. On the other hand, the gene set with high repeat density includes G proteins, tyrosine and serine-threonine kinases, as well as proteins with DHR1, DHR2, FERM and/or EF-hand domains. Moreover, genes with high repeat density belong to signaling pathways, including: MAPK, ErbB, Ras, PI3K-Akt, and cGMP-PKG.

These results emphasize that at the *genomic level* the architecture of genes and the placement of repeat elements is non-random. At the same time, the identified processes have considerable overlap with those shown on Fig. 4. This suggests a coupling between the transcriptomic level, where we uncovered novel interconnections between tRFs and mRNAs, and the architecture of genes at the genomic level.

### tRFs are correlated with mRNAs in a sex-dependent manner

We previously showed that tRF-mRNA correlations capture differences between White and Black/African-American patients with triple-negative breast cancer (19) and PRAD (8). We posited that tRF-mRNA correlations also capture differences between patients of different sex. We are not aware of previous work that examined the possibility of molecular links between sex disparities and tRFs. To this end, we first focused on bladder urothelial carcinoma (BLCA) for which sex disparities with regard to incidence and survival rates have been documented (40). Because BLCA also depends on a patient's race/ethnicity (41), we restricted our analysis to only samples from White patients (Supplemental Table S11).

A first, rather striking observation pertains to the number of correlations per tRNA isoacceptor in patients of different sex (Fig. 6A). The same isoacceptor is associated with markedly different numbers of mRNAs in male (X-axis) and female (Y-axis) patients. In fact, *more isoacceptors* are correlated with *more mRNAs* in females than in males: isoacceptors are labeled if they are associated with  $\geq 2x$  (or  $\leq 0.5x$ ) mRNAs in one sex (Fig. 6A): this difference is statistically very significant (Supplemental Fig. S10A). We also analyzed tRF-mRNA correlations in BLCA focusing on those that are either present in only one sex or change sign between sexes (Supplemental Table S12; Supplemental Fig. S10B). We find that 36% of the tRF-mRNA correlations found in female patients are absent from the male patients; and, 19% of the tRF-mRNA correlations found in male patients are absent from the female patients. Fig. 6B highlights BLCA's sex-dependent differences with the help of cyclin-dependent kinases (CDK) or proteins interacting with CDKs (Supplemental Table S11).

We repeated the same analysis for LUAD (42) as well as for KIRC (43), specifically for subtype ccA (25). In LUAD, the case of *MAP4K4* stands out: its mRNA has the largest number of differential co-expression links (Supplemental Table S11). We examined the MAPK signaling pathway further and found that many of its components and direct interactors are differentially correlated with tRFs between the two sexes (Fig. 6C). Note the presence of critical gene regulatory nodes, like *PTEN*, *CREB1* and *CEPBP*. Interestingly, in

the original TCGA publication on LUAD it was pointed out that mutations could explain only part of the activation of the PI(3)K-MAPK pathway (44).

Similarly, in KIRC-ccA, we identified numerous sex-dependent tRF-mRNA correlation pairs (Supplemental Table S11), including mRNAs from the purine metabolism pathway (Supplemental Table S12). Several mRNAs that encode for enzymes in the biosynthesis of inosinic acid (IMP), the precursor of AMP and GMP, show extensive re-wiring with tRFs as a function of sex (Fig. 6D). The sex-dependent correlation differences extend beyond this pathway. Adenylate cyclase genes (*ADCY3*, *ADCY4* and *ADCY7*) and nucleoside diphosphate kinases (*NME1*, *NME3* and *NME7*) also exhibit sex-dependent differential correlation with tRFs in KIRC-ccA. We note that poor prognosis in KIRC has been associated with alterations in several metabolic pathways, including the pentose phosphate pathway (25).

These results suggest the possibility of tRFs being involved in the molecular events underlying sex disparities in multiple cancer types. They are in complete analogy to our previous reports that the tRFs are linked to race/ethnicity disparities in disease (8,19).

## Discussion

We analyzed 20,722 tRFs that we mined from 10,274 white-listed TCGA datasets representing 32 human cancer types. In concordance with our previous results (5,8), the evidence continues to support the view that tRFs represent a novel and complex layer in post-transcriptional regulation. The structural category of i-tRFs was the richest in terms of the number of distinct tRFs found across the various TCGA datasets. This category, despite having been discovered only recently (5), has been gaining independent computational as well as experimental validations (4,18,45).

First, we evaluated whether base modifications affect our ability to accurately mine tRFs from TCGA. Base modifications in the mature tRNA have been thought to be a consideration when working with datasets generated using standard RNA-seq protocols (46). Through TCGA-wide analyses, we showed that base modifications have a rather limited impact on our ability to mine tRFs from these cancer datasets (Supplemental Fig. S1A-B). This is an expected result when the RNA-seq approach involves ligating both adapters prior to the reverse transcription step, which is the method used by the TCGA. Indeed, had a modification caused the reverse transcriptase to stop, then the corresponding molecule would not have been amplified and therefore would not have appeared among the sequenced reads. We also evaluated the impact of mapping sequenced reads by allowing mismatches. We found that doing so decreases our ability to establish the tRNA provenance (or lack thereof) of the considered molecules (Supplemental Fig. S1C).

We systematically investigated the relationships between tRFs and mRNAs, motivated by earlier work by others and us showing that tRFs affect mRNA abundance by acting like miRNAs (3,5), or by decoying RBPs (18,19). Among the tRF-mRNA pairs that emerge *de novo* from our analyses are several interactions that were validated recently in the literature: i-tRFs from tRNA<sup>Gly</sup> and tRNA<sup>Tyr</sup> were shown linked to be linked to the mRNAs of

*HMGA1*, *CD151*, *CD97* and *TIMP3* through the RNA binding protein YBX1 (18). Additionally, the tRF-mRNA pairs also included several hundred correlations between tRFs and ribosomal proteins (Fig. 4A), as well as more than one thousand significant correlations with aminoacyl-tRNA synthases, including GARS, IARS and MARS: these correlations persist across many cancer types and are concordant with recent findings (1,12,13).

