



# Screening and Comprehensive Analysis of Cancer-Associated tRNA-Derived Fragments

Yiran Zhou<sup>1,2</sup>, Qinghua Cui<sup>1,2</sup> and Yuan Zhou<sup>1,2\*</sup>

<sup>1</sup>MOE Key Lab of Cardiovascular Sciences, Department of Biomedical Informatics, Center for Noncoding RNA Medicine, School of Basic Medical Sciences, Peking University, Beijing, China, <sup>2</sup>MOE Key Lab of Cardiovascular Sciences, Department of Physiology and Pathophysiology, Center for Noncoding RNA Medicine, School of Basic Medical Sciences, Peking University, Beijing, China

## OPEN ACCESS

### Edited by:

Shaveta Kanoria,  
Wadsworth Center, United States

### Reviewed by:

Le Ou-Yang,  
Shenzhen University, China  
Isidore Rigoutsos,  
Thomas Jefferson University,  
United States

### \*Correspondence:

Yuan Zhou  
zhouyuanbioinfo@hsc.pku.edu.cn

### Specialty section:

This article was submitted to  
RNA,  
a section of the journal  
Frontiers in Genetics

Received: 27 July 2021

Accepted: 29 December 2021

Published: 14 January 2022

### Citation:

Zhou Y, Cui Q and Zhou Y (2022)  
Screening and Comprehensive  
Analysis of Cancer-Associated tRNA-  
Derived Fragments.  
Front. Genet. 12:747931.  
doi: 10.3389/fgene.2021.747931

tRNA-derived fragments (tRFs) constitute a novel class of small non-coding RNA cleaved from tRNAs. In recent years, researches have shown the regulatory roles of a few tRFs in cancers, illuminating a new direction for tRF-centric cancer researches. Nonetheless, more specific screening of tRFs related to oncogenesis pathways, cancer progression stages and cancer prognosis is continuously demanded to reveal the landscape of the cancer-associated tRFs. In this work, by combining the clinical information recorded in The Cancer Genome Atlas (TCGA) and the tRF expression profiles curated by MINTbase v2.0, we systematically screened 1,516 cancer-associated tRFs (ca-tRFs) across seven cancer types. The ca-tRF set collectively combined the differentially expressed tRFs between cancer samples and control samples, the tRFs significantly correlated with tumor stage and the tRFs significantly correlated with patient survival. By incorporating our previous tRF-target dataset, we found the ca-tRFs tend to target cancer-associated genes and onco-pathways like ATF6-mediated unfolded protein response, angiogenesis, cell cycle process regulation, focal adhesion, PI3K-Akt signaling pathway, cellular senescence and FoxO signaling pathway across multiple cancer types. And cell composition analysis implies that the expressions of ca-tRFs are more likely to be correlated with T-cell infiltration. We also found the ca-tRF expression pattern is informative to prognosis, suggesting plausible tRF-based cancer subtypes. Together, our systematic analysis demonstrates the potentially extensive involvements of tRFs in cancers, and provides a reasonable list of cancer-associated tRFs for further investigations.

**Keywords:** tRNA-derived fragments, cancer prognosis, cancer subtype, bioinformatics analysis, cancer immunity, small non coding RNAs

## INTRODUCTION

As its name implied, tRNA-derived fragment (tRF), is a novel class of non-coding RNA (ncRNA) cleaved from mature transfer RNAs (tRNA) (Lee et al., 2009; Thompson and Parker, 2009). In early days, tRFs were widely misunderstood as tRNA degeneration byproducts. However, in recent years, extending scope of tRFs' biological functions had been uncovered, bringing tRFs back to researchers' view (Li et al., 2018). For example, tRFs can be loaded onto Argonaute (AGO) family proteins to perform microRNA-like post-transcriptional regulations on target RNAs (Li et al., 2012; Shao et al., 2017). Some tRFs are also found capable of facilitating ribosome biogenesis by interacting with Twi12

to enhance pre-rRNA processing (Couvillion et al., 2012) or accelerating the mRNA translation of ribosomal proteins (Kim et al., 2017). Besides, some other tRFs were also reported to be capable to reduce global translation efficiency (Yamasaki et al., 2009; Ivanov et al., 2011), regulate immuno-functions (Wang et al., 2006) and serve as epigenetic regulators (Chen et al., 2016).

Comparable to tRFs' functions, their dysregulation could be associated with various diseases such as nonalcoholic fatty liver disease (Zhu et al., 2020), Alzheimer's disease (Wu et al., 2021), arterial injury (Zhu et al., 2021) and especially cancers (Balatti et al., 2017; Shao et al., 2017; Falconi et al., 2019; Zhu et al., 2019; Gu et al., 2020; Zhang et al., 2021). Especially, a series of investigations have uncovered novel associations between tRFs and cancers. For instance, Falconi et al. found that a new tRF derived from the 3'-end of tRNA-Glu is significantly down-regulated in breast cancer, and finally validated its tumor repressive role (Falconi et al., 2019). Zhu et al. compared the plasma tRNA levels between liver cancer patients and healthy donors, and rationally determined four tRFs as diagnostic biomarkers (Zhu et al., 2019). However, in most of the above experimental researches on tRF-cancer associations, only few tRFs and cancer types were considered due to the limitation of low-throughput approaches or sample size. Therefore, specific computational screening of cancer-associated tRFs based on high-throughput datasets is continuously in demand to better understand the roles of tRFs in cancer, including but not limited to tRF dysregulation in various cancer types, key oncogenesis pathways targeted by tRFs and the associations between tRF expression pattern and cancer progression and prognosis.

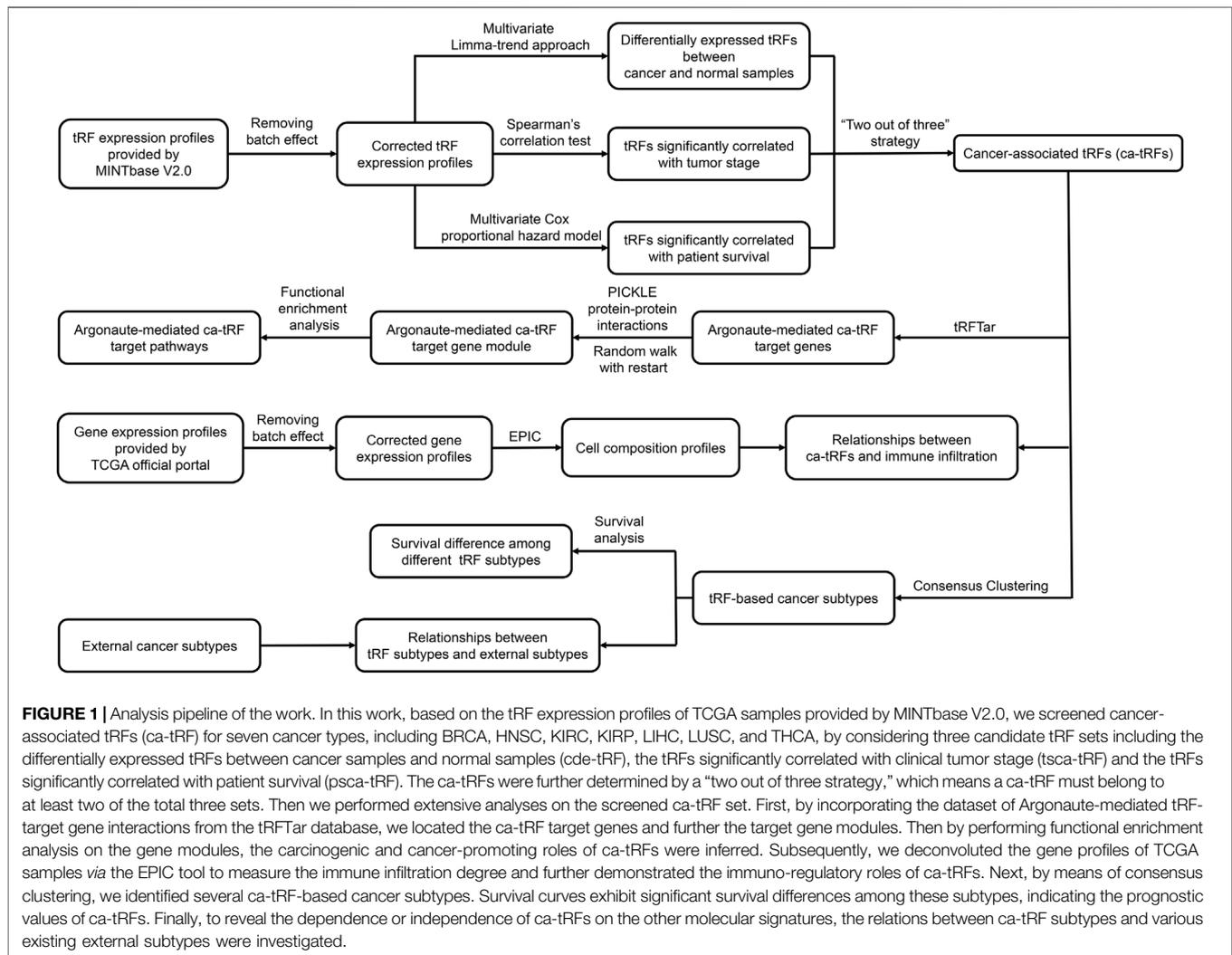
Recently, Rigoutsos lab established a comprehensive database for human tRFs termed MINTbase v2.0 (Pliatsika et al., 2018), which provides detailed annotations of 26,744 tRFs. More importantly, by re-analyzing the small RNA-sequencing library from TCGA project (Cancer Genome Atlas Research et al., 2013), MINTbase v2.0 also provides 10,814 tRF expression profiles. This dataset provides an unprecedented chance for extensive investigating the characteristics of tRFs in cancer. Indeed, based on this dataset, Rigoutsos lab has revealed a lot of biological characteristics of tRFs in various cancer types. For example, in 2015 they firstly investigated the tRF length distribution in the breast cancer dataset and revealed the tRF expression dependence on race (Telonis et al., 2015). Subsequently in 2018, The tRF profiles of prostate cancer were also uncovered (Magee et al., 2018). In the same year, they constructed a complex tRF-miRNA-mRNA co-expression network for triple-negative breast cancer and then discovered many altered mRNA-mRNA co-expression associations depending on disease state and race. More importantly, this heterogeneity could be largely explained by differential tRF-mRNA co-expression, demonstrating tRFs' important biological functions (Telonis and Rigoutsos, 2018). The investigation of tRF-mRNA co-expression network was soon extended to a pan-cancer scale (32 cancer types) (Telonis et al., 2019), where they identified more tRF-involved pathways and additionally found some pathways are regulated by tRFs in a sex-dependent manner, further highlighting the regulatory roles of tRFs. However, on the other hand, their researches tended to depict the general biological characteristics of tRFs (for example, tRF length variation, sex- and race-dependent disparity and tRF-mRNA co-

expression pattern) on the whole set of widely-expressed tRFs, but not focused on the specific screening about tRFs that are associated with oncogenesis, cancer progression and prognosis. Therefore, it is still necessary to investigate the roles of tRFs in cancer by a more specifically designed computational pipeline that combines tRF expression with available clinical data of cancer patients. In this work, based on the annotations from MINTbase v2.0 and the clinical information of TCGA samples, we firstly screened cancer-associated tRFs (ca-tRFs) across seven TCGA cancer types, including breast invasive carcinoma (BRCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung squamous cell carcinoma (LUSC) and thyroid carcinoma (THCA). Then, a series of analyses were conducted to define the functional characteristics of ca-tRFs. In the following text, we will firstly describe our integrative analysis pipeline, and then the analysis results and discussion thereof.

