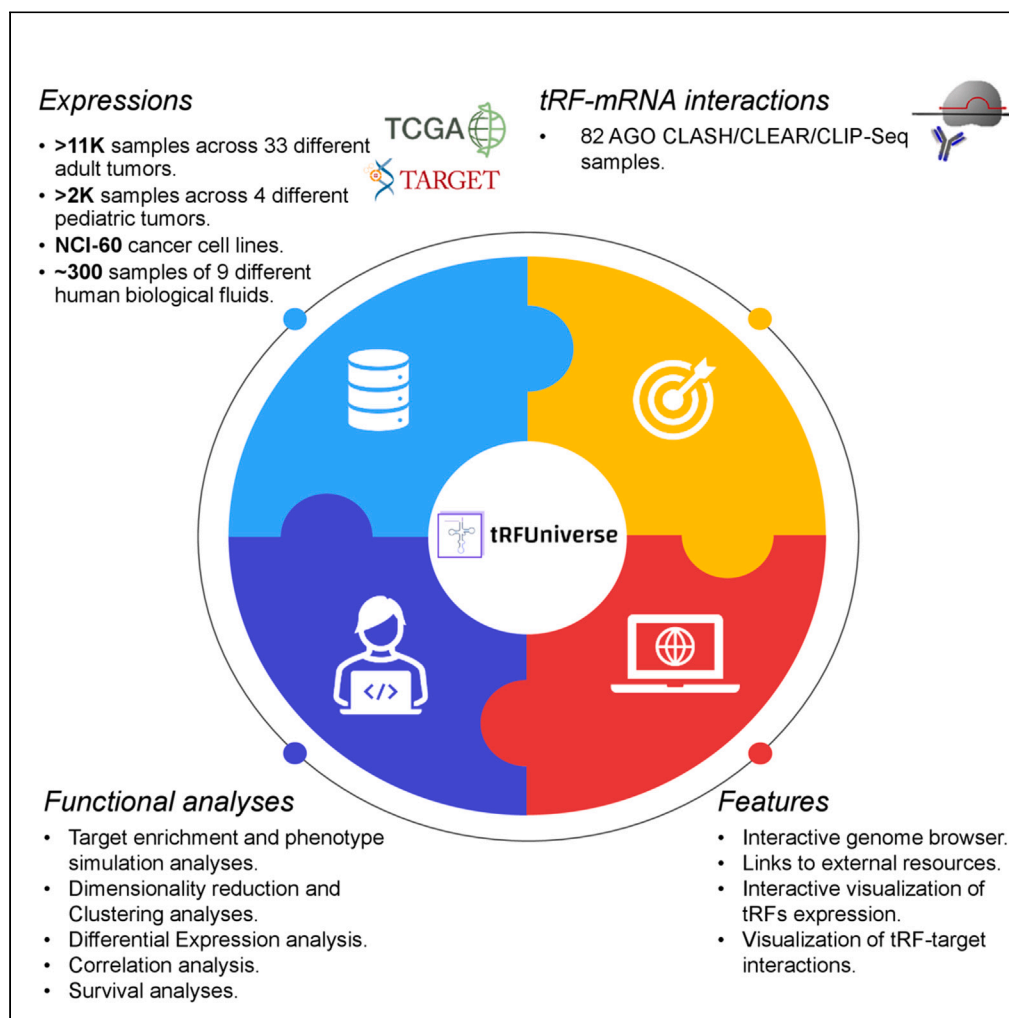


Article

tRFUniverse: A comprehensive resource for the interactive analyses of tRNA-derived ncRNAs in human cancer



Alessandro La Ferlita, Salvatore Alaimo, Giovanni Nigita, ..., Alfredo Ferro, Carlo M. Croce, Alfredo Pulvirenti

alfredo.pulvirenti@unict.it

Highlights

tRNA-derived ncRNAs are active regulators of gene expression involved in cancer

tRFUniverse is a novel online app for analyzing tRNA-derived ncRNAs in cancer

Analyses available in TCGA/TARGET cohorts, NCI60 cell lines, and biological fluids

Targets of tRNA-derived ncRNAs identified in public AGO CLASH/CLIP-Seq data

La Ferlita et al., iScience 27, 108810
February 16, 2024 © 2024 The Author(s).
<https://doi.org/10.1016/j.isci.2024.108810>

Article

tRFUniverse: A comprehensive resource for the interactive analyses of tRNA-derived ncRNAs in human cancer

Alessandro La Ferlita,^{1,4} Salvatore Alaimo,^{2,4} Giovanni Nigita,¹ Rosario Distefano,¹ Joal D. Beane,³ Philip N. Tschlis,¹ Alfredo Ferro,² Carlo M. Croce,¹ and Alfredo Pulvirenti^{2,5,*}

SUMMARY

tRNA-derived ncRNAs are a heterogeneous class of non-coding RNAs recently proposed to be active regulators of gene expression and be involved in many diseases, including cancer. Consequently, several online resources on tRNA-derived ncRNAs have been released. Although interesting, such resources present only basic features and do not adequately exploit the wealth of knowledge available about tRNA-derived ncRNAs. Therefore, we introduce tRFUniverse, a novel online resource for the analysis of tRNA-derived ncRNAs in human cancer. tRFUniverse presents an extensive collection of classes of tRNA-derived ncRNAs analyzed across all the TCGA and TARGET tumor cohorts, NCI-60 cell lines, and biological fluids. Moreover, public AGO CLASH/CLIP-Seq data were analyzed to identify the molecular interactions between tRNA-derived ncRNAs and other transcripts. Importantly, tRFUniverse combines in a single resource a comprehensive set of features that we believe may be helpful to investigate the involvement of tRNA-derived ncRNAs in cancer biology.

INTRODUCTION

With the advent of Next Generation Sequencing (NGS) technologies, the number of identified non-coding RNAs (ncRNAs) classes in eukaryotic cells has dramatically increased.^{1–3} Recently, tRNA-derived ncRNAs, a heterogeneous group of ncRNAs originating from tRNA processing, have been characterized.^{4–6} tRNA biogenesis begins with the transcription of tRNA genes by RNA polymerase III, leading to precursor tRNA (pre-tRNA). Such molecules undergo a maturation process inside the nucleus where 5' leader and 3' trailer sequences are cleaved by ribonuclease P (RNase P) and ribonuclease Z (RNase Z), respectively.^{7–14} Several kinds of tRNA-derived ncRNAs have been discovered in the last few years.^{14,15} However, a universally accepted nomenclature is still missing. A typical grouping of such molecules is based on the location of cleavage sites within the tRNA. tRNA-derived ncRNAs can therefore be divided in: (i) tRNA-derived fragments (tRFs), which derive from the cleavage of either the D- or T-loop of the mature tRNA (tRF-5 and tRF-3, respectively)^{6,11}; (ii) tRNA-halves which are generated by specific cleavage in the anticodon region of the mature tRNA¹⁶ (tRNA-halves can also be produced by an RNase operated cleavage in the anticodon region that occurs under stress condition and, therefore, named stress-induced tRNA fragments (tiRNAs).¹⁷ Note, the term tiRNA was first coined by¹⁸); (iii) tRNA-derived small RNAs (tsRNAs), which derive from the 3' trailer sequence of the pre-tRNA¹⁹ (also named tRF-111); and (iv) 5' leader tRF which are generated from the 5' leader sequence of the pre-tRNA.²⁰ More details about tRNA-derived ncRNA classification and nomenclature can be found in.^{21,22} Importantly, it has been shown that tRNA-derived ncRNAs are not mere byproducts of random tRNA cleavage as reported in.^{11,23} Rather they may actively play roles in several biological phenomena, such as ribosome biogenesis, retrotransposition, virus infections, apoptosis, and cancer pathogenesis.^{6,19,24–33} Furthermore, some tRNA-derived ncRNA classes have been shown to bind PIWI and AGO proteins, potentially acting as pre- or post-transcriptional regulators of gene expression, respectively.^{19,34} Accumulating evidence also indicates the presence of functional tRNA-derived ncRNAs in human biological fluids, such as urine and serum, from patients with cancer.^{7,35–39}

Because of the scientific community's increasing interest in studying the roles of tRNA-derived ncRNAs in cancer biology, several online resources have been released in the last few years. The first resource was a database released in 2014 by Kumar P. et al., named *tRFdb*.⁴⁰ After that, other databases were released, such as *tRF2Cancer*⁴¹ and *MINTbase*.⁴² These databases were just catalogs of tRNA-derived ncRNAs expressed in several tumor types. Therefore, they gave users the opportunity to navigate among these RNA molecules and their features but they did not give them the tools to perform functional analyses.

¹Department of Cancer Biology and Genetics, The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA

²Department of Clinical and Experimental Medicine, Knowmics Lab, University of Catania, Catania, Italy

³Department of Surgery, Division of Surgical Oncology, The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA

⁴These authors contributed equally

⁵Lead contact

*Correspondence: alfredo.pulvirenti@unict.it

<https://doi.org/10.1016/j.isci.2024.108810>



To address the problem, we released in 2019 *tRFExplorer*,²¹ the first online resource that allowed users to search for tRNA-derived ncRNAs and visualize their expression profiles in NCI-60 cell lines and The Cancer Genome Atlas (TCGA) patient cohorts.⁴³ Moreover, through *tRFExplorer*, users could perform differential expression analyses on TCGA samples, correlating tRNA-derived ncRNA expression with several covariates, such as NCI-60 expression data and drug sensitivity, and mRNA/miRNA expression across the TCGA cohorts.²¹ Since then, several newer online resources on tRNA-derived ncRNAs in cancer have been released. Noteworthy examples include *OncotRF*,⁴⁴ *tsRBase*,⁴⁵ *tRFtarget*,⁴⁶ *tRFTar*,⁴⁷ *tRFTars*,⁴⁸ and *tsRFun*⁴⁹ (which is the upgrade of *tRF2cancer*⁴¹). These resources presented some interesting features, such as reporting molecular interactions between tRNA-derived ncRNAs and mRNAs, and performing target enrichment, differential expression, and survival analyses. However, such web applications continue to exhibit limitations: (i) there are missing tRNA-derived ncRNA classes; (ii) the functional analysis tools they provide are limited; and (iii) they fail to provide a single integrated portal that includes most of the features required for a comprehensive analysis.

To address the preceding issues, we developed *tRFUniverse* (<https://trfuniverse.cloud/>), a novel web application for the interactive analysis of tRNA-derived ncRNAs in human cancer. *tRFUniverse* presents the most extensive collection of different classes of tRNA-derived ncRNAs analyzed across all the small RNA sequencing (smRNA-Seq) data available for the TCGA cohorts, NCI-60 cell lines, human biological fluids, and, for the first time, also across all the pediatric tumors available on TARGET. Moreover, several AGO CLASH/CLEAR/CLIP-Seq data have been analyzed to identify the molecular interactions between tRNA-derived ncRNAs and other transcripts. Analysis of such data is allowed to users through a web-based and easy-to-use Graphical User Interface (GUI), that consents (i) to explore the expression of tRNA-derived ncRNAs across all TCGA, TARGET, NCI-60, and biological fluids data; (ii) to perform differential expression analyses on the TCGA and TARGET patient cohorts; (iii) to perform correlation analyses between tRNA-derived ncRNAs and mRNA/miRNA expression in TCGA and TARGET samples; (iv) to navigate through the tRNA-derived ncRNAs-mRNA interactions identified in the AGO CLASH/CLEAR/CLIP-Seq data; (v) to perform pathway enrichment analysis with genes which are either correlated or targeted by tRNA-derived ncRNAs; (vi) to perform survival analyses based on tRNA-derived ncRNA expression in TCGA and TARGET samples as well as many others functional analyses aiming to help researchers to investigate the roles of these small RNA molecules in cancer biology.

RESULTS

Data included in *tRFUniverse*

tRFUniverse is a publicly available online resource developed to provide users with an easy-to-use online resource capable of: (i) enabling the interactive exploration of tRNA-derived ncRNA expression in many different types of human cancer; (ii) and performing several analyses helpful to investigate the biological functions of these small RNA molecules in each analyzed tumor.

In order to achieve our goal, we first analyzed the expression of the tRNA-derived ncRNAs in all the smRNA-Seq data available in TCGA⁴³ and TARGET (<http://ocg.cancer.gov/programs/target>), which accounts for more than 13000 samples spanning across 33 and 4 different tumor types for adult and pediatric tumors, respectively. The number of samples analyzed for each TCGA\TARGET cohort is reported in [Table S1](#). In addition to the TCGA\TARGET samples, we (as in the first version of our resource) provide information about tRNA-derived ncRNA expression in commonly used human cancer cell lines. For this reason, we also analyzed the publicly available smRNA-Seq data of the NCI-60 cell lines, a panel of 60 distinct cancer cell lines spanning nine different types of human tumors ([Table S2](#)). As a result, we identified 44,456 tRNA-derived ncRNAs expressed in TCGA\TARGET cohorts and NCI60 cell lines, and, therefore, we included them in *tRFUniverse*. The numbers of tRNA-derived ncRNAs (for each subclass) expressed in the analyzed datasets are reported in [Table S3](#). Moreover, to implement in *tRFUniverse* the functional analyses described in the next section, clinical information together with gene and microRNA (miRNAs) expression matrices (raw counts) for all the TCGA\TARGET and NCI-60 samples were downloaded from the GDC data portal.

The core of *tRFUniverse* consists of the tRNA-derived ncRNAs expressed in TCGA\TARGET and NCI-60. To extend the current knowledge about such molecules, we also analyzed smRNA-Seq data retrieved from several human biological fluids and AGO CLASH/CLEAR/CLIP-Seq data. This allowed us to both establish their presence in human biological fluids and identify their molecular target. In more detail, 293 samples from several human fluids in physiological conditions, such as seminal fluid, saliva, serum, urine, plasma, stool, ovarian follicular fluid, cerebrospinal fluid, and bile, were analyzed. Concerning the AGO CLASH/CLEAR/CLIP-Seq data, 82 samples were analyzed, providing direct evidence of tRNA-derived ncRNAs-mRNAs interactions in several different biological samples. The lists of analyzed publicly available smRNA-Seq data of biological fluids and AGO CLASH/CLEAR/CLIP-Seq data are reported in [Tables S4](#) and [S5](#), respectively. A schematic representation of the *tRFUniverse* data is shown in [Figure 1](#).

Features available within *tRFUniverse*

tRFUniverse presents several features, and analysis modules developed to provide researchers with information that might help investigate the role of tRNA-derived ncRNAs in cancer biology ([Figure 2](#)). Before describing all the functional analyses implemented in *tRFUniverse*, we first briefly report the crucial features in exploring the tRNA-derived ncRNAs reported in our web application. *tRFUniverse* presents a very intuitive GUI that facilitates the exploration of all the tRNA-derived ncRNAs expressed in the TCGA\TARGET cohorts and NCI-60 cell lines. Users can filter the tRNA-derived ncRNAs from the “search” tab based on their average expression in a specific cohort, the fragment type, the chromosome where they are transcribed, and the anticodon and amino acid of the tRNA from where they are processed. In addition, all the tRNA-derived ncRNAs available in *tRFUniverse* can be explored from the “Browse” tab and filtered accordingly with their name, sequence length, and type. Either way, by selecting a specific tRNA-derived ncRNA, it is possible to navigate among its information using the several features implemented in *tRFUniverse* as described later in [discussion](#).

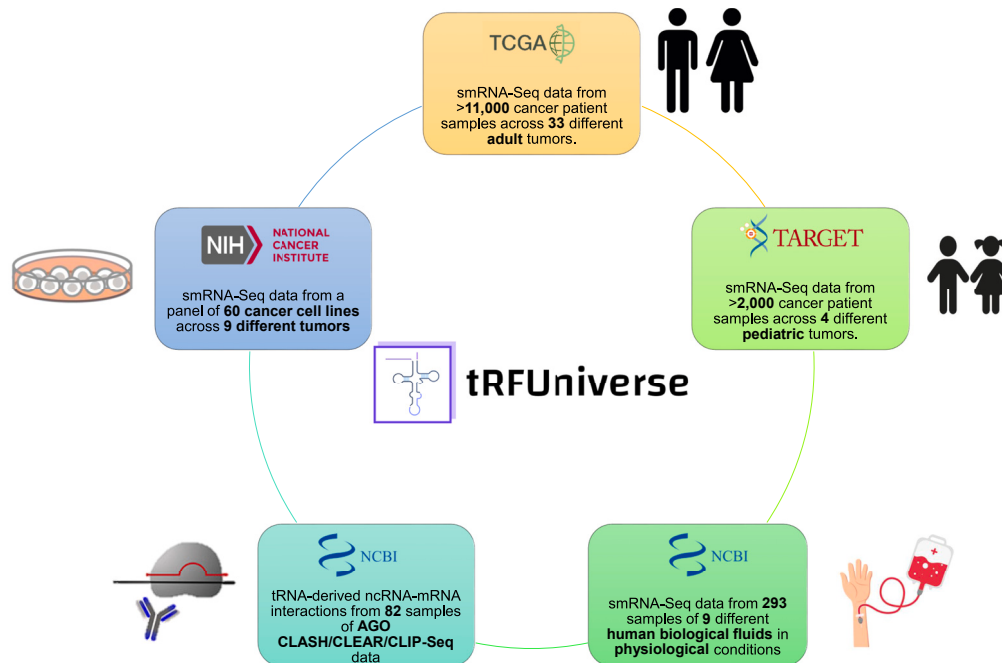


Figure 1. Analyzed data

Schematic representation of the data that has been analyzed for tRFUniverse.

Genome browser

An interactive genome browser is available to visualize both the genomic location where that specific tRNA-derived ncRNA is transcribed together with its sequence. In addition, information such as chromosome, coordinates, strand, amino acid, and anticodon of the tRNA of origin are reported.

Alias

A unique nomenclature for tRNA-derived ncRNAs is still missing. Commonly used IDs for tRNA-derived ncRNAs are reported on tRFdb⁴⁰ and MINTbase.⁴² However, other online databases assign different IDs to these short RNA molecules.^{44,45,49} To avoid confusion, tRFUniverse reports all the aliases of each tRNA-derived ncRNA, linking the external resources.^{44,45,49}

Expression

For each tRNA-derived ncRNA available in tRFUniverse, it is possible to visualize its expression in all the TCGA\TARGET cohorts, NCI-60 cell lines, and biological fluids. This plot can be generated by selecting (i) a cohort/cell line/fluid; (ii) the associated metadata (e.g., sex, stage, survival, type of sample, histological subtypes, molecular subtype); and (iii) how the expression should be presented among some options such as raw counts, Read Per Million (RPM), or normalized counts (with Limma's voom function). Normalized counts can also be corrected for batch effect by selecting among several covariates, such as sex, age, and tumor purity. This latter was calculated as in.⁵⁰ The expression is finally shown as an interactive boxplot that can also be downloaded for the user's convenience.

Targets

A list of mRNA targets is reported for each tRNA-derived ncRNA available in tRFUniverse. Indeed, as it is known from the literature, tRNA-derived ncRNAs, among several other functions, are known to be complexed with the AGO proteins acting, therefore, as post-transcriptional regulators with a miRNA-like manner.^{34,51,52} For this reason, we decided to analyze publicly available AGO CLASH/CLEAR/CLIP-Seq data with the scope to identify the molecular targets of the tRNA-derived ncRNAs identified as expressed in TCGA\TARGET cohorts and NCI-60 cell lines. The analysis of such data revealed that not only the tRFs generated from the mature tRNA were incorporated in the AGO complex but also tsRNAs and 5' leader tRFs were detected. For each tRNA-derived ncRNA, targets are reported in an interactive table that can be filtered by searching for a specific gene using either the gene symbol or ENSEMBL ID and sorted using the number of binding sites present in that specific gene or the Minimum Free Energy (MFE). By selecting a specific target, it is also possible to visualize the transcript IDs, the position of the binding sites (5' UTR, CDS, 3' UTR), the number of evidence, the start and end of the binding site, and the MFE. Moreover, information about the dataset, type of experimental data (CLASH, CLEAR, or CLIP), AGO protein, and sequence of that specific interaction is also reported.

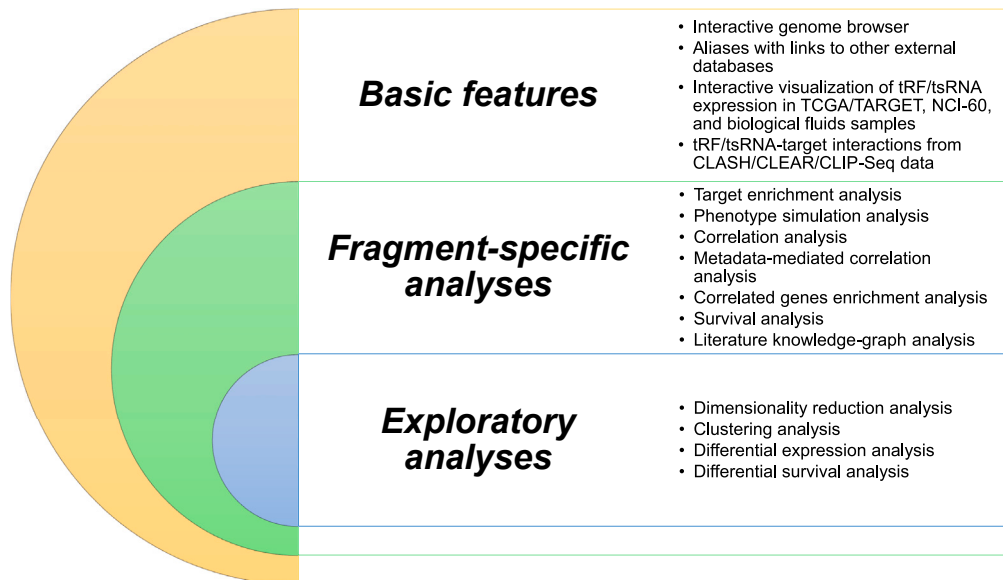


Figure 2. Features and analyses

Classification of all the features and analyses implemented in tRFUniverse.

Analyses of tRFUniverse

In addition to the basic features described in the previous section, several functional analyses (Figure 2) have been implemented to provide users with a tool that might help them investigate tRNA-derived ncRNAs in several human cancers. With this aim, we implemented two major classes of analyses: (i) fragment-specific analyses and (ii) exploratory analyses (Figure 2). The first class of analyses is oriented toward investigating the potential biological functions of a specific tRNA-derived ncRNA (also called “fragment”) in the context of specific or multiple tumor types. In contrast, the second class of analyses is tumor-centric. It aims to identify all the tRNA-derived ncRNAs potentially involved in the development or progression of that tumor type.

Concerning the fragment-specific analyses, these include (i) target enrichment analysis; (ii) phenotype simulation analysis; (iii) correlation analysis; (iv) metadata-mediated correlation analysis; (v) correlated gene enrichment analysis; (vi) survival analysis; and (vii) literature knowledge-graph analysis. These analyses can be run after selecting a specific fragment using either the “search” or “browse” tab from the homepage of tRFUniverse and clicking on the desired analysis as described later in discussion.

Target enrichment analysis

Using the targets identified in the AGO CLASH/CLEAR/CLIP-Seq data, it is possible to perform target enrichment analysis to identify which pathways are enriched when selecting a specific tRNA-derived ncRNA. This analysis may indicate which biological pathways might be regulated by that specific tRNA-derived ncRNA. To run the analysis, users have to select which type of data (CLASH, CLEAR, CLIP, or a combination of them) must be used to get the targets of that specific tRNA-derived ncRNA. Moreover, users can optionally filter the potential targets by selecting only those “active” in a specific dataset. In this case, for active targets, we mean genes potentially regulated by that specific tRNA-derived ncRNA. The activity is calculated by performing a differential expression analysis between samples with high fragment expression (>75 percentile) and samples with low fragment expression (<25 percentile). Then, a p value threshold (defined by the user) is applied to the differential expression analysis results. All targets with a p value below the threshold are considered active in the dataset. This optional filtering is important to select only those targets expressed and potentially regulated in a specific TCGA\TARGET cohort. The results are finally presented as an interactive table that lists all the enriched pathways, with their IDs, descriptions, target genes, and the number of targeted genes for each pathway. The results can also be downloaded as a CSV, Excel, or JSON file.

Phenotype simulation analysis

In addition to the target enrichment analysis, a more complex way to investigate which pathways may be regulated by the user-selected tRNA-derived ncRNA is through a phenotype simulation analysis. In more detail, in tRFUniverse, we have included PHENSIM,^{53,54} a phenotype simulator that relies on the target interactions identified from the AGO CLASH/CLEAR/CLIP-Seq data and the topological information encoded on the KEGG⁵⁵⁻⁵⁷ and REACTOME⁵⁸⁻⁶¹ pathways to identify the metabolic and signaling pathways potentially regulated by a specific tRNA-derived ncRNA and, in turn, its effects on the cellular phenotype. This is important since both tRFs generated from the mature tRNA^{34,51} and the 5' leader and 3' trailer tRFs generated from the precursor of the tRNA¹⁹ have been shown to be bound to AGO proteins, functioning as active regulators of gene expression. However, users should consider that these simulations are limited by data availability. Indeed, since

different tRNA-derived ncRNAs are produced in different tissues,⁶² and we have a limited set of AGO CLASH/CLEAR/CLIP-Seq data in limited experimental conditions, some important biological targets might be missing from the data. However, this is not an inherent limitation of the simulation model that can be expanded as soon as more data are available.

To perform such analysis, as in the previous case, users have to select which type of data (CLASH, CLEAR, CLIP, or a combination of them) has to be used to get the targets of that specific tRNA-derived ncRNA, and optionally, filter the targets by selecting only those that are potentially regulating those targets in a specific dataset as previously described for the target enrichment analysis. Once the analysis is completed, the results are shown in an interactive table that reports all the KEGG and REACTOME pathways found to be regulated by the user-selected tRNA-derived ncRNA with their IDs, activity scores, p values and false discovery rate (FDR). By clicking on a specific pathway, another interactive table is shown, listing all the potentially affected genes included in that pathway. All the results tables can also be downloaded as a CSV, Excel, or JSON file.

Correlation analysis

Looking at the potential direct targets of a specific tRNA-derived ncRNA is not the only way to infer its biological function. Indeed, tRNA-derived ncRNAs may regulate gene expression with various mechanisms, mostly unknown, that go beyond the simple direct targeting of specific mRNAs. Therefore, a more agnostic approach, like a simple correlation analysis, may highlight potentially relevant direct or indirect connections between tRNA-derived ncRNA and gene expression that otherwise would not be possible to detect. For this reason, in *tRFUniverse*, we allow users to perform correlation analyses between tRNA-derived ncRNA and gene/miRNA expression. For performing such analysis, users have to select (i) a TCGA\TARGET cohort or NCI-60 cell line; (ii) the correlation measure (Pearson, Spearman, or Kendall); and (iii) if genes or miRNAs must be correlated with the expression of the selected tRNA-derived ncRNA. In addition to these required parameters, others are optional. These include (i) covariates for batch correction (age, sex, sample type, race, tumor mutational burden, subtypes, tumor purity, and others); (ii) samples' histological or molecular subtypes filter; and (iii) samples' type filter. The results are presented as an interactive table that shows all the correlated genes using their gene symbols and ENSEMBL IDs, the correlation coefficient value, p value, and adjusted p value. Moreover, a downloadable and interactive scatterplot is also shown by clicking on a specific correlated gene. The full results table can also be downloaded as a CSV, Excel, or JSON file.

Metadata-mediated correlation analysis

Even though correlation analyses might help identify direct or indirect connections between tRNA-derived ncRNA expression and regulation of specific genes, the considerable diversity of tumor samples, even within the same molecular or histological subtype, sometimes makes it hard to identify such connections. For this reason, on some occasions, it is essential to look at correlations, considering at the same time additional clinical or molecular variables about the selected subset of samples.⁶³ Such information includes sex, sample type, sample subtype, survival status, potential treatments, race, tumor purity, tumor mutational burden, and so on. Such analysis may identify correlations that otherwise may not be statistically significant.⁶³ For this reason, in *tRFUniverse*, we have implemented the opportunity to perform metadata-mediated correlation analysis. To run the analysis, users must select (i) a TCGA\TARGET cohort or NCI-60 cell line; (ii) the metadata to be used for the correlation correction; and (iii) if genes or miRNAs must be correlated with the expression of the selected tRNA-derived ncRNA. In addition to these mandatory parameters, users might also select (i) the covariates for batch correction (age, sex, sample type, race, tumor mutational burden, subtypes, tumor purity, and others); (ii) the samples' histological or molecular subtypes to be used for the analysis; and also (iii) the samples' type. The results of such analysis are reported as an interactive table that includes the list of correlated genes with their symbols and ENSEMBL IDs, p values, and adjusted p values. The table can also be sorted using the previously mentioned information and downloaded as a CSV, Excel, or JSON file. Moreover, by clicking on a specific correlated gene, an interactive and downloadable scatterplot showing the correlation between the selected tRNA-derived ncRNA and the gene of interest is also shown.

Correlated gene enrichment analysis

As we previously said, identifying correlated genes is another way to look at the possible biological functions of a specific tRNA-derived ncRNA. In fact, the expression of correlated genes may be directly or indirectly affected by the expression of tRNA-derived ncRNAs. At that point, it may be helpful to know which biological pathways are enriched when selecting all the genes correlated with a specific tRNA-derived ncRNA. For this reason, in *tRFUniverse*, we allow users to perform pathway enrichment analysis using as input the list of genes correlated with the selected tRNA-derived ncRNA. In more detail, users must select (i) the dataset to be analyzed; (ii) the correlation measure (Pearson, Spearman, or Kendall); (iii) the minimum absolute correlation threshold; and (iv) if genes or miRNAs must be correlated with the expression of the selected tRNA-derived ncRNA. In addition to these mandatory parameters, others are optional. These include (i) covariates for batch correction (age, sex, sample type, race, tumor mutational burden, subtypes, tumor purity, and others); (ii) filter for samples' histological or molecular subtypes; and (iii) filter for samples' type. After running the analysis, the results are reported as an interactive table that lists all the enriched pathways with their IDs and descriptions, the correlated genes included in each enriched pathway, the gene ratio, p value, and adjusted p value. As for all the analyses implemented in *tRFUniverse*, the results table can be downloaded as a CSV, Excel, or JSON file.

Survival analysis

Another crucial biological aspect to look at in a specific tRNA-derived ncRNA is its potential association with the differences in survival in a given tumor type. Indeed, this may indicate if that fragment may be associated with the development and progression of a specific tumor. For this reason, in *tRFUniverse*, it is possible to perform survival analyses in all TCGA and TARGET cohorts after selecting a tRNA-derived ncRNA.

In more detail, to run the analysis, users must select (i) the dataset (TCGA\TARGET cohorts); (ii) the survival measure (overall survival, disease-free survival, disease-specific survival, or progression-free survival); (iii) the used expression values (counts, RPM, or normalized counts); and (iv) the cutoff to define a sample with a low or high fragment's expression (median, quartile, or custom). In addition to these required parameters, others are optional. These optional parameters allow users to filter the samples to be analyzed based on their (i) histological or molecular subtypes; (ii) types (e.g., primary tumor, metastasis); and (iii) survival range. The results are shown as an interactive and downloadable Kaplan-Meier curve showing the statistical significance (p value) and Hazard Ratio (HR) value.

Literature knowledge-graph analysis

As a final fragment-specific analysis, in *tRFUniverse*, we allow users to build a knowledge graph based on the information available about the selected tRNA-derived ncRNA from the literature. Such a function is powered by NetME,^{64,65} a novel text-mining software that, starting from a set of full texts obtained from PubMed, can extract biological elements from ontological databases and then synthesize a network inferring relations among such elements. In this case, *tRFUniverse* creates a query to NetME using as a biological element the tRNA-derived ncRNA selected from the user with all its potential aliases reported in other databases. At that point, if that tRNA-derived ncRNA has been reported in some research article, NetME will be able to build the fragment-specific knowledge-based graph that may be used to explore its biological functions or association with certain diseases. However, since many tRNA-derived ncRNAs available in *tRFUniverse* have never been studied and, therefore, never mentioned in the scientific literature, for them, such a knowledge-based graph cannot be generated. More details about NetME can be found on its website (<https://netme.click/#/>) and publication.⁶⁴

Concerning the exploratory functional analyses, we implemented (i) dimensionality reduction analysis; (ii) clustering analysis; (iii) differential expression analysis; and (iv) differential survival analysis. These analyses can be run by clicking the "Analysis" tab from the homepage and selecting the desired analysis. After that, a specific web page will open to request that the user select all the required parameters to run the analysis correctly. In what follows, we briefly discuss the exploratory functional analyses implemented in *tRFUniverse*.

Dimensionality reduction

In *tRFUniverse*, we allow users to perform dimensionality reduction analyses to see if it is possible to identify different clusters of samples within a user's selected subset in a two-dimensional space using the expression profile of the tRNA-derived ncRNAs. This type of analysis might be helpful to identify signatures of tRNA-derived ncRNAs that might be used to discriminate between different tumor subtypes or even define a new molecular sub-classification of a specific tumor. The analysis can be easily run by selecting (i) the datasets to be analyzed; (ii) the metadata that must be used to annotate the samples; (iii) the dimensionality reduction algorithm (PCA, MDS, ICA, tSNE, and UMAP); (iv) the covariates for batch correction (age, sex, sample type, race, tumor mutational burden, subtypes, tumor purity, and others); (v) the tumor histological or molecular subtypes; (vi) and the sample types. In addition to all the previously mentioned parameters, specific options are displayed according to the selected dimensionality reduction algorithm for selecting the tRNA-derived ncRNAs to be used for the analysis. The analysis results are finally shown as an interactive dot plot that can also be downloaded for the user's convenience.

Clustering

In addition to the previously described dimensionality reduction analysis, in *tRFUniverse*, we have implemented the opportunity to perform clustering analyses that are visualized in an interactive and customizable heatmap. As for the dimensionality reduction analysis, this type of analysis can be run to identify clusters of samples that show similar tRNA-derived ncRNA expression patterns. The analysis can be run by selecting (i) the dataset to be analyzed; (ii) the metadata that should be used for the samples' annotation; (iii) the covariates for the batch correction; (iv) the tumor subtypes; and (v) samples' type. In addition to these parameters, users must also define the coefficient (mean absolute deviation, variance, absolute median, and absolute mean) to be used to calculate the variability in tRNA-derived ncRNA expression within the selected dataset and the number of the most variable tRNA-derived ncRNAs that should be used for the clustering analysis. As previously said, the results are shown as an interactive and customizable heatmap that can also be downloaded in various formats (PDF, PNG, and SVG).

Differential expression analysis

Another useful analysis implemented in *tRFUniverse* is differential expression analysis. Such analysis is crucial to identify tRNA-derived ncRNAs that have an altered expression in a specific condition and that may be implicated with the development or progression of that malignancy. The differential expression analysis can be run from the "Analysis" tab present on the *tRFUniverse* homepage and selecting all the required parameters. Such parameters include (i) the dataset to be analyzed (TCGA\TARGET cohorts); (ii) the relative metadata; (iii) the covariates for batch correction; (iv) and the contrasts to be performed. In addition to that, users can also define the cutoff for the logFC and q-value for considering a specific tRNA-derived ncRNA as differentially expressed. Moreover, users can also filter lowly expressed tRNA-derived ncRNAs from the analysis by removing all the fragments that have a number of reads under a user's defined cutoff. The analysis results are reported as an interactive table, downloadable as a CSV file. In addition to the results table, an interactive and downloadable volcano plot is also shown.

Differential survival analysis

The last exploratory functional analysis implemented in *tRFUniverse* is the survival analysis. Such analysis allows users to identify all the tRNA-derived ncRNAs associated with statistically significant differences in survival in a specific tumor type. To run the analysis, users have to select

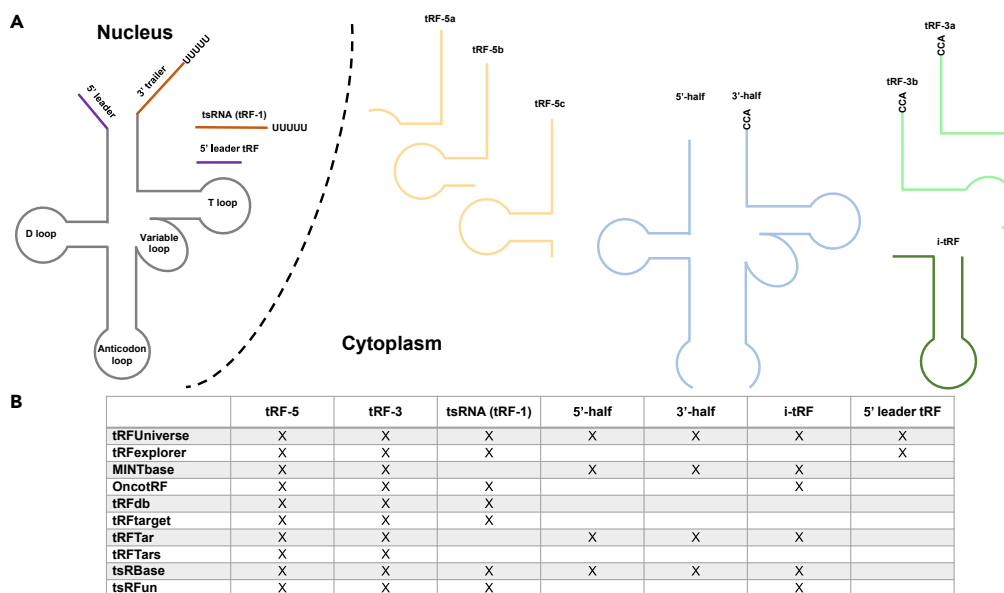


Figure 3. tRNA-derived ncRNA classes and comparison table

(A) Schematic representation of all the different classes of tRNA-derived ncRNAs.

(B) Table showing all the different classes of tRNA-derived ncRNAs reported in tRFUniverse, and in the other compared databases.

(i) the dataset (TCGA/TARGET cohorts); (ii) the survival measure (overall survival, disease-free survival, disease-specific survival, or progression-free survival); (iii) the type of expression value used for the analysis (counts, RPM, or normalized counts); (iv) and the cutoff to define a sample with a low or high fragment's expression (median, quartile, or custom). In addition to the previously required parameters, users can filter samples from the selected TCGA/TARGET cohort by selecting the tumor histological or molecular subtype and the sample type. Moreover, users can optionally select the time range for the survival analysis. As for the other analyses implemented in tRFUniverse, the results are reported as an interactive table that lists all the tRNA-derived ncRNAs that are associated or not with differences in survival with their relative statistics (HR, p value, and q-value). Such a table can also be downloaded as a CSV, Excel, or JSON file. Moreover, by clicking on a specific fragment, an interactive and downloadable Kaplan-Maier curve reporting HR and p value is shown.

Comparison with other online resources

In recent years, several other resources for tRNA-derived ncRNAs have been released, such as MINTbase,⁴² OncotRF,⁴⁴ tRFdb,⁴⁰ tRFtarget,⁴⁶ tRFtar,⁴⁷ tRFtars,⁴⁸ tsRBbase,⁴⁵ and tsRFun⁴⁹ (which is the upgraded version of tRF2cancer⁴¹). In order to show the potential applicability of tRFUniverse, we evaluated our resource with the previously mentioned databases. In more detail, a qualitative evaluation was performed by comparing all the previously mentioned databases in terms of (i) reported classes of tRNA-derived ncRNAs; and (ii) implemented features and analyses.

Concerning the tRNA-derived ncRNA classes available in the compared databases, tRFUniverse is the only one reporting all the different classes so far discovered (Figure 3A). Indeed, databases such as MINTbase and tRFtar report only tRNA-derived ncRNAs generated from the processing of mature tRNA (tRF-5, tRF-3, tiRNA-5, tiRNA-3, and i-tRF) others, such as OncotRF, tRFdb, tRFtarget, and tsRFun, present all tRFs and tsRNAs (tRF-1) but any tiRNAs, while tRFtars report only tRF-5s and tRF-3s (Figure 3B). On the contrary, tsRBbase reports an extensive repertoire of different classes of tRNA-derived ncRNAs with the exception of the 5' leader tRF, also called 5' leader-exon tRF,²⁰ which is uniquely included in tRFUniverse. Although both tRFUniverse and MINTbase report the tRFs from the mature tRNA detected in TCGA using MINTmap, discrepancies between these two databases were detected. The reason for these differences might be due to (i) the different versions of MINTmap used in tRFUniverse (MINTmap v.2) and MINTbase (MINTmap v.1), (ii) and different versions of small RNA-Seq data analyzed (in MINTbase were analyzed the small RNA-seq data of TCGA downloaded in 2015 while we used the 2021 release). All the different classes of tRNA-derived ncRNAs reported in tRFUniverse and the other databases are shown in Figure 3B.

In addition to having an extensive collection of tRNA-derived ncRNAs, one of the major strengths of tRFUniverse is its considerable number of implemented features and analyses. Indeed, tRFUniverse includes, in a single resource, most of the analyses implemented in the other resources so far (Table 1). Moreover, our web application presents several analyses, such as (i) phenotype simulation analysis; (ii) metadata-mediated correlation analyses; (iii) literature knowledge-graph analyses; (iv) dimensionality reduction; and (v) clustering analyses that are uniquely implemented in tRFUniverse. In Table 1, we summarized all the features and functional analyses implemented in tRFUniverse compared with the abovementioned databases.

Table 1. Comparison table of datasets, features, and analyses

	Dataset types	Interactive genome browser	Reporting aliases	Reporting targets	Target enrichment analysis	Phenotype simulation analysis	Correlation analysis	Metadata-mediated correlation analysis	Correlated gene enrichment analysis	Fragment-specific survival analysis	Literature knowledge-graph analysis	Dimensionality reduction analysis	Clustering analysis	Differential Expression Analysis	Differential survival analysis
tRFUniverse	TCGA TARGET NCI-60 Biological Fluids AGO CLASH - Seq AGO CLEAR- Seq AGO CLIP-Seq	X	X	X	X	X	X	X	X	X	X	X	X	X	X
tRFexplorer	TCGA NCI-60	X					X							X	
MINTbase	TCGA														
OncotRF	TCGA	X	X				X		X	X				X	X
tRFdb	Selected smRNA- Seq data from NCBI-SRA														
tRFtarget	–		X	X (predicted)											
tRFTar	AGO CLASH- seq AGO CLIP-Seq			X	X										
tRFTars	AGO CLASH- seq AGO CLEAR- CLIP			X											

(Continued on next page)

Table 1. Continued

	Dataset types	Interactive genome browser	Reporting aliases	Reporting targets	Target enrichment analysis	Phenotype simulation analysis	Correlation analysis	Metadata-mediated correlation analysis	Correlated gene enrichment analysis	Fragment-specific survival analysis	Literature knowledge-graph analysis	Dimensionality reduction analysis	Clustering analysis	Differential Expression Analysis	Differential survival analysis
tsRBase	Selected smRNA-Seq data from NCBI-SRA AGO CLASH-seq AGO CLEAR-CLIP			X											
tsRFun	TCGA AGO CLASH-seq AGO CLEAR-CLIP AGO CLIP-Seq		X	X					X				X		X

Table showing all the datasets, features, and analyses implemented in tRFUniverse and on the other compared databases.

Noteworthy, no other studies have ever analyzed tRNA-derived ncRNAs in pediatric tumors. Indeed, *tRFUniverse* is the first resource reporting the expression of tRNA-derived ncRNAs in all pediatric tumors available from TARGET cohorts (<http://ocg.cancer.gov/programs/target>). Furthermore, all the previously described functional analyses will allow users to investigate, for the first time, the involvement of tRNA-derived ncRNAs in the development of these rare pediatric malignancies.

DISCUSSION

tRNA-derived ncRNAs are recently attracting increasing attention from the scientific community because of their involvement in regulating gene expression through various mechanisms. Indeed, several classes of tRNA-derived ncRNAs have been found to be complex with the AGO proteins regulating gene expression at the post-transcriptional level through a miRNA-like mechanism.^{34,51,52,66,67} Moreover, several studies have linked the dysregulation of these small RNA molecules with the development and progression of several different types of human cancers, suggesting their potential clinical applications as novel diagnostic and prognostic biomarkers or as innovative molecular targets.^{68–71} However, more studies will be required to fully understand the biological functions of the tRNA-derived ncRNAs, especially in complex diseases such as cancer. In this context, bioinformatics and the increasing number of available databases and software for data analysis are providing new ways to study the biological functions of these fascinating regulatory small ncRNAs.

Since the release of the first database about tRNA-derived ncRNA in 2014,⁴⁰ several resources have been released. However, a few focused on human cancer, such as *MINTbase*,⁴² *OncotRF*,⁴⁴ and *tsRFun*.⁴⁹ On the other hand, other online databases were developed to primarily provide the direct targets of the tRNA-derived ncRNAs extracted by analyzing the publicly available AGO CLASH/CLEAR/CLIP-Seq data. Examples of such databases include *tRFtarget*,⁴⁶ *tRFTar*,⁴⁷ *tRFTargets*,⁴⁸ and *tsRBase*.⁴⁵ Although some of these previously mentioned resources present interesting features such as reporting molecular interactions between tRNA-derived ncRNAs and mRNAs, performing target enrichment analyses, differential expression analyses, and survival analyses, they still present some limitations that might affect their applicability in the context of cancer research. Such limitations include (i) missing tRNA-derived ncRNA classes; (ii) lacking some functional analyses; and (iii) the absence of a single integrated portal that includes most of the features already implemented in other resources in a single environment.

For this reason, we decided to implement a novel web application, named *tRFUniverse*, that includes in a single resource most of the features and analyses individually present in some of the pre-existing databases, plus other features and functional analyses that might be helpful to investigate the biological functions of the tRNA-derived ncRNAs in the context of human cancer. Moreover, *tRFUniverse* presents an easy-to-use GUI that allows users with no bioinformatics or computer programming expertise to run all the implemented analyses and easily interpret the results from interactive and downloadable tables and plots.

Noteworthy, in *tRFUniverse*, we reported, for the first time, the expression of tRNA-derived ncRNAs in all the pediatric tumor cohorts available from TARGET, providing, therefore, the first online resource capable of analyzing such molecules in these malignancies.

In conclusion, *tRFUniverse* is a user-friendly web-based application that allows users to perform several functional analyses based on tRNA-derived ncRNA expressions in TCGA, TARGET patient cohorts, NCI-60 cell lines, and human biological fluids to investigate their involvement in user-defined tumor types. To the best of our knowledge, no other online cancer-oriented resource presents, in a single environment, all the features and analyses implemented in *tRFUniverse*. Moreover, the completeness of the different classes of tRNA-derived ncRNAs reported, plus the massive amount of analyzed data, make *tRFUniverse* the most comprehensive online resource to explore tRNA-derived ncRNAs in human cancer. Our web application might be helpful to researchers in investigating the potential molecular functions of tRNA-derived ncRNAs and generating new hypotheses about their biology.

Limitations of the study

This study presents a novel online resource for analyzing tRNA-derived ncRNAs in human cancer called *tRFUniverse*. Despite the extensive analysis implemented, most of their functions and mechanisms remain elusive. Therefore, the features and analyses in *tRFUniverse* cover only some known aspects of their biology. However, *tRFUniverse* also has some limitations that need to be considered. For instance, some lowly expressed tRNA-derived ncRNAs in specific tumor types may not be detected and reported by the current version of *tRFUniverse*. Moreover, *tRFUniverse* does not account for the possible post-transcriptional modifications of tRNA-derived ncRNAs, which may affect their stability, localization, and function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Resource development
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Quantification of tRNA-derived ncRNAs from public smRNA-Seq data

- Analysis of CLASH/CLEAR-Seq data
- Analysis of CLIP-Seq data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.108810>.

ACKNOWLEDGMENTS

The authors thank the Cancer IT Operation Group of The Ohio State University and Mr. Thomas Moore for his valuable technical assistance. AP, SA, and AF have been partially supported by the following research projects: PO-FESR Sicilia 2014-2020 “DiOncoGen: Innovative diagnostics” (CUP G89J18000700007).

AUTHOR CONTRIBUTIONS

AP, SA, and ALF conceived the work. ALF performed the small RNA-Seq and AGO CLASH/CLEAR/CLIP-Seq analyses. GN supervised the small RNA-Seq and AGO CLASH/CLEAR/CLIP-Seq analyses. SA developed the web application. ALF, SA, and AP wrote the article. AP and CMC supervised and coordinated the project. ALF, GN, RD, PNT, and AF deeply tested the usability of the web application. All the authors read and revised the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 28, 2023

Revised: August 2, 2023

Accepted: January 2, 2024

Published: January 5, 2024

REFERENCES

1. Veneziano, D., Nigita, G., and Ferro, A. (2015). Computational Approaches for the Analysis of ncRNA through Deep Sequencing Techniques. *Front. Bioeng. Biotechnol.* 3, 77.
2. Salembhasha, A., and Mishra, S. (2018). Novel molecules lncRNAs, tRFs and circRNAs deciphered from next-generation sequencing/RNA sequencing: computational databases and tools. *Brief. Funct. Genomics* 17, 15–25.
3. Veneziano, D., Di Bella, S., Nigita, G., Laganà, A., Ferro, A., and Croce, C.M. (2016). Noncoding RNA: Current Deep Sequencing Data Analysis Approaches and Challenges. *Hum. Mutat.* 37, 1283–1298.
4. Martens-Uzunova, E.S., Olvedy, M., and Jenster, G. (2013). Beyond microRNA—novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett.* 340, 201–211.
5. Romano, G., Veneziano, D., Acunzo, M., and Croce, C.M. (2017). Small non-coding RNA and cancer. *Carcinogenesis* 38, 485–491.
6. Balatti, V., Pekarsky, Y., and Croce, C.M. (2017). Role of the tRNA-Derived Small RNAs in Cancer: New Potential Biomarkers and Target for Therapy. *Adv. Cancer Res.* 135, 173–187.
7. Li, S., Xu, Z., and Sheng, J. (2018). tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes* 9, 246.
8. Walczyk, D., Göbbringer, M., Rossmannith, W., Zatzepin, T.S., Oretskaya, T.S., and Hartmann, R.K. (2016). Analysis of the Cleavage Mechanism by Protein-Only RNase P Using Precursor tRNA Substrates with Modifications at the Cleavage Site. *J. Mol. Biol.* 428, 4917–4928.
9. Brillante, N., Göbbringer, M., Lindenhofer, D., Toth, U., Rossmannith, W., and Hartmann, R.K. (2016). Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.* 44, 2323–2336.
10. Simos, G., and Hurt, E. (1999). Transfer RNA biogenesis: A visa to leave the nucleus. *Curr. Biol.* 9, R238–R241.
11. 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, 2639–2649.
12. Leung, Y.Y., Kuksa, P.P., Amlie-Wolf, A., Valladares, O., Ungar, L.H., Kannan, S., Gregory, B.D., and Wang, L.-S. (2016). DASHR: database of small human noncoding RNAs. *Nucleic Acids Res.* 44, D216–D222.
13. Chung, I.-F., Chang, S.-J., Chen, C.-Y., Liu, S.-H., Li, C.-Y., Chan, C.-H., Shih, C.-C., and Cheng, W.-C. (2017). YM500v3: a database for small RNA sequencing in human cancer research. *Nucleic Acids Res.* 45, D925–D931.
14. Phizicky, E.M., and Hopper, A.K. (2023). The life and times of a tRNA. *RNA* 29, 898–957.
15. Pekarsky, Y., Balatti, V., and Croce, C.M. (2023). tRNA-derived fragments (tRFs) in cancer. *J. Cell Commun. Signal.* 17, 47–54.
16. Honda, S., Loher, P., Shigematsu, M., Palazzo, J.P., Suzuki, R., Imoto, I., Rigoutsos, I., and Kirino, Y. (2015). Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc. Natl. Acad. Sci. USA* 112, E3816–E3825.
17. Emar, M.M., Ivanov, P., Hickman, T., Dawra, N., Tisdale, S., Kedersha, N., Hu, G.-F., and Anderson, P. (2010). Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J. Biol. Chem.* 285, 10959–10968.
18. Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R.R., Grimmond, S.M., Schroder, K., et al. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* 41, 572–578.
19. Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L.Z., Pass, H.I., Kipps, T.J., Liu, C.-G., and Croce, C.M. (2016). Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. *Proc. Natl. Acad. Sci. USA* 113, 5071–5076.
20. Hanada, T., Weitzer, S., Mair, B., Bernreuther, C., Wainger, B.J., Ichida, J., Hanada, R., Orthofer, M., Cronin, S.J., Komnenovic, V., et al. (2013). CLP1 links tRNA metabolism to progressive motor-neuron loss. *Nature* 495, 474–480.
21. La Ferlita, A., Alaimo, S., Veneziano, D., Nigita, G., Balatti, V., Croce, C.M., Ferro, A., and Pulvirenti, A. (2019). Identification of tRNA-Derived ncRNAs in TCGA and NCI-60 Panel Cell Lines and Development of the Public Database tRFExplorer. *Database* 2019, baz115.
22. Filetti, V., La Ferlita, A., Di Maria, A., Cardile, V., Graziano, A.C.E., Rapisarda, V., Ledda, C., Pulvirenti, A., and Loreto, C. (2022). Dysregulation of microRNAs and tRNA-derived ncRNAs in mesothelial and mesothelioma cell lines after asbestiform fiber exposure. *Sci. Rep.* 12, 9181.
23. Telonis, A.G., Loher, P., Honda, S., Jing, Y., Palazzo, J., Kirino, Y., and Rigoutsos, I. (2015). Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. *Oncotarget* 6, 24797–24822.
24. Kim, H.K., Fuchs, G., Wang, S., Wei, W., Zhang, Y., Park, H., Roy-Chaudhuri, B., Li, P.,

- Xu, J., Chu, K., et al. (2017). A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature* 552, 57–62.
25. Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* 170, 61–71.e11.
 26. Ivanov, P. (2015). Emerging Roles of tRNA-derived Fragments in Viral Infections: The Case of Respiratory Syncytial Virus. *Mol. Ther.* 23, 1557–1558.
 27. Saikia, M., Jobava, R., Parisien, M., Putnam, A., Krokowski, D., Gao, X.-H., Guan, B.-J., Yuan, Y., Jankowsky, E., Feng, Z., et al. (2014). Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. *Mol. Cell Biol.* 34, 2450–2463.
 28. Balatti, V., Nigita, G., Veneziano, D., Drusco, A., Stein, G.S., Messier, T.L., Farina, N.H., Lian, J.B., Tomasello, L., Liu, C.-G., et al. (2017). tsRNA signatures in cancer. *Proc. Natl. Acad. Sci. USA* 114, 8071–8076.
 29. Slack, F.J. (2018). Tackling Tumors with Small RNAs Derived from Transfer RNA. *N. Engl. J. Med.* 378, 1842–1843.
 30. Huang, B., Yang, H., Cheng, X., Wang, D., Fu, S., Shen, W., Zhang, Q., Zhang, L., Xue, Z., Li, Y., et al. (2017). tRF/miR-1280 Suppresses Stem Cell-like Cells and Metastasis in Colorectal Cancer. *Cancer Res.* 77, 3194–3206.
 31. 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, 730–738.
 32. Magee, R., and Rigoutsos, I. (2020). On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Res.* 48, 9433–9448.
 33. Wilson, B., and Dutta, A. (2022). Function and Therapeutic Implications of tRNA Derived Small RNAs. *Front. Mol. Biosci.* 9, 888424.
 34. Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., and Dutta, A. (2018). tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* 24, 1093–1105.
 35. Zhao, C., Tolkach, Y., Schmidt, D., Kristiansen, G., Müller, S.C., and Ellinger, J. (2018). 5'-tRNA Halves are Dysregulated in Clear Cell Renal Cell Carcinoma. *J. Urol.* 199, 378–383.
 36. Yeri, A., Courtwright, A., Reiman, R., Carlson, E., Beecroft, T., Janss, A., Siniard, A., Richholt, R., Balak, C., Rozowsky, J., et al. (2017). Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of Healthy Subjects. *Sci. Rep.* 7, 44061.
 37. Dhahbi, J.M., Spindler, S.R., Atamna, H., Boffelli, D., and Martin, D.I.K. (2014). Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. *Biomarkers Cancer* 6, 37–47.
 38. Godoy, P.M., Bhakta, N.R., Barczak, A.J., Cakmak, H., Fisher, S., MacKenzie, T.C., Patel, T., Price, R.W., Smith, J.F., Woodruff, P.G., and Erle, D.J. (2018). Large Differences in Small RNA Composition Between Human Biofluids. *Cell Rep.* 25, 1346–1358.
 39. Srinivasan, S., Yeri, A., Cheah, P.S., Chung, A., Danielson, K., De Hoff, P., Filant, J., Laurent, C.D., Laurent, L.D., Magee, R., et al. (2019). Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. *Cell* 177, 446–462.e16.
 40. Kumar, P., Mudunuri, S.B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.* 43, D141–D145.
 41. Zheng, L.-L., Xu, W.-L., Liu, S., Sun, W.-J., Li, J.-H., Wu, J., Yang, J.-H., and Qu, L.-H. (2016). tRF2Cancer: A web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic Acids Res.* 44, W185–W193.
 42. Pliatsika, V., Loher, P., Magee, R., Telonis, A.G., Londin, E., Shigematsu, M., Kirino, Y., and Rigoutsos, I. (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, D152–D159.
 43. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.
 44. Yao, D., Sun, X., Zhou, L., Amanullah, M., Pan, X., Liu, Y., Liang, M., Liu, P., and Lu, Y. (2020). OncoRF: an online resource for exploration of tRNA-derived fragments in human cancers. *RNA Biol.* 17, 1081–1091.
 45. Zuo, Y., Zhu, L., Guo, Z., Liu, W., Zhang, J., Zeng, Z., Wu, Q., Cheng, J., Fu, X., Jin, Y., et al. (2021). tsRBbase: a comprehensive database for expression and function of tsRNAs in multiple species. *Nucleic Acids Res.* 49, D1038–D1045.
 46. Li, N., Shan, N., Lu, L., and Wang, Z. (2021). tRFtarget: a database for transfer RNA-derived fragment targets. *Nucleic Acids Res.* 49, D254–D260.
 47. 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.
 48. Xiao, Q., Gao, P., Huang, X., Chen, X., Chen, Q., Lv, X., Fu, Y., Song, Y., and Wang, Z. (2021). tRFtars: predicting the targets of tRNA-derived fragments. *J. Transl. Med.* 19, 88.
 49. Wang, J.-H., Chen, W.-X., Mei, S.-Q., Yang, Y.-D., Yang, J.-H., Qu, L.-H., and Zheng, L.-L. (2022). tsRFun: a comprehensive platform for decoding human tsRNA expression, functions and prognostic value by high-throughput small RNA-Seq and CLIP-Seq data. *Nucleic Acids Res.* 50, D421–D431.
 50. Distefano, R., Tomasello, L., Rampioni Vinciguerra, G.L., Gasparini, P., Xiang, Y., Bagnoli, M., Marceca, G.P., Fadda, P., Laganà, A., Acunzo, M., et al. (2022). Pan-Cancer Analysis of Canonical and Modified miRNAs Enhances the Resolution of the Functional miRNAome in Cancer. *Cancer Res.* 82, 3687–3700.
 51. 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.
 52. 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 β stimulated chondrocytes. *Osteoarthritis Cartilage* 28, 1102–1110.
 53. Alaimo, S., Rapicavoli, R.V., Marceca, G.P., La Ferlita, A., Serebrennikova, O.B., Tschlis, P.N., Mishra, B., Pulvirenti, A., and Ferro, A. (2021). PHENSIM: Phenotype Simulator. *PLoS Comput. Biol.* 17, e1009069.
 54. Maria, N.I., Rapicavoli, R.V., Alaimo, S., Bischof, E., Stasuzzo, A., Broek, J.A.C., Pulvirenti, A., Mishra, B., Duits, A.J., and Ferro, A.; RxCOVEA Framework (2023). Application of the PHENotype SIMulator for rapid identification of potential candidates in effective COVID-19 drug repurposing. *Heliyon* 9, e14115.
 55. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280.
 56. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114.
 57. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361.
 58. Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., et al. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 33, D428–D432.
 59. Sidiropoulos, K., Viteri, G., Sevilla, C., Jupe, S., Webber, M., Orlic-Milacic, M., Jassal, B., May, B., Shamovsky, V., Duenas, C., et al. (2017). Reactome enhanced pathway visualization. *Bioinformatics* 33, 3461–3467.
 60. Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., Wu, G., Stein, L., D'Eustachio, P., and Hermjakob, H. (2018). Reactome graph database: Efficient access to complex pathway data. *PLoS Comput. Biol.* 14, e1005968.
 61. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.
 62. Telonis, A.G., Loher, P., Magee, R., Pliatsika, V., Londin, E., Kirino, Y., and Rigoutsos, I. (2019). tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities. *Cancer Res.* 79, 3034–3049.
 63. Sciacca, E., Alaimo, S., Silluzzo, G., Ferro, A., Latora, V., Pitzalis, C., Pulvirenti, A., and Lewis, M.J. (2023). DEGGs: an R package with shiny app for the identification of differentially expressed gene-gene interactions in high-throughput sequencing data. *Bioinformatics* 39, btad192.
 64. Muscolino, A., Di Maria, A., Alaimo, S., Borzi, S., Ferragina, P., Ferro, A., and Pulvirenti, A. (2021). NETME: On-the-Fly Knowledge Network Construction from Biomedical Literature. In *Complex Networks & Their Applications IX*, R.M. Benito, C. Cherifi, H. Cherifi, E. Moro, L.M. Rocha, and M. Sales-Pardo, eds. (Springer International Publishing), pp. 386–397.
 65. Muscolino, A., Di Maria, A., Rapicavoli, R.V., Alaimo, S., Bellomo, L., Billeci, F., Borzi, S., Ferragina, P., Ferro, A., and Pulvirenti, A. (2022). NETME: on-the-fly knowledge network construction from biomedical literature. *Appl. Netw. Sci.* 7, 1.
 66. Guan, L., Lam, V., and Grigoriev, A. (2021). Large-Scale Computational Discovery of Binding Motifs in tRNA Fragments. *Front. Mol. Biosci.* 8, 647449.
 67. Xie, Y., Yao, L., Yu, X., Ruan, Y., Li, Z., and Guo, J. (2020). Action mechanisms and research

- methods of tRNA-derived small RNAs. *Signal Transduct. Target. Ther.* 5, 109.
68. Gupta, T., Malkin, M.G., and Huang, S. (2022). tRNA Function and Dysregulation in Cancer. *Front. Cell Dev. Biol.* 10, 886642.
 69. Yu, M., Lu, B., Zhang, J., Ding, J., Liu, P., and Lu, Y. (2020). tRNA-derived RNA fragments in cancer: current status and future perspectives. *J. Hematol. Oncol.* 13, 121.
 70. Huang, S.-Q., Sun, B., Xiong, Z.-P., Shu, Y., Zhou, H.-H., Zhang, W., Xiong, J., and Li, Q. (2018). The dysregulation of tRNAs and tRNA derivatives in cancer. *J. Exp. Clin. Cancer Res.* 37, 101.
 71. Veneziano, D., Tomasello, L., Balatti, V., Palamarchuk, A., Rassenti, L.Z., Kipps, T.J., Pekarsky, Y., and Croce, C.M. (2019). Dysregulation of different classes of tRNA fragments in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA* 116, 24252–24258.
 72. Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., and Staudt, L.M. (2016). Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* 375, 1109–1112.
 73. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079.
 74. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
 75. Jackman, J.E., and Phizicky, E.M. (2006). tRNAHis guanylyltransferase adds G-1 to the 5' end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases. *RNA* 12, 1007–1014.
 76. Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 44, D184–D189.
 77. Loher, P., Telonis, A.G., and Rigoutsos, I. (2017). MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Sci. Rep.* 7, 41184.
 78. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
 79. Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665.
 80. Stribling, D., Lei, Y., Guardia, C.M., Li, L., Fields, C.J., Nowialis, P., Opavsky, R., Renne, R., and Xie, M. (2021). A noncanonical microRNA derived from the snaR-A noncoding RNA targets a metastasis inhibitor. *RNA* 27, 694–709.
 81. Moore, M.J., Scheel, T.K.H., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M., and Darnell, R.B. (2015). miRNA–target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* 6, 8864.
 82. Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology* 1, 895–905.
 83. Wu, W.-S., Brown, J.S., Chen, P.-H., Shiue, S.-C., Lee, D.-E., and Lee, H.-C. (2022). CLASH Analyst: A Web Server to Identify In Vivo RNA–RNA Interactions from CLASH Data. *Noncoding. RNA* 8, 6.
 84. Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517.
 85. Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* 9, 263–293.
 86. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
 87. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
tRFUniverse (source code)	GitHub	https://github.com/knowmics-lab/tRFUniverse
TCGA (small RNA-Seq)	NCBI-dbGaP	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v11.p8
TARGET (small RNA-Seq)	NCBI-dbGaP	https://www.cancer.gov/ccg/research/genome-sequencing/target/using-target-data/citing
NCI-60 (PRJNA390643)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA390643
Biological fluids (PRJNA242348)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA242348
Biological fluids (PRJNA415462)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA415462
Biological fluids (PRJNA431173)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA431173
Biological fluids (PRJNA445720)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA445720
Biological fluids (PRJNA505788)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA505788
Biological fluids (PRJNA527257)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA527257
Biological fluids (PRJNA647356)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA647356
AGO CLASH/CLEAR-Seq data (PRJNA217521)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA217521
AGO CLASH/CLEAR-Seq data (PRJNA691465)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA691465
AGO CLASH/CLEAR-Seq data (PRJNA296130)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA296130
AGO CLIP-Seq data (PRJNA129395)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA129395
AGO CLIP-Seq data (PRJNA176418)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA176418
AGO CLIP-Seq data (PRJNA248264)	NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA248264
Software and algorithms		
Samtools (v.1.12)	Ubuntu package	https://manpages.ubuntu.com/manpages/jammy/man1/samtools.1.html
Trim Galore (v.0.6.6)	The Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bowtie 2 (v.2.4.5)	Bioconda	https://anaconda.org/bioconda/bowtie2
tRFExplorer (data)	GitHub	https://github.com/knowmics-lab/tRFExplorer
MINTmap (v.2)	Jefferson	https://cm.jefferson.edu/mintmap/
Bedtools (v.2.30.0)	Bioconda	https://anaconda.org/bioconda/bedtools
Flexbar (v.2.5)	Ubuntu package	https://manpages.ubuntu.com/manpages/jammy/man1/flexbar.1.html
CLASH Analyst	GitHub	https://github.com/t50504/CLASHAnalyst
RNAhybrid	Ubuntu package	https://manpages.ubuntu.com/manpages/trusty/man1/RNAhybrid.1.html
MACS2 (v.2.2.7.1)	PyPI	https://pypi.org/project/MACS2/
Laravel (v.9.19)	Composer	https://packagist.org/packages/laravel/framework
Next.js (v.12.2.5)	NPM	https://www.npmjs.com/package/next
JBrowse 2 (v.2.1.5)	NPM	https://www.npmjs.com/package/@jbrowse/react-linear-genome-view
R (v.4.2)	Ubuntu package	https://cloud.r-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, Alfredo Pulvirenti (alfredo.pulvirenti@unict.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data are publicly available at the accession numbers listed in the [key resources table](#) and on GitHub at <https://github.com/knowmics-lab/tRFUniverse>. tRFUniverse is available at <https://trfuniverse.cloud/>.
- The source code is available on GitHub at <https://github.com/knowmics-lab/tRFUniverse>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS

Resource development

The *tRFUniverse* web application has been developed using a microservice-based architecture for improved performance and availability (Figure S1). All services that make up the database are deployed in Docker containers. Container monitoring and orchestration are delegated to the Kubernetes platform.

The *tRFUniverse* web application comprises two primary services: a frontend service and a backend service. Both services employ a Kubernetes ingress service for external availability and a Kubernetes load balancer to dynamically spawn new on-demand instances based on the resource requested by the users.

The frontend service has been implemented as a stateless web app built with the Next.js framework to exploit its static site generation (SSG) and incremental static regeneration (ISR) capabilities. SSG is a technique that creates a snapshot of a dynamic web page at build time to remove the burden of computing the page for each user request. Dynamic content is automatically hydrated on request directly on the user's browser, reducing the load on the server. ISR is a dynamic extension of SSG where the snapshot of a web page is computed only on the first user request and cached for further access by other users. These two techniques allow us to dramatically reduce the computational resources needed to provide a good user experience.

The backend service has been implemented as a stateful API (Application Programming Interface) with the Laravel framework. The backend service manages all analyses requested by the user, input validation, and coordinates access to the database containing all *tRFUniverse* data stored within a MySQL database. Furthermore, the backend service uses a separate Job Service that is spawned on demand to perform the analyses a user requests. Currently, ten job services can be activated simultaneously, each managing up to three concurrent computations.

A Redis service, an in-memory key-value store database, is used for job queue management. When a user requests a new analysis, a record containing all analysis parameters is added to the Redis-based queue. Once a free job service instance is available, the first job is picked from the front of the queue and processed. Kubernetes monitors each instance continuously and spawns new ones when more resources are needed. The Redis service is also employed for caching and user session management.

Finally, communication between the frontend and the backend is achieved via REST (REpresentational State Transfer) HTTP requests and WebSocket-based real-time channels. Real-time channels are maintained by a Soketi service, a fast and resilient open-source WebSocket server.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of tRNA-derived ncRNAs from public smRNA-Seq data

smRNA-Seq data in BAM format from TCGA (<https://cancer.gov/tcga>) and TARGET (<http://ocg.cancer.gov/programs/target>) cohorts were downloaded using Genomic Data Commons (GDC) Data Transfer Tool⁷² after obtaining authorization from the data access committee (DBGap Project IDs: #11332 and #73394 for TCGA; #22219 for TARGET). Subsequently, the BAM files were converted in FASTQ format using Samtools (bam2fq) (v.1.12).⁷³ Instead, smRNA-Seq data for the NCI-60 (BioProject: PRJNA390643) and biological fluids (BioProject: PRJNA242348, PRJNA415462, PRJNA431173, PRJNA445720, PRJNA505788, PRJNA527257, PRJNA647356) samples were downloaded in FASTQ format from NCBI-SRA. All downloaded smRNA-Seq data in FASTQ format were quality trimmed (-q 20), and adapters were removed using Trim Galore (v.0.6.6) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were then mapped by Bowtie2 without allowing mismatches, insertions or deletions (-L 10 -N 0) (v.2.4.5)⁷⁴ on a custom transcriptome generated from the tsRNA (average length 23 nts) and 5' leader tRF (5' leader region of the pre-tRNA - 20 nts length) sequences retrieved from Pekarsky Y et al.,¹⁹ and *tRFexplorer*,²¹ respectively (tRFs with post-transcriptional additions such as that of the nuclear tRNA HisGTG⁷⁵ are not analyzed). In both cases, the tRNA sequences and their genomic coordinates used originally for making the tsRNA and 5' leader tRF annotation were downloaded from GtRNAdb.⁷⁶ This step allowed us to filter tsRNAs and 5' leader tRFs from tRFs and tiRNAs since MINTmap does not analyze tRNA-derived ncRNAs generated from the pre-tRNA. After that, the reads that were not mapped to tsRNAs and 5' leader tRFs (by Bowtie2)

were then extracted in FASTQ format using the Bowtie2's parameter “-un” and used as input for MINTmap (v.2)⁷⁷ in order to quantify tRFs and tiRNAs (exclusive counts). On the other hand, the reads that mapped on the tsRNA\5' leader tRF transcriptome (in SAM format) were extracted (samtools view -h -F 4), converted in BAM format (samtools view), sorted by coordinates (samtools sort), and indexed (samtools index) using Samtools (v.1.12).⁷³ Subsequently, tsRNAs and 5' leader tRFs were quantified using Bedtools (multicov) (v.2.30.0).⁷⁸ Finally, the raw count matrices generated by MINTmap for tRFs and tiRNAs and by Bedtools for tsRNAs and 5' leader tRFs were all merged to have a single matrix for each data project and included in *tRFUniverse*. A schematic representation of the analysis pipeline is shown in [Figure S2](#). Only the tRNA-derived ncRNAs with a length >14nt and an average expression >0.1 Read Per Million (RPM) for each cohort were selected and included in *tRFUniverse*. Note, the impact of such a threshold might have been different in TCGA/TARGET cohorts with few samples compared to other cohorts with more samples, potentially affecting the total number of expressed tRNA-derived ncRNAs per cohort.

Analysis of CLASH/CLEAR-Seq data

AGO CLASH/CLEAR-Seq data (BioProject: PRJNA217521, PRJNA691465, PRJNA296130)^{79–81} in FASTQ format were downloaded from NCBI-SRA. Low-quality reads and adapters were removed using Trim Galore (v.0.6.6) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), while potential 5' barcodes were removed using Flexbar (v.2.5).⁸² After that, the identification of the interactions between tRNA-derived ncRNAs and mRNAs was performed by the CLASH Analyst pipeline⁸³ using as input (i) the processed AGO CLASH/CLEAR-Seq data (in FASTQ format), (ii) the sequences of all the expressed tRNA-derived ncRNAs included in *tRFUniverse*, (iii) and the transcripts available on Gencode (v. 40) in FASTA format. The secondary structure of binding sites of the identified interactions were finally shown using RNAhybrid⁸⁴ as indicated in.⁷⁹

Analysis of CLIP-Seq data

AGO CLIP-Seq data (BioProject: PRJNA129395, PRJNA176418, PRJNA248264) in FASTQ format were retrieved from NCBI-SRA. Replicates were merged to improve sequencing depth and analyzed by adapting the protocol published by Moore et al.⁸⁵ In more detail, low-quality reads and adapters were removed using Trim Galore (v.0.6.6) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). After that, trimmed reads were mapped to the human genome (HG38) using Bowtie2 (v.2.4.5) (46), and then aligned reads in SAM format were converted to BAM, sorted and indexed using Samtools (v.1.12) (45). The peak calling step using MACS2 (v.2.2.7.1)⁸⁶ was successively performed to identify the gene regions immunoprecipitated with the AGO proteins potentially containing the binding sites of the tRNA-derived ncRNAs. With this purpose, all the identified peaks were then filtered using the GenomicRanges R package⁸⁷ in order to remove intergenic and intronic peaks and annotated following the genomic coordinates present in the GTF file downloaded from Gencode (v.40). The sequences of the filtered peaks were finally retrieved using the *getfasta* function of Bedtools (v.2.30.0).⁷⁸ Concerning the tRNA-derived ncRNAs, the previously described pipeline used to quantify the tRNA-derived ncRNAs from the smRNA-Seq data was also used on the AGO CLIP-Seq data to identify the immunoprecipitated tRNA-derived ncRNAs. After that, their full sequences were collected and formatted in FASTA. Finally, the FASTA files containing the binding regions of the targets and the sequences of the regulators (tRNA-derived ncRNAs) were used as input for RNAhybrid⁸⁴ to predict their interactions.