Among the mRNAs that are correlated with tRFs, we observed several pathways that are consistently present across multiple cancer types (Fig. 4B, and Supplemental Tables S4-S5). In addition, we identified other pathways that are unique to individual cancer types. Not surprisingly, our analyses show a tissue-specific contribution to the discovered correlation patterns. This is in agreement with our previously reported findings of context- and tissue-specific correlations between miRNAs and mRNAs during cancer development (47) as well as other related work on additional levels of biological function (48). Specifically for glycolysis (Fig. 4), while the Warburg effect is a hallmark of cancer metabolism, there is increasing evidence in support of cancer-type- and context-specific metabolic signatures (49). From this perspective, the identified links to tRFs could shed light on the molecular events behind their regulation and functions in different tissues and cancer types.

MT tRFs featured prominently in all of our findings. Studies of the tRFs' sub-cellular distributions are in their early stages (3). From this standpoint, the correlation of MT tRFs with processes that are not MT-specific suggests possible biogenesis from either MT tRNAs that exit from the mitochondrion (50) or from the transcripts of the MT "tRNA-lookalikes" we reported in nuclear chromosomes (33). It is worth recalling that the mitochondria have manifested roles in cancer, at multiple levels between the genetic (48) and the metabolic (51), and that mitochondrial processes can signal and affect the nuclear genome's state (52). Thus, in light of our results, we postulate that tRFs act as mediators of an information exchange between the MT and the nucleus. For example, the localization patterns of Fig. 3B suggest possible roles of tRFs as regulators marshaling the communication exchange between different cell compartments.

Sex disparities are attracting increasing scientific interest. We examined three cancer types, BLCA, LUAD and KIRC, with documented disparities in TCGA (53) using a differential co-expression approach that provides deeper insights than differential expression (19,54). In all cases, we identified pathways that are integral components of the molecular biology of each cancer type as well as responsive to the hormone environment. Sex hormones are arguably important contributors to sex disparities in the disease context. In fact, sex hormones have been shown to regulate cell proliferation in the context of BLCA (55), signaling cascades, including the MAPK pathway, in lung cancer (56) as well as the pentose phosphate pathway (57). In addition, some tRFs have sex-hormone-dependencies (58). We posit that the networks of Fig. 6 depict components of the mechanistic contributions to sex disparities in cancer by tRFs.

We also identified links between tRFs and genomic features. Two striking results are the bias in the length of the mRNAs that participate in correlations with tRFs (Fig. 3A), and the enrichment of their genomic span in specific classes of repeats (Fig. 5). We note that the repeat class of 'tRNA' was not found enriched among the exons that comprise the mRNAs



exhibiting correlations with tRFs (Supplemental Table S9). This indicates that the observed tRF-mRNA correlations are not due to an enrichment of exons in tRFs.

In our earlier work, we described the non-random placement of repeat elements within the exons of genes (35), and across the genome (36), as well as outlined conspicuous interconnections with short RNAs (piRNAs) produced by exonic regions (59). In parallel studies of genome-wide methylation, we showed that repeat elements become demethylated as stem cell differentiation progresses (60). These previous findings collectively suggested potentially important roles for repeat elements.

More recent work highlighted the potential for regulation of repeat elements by tRFs (6,37,61). Consequently, the roles of repeat elements in gene expression regulation have been attracting attention (36,62). In the context of cancer, repeat elements continue to emerge as important components in genomic rearrangements (63) as well as potent regulators of gene expression (64), and as determinants of overall survival (65). As repeat elements are also associated with chimeric transcripts involving protein-coding exons (66), it is conceivable that tRFs can interact with such junctions just like we previously reported for miRNAs (67). Our results also suggest that, in addition to interacting with repeat elements to protect the genome's integrity (61), tRFs are also involved in complex gene regulation. Indeed, short gene length has been associated with highly expressed genes and a proliferative phenotype (39). The state of proliferation is also molecularly unique at additional levels of cellular function. The metabolism of proliferative cells has the characteristic signature of the Warburg effect, whereas mature tRNA abundance profiles are distinct in proliferative cells as compared to differentiated ones (68). In addition, based on our genome-level analysis (Supplemental Table S10), pathways promoting proliferation include genes dense in repeats whereas differentiation include genes with no, or little, repeat content. These results have been discussed in the literature (36). Our analyses place the tRFs at crucial junctions of the multi-dimensional and complex process of cell proliferation.

In conclusion, our analyses reveal a dazzling array of complex relationships between tRFs and protein-coding mRNAs. These associations suggest the existence of numerous molecular interactions that await discovery and characterization. The presence of multiple families of repeats in the introns (and exons) of mRNAs with which the tRFs are correlated adds yet another level of complexity to these complex relationships.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are indebted to the National Institutes of Health for making the TCGA data publicly available. We thank Dr. Megumi Shigematsu, Dr. Takuya Kawamura, and the other members of the Center for discussions and input on this manuscript. The work was supported partially by a William M. Keck Foundation grant (IR), by NIH/NCI R21-CA195204 (IR), and by Institutional Funds.

## REFERENCES

1. Keam SP, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life (Basel)* 2015;5:1638–51 [PubMed: 26703738]
2. Shigematsu M, Kirino Y. tRNA-Derived Short Non-coding RNA as Interacting Partners of Argonaute Proteins. *Gene Regul Syst Bio* 2015;9:27–33
3. Kumar P, Anaya J, Mudunuri SB, Dutta A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol* 2014;12:78 [PubMed: 25270025]
4. 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:D152–D9 [PubMed: 29186503]
5. 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 [PubMed: 26325506]
6. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 2016;351:391–6 [PubMed: 26721685]
7. Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* 2014;17:667–9 [PubMed: 24728267]
8. Magee RG, Telonis AG, Loher P, Londin E, Rigoutsos I. Profiles of miRNA Isoforms and tRNA Fragments in Prostate Cancer. *Sci Rep* 2018;8:5314 [PubMed: 29593348]
9. Selitsky SR, Baran-Gale J, Honda M, Yamane D, Masaki T, Fannin EE, et al. Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C. *Sci Rep* 2015;5:7675 [PubMed: 25567797]
10. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 2016;351:397–400 [PubMed: 26721680]
11. Gebetsberger J, Wyss L, Mleczko AM, Reuther J, Polacek N. A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol* 2017;14:1364–73 [PubMed: 27892771]
12. Keam SP, Sobala A, Ten Have S, Hutvagner G. tRNA-Derived RNA Fragments Associate with Human Multisynthetase Complex (MSC) and Modulate Ribosomal Protein Translation. *J Proteome Res* 2017;16:413–20 [PubMed: 27936807]
13. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* 2011;43:613–23 [PubMed: 21855800]
14. Honda S, Morichika K, Kirino Y. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat Protoc* 2016;11:476–89 [PubMed: 26866791]
15. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, et al. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol* 2011;8:158–77 [PubMed: 21282978]
16. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U S A* 2013;110:1404–9 [PubMed: 23297232]
17. Kuscuc C, Kumar P, Kiran M, Su Z, Malik A, Dutta A. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* 2018;24:1093–105 [PubMed: 29844106]
18. 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 [PubMed: 25957686]
19. Telonis AG, Rigoutsos I. Race disparities in the contribution of miRNA isoforms and tRNA-derived fragments to triple-negative breast cancer. *Cancer Res* 2017