## MATERIALS AND METHODS

### Screening of ca-tRFs

The overall computational pipeline of this work is depicted in **Figure 1**. Datasets and key source code used in this work were also uploaded to the GitHub (<https://github.com/Load-Star/ca-tRF/>). In this work, we screened ca-tRFs by integrating the tRF expression profiles and the clinical information of the TCGA samples, which were obtained from the MINTbase V2.0 (Pliatsika et al., 2018) database and the TCGA official portal (Cancer Genome Atlas Research et al., 2013), respectively. To adapt downstream analyses, the tRF expression data was log<sub>2</sub>-transformed and processed with batch effect correction by R package *limma* (Ritchie et al., 2015), where the tissue source site and plate ID of each sample were modeled as the correction covariates. In the following step for ca-tRF screening, we considered three candidate tRF sources: 1) the differentially expressed tRFs between cancer samples and normal samples (referred as cde-tRF hereafter) identified by multivariate *limma*-trend approach (by R package *limma*), where the tissue type (cancer or normal), patient sex and race were modeled as dummy variables while patient age was modeled as numeric variable; 2) the tRFs significantly correlated with clinical tumor stage (referred as tsca-tRF hereafter) identified by Spearman's rank correlation test (by python package *Scipy*); 3) the tRFs significantly correlated with patient survival (referred as psca-tRF hereafter) detected by multivariate Cox proportional hazards model (by python package *lifelines*), where the patient sex and race were modeled as dummy variables while tRF expression abundance and patient age were modeled as numeric variables. The *p*-values were corrected by Storey's *q*-value approach. To ensure the robustness of the results, here we only considered the cancer types with sufficient samples and clinical information about tumor stage and survival, including BRCA, HNSC, KIRC, KIRP, LIHC, LUSC, and THCA. Besides, only the tRFs detectable (with an expression threshold of no less than 1 RPM) in no less than 10% samples were retained. In addition, we applied a "two out of three" strategy for further false positive control. More specifically, only if one tRF passed at least two of the total three tests (namely the above-mentioned multivariate *limma*-trend



approach, Spearman’s rank correlation test and multivariate Cox proportional hazards model), it could be listed as a ca-tRF of the corresponding cancer type. According to our observations, the amounts of tRFs overrepresented by Spearman’s correlation test and Cox regression model are much smaller than those overrepresented by limma-trend approach. Therefore, to screen enough ca-tRFs for downstream analyses, a relaxed but acceptable q-value cutoff 0.1 was used during the ca-tRF screening. The integrated list of screened ca-tRFs were shown in **Supplementary Data S1**.

## Identification and Functional Investigation of Gene Modules Targeted by ca-tRFs

Like microRNAs, some tRFs are also capable of binding AGO-family proteins and regulate target genes. Recently, we identified a considerable number of tRFs and genes from AGO-associated crosslinking-immunoprecipitation and high-throughput sequencing (AGO-CLIP) libraries and strictly screened AGO-mediated tRF-gene interactions by computationally simulating

the base pairing (more specifically, annealing processes) between tRFs and the regions in mRNAs/lncRNAs covered by AGO-CLIP. The interactions satisfying tight base pairing (normalized free energy less than  $-1.25$  kcal/mol) had been deposited into a database named tRFTar (Zhou et al., 2021). Here in order to understand the biological functions that the ca-tRFs may participate in, ca-tRF target genes were retrieved from tRFTar. To control the false positive rate, only the interactions supported by tRF-gene co-expression in TCGA samples were retained. Our hypothesis is that tRFs that 1) show tight base-pairing with potential target genes on Argonaute-covered target regions and 2) show significant expression correlation with potential target genes potentially act as regulatory targeting factors. It is also likely that tRFs themselves could be served as the targets of other regulatory molecules like miRNAs but we here only focused on tRF-gene interactions to avoid much complicated reasoning that involves tRF-miRNA-gene triplexes. Then for each cancer type, taking the ca-tRF target genes as seed genes, an extended gene module targeted by ca-tRFs was detected by the algorithm of random walk with restart (RWR) on the protein-protein

interaction (PPI) network. The PPI network was constructed from 214,666 cross-validated interactions in the PICKLE 3.0 database (Dimitrakopoulos et al., 2020). The RWR algorithm could be described as the following steps:

- (1) Construct an adjacent square matrix  $A$  to represent whether the gene  $i$  (the  $i$ -th row) and gene  $j$  (the  $j$ -th column) are directly adjacent in the network:

$$A_{i,j} = A_{j,i} = \begin{cases} 1, & \text{gene } i \in \text{Neighbor}(\text{gene } j) \\ 0, & \text{else} \end{cases} \quad (1)$$

where  $\text{Neighbor}(\text{gene } j)$  represents the gene set directly adjacent to gene  $j$  in the network, including gene  $j$  itself.

- (2) Construct a transition matrix  $T$  to represent the probabilities from gene  $i$  (the  $i$ -th row) to gene  $j$  (the  $j$ -th column) in an iteration:

$$T_{i,j} = \frac{A_{i,j}}{\sum_k A_{i,k}} \quad (2)$$

where  $n_{\text{gene}}$  represents the total number of genes in the network.

- (3) Construct a column vector  $e$  as the input to represent the initial probabilities of seed genes (ca-tRF target genes here):

$$e_i = \begin{cases} 1/n_{\text{seed}} & , \text{ gene } i \in \{\text{seed genes}\} \\ 0 & , \text{ else} \end{cases} \quad (3)$$

where  $n_{\text{seed}}$  represents the total number of input seed genes.

- (4) Perform iterations according to the following formula until the  $s$  achieves convergent:

$$s^{n+1} = \begin{cases} e & , n = 1 \\ c \cdot e + (1 - c) \cdot (T^T s^n) & , n \neq 1 \end{cases} \quad (4)$$

where  $n$  represents the iteration number and  $c$  is exactly the restart probability. If  $c = 1$ , then the  $s$  will be constantly equal to  $e$ ; if  $c = 0$ , then the RWR model will reduce to traditional random walk model. In this work, a moderate  $c = 0.5$  was selected to balance the local properties of the input and the global properties of the network.

- (5) Determine the genes with top probabilities in the convergent  $s$  for downstream analyses, here we selected the top genes whose cumulative probability reach at 0.25 as the resulting gene module.

Next, we conducted gene functional enrichment analyses (by Fisher's exact test) for the gene modules so that potential functional roles of ca-tRFs playing in cancers could be inferred. Firstly, according to a list of 711 cancer-associated genes provided by the NCG v6.0 database (Repana et al., 2019), we evaluated the enrichment degree of cancer-associated genes in the modules. Then, the R software package *clusterProfiler* (Yu et al., 2012) was used to perform enrichment

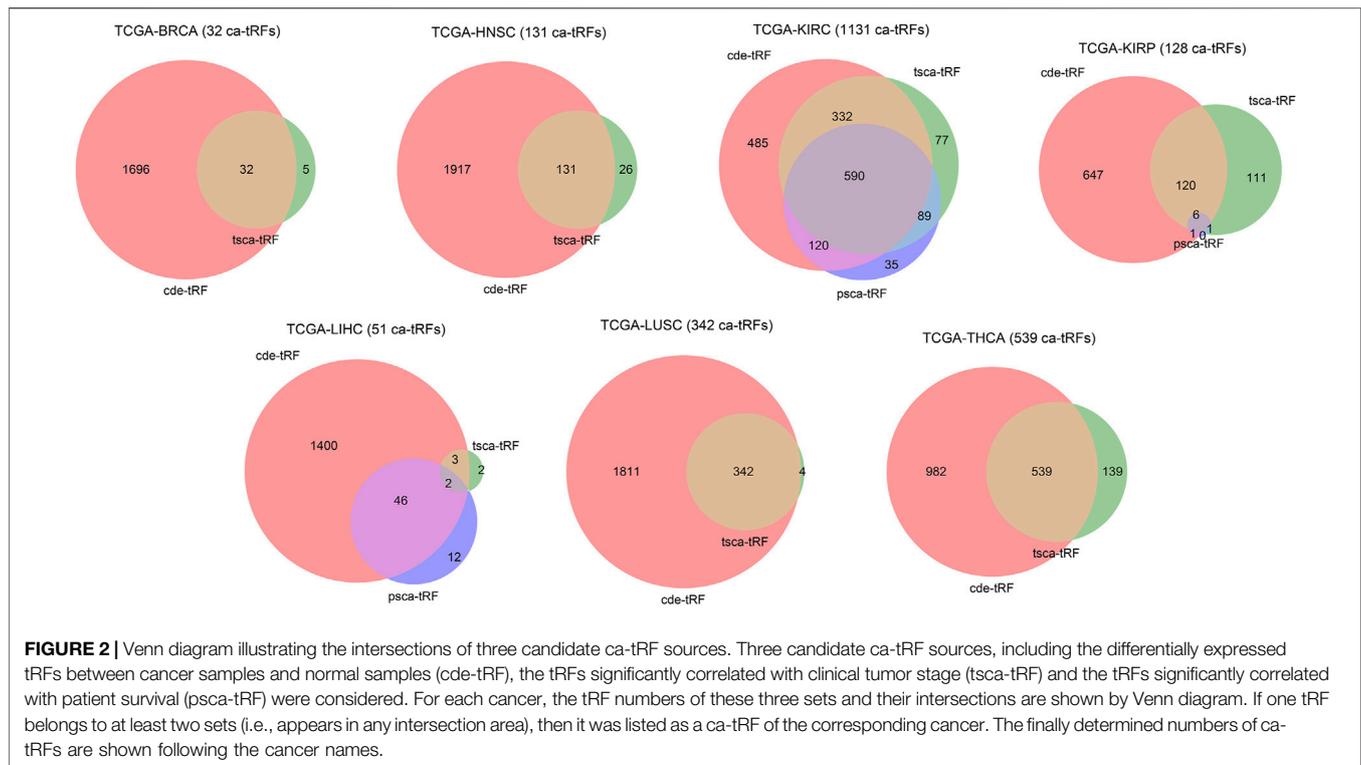
analyses of Gene Ontology Biological Process (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) to investigate the biological processes and pathways that the gene modules are involved in. After semantic deduplication, the terms with  $q$ -value no more than 0.05 were deemed as the statistically significant results.

## Estimation of Immune Infiltration by Gene Expression Deconvolution

We downloaded tumor samples' gene expression data from the TCGA official portal and then deconvolute them into quantitative cell composition lists by the EPIC tool (Racle et al., 2017). EPIC supposes different cell types prominently express their own cell type marker genes. Given the built-in marker genes' abundance of several pre-defined cell types (T cell, B cell, macrophage, natural killer cell, cancer-associated fibroblast and endothelial cell), the absolute proportions of these cell types inherent in the inputted TCGA bulk gene expression profiles can be solved by constrained linear model. Here the immune infiltration degree was quantified by the proportions of T cells.