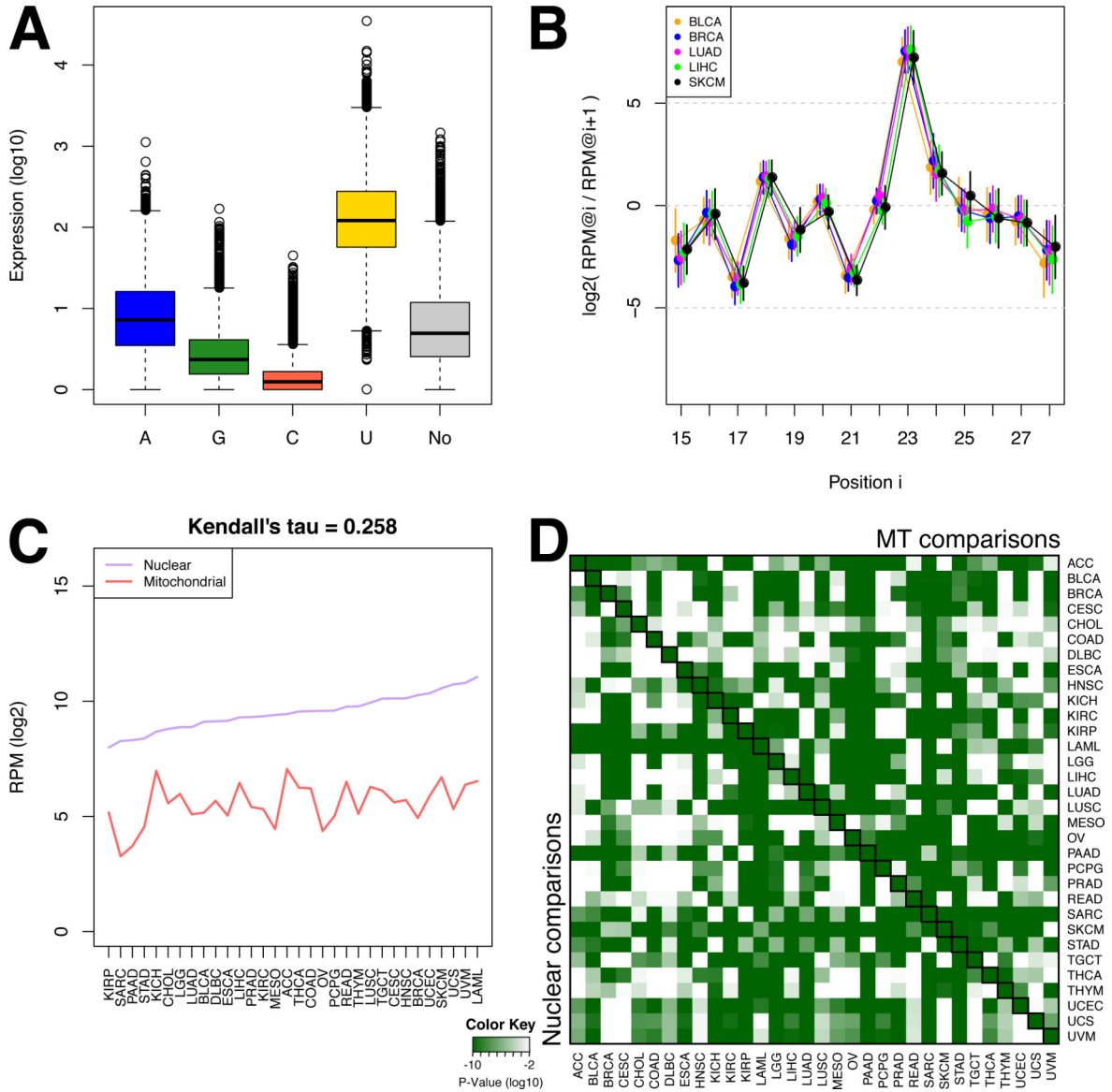
20. Loher P, Telonis AG, Rigoutsos I. MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Sci Rep* 2017;7:41184 [PubMed: 28220888]
21. Loher P, Telonis AG, Rigoutsos I. Accurate Profiling and Quantification of tRNA Fragments from RNA-Seq Data: A Vade Mecum for MINTmap. *Methods Mol Biol* 2018;1680:237–55 [PubMed: 29030853]
22. 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 [PubMed: 28203700]
23. Pliatsika V, Loher P, Telonis AG, Rigoutsos I. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics* 2016;32:2481–9 [PubMed: 27153631]
24. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* 2018;46:D303–D7 [PubMed: 29106616]
25. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013;499:43–9 [PubMed: 23792563]
26. 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 [PubMed: 29023571]
27. 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 [PubMed: 19131956]
28. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011;6:e21800 [PubMed: 21789182]
29. The UniProt C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;45:D158–D69 [PubMed: 27899622]
30. Chu A, Robertson G, Brooks D, Mungall AJ, Birol I, Coope R, et al. Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res* 2016;44:e3 [PubMed: 26271990]
31. Shigematsu M, Honda S, Loher P, Telonis AG, Rigoutsos I, Kirino Y. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* 2017;45:e70 [PubMed: 28108659]
32. Honda S, Kawamura T, Loher P, Morichika K, Rigoutsos I, Kirino Y. The biogenesis pathway of tRNA-derived piRNAs in *Bombyx* germ cells. *Nucleic Acids Res* 2017;45:9108–20 [PubMed: 28645172]
33. Telonis AG, Kirino Y, Rigoutsos I. Mitochondrial tRNA-lookalikes in nuclear chromosomes: could they be functional? *RNA Biol* 2015;12:375–80 [PubMed: 25849196]
34. Reznik E, Miller ML, Senbabaoglu Y, Riaz N, Sarungbam J, Tickoo SK, et al. Mitochondrial DNA copy number variation across human cancers. *Elife* 2016;5
35. Rigoutsos I, Huynh T, Miranda K, Tsirigos A, McHardy A, Platt D. Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. *Proc Natl Acad Sci U S A* 2006;103:6605–10 [PubMed: 16636294]
36. Tsirigos A, Rigoutsos I. Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput Biol* 2009;5:e1000610 [PubMed: 20019790]
37. Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* 2017;170:61–71 e11 [PubMed: 28666125]
38. Deininger P. Alu elements: know the SINES. *Genome Biol* 2011;12:236 [PubMed: 22204421]
39. Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. *Bioessays* 2015;37:148–54 [PubMed: 25400101]
40. Burge F, Kockelbergh R. Closing the Gender Gap: Can We Improve Bladder Cancer Survival in Women? - A Systematic Review of Diagnosis, Treatment and Outcomes. *Urol Int* 2016;97:373–9 [PubMed: 27595416]

41. Hollenbeck BK, Dunn RL, Ye Z, Hollingsworth JM, Lee CT, Birkmeyer JD. Racial differences in treatment and outcomes among patients with early stage bladder cancer. *Cancer* 2010;116:50–6 [PubMed: 19877112]
42. Townsend EA, Miller VM, Prakash YS. Sex differences and sex steroids in lung health and disease. *Endocr Rev* 2012;33:1–47 [PubMed: 22240244]
43. Huang Q, Sun Y, Ma X, Gao Y, Li X, Niu Y, et al. Androgen receptor increases hematogenous metastasis yet decreases lymphatic metastasis of renal cell carcinoma. *Nat Commun* 2017;8:918 [PubMed: 29030639]
44. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543–50 [PubMed: 25079552]
45. Jackowiak P, Hojka-Osinska A, Philips A, Zmienko A, Budzko L, Maillard P, et al. Small RNA fragments derived from multiple RNA classes - the missing element of multi-omics characteristics of the hepatitis C virus cell culture model. *BMC Genomics* 2017;18:502 [PubMed: 28666407]
46. Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* 2015;12:879–84 [PubMed: 26237225]
47. 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 [PubMed: 28206648]
48. Feeley KP, Bray AW, Westbrook DG, Johnson LW, Kesterson RA, Ballinger SW, et al. Mitochondrial Genetics Regulate Breast Cancer Tumorigenicity and Metastatic Potential. *Cancer Res* 2015;75:4429–36 [PubMed: 26471915]
49. Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, et al. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol* 2013;31:522–9 [PubMed: 23604282]
50. Maniataki E, Mourelatos Z. Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. *RNA* 2005;11:849–52 [PubMed: 15872185]
51. Wallace DC. Mitochondria and cancer. *Nat Rev Cancer* 2012;12:685–98 [PubMed: 23001348]
52. Lozoya OA, Martinez-Reyes I, Wang T, Grenet D, Bushel P, Li J, et al. Mitochondrial nicotinamide adenine dinucleotide reduced (NADH) oxidation links the tricarboxylic acid (TCA) cycle with methionine metabolism and nuclear DNA methylation. *PLoS Biol* 2018;16:e2005707 [PubMed: 29668680]
53. Yuan Y, Liu L, Chen H, Wang Y, Xu Y, Mao H, et al. Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell* 2016;29:711–22 [PubMed: 27165743]
54. Lopes-Ramos CM, Kuijjer ML, Ogino S, Fuchs CS, DeMeo DL, Glass K, et al. Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism. *Cancer Res* 2018;78:5538–47 [PubMed: 30275053]
55. Bahk JY, Kim MO, Park MS, Lee HY, Lee JH, Chung BC, et al. Gonadotropin-releasing hormone (GnRH) and GnRH receptor in bladder cancer epithelia and GnRH effect on bladder cancer cell proliferation. *Urol Int* 2008;80:431–8 [PubMed: 18587256]
56. Chakraborty S, Ganti AK, Marr A, Batra SK. Lung cancer in women: role of estrogens. *Expert Rev Respir Med* 2010;4:509–18 [PubMed: 20658912]
57. Jiang P, Du W, Wu M. Regulation of the pentose phosphate pathway in cancer. *Protein Cell* 2014;5:592–602 [PubMed: 25015087]
58. 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:E3816–25 [PubMed: 26124144]
59. Rigoutsos I Short RNAs: how big is this iceberg? *Curr Biol* 2010;20:R110–3 [PubMed: 20144771]
60. Laurent L, Wong E, Li G, Huynh T, Tsigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. *Genome Res* 2010;20:320–31 [PubMed: 20133333]
61. Martinez G, Choudury SG, Slotkin RK. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res* 2017;45:5142–52 [PubMed: 28335016]

62. Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know 'bout retroelements in cancer. *Semin Cancer Biol* 2010;20:200–10 [PubMed: 20600922]
63. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res* 2014;24:1053–63 [PubMed: 24823667]
64. Cruickshanks HA, Vafadar-Isfahani N, Dunican DS, Lee A, Sproul D, Lund JN, et al. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. *Nucleic Acids Res* 2013;41:6857–69 [PubMed: 23703216]
65. 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 [PubMed: 28535802]
66. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 2016;7:24 [PubMed: 27980689]
67. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 2008;455:1124–8 [PubMed: 18806776]
68. Gingold H, Tehler D, Christoffersen NR, Nielsen MM, Asmar F, Kooistra SM, et al. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 2014;158:1281–92 [PubMed: 25215487]

**Significance**

Across 32 TCGA cancer contexts, nuclear and mitochondrial tRNA fragments exhibit associations with mRNAs that belong to concrete pathways, encode proteins with particular destinations, have a biased repeat content, and are sex-dependent.