## Identification and Analysis of ca-tRF-Based Cancer Subtypes

By using the consensus clustering computational framework provided by the R package *ConsensusClusterPlus* (Wilkerson and Hayes, 2010), we clustered tumor samples' ca-tRF expression profiles and assigned subtype labels to each tumor sample according to the clustering result (**Supplementary Data S2**). Consensus clustering is a widely used algorithm in biological clustering problems (Niu et al., 2016; Lu and Leong, 2018; Zhang et al., 2018). For a consensus clustering task, a basic clustering method needs to be assigned and here the PAM clustering algorithm, an improved version of K-mean clustering which is more robust to noises and outliers, was adopted. The principle of consensus clustering is measuring the distances within every sample-sample pairs by their probabilities of being clustered into the same group (called "consensus index" below) in several times of sub-sampling processes. Specifically, a sub-sampling proportion  $p$  (0.8 here) and times  $T$  (100 here) are required to be pre-defined. Then for each time, a subset of samples will be randomly selected according to  $p$  and then clustered by the PAM clustering algorithm. The sub-sampling process disturbs the original data structure to an extent, thus if two samples are objectively similar, they will resist this disturbance and present a high consensus index. On the contrary, two dissimilar samples will exhibit low consensus indexes. Therefore, if the clustering performed well, the overall consensus indexes of sample-sample pairs will be concentrated at 0 (for samples belonging to different groups) or 1 (for samples belonging to the same group). According to this property, we rationally determined the cluster number  $k$  for each cancer, namely the number of ca-tRF-based subtypes, where a smaller slope of the cumulative distribution curve between the interval of (0.2, 0.8) was preferred. In case the slopes are comparable among different  $k$  choices, the area under the cumulative distribution curve, which is usually approximating to the maximum when the  $k$  saturated, was considered as the secondary metric for the optimization of  $k$ .



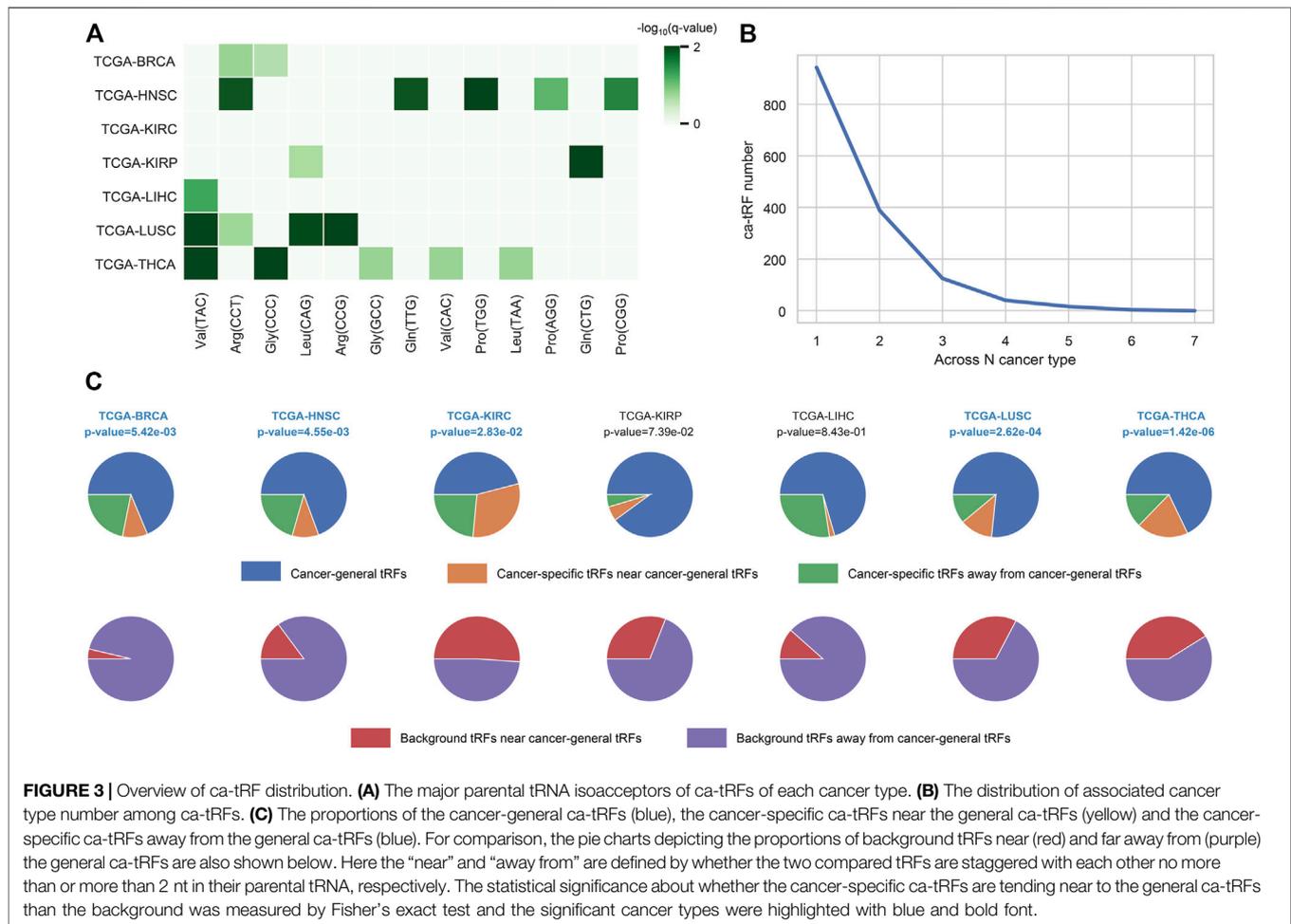
After clustering, the timeline-dependent survival rates of distinct subtypes were visualized by Kaplan-Meier plot for each cancer type using the R packages *survival* and *survminer*. Meanwhile the overall and specific inter-subtype survival differences were evaluated by log-rank test (**Supplementary Data S3**). For KIRC, KIRP, and LIHC, we also collected previous subtyping results based on the other indicators such as mRNA, microRNA, methylation level and so on (**Supplementary Data S4**) (Cancer Genome Atlas Research Network. Electronic address and Cancer Genome Atlas Research, 2017; Ricketts et al., 2018) to check their accordance and discrepancy with ca-tRF-based subtypes. The overlapping significance between ca-tRF subtypes and the other subtypes were measured by Fisher's exact test. We further calculated enrichment scores (ES) for LIHC samples in 236 biological pathways from KEGG and Molecular Signatures Database Hallmark (MSigDB Hallmark) (**Supplementary Data S5**), by single sample gene set enrichment analysis (ssGSEA) (Barbie et al., 2009). The ESs among different cancer subtypes were compared by Kruskal-Wallis test (for multi-group comparison) or Mann-Whitney test (for two-group comparison) to measure the differences of pathway activities. For all above statistical tests, a *p*-value cutoff 0.05 was used.

## RESULTS

### Overview of ca-tRFs Across Seven TCGA Cancer Types

In this work, we screened ca-tRFs for seven TCGA cancer types with sufficient samples and clinical information, including BRCA, HNSC, KIRC, KIRP, LIHC, LUSC, and THCA. Three tRF sets,

including the differentially expressed tRFs between cancer samples and normal samples (cde-tRF), the tRFs significantly correlated with clinical tumor stage (tsca-tRF) and the tRFs significantly correlated with patient survival (psca-tRF), were firstly constructed as candidate sources of ca-tRFs (See Materials and Methods). If one tRF is detectable (no less than 1 RPM) in no less than 10% samples and exists in at least two sets, then it was determined as a ca-tRF (**Figure 2**; **Supplementary Data S1**). We firstly investigated the major parental tRNA isoacceptors from which these ca-tRFs were derived by using Fisher's exact test. As the result, the ca-tRFs are found widely originated from the tRNA Val (TAC) and Arg (CCT) while the tRNA Leu (CAG) and Gly (CCC) could serve as secondary sources of ca-tRFs (**Figure 3A**). Then we defined cancer-specific tRFs and cancer-general tRFs as the ca-tRFs identified in only one cancer and those identified in no less than two cancers, respectively (**Figure 3B**). For all cancer types, a considerable proportion of cancer-general tRFs were observed. For comparison purpose, we also defined the tRFs staggered no more than 2 nt with the general ca-tRFs in their parental tRNAs as the tRFs near cancer-general tRFs. In the same way, tRFs staggered more than 2 nt with the general ca-tRFs in their parental tRNAs were defined as the tRFs far away from the cancer-general tRFs. By procedure, we found that for most cancers the cancer-specific tRFs significantly prone to be derived near the general ones in comparison with background (Fisher's exact test; **Figure 3C**), outlining the sequence similarity and therefore the potential functional similarity between cancer-specific and cancer-general tRFs. Inspired by this point, we visualized the cleavage site distribution of tRFs in mature



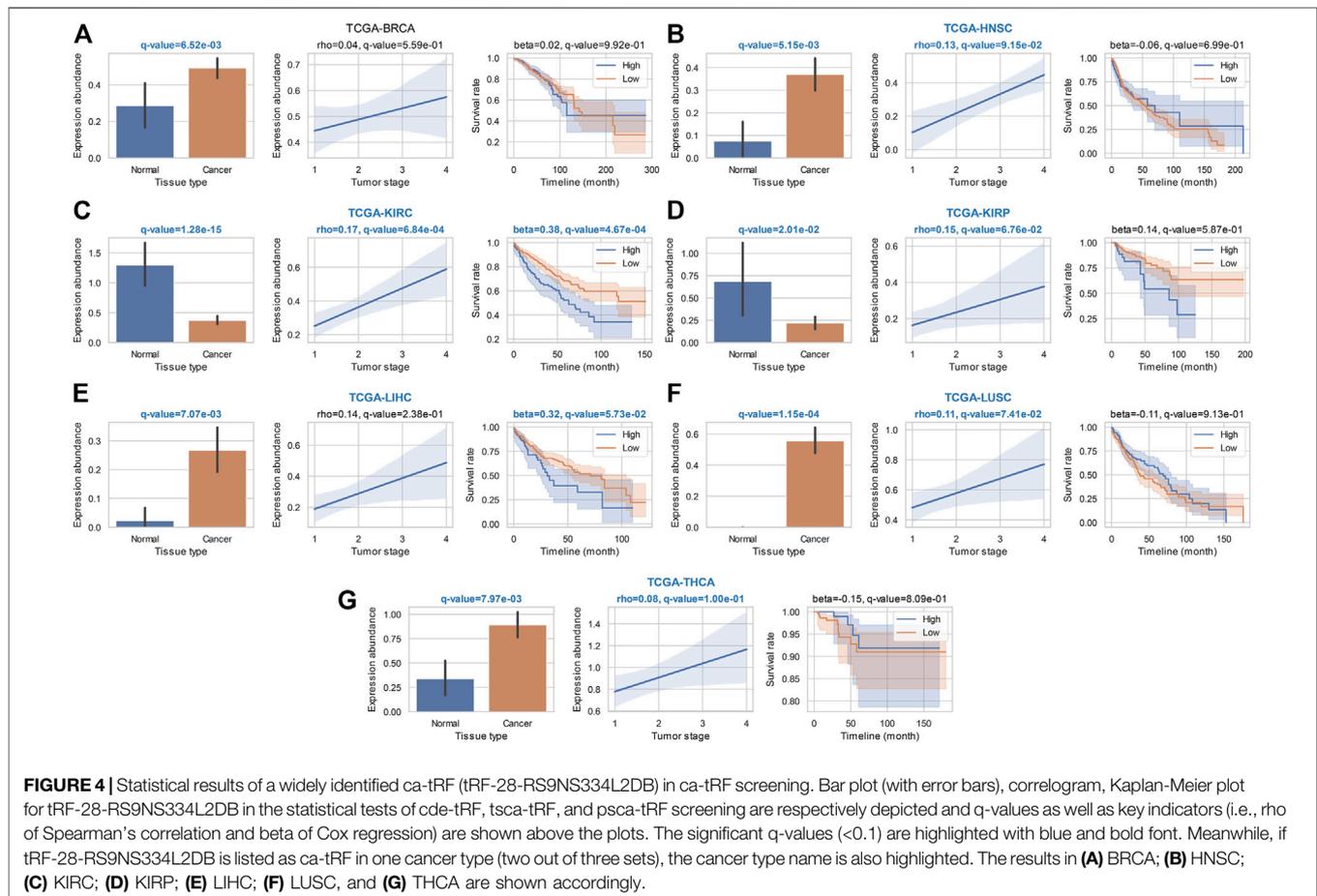
tRNAs (**Supplementary Figure S1**). Firstly, the background tRFs sourced from the middle or middle-latter sections of tRNAs are relatively less than those from the other sections, which is accordant with previous reports (Telonis et al., 2019). Further, in comparison with the background, the ca-tRFs of BRCA, HNSC, KIRC are found more likely from the 5' half parts of mature tRNAs whereas the ca-tRFs of LUSC and LIHC prone to come from the 3' half parts. That is to say, the cancer-associated tRFs may exhibit more prominent location bias in comparison with background tRFs for several cancer types.