**Figure 1. The noteworthy case of tRNA<sup>HisGTG</sup>.**

(A) Barplot showing the relative expression of the 5'-tRFs of tRNA<sup>HisGTG</sup> grouped based on their starting nucleotide at the -1 position (see text). "No" corresponds to 5'-tRFs that begin at position +1 and have no post-transcriptional additions. (B) Ratios of abundances between His(-1U) 5'-tRFs that end at positions *i* and *i+1* respectively of tRNA<sup>HisGTG</sup>, for primary tumors from selected cancer types. The X axis represents ending positions *i* within the mature tRNA<sup>HisGTG</sup>. Vertical bars represent standard deviation. The ratios of abundances for all 32 cancer types as well as normal tissues are included in Supplemental Table S2 and Supplemental Fig. S2. (C) The median abundance of the nuclear and the MT tRNA<sup>HisGTG</sup> genes. The abundance is calculated as the sum of the abundances of the tRFs each tRNA produces. This is a simplified version of the bar-plots of Supplemental Fig. S1E-F. Cancer types are sorted based on the abundance of the nuclear tRNA. (D) Heatmap representing the P values (Mann-Whitney U-test) when comparing the abundances of the nuclear- and MT

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

tRNA<sup>HisGTG</sup>-derived fragments within the same cancer type (the diagonal of the matrix), or when comparing the abundance of the MT (upper triangle) or the nuclear (bottom triangle) tRNA<sup>HisGTG</sup> among cancer types. P values are log<sub>10</sub>-scaled.

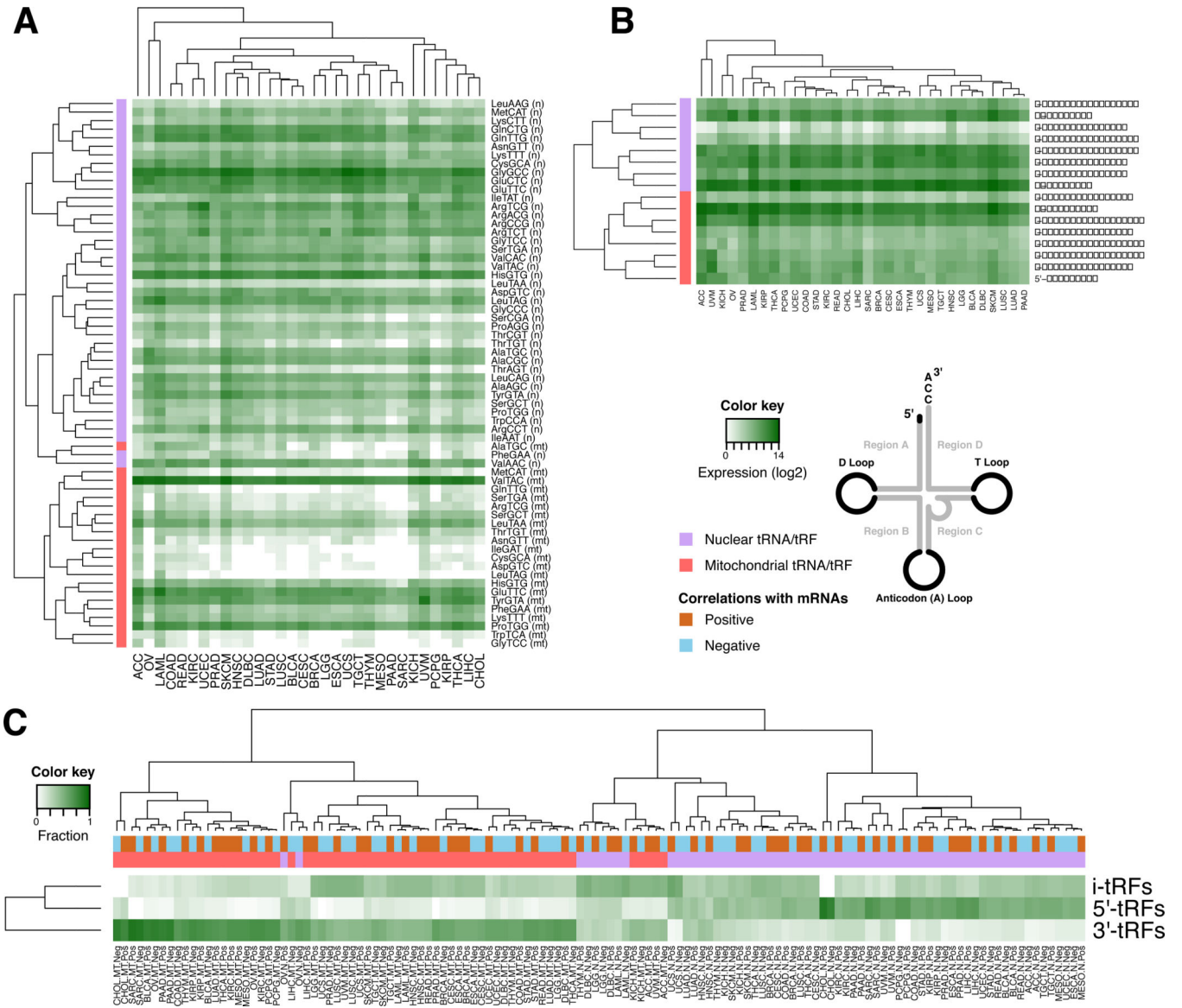
Author Manuscript

Author Manuscript

Author Manuscript

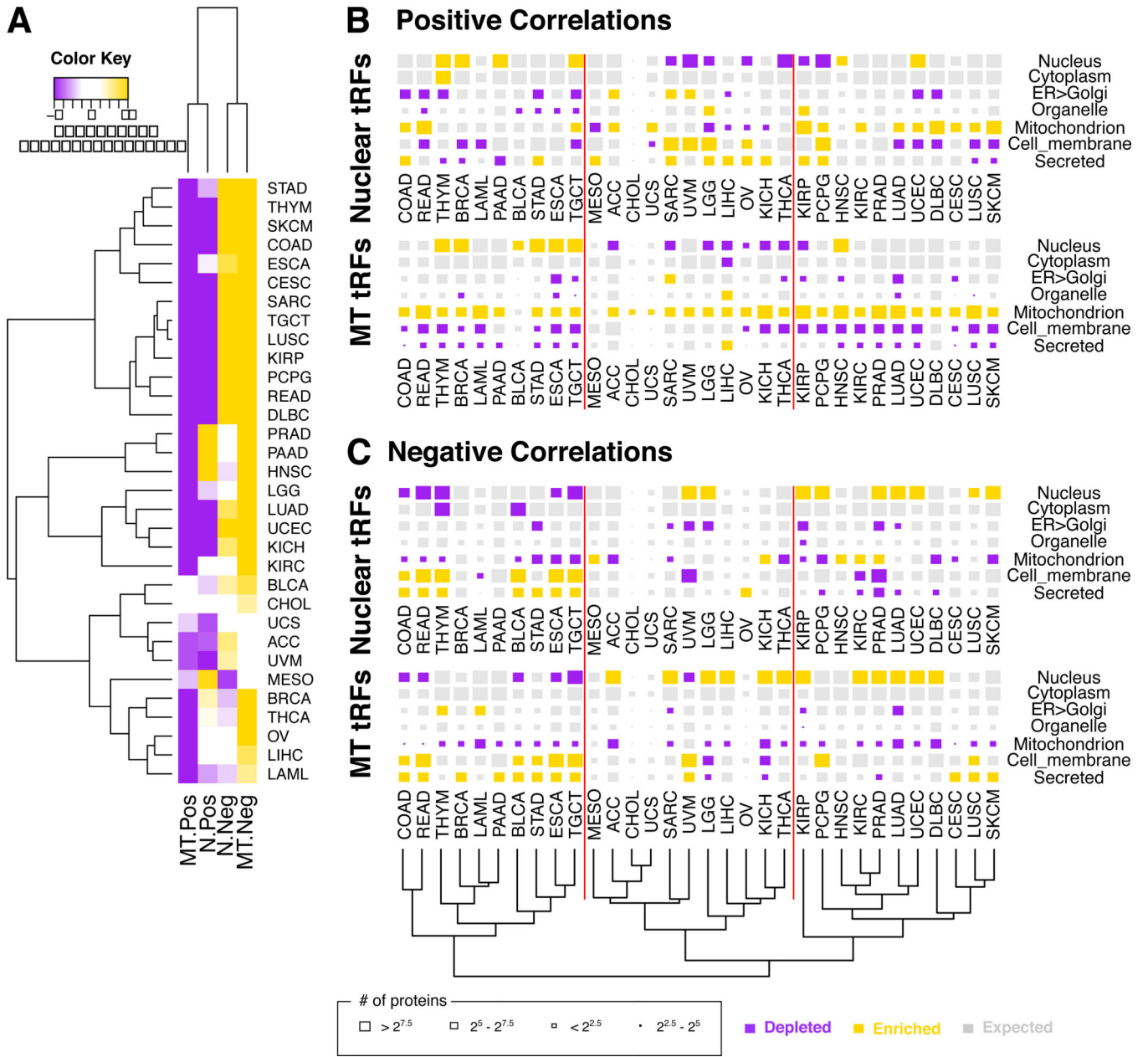
Author Manuscript





**Figure 2. Distinct characteristics between nuclear and mitochondrial tRFs.**

(A) Heatmap showing the mean isoacceptor abundance (see Methods). Hierarchical clustering (metric: Kendall’s tau distance) groups mitochondrial (MT) and nuclear (N) isoacceptors into separate clusters. (B) Heatmaps and hierarchical clustering (metric: Kendall’s tau distance) of the mean abundance of each structural category per genome (nuclear or MT). The i-tRFs are split into sub-categories based on their the location of the 5’ terminus. Note the separation of nuclear and MT tRFs. (C) Heatmap and hierarchical clustering (metric: Euclidean distance) of the distribution of tRFs participating in correlations with mRNAs, for three structural categories. The values represent number of tRFs normalized to the number of i-tRFs 5’-tRFs and 3’-tRFs.

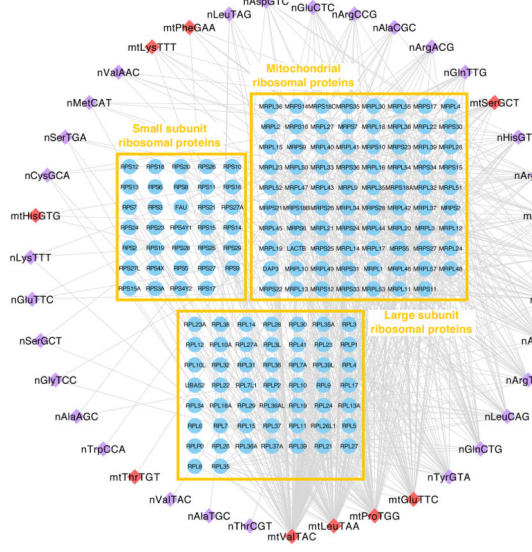


**Figure 3. tRFs are preferentially positively correlated with shorter mRNAs and context-specific cellular destinations of the encoded proteins.**

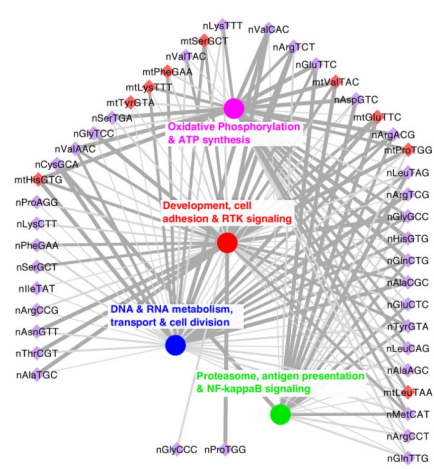
(A) Heatmap and hierarchical clustering (metric: Euclidean distance) on the Z-scores of the mean length of a gene’s mRNA-space (i.e. the union of the exons) for mRNAs participating in tRF-mRNA correlations, compared to the observed length distribution of the transcribed mRNAs. Purple color indicates statistically significant depletion whereas gold means statistically significant enrichment. (B-C) The localization of the protein products whose mRNAs are statistically-significantly correlated either positively (B), or negatively (C), with nuclear (top row in each group) and mitochondrial (MT) (bottom row in each group) tRFs. The size of the shown rectangles corresponds to the number of protein products that localize in the shown compartment. The color of the block represents enrichment (gold) or depletion

(purple) compared to the expected distribution ( $P < 0.001$ ;  $\chi^2$  test). The shown dendrogram results from the hierarchical clustering (metric: Euclidean distance) of cancer types on the residual scores, as computed by the  $\chi^2$  test, of all panels. The vertical red lines separate the three main cancer groupings as defined by the dendrogram and serve as visual reference points within the figure.