We also noticed 19 most widely identified general ca-tRFs that are presented in at least five cancer types (**Supplementary Data S1**). As an instance, the statistical results of tRF-28-RS9NS334L2DB, a ca-tRF across six cancer types, are depicted in detail in **Figure 4**. Specifically, tRF-28-RS9NS334L2DB is significantly differential expressed in all considered seven cancer types when comparing cancer samples with normal control, and in most cancers (i.e., BRCA, HNSC, LIHC, LUSC, and THCA), an up-regulation was observed (bar plots in **Figure 4**). tRF-28-RS9NS334L2DB expression abundance was also found significantly positively correlated with clinical tumor stage in 5 cancers (HNSC, KIRC, KIRC, LUSC and THCA; correlograms in **Figure 4**) and poor prognosis in 2 cancers

(KIRC and LIHC; Kaplan-Meier plots in **Figure 4**). Therefore, tRF-28-RS9NS334L2DB may be positively indicative in a noticeable range of cancer onset and development.

## ca-tRFs Tend to Target Cancer-Related Pathways and Associate With Tumor Immune Infiltration

After the ca-tRF screening, we further investigated their potential functions. In consideration of tRFs' microRNA-like, AGO-dependent mRNA targeting ability, we firstly asked whether the target genes of ca-tRFs could participate in important biological processes related to cancer. According to the AGO-mediated tRF-gene interactions recorded in the tRFTar database (Zhou et al., 2021), for each cancer type we detected a significant gene module targeted by ca-tRFs by means of the RWR model on the PPI network (See Materials and Methods). We first noted that the ca-tRF target modules are enriched with the curated cancer-associated genes (**Figure 5A**), demonstrating the dysregulation of ca-tRFs is likely to contribute to oncogenesis and cancer development. Subsequently, referring to the gene sets of GO-BP and KEGG, we performed functional enrichment analysis for each ca-tRF target module (**Figure 5B–C**). Extensive terms well-



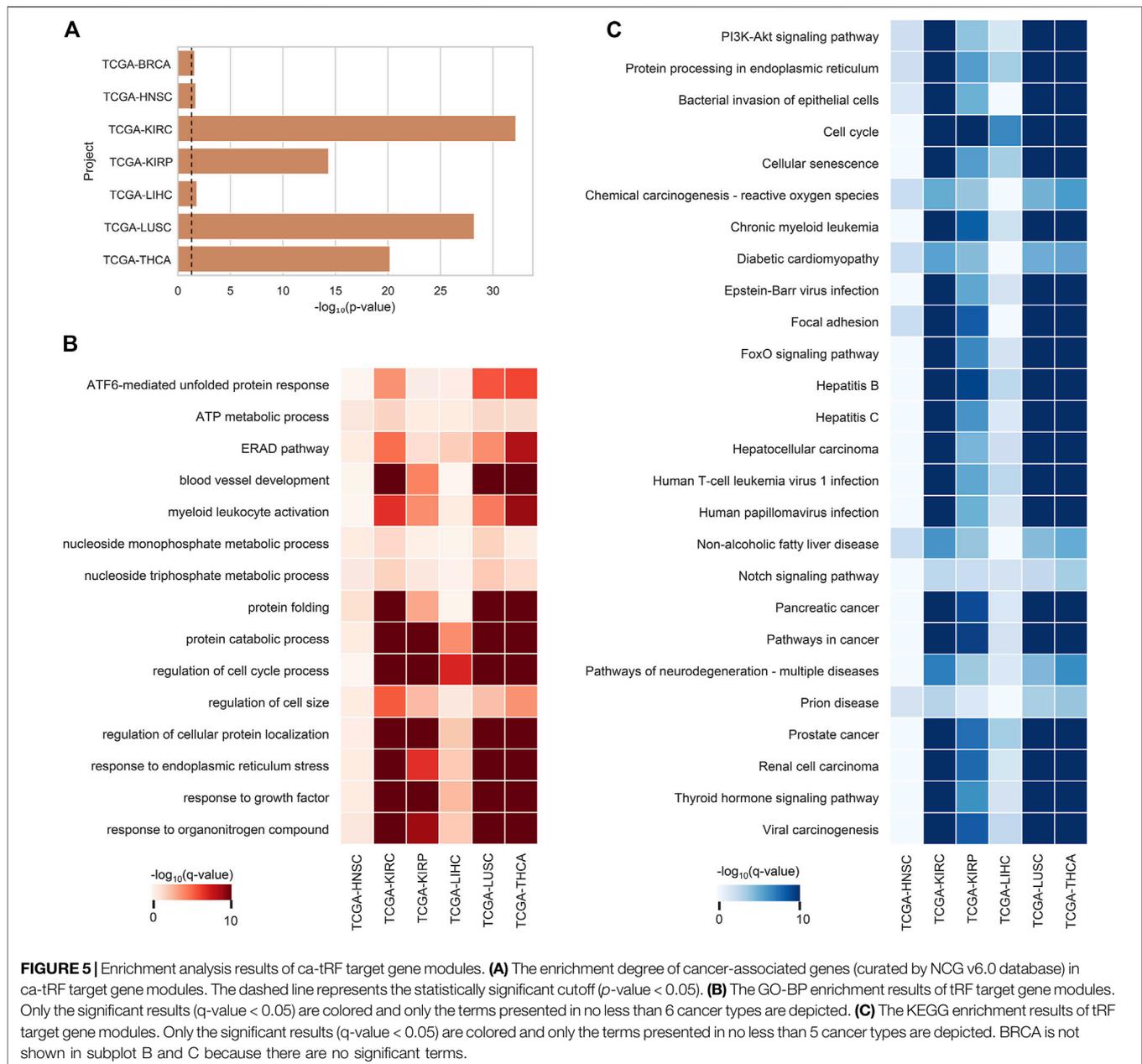
known to be associated with cancer such as ATF6-mediated unfolded protein response (Lin et al., 2021), blood vessel development (Carmeliet and Jain, 2011), regulation of cell cycle process (Evan and Vousden, 2001), focal adhesion (Eke and Cordes, 2015), PI3K-Akt signaling pathway (Martini et al., 2014), cellular senescence (Campisi, 2013) and FoxO signaling pathway (Farhan et al., 2017) are presented across multi-cancer types, further supporting the associations of ca-tRFs with the cancer-related functions. It should be noted that among these terms, blood vessel development, regulation of cell cycle process, focal adhesion and PI3K-Akt signaling pathway are also identified by Telonis et al. as the pathways universally regulated by tRFs (Telonis et al., 2019). But on the other hand, there are still novel tRF target pathways that are not previously overrepresented such as ATF6-mediated unfolded protein response, cellular senescence and FoxO signaling pathway. This result also highlights the importance of specific screening the pathways that are likely disturbed by tRFs through the extensive tRF-target analysis.

The functional enrichment results also revealed a few immuno-pathways and infectious processes like myeloid leukocyte activation, hepatitis, human T-cell leukemia virus 1 infection and bacterial invasion of epithelial cells, implying plausible immuno-regulatory roles of ca-tRFs. To further explore this point, we estimated the T-cell infiltration degree

of tumor samples by gene expression deconvolution (See Materials and Methods) and further studied its correlation with ca-tRF expression abundance for each cancer type. As expected, ca-tRFs exhibit significantly higher correlations with T-cell infiltration than other tRFs (**Figure 6**). Considering the significance of T-cell infiltration in cancer prognosis (Gu-Trantien et al., 2013; Ge et al., 2019; Zhang et al., 2019), this result indicates the closer link between ca-tRFs and tumor immunity.

## ca-tRF-Based Cancer Subtypes are Informative to Prognosis

Given the above-mentioned associations of ca-tRFs with cancer-related pathways and tumor immune infiltration, we further tested whether the ca-tRFs, if collectively investigated, could serve as an informative indicator of cancer prognosis. According to the ca-tRF expression profiles, we clustered the tumor samples under the consensus clustering framework (See Materials and Methods). Consequently, we identified five subtypes for HNSC, LIHC, and LUSC, six subtypes for KIRC, seven subtypes for BRCA and THCA and nine subtypes for KIRP (panels A–B of **Figures 7, 8** and **Supplementary Figures S2–S6; Supplementary Data S2**). Among these cancers, HNSC, KIRC, KIRP, and LIHC exhibit significant survival differences among