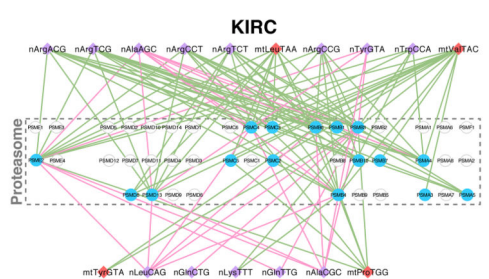
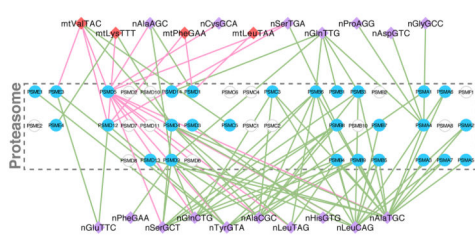
**A Ribosomal proteins**



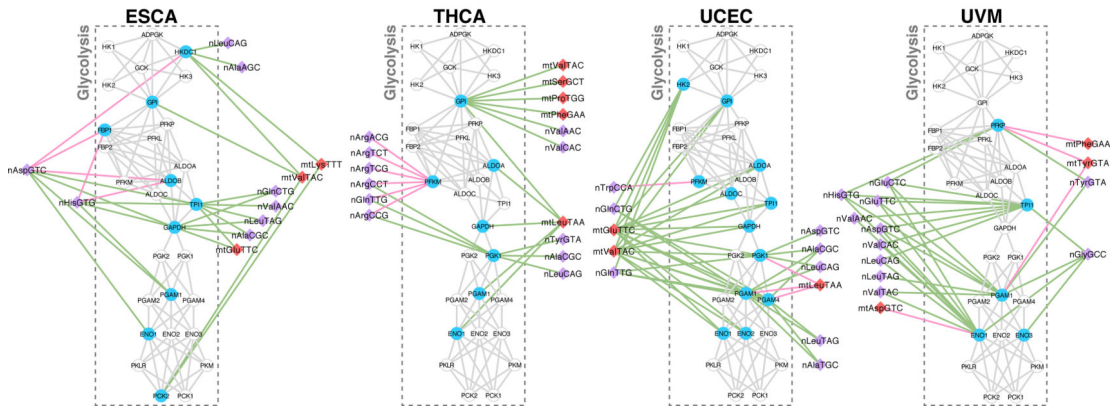
**B Common pathways**



**C Proteasome**



**D Glycolysis**



**Figure 4. tRFs correlate with universal processes in a context-dependent manner.**

(A) Ribosomal proteins as an example of a core pathway comprising genes whose mRNAs are correlated with tRFs in at least three different cancer types. The mRNAs are grouped based on the complexes in which the encoded proteins participate. (B) Network of tRFs and groups of enriched biological processes are linked if they appear in at least 10 cancer types. The thickness and gray tone of the edge is proportional to the number of average correlations of the tRF-mRNA pairs across cancers. The GO terms and their groupings are shown in Supplemental Fig. S8-S9. (C-D) Examples of context-specific wiring of core

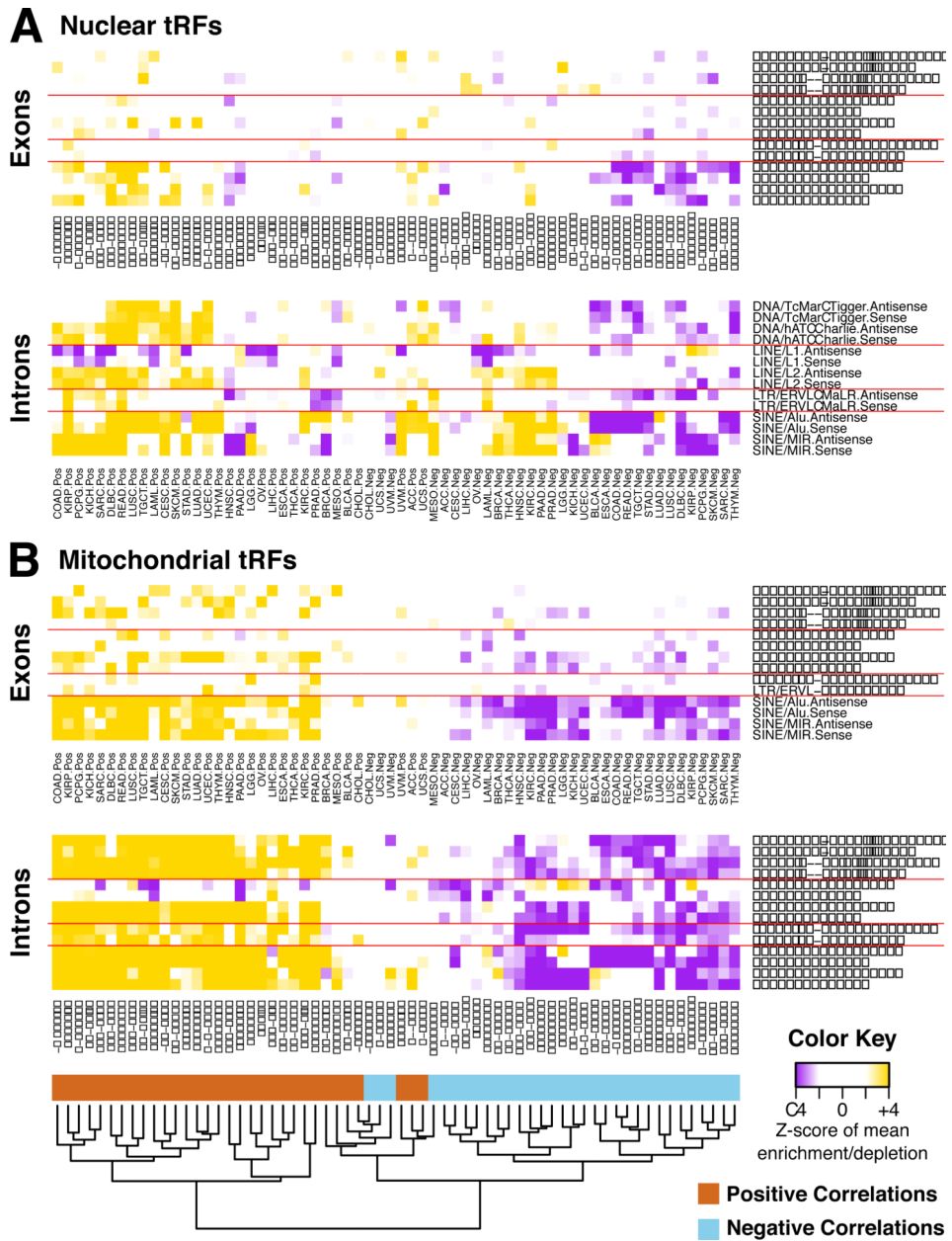
pathways with nuclear and mitochondrial tRFs. The proteasome (C) genes are grouped based on subunit identity. For the glycolysis network (D), we connected genes if the encoded enzymes catalyze consecutive reactions. Gene nodes are colored cyan if they are correlated with tRFs in that cancer, otherwise they are shown as cyclical contours. For both the proteasome (C) and the glycolysis (D) networks, the mRNAs are arranged in exactly the same manner: note how, in different cancer types, the tRFs are correlated with different mRNAs within these networks.