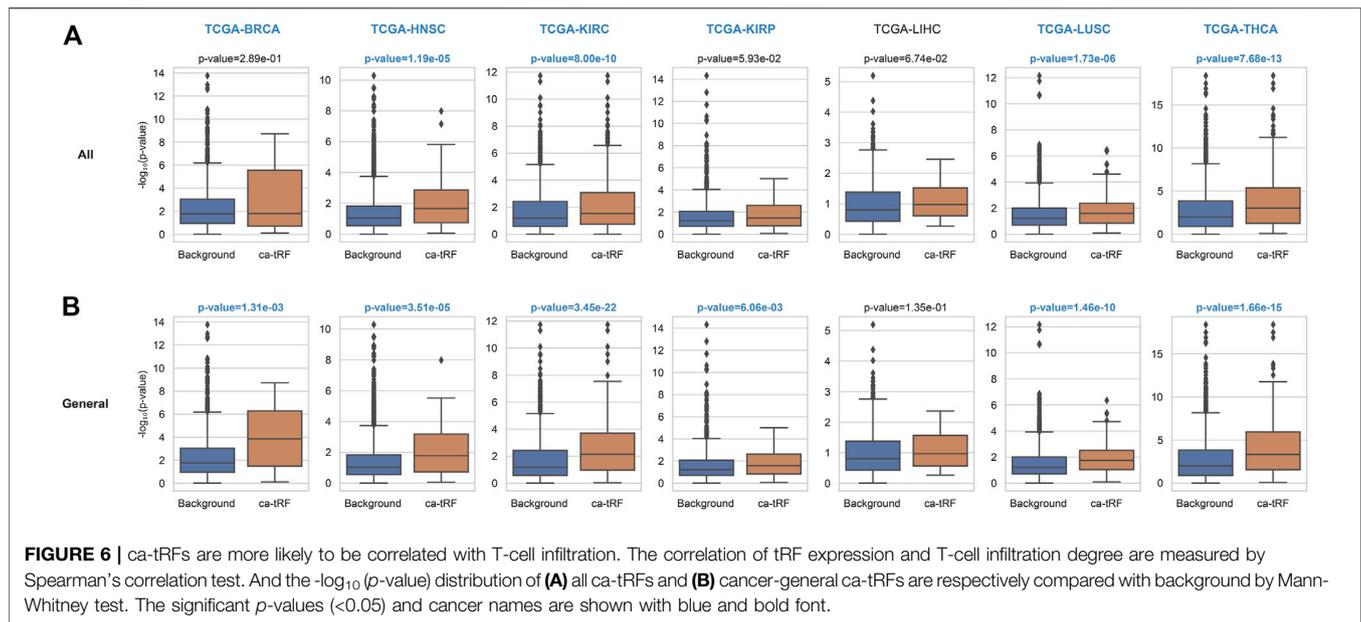


subtypes in overall, and the differences are especially obvious in KIRC and LIHC ( $p\text{-value}$  less than 0.0001) (panel C of **Figures 7, 8** and **Supplementary Figures S2–S6**), suggesting the indicative role of ca-tRF-based subtyping in cancer prognosis. By contrast, the prognostic divergence is not significant in BRCA, LUSC, and THCA. Previous researches had reported the phenomenon of tRF expression disparity depending on sex and race (Telonis et al., 2015; Magee et al., 2018; Telonis and Rigoutsos, 2018; Magee and Rigoutsos, 2020). Inspired by this point, we further grouped the ca-tRF subtypes by sex and race (only the sex and race with sufficient samples considered) to investigate whether the disparity is also presented in survival rate (panel D of **Figures 7, 8** and **Supplementary Figures S3–S4**). BRCA, LUSC, and THCA were

not included in this analysis because their original ca-tRF subtypes are failed to show decent survival differences. As a result, the patterns of survival curves among different subtypes are still maintained when considering specific sex and race, indicating the robustness of ca-tRF subtyping.

### Comparison of ca-tRF-Based Subtypes With the External Subtypes

Next, we further surveyed whether the ca-tRF-based subtypes could be linked or independent with the previous subtyping methods. In this comparison, only cancer types with sufficient external subtype data and showing significant prognostic

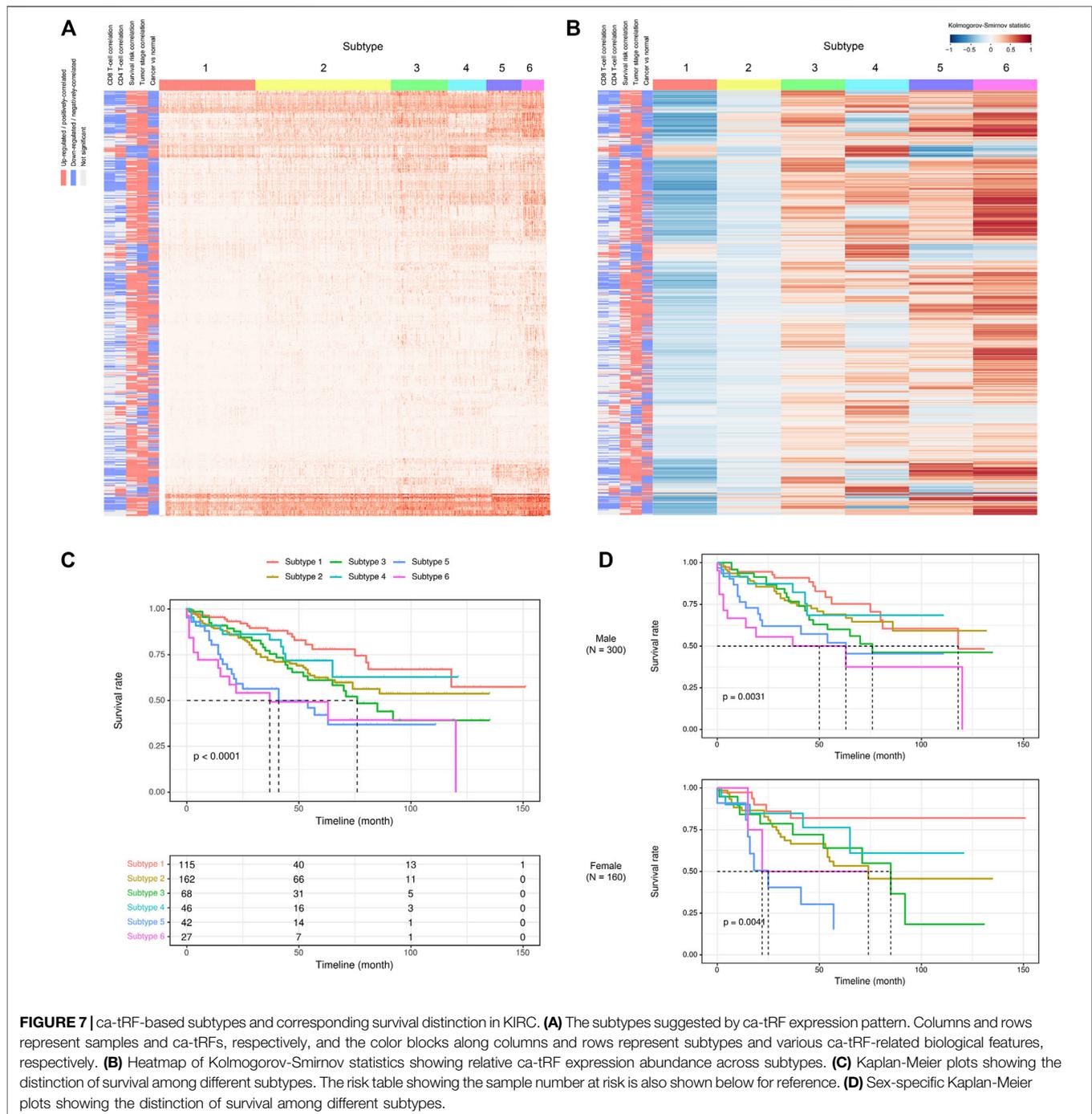


divergence in ca-tRF-based subtyping were considered, containing LIHC (**Figure 9**), KIRC (**Supplementary Figure S7**) and KIRP (**Supplementary Figure S8**). For these three cancer types, various existing external subtypes based on mRNA, miRNA, methylation level and so on were covered in this comparison (**Supplementary Data S4**), including but not limited to the subtypes delineated by TCGA Consortia and identified by non-negative matrix factorization (Cancer Genome Atlas Research Network. Electronic address and Cancer Genome Atlas Research, 2017; Ricketts et al., 2018). We found most ca-tRF-based subtypes could decently overlap with at least one external subtype. However, as for subtype 2 of KIRC as well as subtype 2 and 6 of KIRP, no matched external subtype was found. Moreover, for all the surveyed cancers, although there are wide overlaps between ca-tRF-based subtypes and external subtypes, few ca-tRF-based subtypes could be fully represented by another external subtype. In other words, there are also detectable independence between ca-tRF subtypes and external subtypes, indicating tRFs' potential to serve as novel prognostic factors that could supplement known subtyping.

To better demonstrate the associations and independence between the ca-tRF-based subtypes and the external subtypes, we further investigated LIHC subtypes in details because 1) ca-tRF-based subtypes show prominent survival differences in LIHC and 2) there are the largest number of external subtypes for reference in LIHC (8 kinds of subtypes in total), which could maximally reflect the relationships between the subtypes based on ca-tRFs and the other indicators. As depicted in **Figure 8C**, in LIHC, ca-tRF subtype 2 is of the best survival (low-risk group), ca-tRF subtypes 1 and 3 are with medium risks (medium-risk group) and ca-tRF subtypes 4 and 5 are the most malignant (high-risk group). Besides, the ca-tRF subtype 2, 4, and 5 significantly overlap with five, four, and six external subtypes, respectively. More importantly, the subtype 2, 4, and 5 were found significantly

overlapped with iCluster subtypes (**Figure 9H**), which comprehensively integrate multi-dimensional data including transcriptome, miRNA expression pattern, methylome, proteome and genomic variation. Unlike ca-tRF subtype 2, 4, and 5 that show decent overlap with external subtypes, ca-tRF subtype 1 and 3 only match three and two external subtypes, respectively, and show no significant overlap with any iCluster subtypes, indicating they are relatively independent to the external subtypes (**Figure 9**).

We further noted that samples assigned to different ca-tRF-based subtypes would be associated with different pathways, even in the same iCluster subtype group. To compare the molecular signatures of different ca-tRF subtypes, we evaluated the activities of 236 biological pathways for LIHC samples (**Supplementary Data S5**) by ssGSEA enrichment scores (ES) and further screened 50 most variable pathways among ca-tRF subtypes by Kruskal-Wallis test (see Materials and Methods for more details). The screened pathways could be clustered into four pathway sets: the set 1 involves cell cycle, DNA repair, RNA process and apoptosis; the set 2 involves glycometabolism, mTOR signaling, ROS signaling and unfolded protein response; the set 3 includes many metabolism- and degradation-associated pathways; and the few other pathways were assigned to the pathway set 4 (**Figure 10A**). As expected, further pairwise comparisons demonstrate the pathway activities among ca-tRF subtypes are significantly different and meanwhile the low- and high-risk group exhibit more distinctive signatures than the medium-risk group (**Figure 10A**). More interestingly, we noted that the ES differences in pathway set 1 and 2 are still maintained when specifically considering iCluster subtype 1 (**Figure 10B**) and 3 (**Figure 10C**). That is to say, although the samples within the iCluster subtype 1 or 3 are of similar multi-omics characteristics, they will be distinguishable in many important pathway activities when further considering ca-tRF expression patterns. Such observation indicates that ca-tRF-based subtyping would

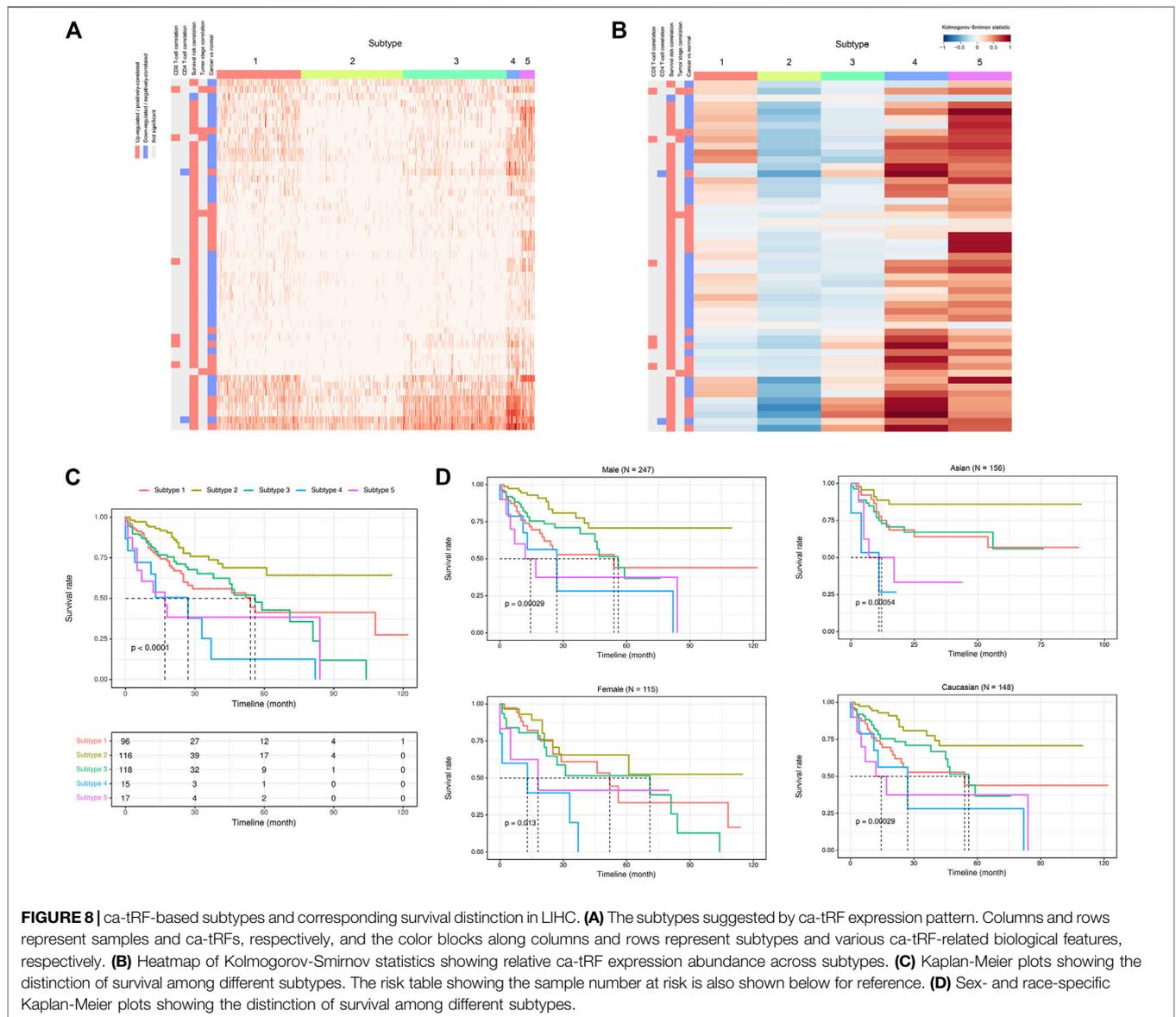


specifically reveal the functional characteristics between different sample groups that are independent to the external multi-omics characteristics-based subtyping.

## DISCUSSION

As a novel class of non-coding RNA, tRF has been getting increasing attention in recent years. Experimental researches

have revealed that some tRFs could serve as biomarkers in some cancers (Falconi et al., 2019; Zhu et al., 2019). However, these low throughput experimental researches are not sufficient to explore the landscape of dysregulated tRFs at the pan-cancer level. In this work, according to TCGA samples' tRF expression data provided by MINTbase v2.0, we screened ca-tRFs for seven TCGA cancer types with a "two out of three" strategy. From the pan-cancer perspective, the ca-tRFs are found significantly derived from the tRNA Val (TAC) and Arg (CCT). Moreover,

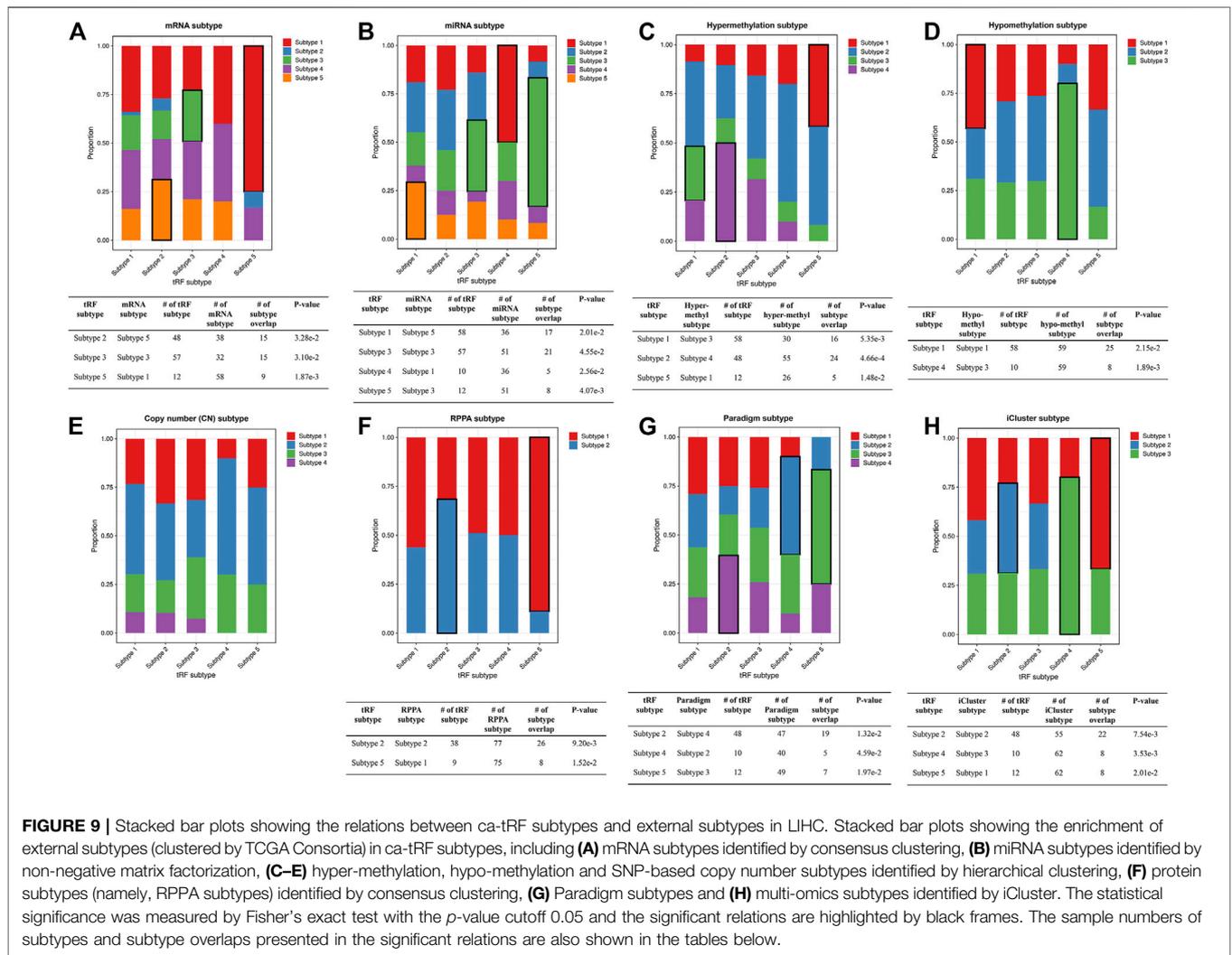


we observed that most identified ca-tRFs are presented in multiple cancer types, and more interestingly, a considerable fraction of cancer-specific ca-tRFs are actually derived from the proximal region of cancer-general ca-tRFs on tRNAs. Unlike the biogenesis of canonical RNAs which are directly transcribed from the genome, tRFs are originated from the cleavage of mature tRNAs. Our results indicate the distribution of ca-tRFs on tRNAs are not random, and there are likely “hotspots” in tRNAs to produce ca-tRFs in tumors.

We also noticed that the variation tendency of ca-tRFs in three candidate tRF sets (i.e., cde-tRFs, tsca-tRFs, and psca-tRFs) are not always unidirectional. For example, a tRF may be significantly down-regulated in cancer samples in comparison with normal control but positively correlated with clinical tumor stage (**Supplementary Data S1**). This phenomenon demonstrates some tRFs may play opposite roles in oncogenesis and cancer

progression. Meanwhile, it also reflects the advantages of our multi-view screening to capture comprehensive features of ca-tRFs.

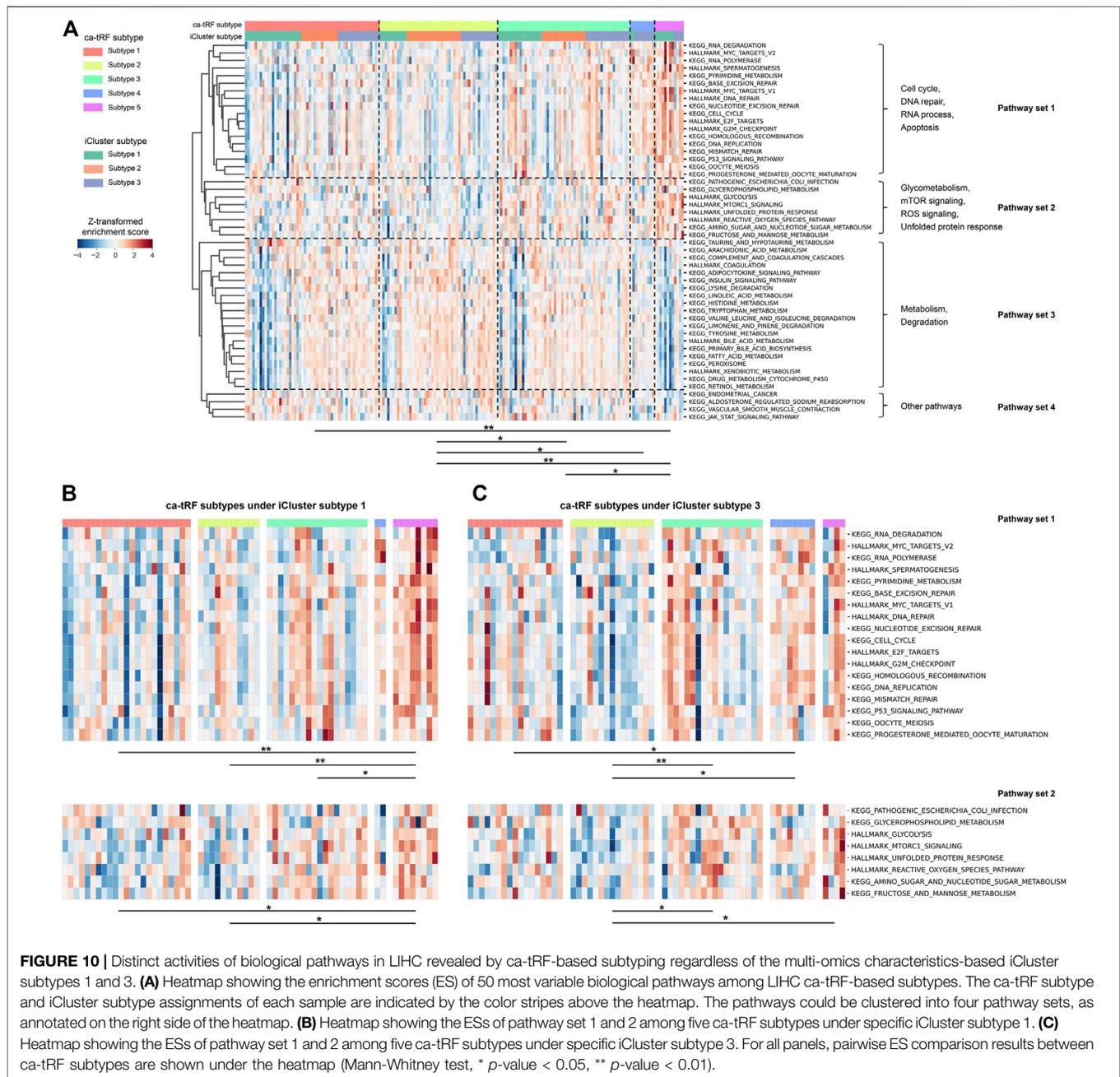
Subsequent functional analysis reveals that the ca-tRF target gene modules participate in many oncogenesis and tumor progression-related processes such as ATF6-mediated unfolded protein response, angiogenesis, cell cycle process regulation, focal adhesion, PI3K-Akt signaling pathway, cellular senescence and FoxO signaling pathway. Meanwhile, the ca-tRFs also tend to be correlated with T-cell infiltration in comparison with other tRFs. Both results imply that ca-tRFs could play critical roles in cancer development and thus be of prognostic values. Indeed, cancer subtyping based on the ca-tRF expression pattern exhibit significant differences in survival and the differences are especially obvious in KIRC and LIHC. In addition, the survival patterns are still robust when considering specific sex



and race, therefore ca-tRFs are potential to serve as universal prognostic factors. We also found most ca-tRF subtypes could decently link with external subtypes. But on the hand, the decent but not prominent overlaps also imply detectable independence between ca-tRF subtypes and external subtypes, indicating ca-tRFs would provide novel predictive clues for cancer prognosis. As a demonstration, we deeply investigated the relationships between ca-tRF subtypes and iCluster subtypes in LIHC, by means of ssGSEA. We found the ca-tRF expression patterns could associate with many pivotal biological pathways such as cell cycle, apoptosis, mTOR signaling pathway and so on regardless of the multi-omics characteristics of iCluster subtype 1 and 3, underlining the important regulatory roles of ca-tRFs.

Our analysis has systematically uncovered the potential roles of tRFs in oncogenesis and development and provided a reasonable ca-tRF list for future researches. However, there are still limitations in current work. The first limitation is in the step of investigating ca-tRF functions by the tRF-target interaction analysis, where we only considered the AGO-mediated tRF-gene

interactions based on AGO-CLIP datasets. However, as we describe before, beyond binding with AGO-family proteins, tRFs also have various other functions which may involve some other RNA-binding proteins (RBP) (Couvillion et al., 2012; Kim et al., 2017). But to our best knowledge, so far there has not been tRF dataset focusing on the other RBPs because high-throughput techniques for profiling RBP interactions like CLIP-seq are mainly designed for mRNAs rather than tRFs. Suffering from this limitation, our results could only reveal partial ca-tRF functions in cancer. On the other hand, although there are evidences that the functions of some tRFs are indeed AGO-dependent, AGO-association presented in CLIP data do not necessarily indicate AGO-dependent mechanism-of-action. Therefore, in comparison with some technologies which can directly capture AGO-mediated small RNA-gene interactions (for example, CLASH) (Helwak et al., 2013), CLIP-derived interactions should contain more false positives. However, in the current stage, such CLASH-based datasets like are insufficient (usually identifying ~1,000 tRF-gene pairs) to profile large-scale tRF-gene



interactions. Therefore, we still adopted the more widely-used CLIP datasets, and to reduce false positives, we constrained the results with tight base pairing. In the previous work (Zhou et al., 2021), we validated the tRFs presented in the screened tRF-gene duplexes exhibit much better complementarily pairing abilities with CLIP-peaks than randomly generated small RNAs. Besides, the screened tRF-gene pairs are more prone to be co-expressed than background in TCGA samples, indicating the regulatory roles of these tRFs. Moreover, in accordance with the fact that AGO proteins are prone to bind smaller RNAs, ~89% of the screened interactions involve tRFs less than 24 nt. When filtering the interactions with consistent co-expression in TCGA samples

(i.e., the tRF-gene interaction set we used in this work), the ratio of interactions involving smaller tRFs will be dominating (~98%). Overall, it could be rationally inferred that these interactions are with decent reliability. Another thing should be noted that Kumar et al. found the abundance of tRFs loaded on AGO2 are much lower than AGO1, AGO3 and AGO4 (Kumar et al., 2014), based on Hafner et al.'s HEK293 datasets (Hafner et al., 2010), plausibly indicating tRFs' poor AGO2-binding ability. However, we found this phenomenon is not very repeatable in some other datasets such as Vongrad et al.'s human macrophage datasets (Vongrad et al., 2015), Benway et al.'s HK-2 datasets (Benway and Iacomini, 2018) and Hamilton et al.'s DU145 datasets (Hamilton et al.,

2016). For example, in the AGO2-specific PAR-CLIP data of human macrophages generated by Vongrad et al., many tRFs are of hundreds of RPMs (median 150.13 RPMs of top 30 tRFs) and more surprisingly, much more abundant than miRNAs (median 5.80 RPMs of top 30 miRNAs). Higher abundances of tRFs relative to miRNAs were also observed in Benway et al.'s AGO2 PAR-CLIP data (median 32.86 RPMs of top 30 tRFs versus median 1.08 RPMs of top 30 miRNAs) and Hamilton et al.'s AGO2 PAR-CLIP data of DU145 cell line. What's more, some publications have experimentally validated some tRFs should rely on AGO2 protein to perform downstream regulation (Li et al., 2012; Luo et al., 2018; Green et al., 2020). Therefore, the AGO2 protein is still considered in our pipeline.

The second limitation is due to the extreme scarcity of small RNA-sequencing data in public databases, we could not validate the ca-tRF-based subtyping in a totally independent dataset. Therefore, we could only perform validation on the original TCGA sample set. Specifically, for each cancer type, we randomly split the original set into training set and testing set with the ratio of 1:1, and then re-screened ca-tRFs on the training set and re-clustered tumor samples on both training and testing set based on the re-screened ca-tRFs. Subsequently, we measured the set similarity between the original ca-tRF subtypes and the re-clustered ca-tRF subtypes by Ochiai coefficient. As a result, the re-clustered subtypes exhibit decent consistency with the original subtypes, where most cancer types can reach at an Ochiai coefficient of 0.5 in average (**Supplementary Figure S9**). In other words, each original subtype could expectedly match with a highly overlapped re-clustered subtype, reflecting the robustness of ca-tRF-based subtyping. Nevertheless, the ca-tRF subtypes are still required to be further validated when the external resource of small RNA-sequencing is abundant enough in the future.

The third limitation worth discussing is, in this work we mainly surveyed tRF-cancer associations from the perspective of expression pattern. However, in recent years, many computational approaches based on machine learning and graphic theory are proposed to predict the associations between non-coding RNAs and diseases (Chen et al., 2018; Zeng et al., 2018), which provides another approach to discover novel tRF-cancer associations. However, an obligate prerequisite of these

algorithms is a sizable training dataset derived from prior experimental knowledge. And to our best knowledge, existing publications about tRF and cancer are extremely scarce (actually less than 200, by querying the keywords “tsRNA cancer” and “tRNA-derived fragment cancer” in PubMed). Therefore, in current stage, existing data have not been able to support these algorithms, but we believe it would be feasible with the accumulation of experimental data in the future.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

YuZ, YiZ, and QC contributed to conception and design of the study. YiZ performed the analysis and drafted the manuscript. YuZ supervised the study and revised the manuscript. All authors approved the submitted version.

## FUNDING

This study was supported by the grants from National Natural Science Foundation of China (31801099 and 32070658, to YuZ).

## ACKNOWLEDGMENTS

We thank Chunmei Cui for her assistance in data downloading.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.747931/full#supplementary-material>

## REFERENCES

- Balatti, V., Nigita, G., Veneziano, D., Drusco, A., Stein, G. S., Messier, T. L., et al. (2017). tsRNA Signatures in Cancer. *Proc. Natl. Acad. Sci. USA* 114 (30), 8071–8076. doi:10.1073/pnas.1706908114
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460
- Benway, C. J., and Iacomini, J. (2018). Defining a microRNA-mRNA Interaction Map for Calcineurin Inhibitor Induced Nephrotoxicity. *Am. J. Transpl.* 18 (4), 796–809. doi:10.1111/ajt.14503
- Campisi, J. (2013). Aging, Cellular Senescence, and Cancer. *Annu. Rev. Physiol.* 75, 685–705. doi:10.1146/annurev-physiol-030212-183653
- Cancer Genome Atlas Research Network. Electronic address and Cancer Genome Atlas Research (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169 (7), 1327–1341. e1323. doi:10.1016/j.cell.2017.05.046
- Carmeliet, P., and Jain, R. K. (2011). Molecular Mechanisms and Clinical Applications of Angiogenesis. *Nature* 473 (7347), 298–307. doi:10.1038/nature10144
- Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., et al. (2016). Sperm tsRNAs Contribute to Intergenerational Inheritance of an Acquired Metabolic Disorder. *Science* 351 (6271), 397–400. doi:10.1126/science.aad7977
- Chen, X., Gong, Y., Zhang, D.-H., You, Z.-H., and Li, Z.-W. (2018). DRMDA: Deep Representations-Based miRNA-Disease Association Prediction. *J. Cel. Mol. Med.* 22 (1), 472–485. doi:10.1111/jcmm.13336
- Couvillion, M. T., Bounova, G., Purdom, E., Speed, T. P., and Collins, K. (2012). A Tetrahymena Piwi Bound to Mature tRNA 3' Fragments Activates the Exonuclease Xrn2 for RNA Processing in the Nucleus. *Mol. Cel* 48 (4), 509–520. doi:10.1016/j.molcel.2012.09.010
- Dimitrakopoulos, G. N., Klapa, M. I., and Moschonas, N. K. (2020). PICKLE 3.0: Enriching the Human Meta-Database with the Mouse Protein Interactome

- Extended *via* Mouse-Human Orthology. *Bioinformatics* 37, 145–146. doi:10.1093/bioinformatics/btaa1070
- Eke, I., and Cordes, N. (2015). Focal Adhesion Signaling and Therapy Resistance in Cancer. *Semin. Cancer Biol.* 31, 65–75. doi:10.1016/j.semcancer.2014.07.009
- Evan, G. I., and Vousden, K. H. (2001). Proliferation, Cell Cycle and Apoptosis in Cancer. *Nature* 411 (6835), 342–348. doi:10.1038/35077213
- Falconi, M., Giangrossi, M., Zabaleta, M. E., Wang, J., Gambini, V., Tilio, M., et al. (2019). A Novel 3'-tRNA<sup>Glu</sup>-Derived Fragment Acts as a Tumor Suppressor in Breast Cancer by Targeting Nucleolin. *FASEB J.* 33 (12), 13228–13240. doi:10.1096/fj.201900382RR
- Farhan, M., Wang, H., Gaur, U., Little, P. J., Xu, J., and Zheng, W. (2017). FOXO Signaling Pathways as Therapeutic Targets in Cancer. *Int. J. Biol. Sci.* 13 (7), 815–827. doi:10.7150/ijbs.20052
- Ge, P., Wang, W., Li, L., Zhang, G., Gao, Z., Tang, Z., et al. (2019). Profiles of Immune Cell Infiltration and Immune-Related Genes in the Tumor Microenvironment of Colorectal Cancer. *Biomed. Pharmacother.* 118, 109228. doi:10.1016/j.biopha.2019.109228
- Green, J. A., Ansari, M. Y., Ball, H. C., and Haqqi, T. M. (2020). tRNA-derived Fragments (tRFs) Regulate post-transcriptional Gene Expression *via* AGO-dependent Mechanism in IL-1 $\beta$  Stimulated Chondrocytes. *Osteoarthritis and Cartilage* 28 (8), 1102–1110. doi:10.1016/j.joca.2020.04.014
- Gu, X., Wang, L., Coates, P. J., Boldrup, L., Fähræus, R., Wilms, T., et al. (2020). Transfer-RNA-Derived Fragments Are Potential Prognostic Factors in Patients with Squamous Cell Carcinoma of the Head and Neck. *Genes* 11 (11), 1344. doi:10.3390/genes11111344
- Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., et al. (2013). CD4+ Follicular Helper T Cell Infiltration Predicts Breast Cancer Survival. *J. Clin. Invest.* 123 (7), 2873–2892. doi:10.1172/JCI67428
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide Identification of RNA-Binding Protein and microRNA Target Sites by PAR-CLIP. *Cell* 141 (1), 129–141. doi:10.1016/j.cell.2010.03.009
- Hamilton, M. P., Rajapakshe, K. I., Bader, D. A., Cerne, J. Z., Smith, E. A., Coarfa, C., et al. (2016). The Landscape of microRNA Targeting in Prostate Cancer Defined by AGO-PAR-CLIP. *Neoplasia* 18 (6), 356–370. doi:10.1016/j.neo.2016.04.008
- Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* 153 (3), 654–665. doi:10.1016/j.cell.2013.03.043
- Ivanov, P., Villen, J., Gygi, S. P., Anderson, P., and Anderson, P. (2011). Angiogenin-induced tRNA Fragments Inhibit Translation Initiation. *Mol. Cell* 43 (4), 613–623. doi:10.1016/j.molcel.2011.06.022
- Kim, H. K., Fuchs, G., Wang, S., Wei, W., Zhang, Y., Park, H., et al. (2017). A Transfer-RNA-Derived Small RNA Regulates Ribosome Biogenesis. *Nature* 552 (7683), 57–62. doi:10.1038/nature25005
- Kumar, P., Anaya, J., Mudunuri, S. B., and Dutta, A. (2014). 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.* 12, 78. ARTN 78. doi:10.1186/s12915-014-0078-0
- Lee, Y. S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A Novel Class of Small RNAs: tRNA-Derived RNA Fragments (tRFs). *Genes Dev.* 23 (22), 2639–2649. doi:10.1101/gad.1837609
- Li, S., Xu, Z., and Sheng, J. (2018). tRNA-Derived Small RNA: A Novel Regulatory Small Non-coding RNA. *Genes* 9 (5), 246. doi:10.3390/genes9050246
- Li, Z., Ender, C., Meister, G., Moore, P. S., Chang, Y., and John, B. (2012). Extensive Terminal and Asymmetric Processing of Small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res.* 40 (14), 6787–6799. doi:10.1093/nar/gks307
- Lin, H., Wang, J., Wang, T., Wu, J., Wang, P., Huo, X., et al. (2021). The LncRNA MIR503HG/miR-224-5p/TUSC3 Signaling Cascade Suppresses Gastric Cancer Development *via* Modulating ATF6 Branch of Unfolded Protein Response. *Front. Oncol.* 11, 708501. doi:10.3389/fonc.2021.708501
- Lu, B., and Leong, H. W. (2018). GI-cluster: Detecting Genomic Islands *via* Consensus Clustering on Multiple Features. *J. Bioinform. Comput. Biol.* 16 (3), 1840010. doi:10.1142/S0219720018400103
- Luo, S., He, F., Luo, J., Dou, S., Wang, Y., Guo, A., et al. (2018). Drosophila tsRNAs Preferentially Suppress General Translation Machinery *via* Antisense Pairing and Participate in Cellular Starvation Response. *Nucleic Acids Res.* 46 (10), 5250–5268. doi:10.1093/nar/gky189
- Magee, R. G., Telonis, A. G., Loher, P., Londin, E., and Rigoutsos, I. (2018). Profiles of miRNA Isoforms and tRNA Fragments in Prostate Cancer. *Sci. Rep.* 8 (1), 5314. doi:10.1038/s41598-018-22488-2
- Magee, R., and Rigoutsos, I. (2020). On the Expanding Roles of tRNA Fragments in Modulating Cell Behavior. *Nucleic Acids Res.* 48 (17), 9433–9448. doi:10.1093/nar/gkaa657
- Martini, M., De Santis, M. C., Braccini, L., Gulluni, F., and Hirsch, E. (2014). PI3K/AKT Signaling Pathway and Cancer: an Updated Review. *Ann. Med.* 46 (6), 372–383. doi:10.3109/07853890.2014.912836
- Niu, Z., Chasman, D., Eisfeld, A. J., Kawaoka, Y., and Roy, S. (2016). Multi-task Consensus Clustering of Genome-wide Transcriptomes from Related Biological Conditions. *Bioinformatics* 32 (10), 1509–1517. doi:10.1093/bioinformatics/btw007
- Pliatsika, V., Loher, P., Magee, R., Telonis, A. G., Londin, E., Shigematsu, M., et al. (2018). 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.* 46 (D1), D152–D159. doi:10.1093/nar/gkx1075
- Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *Elife* 6, e26476. doi:10.7554/eLife.26476
- Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., et al. (2019). The Network of Cancer Genes (NCG): a Comprehensive Catalogue of Known and Candidate Cancer Genes from Cancer Sequencing Screens. *Genome Biol.* 20 (1), 1. doi:10.1186/s13059-018-1612-0
- Ricketts, C. J., De Cubas, A. A., Fan, H., Smith, C. C., Lang, M., Reznik, E., et al. (2018). The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* 23 (1), 313–326. e5. doi:10.1016/j.celrep.2018.03.07510.1016/j.celrep.2018.06.032
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Shao, Y., Sun, Q., Liu, X., Wang, P., Wu, R., and Ma, Z. (2017). tRF-Leu-CAG Promotes Cell Proliferation and Cell Cycle in Non-small Cell Lung Cancer. *Chem. Biol. Drug Des.* 90 (5), 730–738. doi:10.1111/cbdd.12994
- Telonis, A. G., Loher, P., Honda, S., Jing, Y., Palazzo, J., Kirino, Y., et al. (2015). Dissecting tRNA-Derived Fragment Complexities Using Personalized Transcriptomes Reveals Novel Fragment Classes and Unexpected Dependencies. *Oncotarget* 6 (28), 24797–24822. doi:10.18632/oncotarget.4695
- Telonis, A. G., Loher, P., Magee, R., Pliatsika, V., Londin, E., Kirino, Y., et al. (2019). tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities. *Cancer Res.* 79 (12), 3034–3049. doi:10.1158/0008-5472.Can-19-0789
- Telonis, A. G., and Rigoutsos, I. (2018). Race Disparities in the Contribution of miRNA Isoforms and tRNA-Derived Fragments to Triple-Negative Breast Cancer. *Cancer Res.* 78 (5), 1140–1154. doi:10.1158/0008-5472.CAN-17-1947
- Thompson, D. M., and Parker, R. (2009). Stressing Out over tRNA Cleavage. *Cell* 138 (2), 215–219. doi:10.1016/j.cell.2009.07.001
- Vongrad, V., Imig, J., Mohammadi, P., Kishore, S., Jaskiewicz, L., Hall, J., et al. (2015). HIV-1 RNAs Are Not Part of the Argonaute 2 Associated RNA Interference Pathway in Macrophages. *Plos One* 10 (7), e0132127. doi:10.1371/journal.pone.0132127
- Wang, Z., Xiang, L., Shao, J., and Yuan, Z. (2006). The 3' CCACCA Sequence of tRNA Ala (UGC) Is the Motif that Is Important in Inducing Th1-like Immune Response, and This Motif Can Be Recognized by Toll-like Receptor 3. *Clin. Vaccin. Immunol.* 13 (7), 733–739. doi:10.1128/0011-0966-00019-06
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170
- Wu, W., Lee, I., Spratt, H., Fang, X., and Bao, X. (2021). tRNA-Derived Fragments in Alzheimer's Disease: Implications for New Disease Biomarkers and Neuropathological Mechanisms. *J. Alzheimer's Dis.* 79 (2), 793–806. doi:10.3233/JAD-200917

- Yamasaki, S., Ivanov, P., Hu, G.-f., and Anderson, P. (2009). Angiogenin Cleaves tRNA and Promotes Stress-Induced Translational Repression. *J. Cel Biol* 185 (1), 35–42. doi:10.1083/jcb.200811106
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118
- Zeng, X., Liu, L., Lü, L., and Zou, Q. (2018). Prediction of Potential Disease-Associated microRNAs Using Structural Perturbation Method. *Bioinformatics* 34 (14), 2425–2432. doi:10.1093/bioinformatics/bty112
- Zhang, S., Wang, Y., Gu, Y., Zhu, J., Ci, C., Guo, Z., et al. (2018). Specific Breast Cancer Prognosis-Subtype Distinctions Based onDNAmethylation Patterns. *Mol. Oncol.* 12 (7), 1047–1060. doi:10.1002/1878-0261.12309
- Zhang, S., Zhang, E., Long, J., Hu, Z., Peng, J., Liu, L., et al. (2019). Immune Infiltration in Renal Cell Carcinoma. *Cancer Sci.* 110 (5), 1564–1572. doi:10.1111/cas.13996
- Zhang, Y., Bi, Z., Dong, X., Yu, M., Wang, K., Song, X., et al. (2021). tRNA -derived Fragments: tRF-Gly-CCC -046, tRF-Tyr-GTA -010 and tRF-Pro-TGG -001 as Novel Diagnostic Biomarkers for Breast Cancer. *Thorac. Cancer* 12, 2314–2323. doi:10.1111/1759-7714.14072
- Zhou, Y., Peng, H., Cui, Q., and Zhou, Y. (2021). tRFTar: Prediction of tRF-Target Gene Interactions via Systemic Re-analysis of Argonaute CLIP-Seq Datasets. *Methods* 187, 57–67. doi:10.1016/j.ymeth.2020.10.006
- Zhu, J., Cheng, M., and Zhao, X. (2020). A tRNA-Derived Fragment (tRF-3001b) Aggravates the Development of Nonalcoholic Fatty Liver Disease by Inhibiting Autophagy. *Life Sci.* 257, 118125. doi:10.1016/j.lfs.2020.118125
- Zhu, L., Li, J., Gong, Y., Wu, Q., Tan, S., Sun, D., et al. (2019). Exosomal tRNA-Derived Small RNA as a Promising Biomarker for Cancer Diagnosis. *Mol. Cancer* 18 (1), 74. doi:10.1186/s12943-019-1000-8
- Zhu, X.-L., Li, T., Cao, Y., Yao, Q.-P., Liu, X., Li, Y., et al. (2021). tRNA-derived Fragments tRFGlnCTG Induced by Arterial Injury Promote Vascular Smooth Muscle Cell Proliferation. *Mol. Ther. - Nucleic Acids* 23, 603–613. doi:10.1016/j.omtn.2020.12.010

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhou, Cui and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.