Author Manuscript

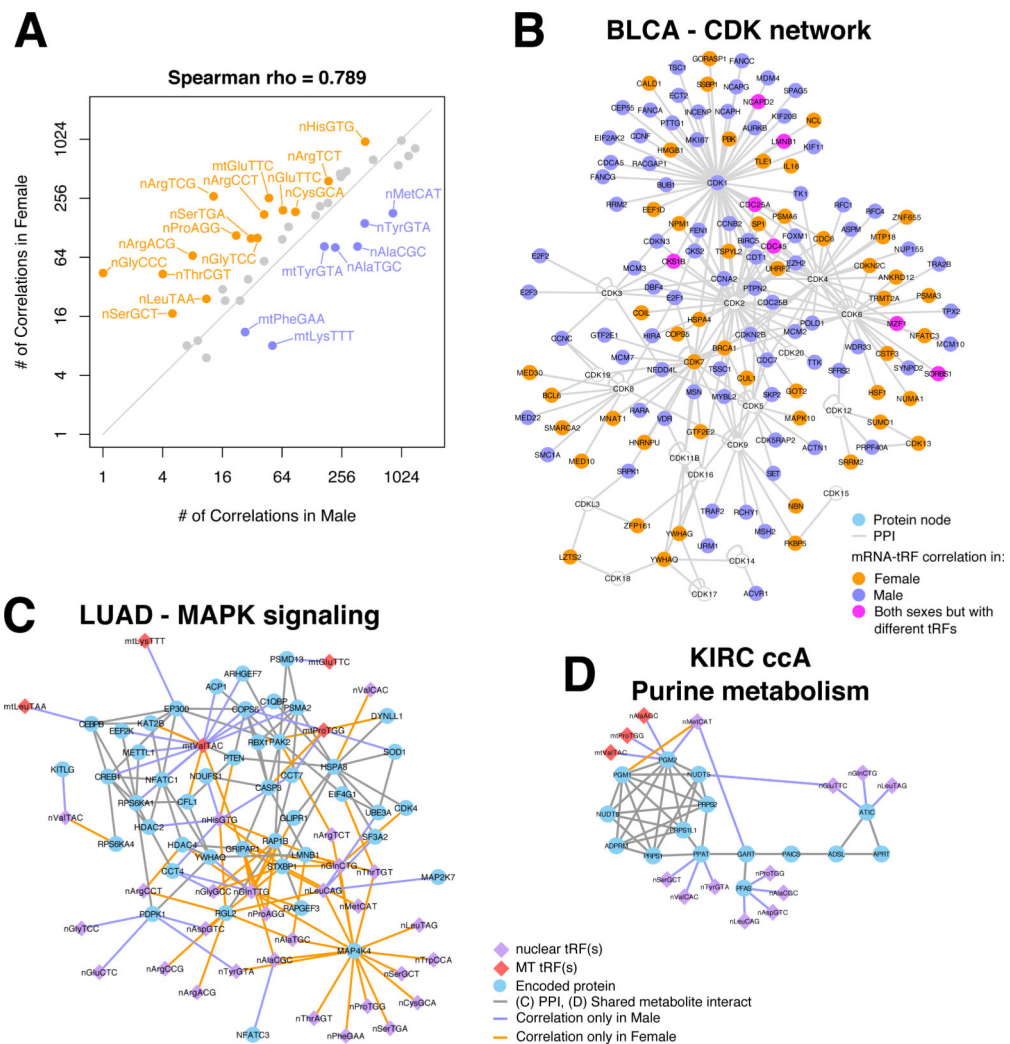
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5. tRFs are correlated with genes of specific repeat element content.** Heatmaps of the Z-scores of the mean density of each repeat element category with reference to the background density distribution of mean repeat content in genes correlated with nuclear (A) and MT (B) tRFs. The enrichments/depletions were calculated separately for the exons (top panel) and the introns (bottom panel) of the genes whose mRNAs are correlated with the tRFs. The repeat categories (rows) are ordered in the same way for all four panels. The shown dendrogram at the bottom of the figure results from the hierarchical clustering (metric: Manhattan distance) on the matrix of the Z-scores of all shown panels. Details about the overlap of repeat families with the genes whose mRNAs are correlated with tRFs can be found at Supplemental Table S8 for each of the cancer types.



**Figure 6. Sex disparities in the correlation of tRFs with mRNAs in bladder, lung and kidney cancers.**

(A) Plot showing the number of correlations that the tRFs from different isoacceptors have with mRNAs in each sex in primary tumors of BLCA. Isoacceptors are colored and labeled if the tRFs that originate in them participate in correlations with mRNAs that are *at least* twice as many in one of the two sexes compared to the other. Note that this is a  $\log_2$ - $\log_2$  plot. (B) Protein-Protein interaction network of CDKs and the proteins that interact with CDKs. Nodes are colored based on the mRNAs' sex-specific correlation patterns in BLCA. Specifically, nodes are colored green if the mRNA is correlated with tRFs exclusively in male subjects and orange if it is respectively found in female subjects only. If the mRNA is differentially co-expressed with different tRFs in each sex, then the node is colored magenta. CDKs that are not differentially co-expressed with tRFs are colored cyan. (C) Protein-protein interaction network of the MAPK signaling network and the proteins that interact with them. The nodes are connected to isoacceptors if the corresponding mRNAs are correlated with the corresponding tRFs in only one sex. (D) Metabolic network of IMP biosynthesis. The nodes are connected if the encoded proteins catalyze consecutive reactions in purine metabolism. The nodes are connected to isoacceptors if the corresponding mRNAs

are correlated with the corresponding tRFs in only one sex. All analyzed samples in this Figure correspond to donors of one race/ethnicity (White).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript