

tRFTars: predicting the targets of tRNA-derived fragments

Qiong Xiao

The First Affiliated Hospital of China Medical University

Peng Gao

The First Affiliated Hospital of China Medical University

Xuanzhang Huang

The First Affiliated Hospital of China Medical University

Xiaowan Chen

The First Affiliated Hospital of China Medical University

Quan Chen

The First Affiliated Hospital of China Medical University

Xinger Lv

The First Affiliated Hospital of China Medical University

Yu Fu

The First Affiliated Hospital of China Medical University

Yongxi Song

The First Affiliated Hospital of China Medical University

Zhen-Ning Wang (✉ josieon826@sina.cn)

The First Affiliated Hospital of China Medical University <https://orcid.org/0000-0003-0557-3097>

Research

Keywords: tRNA derived fragments, crosslinking, ligation, and sequencing of hybrids, features of tRF targeting, Support Vector Machine, the first tRF target predicting tool

Posted Date: January 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-52933/v3>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 25th, 2021. See the published version at <https://doi.org/10.1186/s12967-021-02731-7>.

Abstract

Background: tRNA-derived fragments (tRFs) are 14–40-nucleotide-long, small non-coding RNAs derived from specific tRNA cleavage events with key regulatory functions in many biological processes. Many studies have shown that tRFs are associated with Argonaute (AGO) complexes and inhibit gene expression in the same manner as miRNAs. However, there are currently no tools for accurately predicting tRF target genes.

Methods: We used tRF-mRNA pairs identified by crosslinking, ligation, and sequencing of hybrids (CLASH) and covalent ligation of endogenous AGO-bound RNAs (CLEAR)-CLIP to assess features that may participate in tRF targeting, including the sequence context of each site and tRF-mRNA interactions. We applied genetic algorithm (GA) to select key features and support vector machine (SVM) to construct tRF prediction models.

Results: We first identified features that globally influenced tRF targeting. Among these features, the most significant were the minimum free folding energy (MFE), position 8 match, number of bases paired in the tRF-mRNA duplex, and length of the tRF, which were consistent with previous findings. Our constructed model yielded an area under the receiver operating characteristic (ROC) curve (AUC) = 0.980 (0.977-0.983) in the training process and an AUC = 0.847 (0.83-0.861) in the test process. The model was applied to all the sites with perfect Watson-Crick complementarity to the seed in the 3' untranslated region (3'-UTR) of the human genome. Seven of nine target/nontarget genes of tRFs confirmed by reporter assay were predicted. *We also validated the predictions via quantitative real-time PCR (qRT-PCR). Thirteen potential target genes from the top of the predictions were significantly down-regulated at the mRNA levels by overexpression of the tRFs (tRF-3001a, tRF-3003a or tRF-3009a).*

Conclusions: Predictions can be obtained online, tRFTars, freely available at <http://trftars.cmuzhenninglab.org:3838/tar/>, which is the first tool to predict targets of tRFs in humans with a user-friendly interface.

1. Background

tRNA-derived fragments (tRFs) are small non-coding RNAs derived from tRNAs with lengths of 14-40 nucleotides (nts). They have been identified at high abundances in many species (1-3) and can be divided into five categories: i) tRF-5s, from the 5' ends of mature tRNAs; ii) tRF-3s, from the 3' ends of mature tRNAs with 3'-CCA termini; iii) i-tRFs, from the internal cleavage of mature tRNAs; iv) tRF-1s (3'U tRFs), from the 3' trailing sequences of pre-tRNAs with poly-U residues; and v) tiRNAs, tRNA halves from cleavage at the anticodon of mature tRNAs (2-4). Accumulating evidence has shown that tRFs are derived from specific tRNA cleavage events catalyzed by enzymes such as angiogenin, tRNase Z, RNase P and Dicer (not necessary) rather than random tRNA degradation (5). Since databases for identifying and storing tRF sequences such as tRFdb and tRF2Cancer have been published, many studies have reported the roles of tRFs in humans (6-10). Deep sequence analysis of small RNAs associated with Argonaute

(AGO) complexes, the main components of the RNA-induced silencing complex (RISC), has allowed a large number of reads to be mapped to fragments of tRNA (11-13). Kumar further mined AGO PAR-CLIP data and showed that tRF-3s and tRF-5s could associate with target mRNAs by their 5' seed sequence (tRF nts 2–7) in a manner similar to that of miRNAs (11, 12, 14, 15), with hexamers complementary to the seed referred to as “seed matches”. Moreover, many studies using reporter assays have confirmed that tRFs show activity in regulating the expression of protein-coding genes through complementary pairing between the seed sequence and the 3' untranslated region (3'-UTR) of the target mRNA (3, 11, 16). Although seed pairing is commonly thought to function in gene expression regulation (17, 18), studies on specific factors that affect tRF targeting are limited.

Identification of the targets of tRFs is central for characterizing the functional roles of tRFs. However, there have been few investigations of the relationships between tRFs and mRNA, and it is unrealistic to identify all tRF targets by experiments. Researchers have no choice but to rely on algorithms that predict the targets of miRNAs (16, 19, 20). For example, Maute et al. used TargetScan to predict the targets of tRFGlyGCC, but only one in three was repressed by the tRF (16). Similarly, Zhang et al. validated only one in five mRNAs predicted by miRanda and RNAhybrid using real-time PCR (21). The accuracy of such approaches has been poor, and there are currently no better methods for predicting the targets of tRFs in humans. Therefore, a computation tool for predicting the targets of tRFs is urgently needed.

In the present study, we screened features that influenced tRF targeting. Then, we used a support vector machine (SVM) to build models with key features selected by a genetic algorithm (GA) using the pairs identified by crosslinking, ligation, and sequencing of hybrids (CLASH) and covalent ligation of endogenous AGO-bound RNAs (CLEAR)-CLIP to achieve relatively high accuracy in both the training and validation processes. We developed the computational tool tRFTars, available at <http://trftars.cmuzhenninglab.org:3838/tar/> (mirror site at <http://trftar.cmuzhenninglab.org:3838/tar/>), which is the first database for predicting the potential targets of tRFs.

2. Methods

2.1 Data preparation and preprocessing

We obtained mRNA sequences from UCSC 2019 (22). The sequences were annotated according to the human genome (hg19) in RefSeq, and only “NM_” transcripts were retained. tRF-3 and tRF-5 sequences were downloaded from tRFdb (6).

We identified tRF-mRNA pairs in HEK293 cells by CLASH and in Huh-7.5 cells by CLEAR-CLIP (23, 24), which connect AGO-bound small RNAs and target RNAs as chimeric reads in the same RISC complex. After removing adaptors and PCR duplicates, we mapped the reads to the ends of tRNAs as long as possible until mismatches or bulges appeared with blastn (e-value < 0.01) (15). Only tRFs in tRFdb (6) were selected for further study to avoid including tRNA degradation products. After removing part of the sequence mapped to each tRNA, the remaining fragment of the read was mapped to the hg19 3'-UTR. Only pairs with perfect Watson-Crick complementarities between the tRF seed and 3'-UTR sequence were

kept in the positive group (the 3'-UTRs were required to have at least 6 contiguous bases paired with 2-7 bases of the 5' end of tRFs). After removing the transcripts from the positive group, we searched the remaining 3'-UTRs for segments with perfect seed pairing to the tRFs in CLASH as the background (25), which reflects average levels of the features for all possible seed pairing.

We obtained data from poly(A)-position profiling by sequencing (3P-seq) with TargetScan 7.2 (26, 27), which measured 3'-UTR isoform quantifications. When multiple 3'-UTRs mapped to the same genomic region, those with the most 3P-seq tags in the corresponding cell line were chosen. The pipeline of the process is presented in Supplementary Figure 1.

2.2 Computational features

After identifying the interactions as described above, we analyzed the features of the positive pairs in CLASH and the background. We mainly focused on potential target sites with seed pairing and considered the pairing type, location, and base identity around them. Then, we assessed the sequence properties of the whole transcripts and tRFs involved in pairing to view their effects on the target sites. Moreover, we selected the local 3'-UTR regions of the seed matches and considered their individual secondary structures and stabilities after binding to tRFs. The collected information is listed in Supplementary Table 1.

Notably, when considering the effect of base components near the target site, we assumed that the identity of bases surrounding the seed matches with different distances had different weights. The score for AU bases was computed following the rubric below (28):

$$S_{AU} = \sum_{i=1}^{i=n} 1/d_{iAU}$$

where d_{iAU} denotes the distance of an A or a U to the seed match within a particular range. We separately computed scores of 35 nts upstream or 15 nts downstream of the site. Considering more bases tended to pair with tRF beyond seed pairing, we therefore also computed the score excluding 10 nts immediately upstream of the seed match.

The main secondary structure of the target region was computed by the "RNAfold" program in the ViennaRNA 2.0 package (Lorenz et al., 2011, <http://www.tbi.univie.ac.at/RNA>) with sequences including 40 nts upstream and 40 nts downstream of the seed match. The minimum free folding energy (MFE) of the duplex was computed by the "RNAup" program in the ViennaRNA 2.0 package (Lorenz et al., 2011, <http://www.tbi.univie.ac.at/RNA>). All features listed were compared between the positive group and the background.

2.3 The model constructed by the GA and SVM

We used SVM to construct a model for predicting the targets of tRFs. To balance the numbers of pairs in the positive and negative groups during training, we randomly chose 2,000 of the tRF-mRNA pairs with

the most 3P-seq tags (five times the number in the positive group) from the background as the negative group. To reduce the risks of overfitting and selection bias and to facilitate parameter optimization, 5-fold cross-validation was conducted to build the model. All the pairs (pairs identified by CLASH or CLEAR-CLIP and negative pairs) were randomly divided into five subsamples of equal size, with four subsamples used for training and the remaining subsample used for testing each time. This process was repeated five times, with each subsample tested exactly once. To improve training efficiency and retain appropriate features, the features mentioned above were selected by GA reflecting the process of natural selection. We set the number of iterations of the GA to 10,000, with possibilities of crossing and mutation of 10% and 30%, respectively. Finally, we averaged the positive probabilities of individual target sites for each fold. We assumed that multiple sites typically acted independently from each other. The positive probability for each potential target transcript with seed matches was computed as follows:

$$P=1-\prod_{i=1}^n p_i$$

where n denotes the number of complementary sites in the transcripts and p_i denotes the possibility of an individual site being predicted as functional by the SVM model. The pairs with a positive probability > 0.5 were selected as targets of a specific tRF.

2.4 The probabilistic model for potential target transcripts

We computed the probability of a specific seed match appearing at a potential 3'-UTR target as P by Markov model (MM) (order 1), depending on the base composition of the sequence. For a specific tRF-3'-UTR pair, we performed a binomial test to compute as follows:

$$P_s = \sum_{i=f}^{l-k+1} \binom{l-k+1}{i} p^i (1-p)^{l-k+1-i}$$

where l represents the length of the target sequence, k represents the number of bases in the seed (k=6), and f represents the frequency of seed matches in the potential target transcripts. We used this formula to calculate for all tRF-3'-UTR pairs, adjusted the calculated values using the Benjamini-Hochberg procedure and evaluated the false discovery rate (FDR).

2.5 Conservation of the target sites

To support our analysis, we used the Bioconductor package GenomicScores, which conveniently provided genome-wide position-specific scores (29). By loading a Bioconductor annotation package (phastCons100way.UCSC.hg19), we obtained phastCons conservation scores, which were based on a two-state phylogenetic hidden MM (phylo-HMM) and multiple alignments of the human genome (hg19) to the genomes of 99 other vertebrate species (30). The scores of each seed region, 35 nts upstream, 15 nts downstream, and the whole 3'-UTR were compared between the target and background.

2.6 Database organization and web interface

The tRF target prediction workflow is shown in Figure 1. tRFTars was implemented by the R package Shiny and hosted on a Linux server (Centos 7.5) with MySQL 5.7.18 as its database engine. The web layout was built by the R package shinydashboard, with results shown in interactive tables by the R package DT.

2.7 Algorithm evaluation and comparison with miRNA target prediction models

To evaluate the universal applicability of tRFTars, the SVM-GA model was compared with the commonly used miRNA target predicting algorithms TargetScan (27) and miRanda (31) with the default parameters. The conservative model and the probabilistic model were also considered as independent prediction criteria to be tested. We compared the receiver operation characteristic (ROC) curves of these methods in predicting pairs from CLASH/CLEAR-CLIP. Moreover, the sensitivity, specificity and Matthews correlation coefficient (MCC) of various methods were assessed using the following formulas (32):

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{MCC} = \frac{TP*TN - FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

where TP, TN, FP and FN denote the numbers of true positives, true negatives, false positives and false negatives, respectively, which reflect the consistency of prediction and the experimental results. MCC values range between -1 and 1, indicating the correlations between predictions and experimental observations. The experimentally tested pairs identified in the reporter assay were employed to further evaluate the performance of our models.

2.8 Validation of prediction results with expression profiles

We quantified the expression of tRFs, miRNAs and mRNAs in 20 gastric tumors and matched adjacent normal tissues from our institution. The steps of sample preparation, sequencing and microarray analysis are presented in the Supplementary methods. We calculated the Spearman correlation coefficient between mRNA and tRF/miRNAs expression levels. All the pairs with seed matches and negative expression correlation coefficients were selected for further analysis. We defined the pairs with correlation coefficients < -0.3 as negatively related pairs (33). Moreover, the negatively related tRF-mRNA pairs with seed matches in TCGA database were downloaded (33). Considering the complexity of expression-related regulation in tissues, we chose only the tRFs in tRFdb with CLASH reads. The most relevant pairs in each cancer with correlation coefficients < -0.3 in TCGA ($P < 5E-06$) were retained. To eliminate the influence of the total number of target genes predicted by the different tools due to the different cut-offs, we used a chi-square test to calculate whether the predicted targets of the SVM-GA

model were more likely to be downregulated and to compare our predictions with the results from TargetScan and miRanda.

We selected potential target genes of tRF-3001a, tRF-3003a and tRF-3009a from the top of the predictions. We measured the potential target gene expression levels after transfection the tRFs through quantitative real-time PCR (qRT-PCR) in MGC-803 cells. We mutated part of the tRFs seed sequence to the complementary form and observed the expression of the potential target genes (Supplementary methods).

2.9 Statistical Programs and Software

Statistical analyses of our results were conducted with R version 3.5.3 (<https://www.r-project.org/>). Statistical significance (P-value) for the features between positive group and the background was calculated with Student's t-test and a threshold value of $P < 0.05$ was considered statistically significant (with exceptional circumstances explained individually). The SVM algorithm was built with the LIBSVM program (Chang et al, 2018, <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>). The SVM and GA algorithm was coded with MATLAB 2016a (MathWorks, Natick, MA, USA). The source code is freely available at Github (https://github.com/cmuxiaoqiong/SVM_GA_tRF_targets).

3. Results

3.1 Identification of tRF targets from CLASH data

Based on CLASH data from HEK293 cells and CLEAR-CLIP data from Huh-7.5 cells, we obtained 547 tRF-mRNA pairs (532 from CLASH and 15 from CLEAR-CLIP) involving 28 tRFs (20 tRF-3s and eight tRF-5s) in CLASH and 15 tRFs (six tRF-3s and nine tRF-5s) in CLEAR-CLIP. A total of 115,418 seed matches paired with the tRFs were used as the background.

3.2 Characterizing target recognition features

The P-values of features considered in this study that may be related to tRF-mRNA interactions are listed in Table 1. We observed significant differences in most sequence features of individual seed matches, transcripts, or tRFs and the features of the tRF-mRNA duplex. The features most significantly different between the positive group and background are shown in Figure 2.

Table 1. The P-value of features considered in this study that can contribute to tRF-mRNA interaction

Category	Feature	Including feature	P value	
Sequence feature of target sites	type of seed match	position 8 match	2.87E-28	
		position 1 match	7.28E-13	
		position 1 A	0.34	
	bases identity immediately flanking seed match (2nt)	base identity of position 2 upstream		0.80
				2.88E-03
		base identity of position 1 downstream		7.67E-07
				0.43
	bases component in the vicinity of seed match	GC percentage 35nt upstream		5.89E-13
				4.28E-14
		GC percentage 25nt upstream(excluding 10 nts immediately upstream)		2.18E-11
				7.51E-03
		score 35nt upstream		2.61E-16
				7.49E-12
		score 25nt upstream(excluding 10 nts immediately upstream)		See Supplementary Table
				See Supplementary Table
		base component 35nt upstream		See Supplementary Table
				See Supplementary Table
	base component 15nt downstream		See Supplementary Table	
			See Supplementary Table	
	Sequence feature of transcripts	distance to the end of tRF	distance to the 5' end	2.37E-08
distance to the 3' end			1.90E-03	
3'-UTR properties		distance to the nearest end	2.12E-03	
		length of 3'-UTR	4.80E-08	
		GC percentage of 3'-UTR	5.40E-06	
frequency of seed matches in 3'-UTR			0.08	
		base component(AGCT) of 3'-UTR	See Supplementary Table	
		dinucleotide component of 3'-UTR	See Supplementary Table	
CDS properties		length of CDS	1.95E-04	
		GC percentage of CDS	2.33E-03	
		frequency of seed matches in CDS	6.51E-06	
5'-UTR properties		length of 5'-UTR	0.40	
	GC percentage of 5'-UTR	0.08		
	frequency of seed matches in 5'-UTR	0.02		
Sequence feature of tRFs	tRF sequence properties	length of tRF	2.77E-21	
		GC percentage of tRF	2.46E-11	
	TA	target abundance in genome	7.19E-03	
		target abundance in 3'-UTR	0.09	
	SPS	GC percentage of seed	8.69E-18	

Stability and thermodynamics of tRF-mRNA interaction	secondary structure of target mRNA	nucleotides exposed at seed match	7.49E-03
		nucleotides exposed surrounding seed match	0.99
		energy to free base-pairing interactions of target site	3.51E-06
	secondary structure of tRF-mRNA	number of bases paired in duplex	1.53E-20
		MFE	1.36E-57

Sequence features of target sites

Seed match types and surrounding base properties determined the potential of a tRF target site. Position 8 matches ($P=2.87E-28$) and position 1 matches ($P=7.28E-13$) were more common in the positive group than in the background, while no significant difference was observed at position 1 A ($P=0.34$). The number of different kinds of seed types is provided in Supplementary Table 2. Consistent with the finding of PAR-CLIP (14, 15), 7-mer-m8 sites (binding to positions 2–8 in the tRFs) were most enriched in the positive group. The bases flanking the functional sites were more likely to be A or U compared with the bases in the background. The differences in nucleotide identity are listed in Supplementary Table 3 and Supplementary Table 4. The GC percentages were significantly lower for bases upstream ($P=5.89E-13$) and downstream ($P=2.18E-11$). When using the formula to incorporate distances to seed matches with different weights, the scores in the positive group were significantly higher for all three kinds of surrounding regions, including upstream ($P=7.51E-03$), upstream excluding 10 nts ($P=2.61E-16$), and downstream ($P=7.49E-12$).

The locations of seed matches in the 3'-UTRs were related to the tRF binding activity of the UTRs. The cumulative distribution curves of the distance to the 3'-UTR ends are shown in Supplementary Figure 2. We observed that effective sites avoided appearing immediately downstream of the stop codon, while tended to reside adjacent to the ends within the rest of the 3'-UTR in the positive group ($P=2.12E-03$), especially the 5' end ($P=2.37E-08$).

Sequence features of the transcripts

When focusing on the whole-sequence features to assess their impacts on the efficacies of sites, we found that the positive group had a significantly lower global GC content than the background in the 3'-UTR ($P=5.40E-06$), but this trend was not as significant as that detected in local comparisons near seed matches. In agreement with the finding that more sites were preferentially adjacent to the ends, the target sites were selectively located in the shorter 3'-UTRs ($P=4.80E-08$).

Sequence features of tRFs

We investigated whether some tRF sequences might be intrinsically more capable of targeting than others. In tRFdb, tRF-3 and tRF-5 sequences could be categorized into tRF-3a (18 nts), tRF-3b (22 nts), tRF-5a (15 nts), tRF-5b (22 nts), and tRF-5c (31 nts) sequences according to their lengths. Compared to

those in the background, we observed significant differences in the lengths of tRFs (lengths in tRFdb) ($P=2.77E-21$), consistent with the finding that repression was mediated by tRF-3as instead of tRF-3bs derived from the 3' end of the same tRNA (11). tRFs in the positive group had a higher GC percentage for both whole sequences ($P=2.46E-11$) and seed sites ($P=8.69E-18$), which could contribute to pairing stability, especially for seed regions (34). In addition, we detected significant differences in target site abundance (TA) in the genome ($P=7.19E-03$), consistent with the finding that extensive pairing could decrease the function of sRNA pairing to their authentic target sites (35-37).

Structure and thermodynamic properties of the tRF-mRNA duplex

In addition to the sequence features of tRFs and their targets, the secondary structures of target mRNAs likely contribute to target recognition. We showed that nucleotides in the positive group were more exposed than those in the background at the seed matches ($P=7.49E-03$), while less energy was needed to free base-pairing interactions within the secondary structures of target mRNAs ($P=3.51E-06$). These features indicated that effective target sites were more accessible for tRF binding.

Thermodynamic features of the tRF-mRNA duplex were then analyzed. Compared with that in the background, the tRF seed-target binding in the positive group was usually more stable, as revealed by a lower MFE ($P=1.36E-57$), which represented the most stable structure of the helix. Moreover, additional pairing beyond seed matches contributed to target recognition in the positive group ($P=1.53E-20$), consistent with the experimental discovery of Kuscu et al. (11). These features were considered in our models.

3.3 The target prediction model established by SVM and GA

A total of 547 positive pairs (including 489 tRF-3 pairs and 58 tRF-5 pairs) and 2,000 negative pairs (including 1,596 tRF-3 pairs and 404 tRF-5 pairs) were retained according to our filtering criteria. The features with significant differences between the positive group and negative group are listed in Supplementary Table 5, with no significant differences observed for the vast majority of features between CLASH and CLEAR-CLIP. We considered the 96 features listed in Table 1 as potentially informative in relation to tRF targeting, while 51 features were identified by the GA after parameter optimization (Figure 3). These features were modeled in an SVM framework to determine their individual contributions for model implementation. The result of each fold was stable, with an area under the ROC curve (AUC) = 0.980 (from 0.977 to 0.983) during the training process and an AUC = 0.847 (from 0.83 to 0.861) during the validation process (Supplementary Table 6). Fourteen of 15 pairs detected by CLEAR-CLIP and 455 of 532 detected by CLASH were predicted, which indicated the efficiency of our model in different cell lines.

3.4 Predicting the targets with the probabilistic model

Functional tRF-target interactions, which account for a small proportion of seed pairings, derive directly from coevolution of the tRF and its target. Potential functional 3'-UTR targets should contain more complementary sites overrepresented relative to a random background, which was measured by Ps for

each tRF-3'-UTR pair. P_s was significantly lower in the positive group than in the background ($P=2.94E-10$). We generated the final predictions by ranking adjusted P_s , showing that 26,380 tRF-3 targets and 8,670 tRF-5 targets were overrepresented. This method could be used to independently predict the targets of tRFs by simulating miRNA target prediction tools such as PicTar and PACMIT (38, 39).

3.5 Predicting the targets with conserved seed match properties

Biologically functional target sites tend to be located in conserved tRF-pairing motifs within 3'-UTRs (18). We found that target sites were significantly more conserved than the background ($P=3.68E-08$), as determined by phastCons scores from comparative sequence analyses of the human genome (hg19) with the genomes of 99 other vertebrate species. Similar levels of performance were observed immediately upstream ($P=2.31E-10$), immediately downstream ($P=8.29E-10$), and in the whole transcript ($P=3.66E-13$). We applied this measure of performance as an independent factor to predict the targets of tRFs (25, 30, 40). A total of 111,874 tRF-3 target sites and 152,464 tRF-5 target sites were predicted with 0.5 as a cut-off. *We performed gene ontology enrichment analysis on the target genes predicted by the conservation analysis. The results were listed in Supplementary Table 7.*

3.6 Algorithm evaluation of tRF target prediction models

The prediction abilities of tRFTars and common miRNA target prediction programs were assessed by comprehensively comparing their identified pairs with the pairs identified by CLASH/CLEAR-CLIP or reporter assays. The performance of our model was evaluated with ROC curve analysis, yielding an AUC = 0.980 in the training process and an AUC = 0.847 in the validation process (Figure 4A), better than commonly used miRNA target prediction models (intersection of TargetScan and miRanda) (AUC = 0.743, $P < 0.0001$). Moreover, five of seven positive pairs and two of two negative pairs were predicted by the SVM-GA model (Table 2) (11, 16, 41), while only three of seven positive pairs were predicted by miRNA models. The sensitivity, specificity and MCC are listed in Supplementary Table 8. Both lines of evidence suggested that the SVM-GA model was the most effective tool for distinguishing the targets of tRFs with a relatively high accuracy. In addition, we searched the pairs confirmed by reporter assay with a clear tRF sequence beyond those in tRFdb, and KLF12 was predicted to be the target of tRFGluTTC (positive probability=0.76) (42), which further proved the effectiveness of our model.

Table 2. Validation of the models with tRF-mRNA pairs reported by reporter assay

tRF	gene	Whether target of tRF	References	Prediction of SVM-GA model	Prediction of conservative model	Prediction of probabilistic model	Prediction of miRNA model
3027b	RPA1	Yes	Maute et al	0.58(Yes)	0(No)	0.06(No)	Yes
3009a	DGCR2	Yes	Kuscu et al	0.26(No)	0(No)	0.72(No)	No
3009a	SLC6A9	Yes	Kuscu et al	0.89(Yes)	0(No)	0.30(No)	No
3009a	SMAD1	Yes	Kuscu et al	0.73(Yes)	0.62(Yes)	0.03(No)	No
3009a	TBL1X	Yes	Kuscu et al	0.10(No)	0(No)	0.32(No)	No
3009a	FER	Yes	Kuscu et al	0.86(Yes)	0(No)	0.05(No)	Yes
5030c	LRP8	Yes	Deng et al	0.89(Yes)	0.47(No)	0.17(No)	Yes
3027b	STAG2	No	Maute et al	0.22(No)	1(Yes)	0.01(No)	No
3027b	NSD3	No	Maute et al	0.21(No)	1(Yes)	0.02(No)	No

Moreover, we investigated the relationships between different tRF target prediction models. We discovered that most of the features for effective target site prediction were correlated with the probabilistic model or conservative model (Figure 4B). The intersections of three algorithms and miRNA target prediction models (tRFs found in CLASH) are displayed in Figure 4C, and the intersections are listed in Supplementary Table 9. The proportion of targets predicted by each model is presented in Figure 4D.

3.7 Validation of prediction results with expression correlations

We applied the filtering criteria and selected 1226 tRF-mRNA pairs with negative expression correlations in TCGA database (33). The pairs predicted by the SVM-GA model were more likely to have negative expression correlations ($P=8.97E-04$), and the model performed better than TargetScan ($P=0.79$) (Supplementary Table 10). *We have listed the number of pairs in different kinds of tumors in Supplementary Table 11.* Moreover, by analyzing tsRNA sequencing (tsRNA-Seq) and lncRNA + mRNA microarray data from our institution, 34 tRFs with adequate variability and their expression-related transcripts were chosen. Compared with those of TargetScan and miRanda, the predictions of the SVM-GA model were more enriched in pairs with negative correlations (SVM-GA model: $P=3.09E-03$; TargetScan and miRanda: $P=0.09$) (Supplementary Table 12). Additionally, we observed significant consistency between the correlation coefficients of the pairs and the likelihood of a positive result from the SVM-GA model (for tRF-5s: $P=1.87E-10$; for tRF-3s: $P=2.77E-05$). This finding indicates that our predicted tRF targets with higher scores tend to be more downregulated. *We used Multiple Linear Regression to exclude the influence of miRNA in the sequencing and microarray data from our institution. We chose the pairs with mRNAs regulated by multiple highly expressed miRNAs, and 96 of 397 tRF-mRNA pairs still had significant negative correlation, listed at Supplementary Table 13. It showed that nearly 1/4 tRF-mRNA pairs played a most vital role in gene expression regulation. We showed the correlation heatmap of 20 tRFs/miRNAs and mRNAs with the highest frequency among the target gene predictions (Supplementary Figure 3).*

3.8. Validation the predictions by qRT-PCR

We selected potential target genes of tRF-3001a (ELAVL1, SOCS7, ATF6B, RINL, PRR11, and ZNF268), potential target genes of tRF-3003a (CBX5, EIF4E, PRKAA1, TFDP2, SH3TC2, and PDE12) and potential target genes of tRF-3009a (ATF6B, ARF3, CDS2, CLN8, MAP2K7, and SNX12) from the top of the predictions for further confirmation using qRT-PCR (Supplementary methods). We evaluated potential target gene expression levels after transfection with tRF mimic or the mutated tRF mimic in MGC-803 cells. The expression of 13 target genes was markedly reduced at the mRNA levels by transfecting tRF-mimic compared to the corresponding NC group in MGC-803 cells (Supplementary Table 14) (Supplementary Figure 4).

3.9 Website

The predicted targets are available online from tRFTars (<http://trftars.cmuzhenninglab.org:3838/tar/>) (mirror site at <http://trftar.cmuzhenninglab.org:3838/tar/>) (Figure 5). Strict matches to the official gene symbols, RefSeq IDs, tRFs in tRFdb or anticodon/amino acids of the source tRNA are necessary as input. Users can choose to search the specified target sites or transcripts according to their needs. Candidates can be ranked according to the likelihood of a positive result assigned by the SVM-GA model, the conservation score or Ps by the probabilistic model, although we recommend adopting the SVM-GA model. Users can choose to comprehensively view the information for all seed pairings. Details of the prediction results can be found on the statistics page of the website.

4. Discussion

In this study, we analyzed the results of experiments in which tRFs were ligated to target RNA in purified AGO complexes (CLASH in HEK293 cells and CLEAR-CLIP in Huh-7.5 cells) to identify tRF-mRNA interactions and extended the discovery to all tRFs. We first comprehensively assessed features for effective tRF-mRNA interaction by statistical analysis and discovered that tRFs could indeed act similarly to miRNAs, with most features involved in miRNA targeting significantly different from those in the background. Features with the most significant variance between the positive group and background were the MFE, position 8 match, number of bases paired in the tRF-mRNA duplex, and length of the tRF, which were consistent with previous findings (11, 14). The intrinsic mechanisms that influenced the interactions can be explained by binding affinity and target site accessibility (28). SVM was used to incorporate all these features and to determine the contributions of individual features to target predictions.

We validated the effectiveness of our SVM-GA model, obtaining an AUC = 0.847 during the internal validation process. Furthermore, seven of nine target/nontarget genes confirmed by reporter assay were predicted successfully, far greater than the number obtained with miRNA target prediction algorithms. For example, Maute et al. demonstrated that 3027b (CU1276) (human) overexpression by both tRNA and tRF hairpin transfections significantly repressed RPA1 in a B-cell lymphoma line, as determined using 3'-UTR reporter assays, western blotting, and qRT-PCR (16). NSD3 (WHSC1L1) and STAG2, predicted by

TargetScan, showed almost no response to 3027b overexpression. This is consistent with our predictions. Similarly, SMAD1, SLC6A9, and FER (FER1, matching the 3'-UTR sequence of FER in the NCBI database), which were validated to be targets of 3009a (human) using the luciferase reporter assay and RNA-seq by Kuscu et al., were also predicted. Conversely, although Maute et al. found that TargetScan-predicted genes were significantly enriched in downregulated genes by 3027b, the accuracy rate in the study was less than 10% for both tRNA and hairpin transfections (16). Although this was not a gold standard method, the prediction accuracy was still unsatisfying.

It is undeniable that this is an unfair comparison because these interactions were chosen for validation after prediction with algorithms for miRNAs. However, our SVM-GA model still outperformed the intersection of miRNA target prediction models. We aimed to understand the mechanism underlying this accuracy difference. Though our model and other miRNA target prediction algorithms considered sequence complementarity, thermodynamic calculations of duplex formation, and evolutionary properties to rank the potential target candidates, the detailed contributions of individual features were different. For example, different from the relatively fixed length of 22 nts of miRNAs, tRFs have a variable length from 14 nts to 40 nts. They may act differently from miRNAs, whose 13-16 bases play an essential role in targeting in addition to seed pairing. Moreover, conservation may not be a determinant of tRF targeting. Prediction by a probabilistic model and conservative model will miss some cases because of relatively low sensitivity. While, additional complementarity downstream from the seed sequence is a more essential feature for tRF targeting. Accounting for such complementarity can lead to relatively good model performance by considering the secondary structure of the tRF-mRNA duplex.

However, we investigated why the system missed predicting targets of 3009a, TBL1X and DGCR2 in two cases. Both target sequences were located in the middle of long 3'-UTRs with a relatively high GC content near the seed matches, which was in conflict with the requirements for site accessibility. Although TBL1X had seven additional bases paired with 3009a, it was not the most stable structure because of intrinsic pairing within secondary structures near the seed match, as predicted by RNAup. However, we believe that extensive base pairing can occur when the interacting molecules are present at high concentrations, which could contribute to effective 3009a-TBL1X interactions. Furthermore, although DGCR2 had five additional bases paired with 3009a, the resulting duplex had a higher MFE, which was contrary to the expectation of site affinity.

This study had some limitations. First, we did not allow any mismatches or bulges when mapping reads to tRNAs or 3'-UTRs. To avoid influences of the RNase step in the CLASH protocol and the inclusion of tRFs from random tRNA degradation products, we included only the tRFs in tRFdb in our model. We adopted stringent inclusion criteria at the expense of excluding some effective pairs in the positive group. Second, it was difficult to conduct large-scale precision validation of tRF target genes. *We observed a tendency for predicted targets with higher positive probability in SVM-GA model to be more suppressed in mRNA levels.* More experiments that quantify tRF repression strength by transfection or knockdown of a particular tRF will be conducted in different cells and under different conditions to improve our model. Third, the hg38 sequence information of tRFs is not updated in public databases. Additionally, tiRNAs

and tRFs not included in tRFdb, such as 5'-tiRNAVal and tRF5-Glu, which have been reported to repress mRNA by seed pairing (18, 43, 44), were not considered in this version. Fourth, exceptional cases, such as contiguous pairing of the 3'-end and non-canonical binding, need to be further investigated (41). We will also investigate whether tRF-specific features act in tRF targeting.

5. Conclusions

The tRFTars is the first website to predict the target of tRF in humans. We intend to update predictions of the website as more tRFs identified to be functional, and plan to extend our model to more species. This website provides convenience to identify potential human targets of particular tRFs for experimental confirmation. Furthermore, it will greatly facilitate our understanding of gene regulation and the functions of tRFs.

6. List Of Abbreviations

tRFs: tRNA-derived fragments, CLASH: crosslinking, ligation, and sequencing of hybrids, GA: genetic algorithm, SVM: support vector machine, ROC: receiver operating characteristic curve, AUC: the area under the receiver operating characteristic curve, RISC: RNA-induced silencing complex, 3'-UTR: 3' untranslated regions, CLEAR-CLIP: covalent ligation of endogenous Argonaute-bound RNAs, 3P-seq: poly(A)-position profiling by sequencing, MFE: minimum free folding energy, MM: Markov Model, FDR: false discovery rate, phylo-HMM: phylogenetic hidden Markov model, MCC: Matthews Correlation Coefficient, TA: target-site abundance, GEO: Gene Expression Omnibus, TCGA: The Cancer Genome Atlas, qRT-PCR: quantitative real-time PCR

7. Declarations

Ethics approval and consent to participate

We obtained the patients' written informed consent based on ethical guidelines. The study was approved by the research ethics committee of the First Affiliated Hospital of China Medical University and complied with the Declaration of Helsinki.

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed in this study are available in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) (GES50452, GSE73059). The data and source code of prediction model is freely available at Github (https://github.com/cmuxiaoqiong/SVM_GA_tRF_targets).

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the National Key R&D Program of China [MOST-2017YFC0908300, MOST-2016YFC1303200]; Major Scientific and Technological Special Project of Liaoning Province of China (2019020176-JH1/103, 2019JH1/10300007); National Natural Science Foundation of China (82002599); Natural Science Foundation of Liaoning Province of China [2019-MS-390]; China Postdoctoral Science Foundation Grant [2018M641746]; and The Natural Science Foundation Medical and Health Joint Fund Project of Liaoning Province [20180530006].

Authors' contributions

QX., PG. contributed equally to this work. ZNW. participated in the conception and design of the study; XW.C., QC., XE.L., and YF. have participated in collection, analysis and interpretation of data. QX., PG., XZ.H. YX.S. and ZN.W. were responsible for drafting or revising the manuscript, QX. and PG. were responsible for the website design and database maintenance. All authors have read and approved the version to be published.

Acknowledgements

We thank the Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education (China Medical University, Shenyang, China) for technical assistance.

8. References

1. Soares AR, Fernandes N, Reverendo M, Araujo HR, Oliveira JL, Moura GM, et al. Conserved and highly expressed tRNA derived fragments in zebrafish. *BMC molecular biology*. 2015;16:22.
2. Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & development*. 2009;23(22):2639-49.
3. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA (New York, NY)*. 2010;16(4):673-95.
4. Kumar P, Kuscu C, Dutta A. Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends in biochemical sciences*. 2016;41(8):679-89.
5. Guo Y, Bosompem A, Mohan S, Erdogan B, Ye F, Vickers KC, et al. Transfer RNA detection by small RNA deep sequencing and disease association with myelodysplastic syndromes. *BMC genomics*. 2015/09/25 ed2015. p. 727.
6. Kumar P, Mudunuri SB, Anaya J, Dutta A. tRFdb: a database for transfer RNA fragments. *Nucleic acids research*. 2015;43(Database issue):D141-5.
7. Pliatsika V, Loher P, Telonis AG, Rigoutsos I. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics (Oxford, England)*. 2016;32(16):2481-9.

8. Pliatsika V, Loher P, Magee R, Telonis AG, Londin E, Shigematsu M, et al. MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic acids research*. 2018;46(D1):D152-D9.
9. Selitsky SR, Sethupathy P. tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. *BMC bioinformatics*. 2015;16:354.
10. Zheng LL, Xu WL, Liu S, Sun WJ, Li JH, Wu J, et al. tRF2Cancer: A web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic acids research*. 2016;44(W1):W185-93.
11. Kuscu C, Kumar P, Kiran M, Su Z, Malik A, Dutta A. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA (New York, NY)*. 2018;24(8):1093-105.
12. Burroughs AM, Ando Y, de Hoon ML, Tomaru Y, Suzuki H, Hayashizaki Y, et al. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA biology*. 2014;8(1):158-77.
13. Li Z, Ender C, Meister G, Moore PS, Chang Y, John B. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic acids research*. 2012;40(14):6787-99.
14. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010;141(1):129-41.
15. Kumar P, Anaya J, Mudunuri SB, Dutta A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC biology*. 2014;12:78.
16. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(4):1404-9.
17. Karaiskos S, Naqvi AS, Swanson KE, Grigoriev A. Age-driven modulation of tRNA-derived fragments in *Drosophila* and their potential targets. *Biology direct*. 2015;10:51.
18. Luo S, He F, Luo J, Dou S, Wang Y, Guo A, et al. *Drosophila* tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. *Nucleic acids research*. 2018;46(10):5250-68.
19. Cui Y, Huang Y, Wu X, Zheng M, Xia Y, Fu Z, et al. Hypoxia-induced tRNA-derived fragments, novel regulatory factor for doxorubicin resistance in triple-negative breast cancer. *J Cell Physiol*. 2019;234(6):8740-51.

20. Huang Y, Ge H, Zheng M, Cui Y, Fu Z, Wu X, et al. Serum tRNA-derived fragments (tRFs) as potential candidates for diagnosis of nontriple negative breast cancer. *J Cell Physiol*. 2019.
21. Zhang M, Li F, Wang J, He W, Li Y, Li H, et al. tRNA-derived fragment tRF-03357 promotes cell proliferation, migration and invasion in high-grade serous ovarian cancer. *OncoTargets and Therapy*. 2019;Volume 12:6371-83.
22. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic acids research*. 2019;47(D1):D853-D8.
23. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 2013;153(3):654-65.
24. Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, Nishiuchi E, et al. miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature communications*. 2015;6:8864.
25. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*. 2003;115(7):787-98.
26. Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*. 2010;469(7328):97-101.
27. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4.
28. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*. 2007;27(1):91-105.
29. Puigdevall P, Castelo R. GenomicScores: seamless access to genomewide position-specific scores from R and Bioconductor. *Bioinformatics (Oxford, England)*. 2018;34(18):3208-10.
30. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*. 2009;19(1):92-105.
31. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic acids research*. 2008;36(Database issue):D149-53.
32. Fan X, Kurgan L. Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Brief Bioinform*. 2015;16(5):780-94.
33. Telonis AG, Loher P, Magee R, Pliatsika V, Londin E, Kirino Y, et al. tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities. *Cancer research*. 2019;79(12):3034-49.
34. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *Isy-6* and other microRNAs. *Nature structural & molecular biology*. 2011;18(10):1139-46.
35. Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods*. 2007;4(9):721-6.

36. Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular systems biology*. 2010;6.
37. Anderson EM, Birmingham A, Baskerville S, Reynolds A, Maksimova E, Leake D, et al. Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA (New York, NY)*. 2008;14(5):853-61.
38. Marin RM, Vanicek J. Efficient use of accessibility in microRNA target prediction. *Nucleic acids research*. 2011;39(1):19-29.
39. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. *Nature genetics*. 2005;37(5):495-500.
40. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120(1):15-20.
41. Deng J, Ptashkin RN, Chen Y, Cheng Z, Liu G, Phan T, et al. Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. *Molecular therapy : the journal of the American Society of Gene Therapy*. 2015;23(10):1622-9.
42. Shen L, Tan Z, Gan M, Li Q, Chen L, Niu L, et al. tRNA-Derived Small Non-Coding RNAs as Novel Epigenetic Molecules Regulating Adipogenesis. *Biomolecules*. 2019;9(7).
43. Mo D, Jiang P, Yang Y, Mao X, Tan X, Tang X, et al. A tRNA fragment, 5'-tiRNA(Val), suppresses the Wnt/beta-catenin signaling pathway by targeting FZD3 in breast cancer. *Cancer letters*. 2019;457:60-73.
44. Zhou K, Diebel KW, Holy J, Skildum A, Odean E, Hicks DA, et al. A tRNA fragment, tRF5-Glu, regulates BCAR3 expression and proliferation in ovarian cancer cells. *Oncotarget*. 2017;8(56):95377-91.

Figures

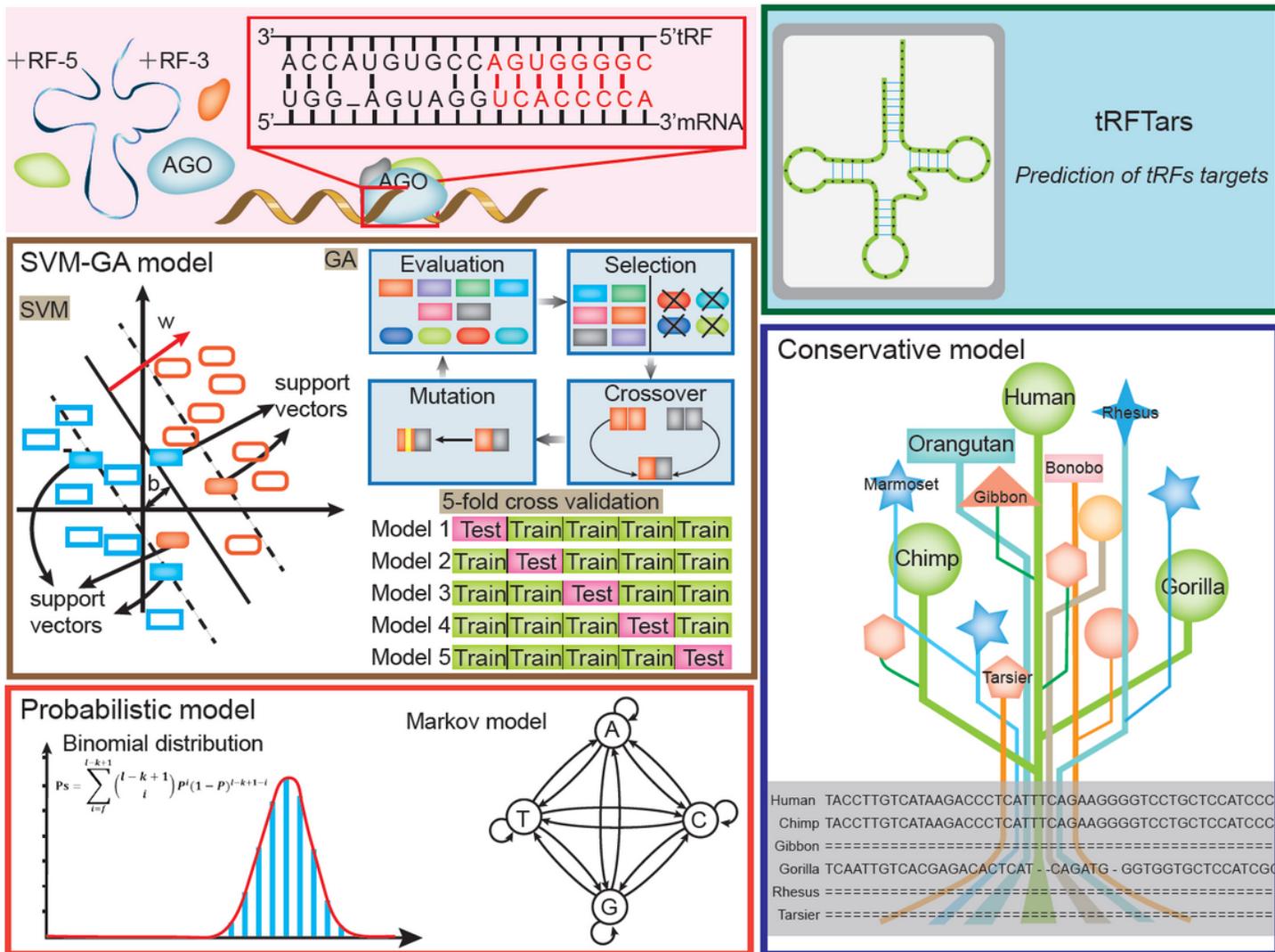


Figure 1

The workflow of the target prediction pipeline of tRFTars.

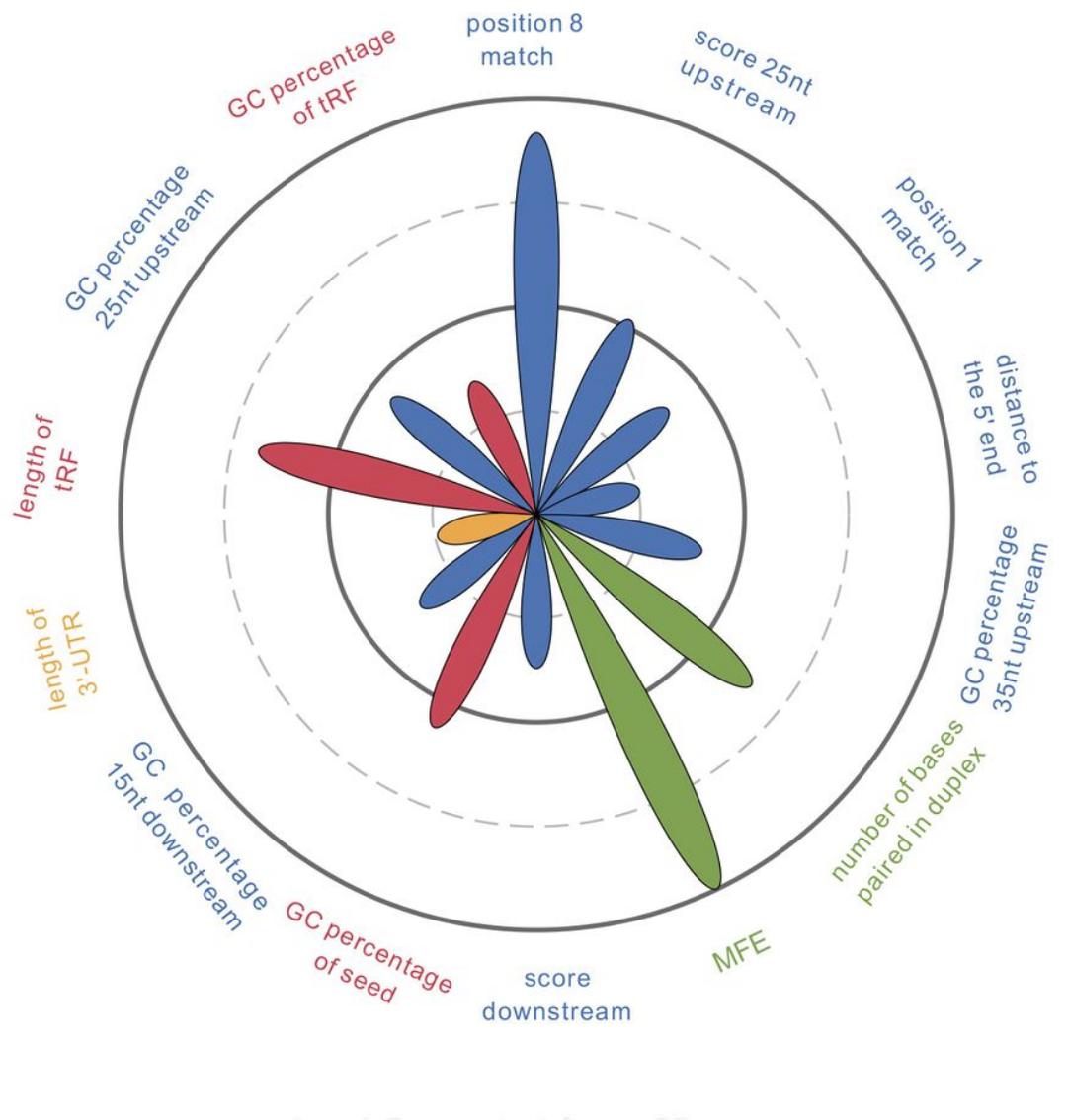


Figure 2

Features that influence tRF targeting. Features most significantly different between the positive group and background are displayed. The size of the petal represents $-\log(P \text{ value})$, which indicates the degree of significance.

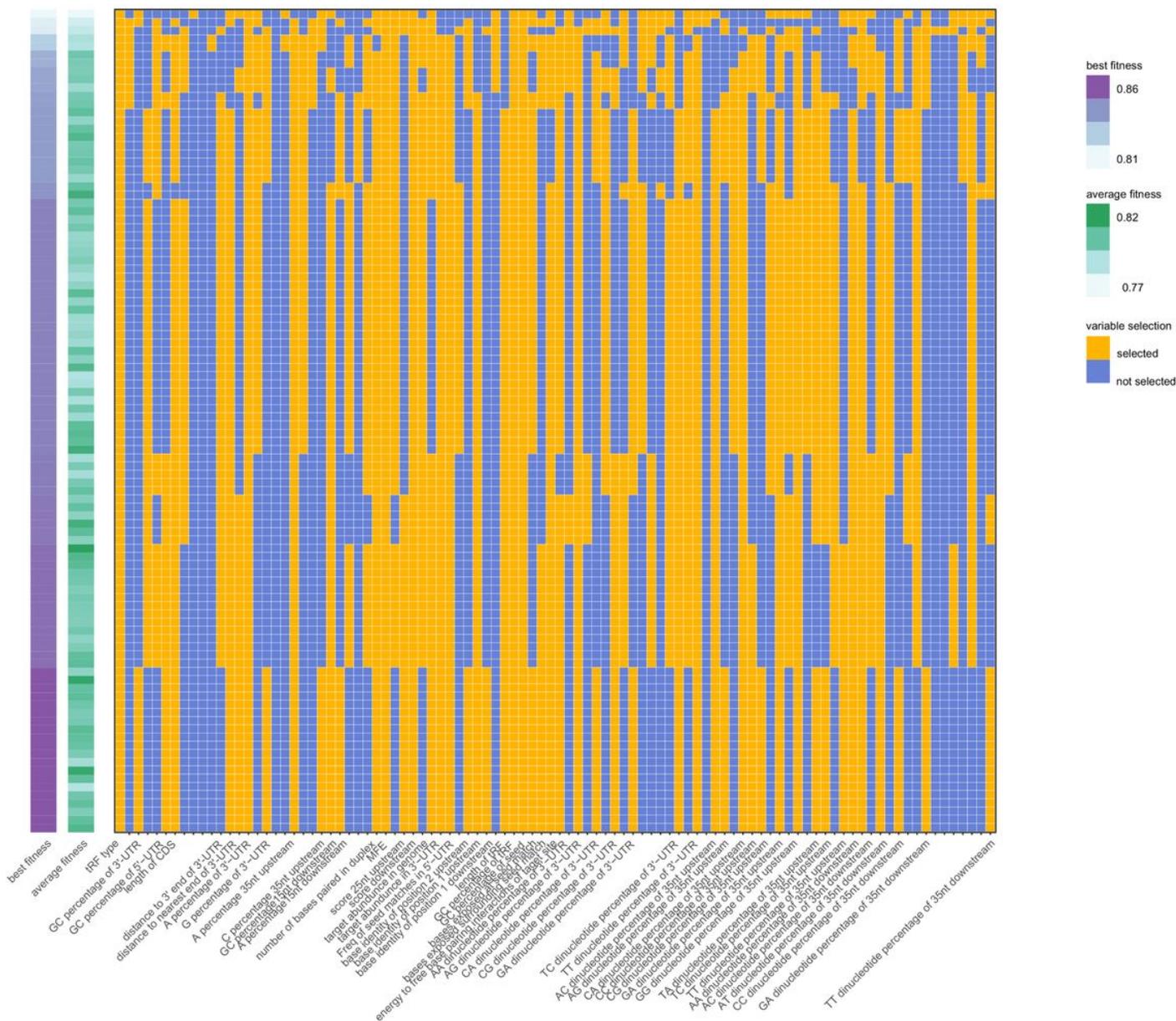


Figure 3

The result of the genetic algorithm (GA) during establishment of the SVM-GA model using the training cohort. The iterations of each variable in the GA are presented in the longitudinal axes and the selected variables in the SVM-GA model are presented in the transverse axes. The features selected are labeled.

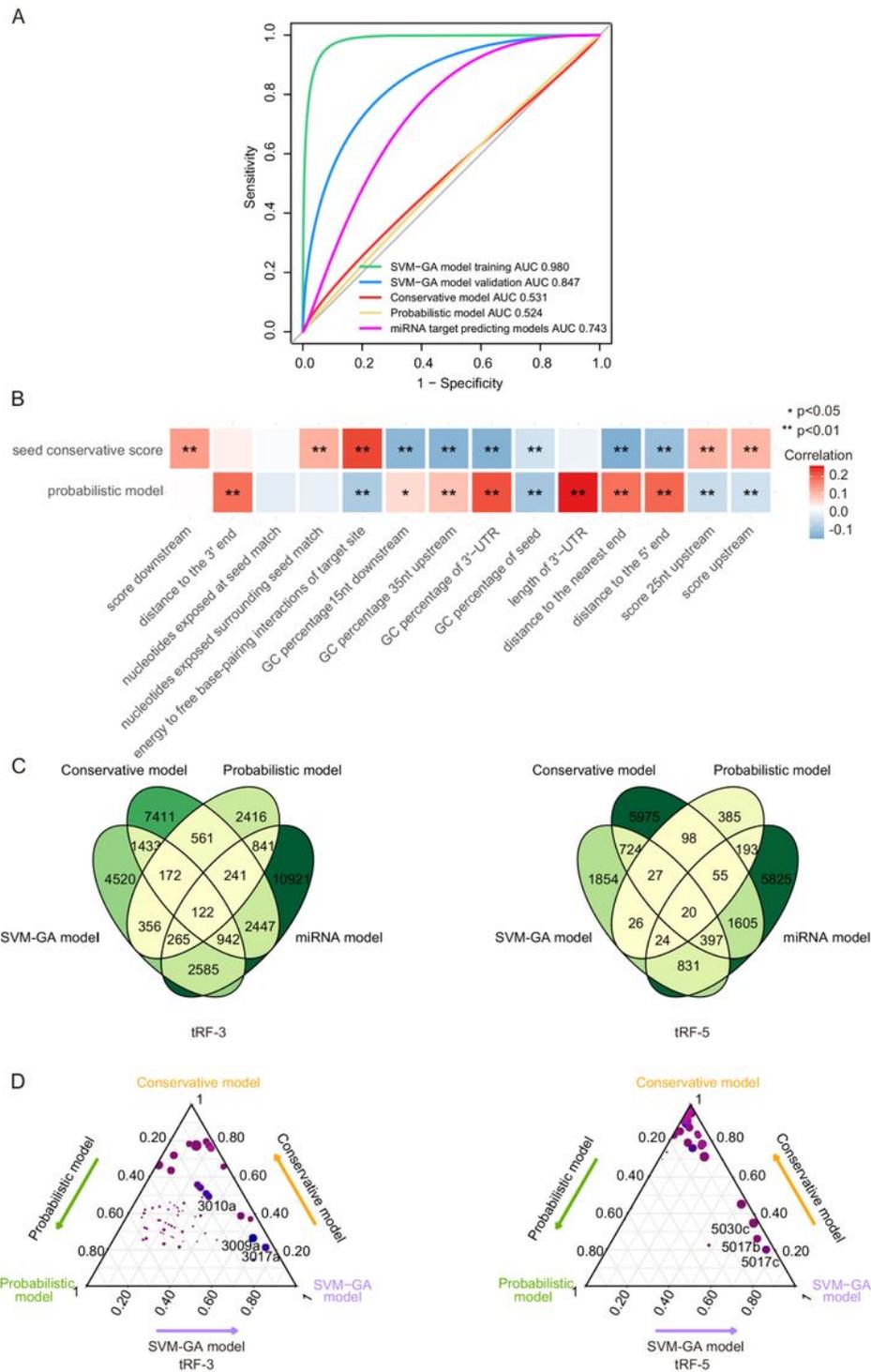


Figure 4

Comparison of tRF target predicting models. A. The receiver operating characteristic (ROC) curve for classification of the pairs for model establishment, including SVM-GA model, conservation model, probabilistic model and intersection of miRNA target predicting models (TargetScan and miRanda). B. The relationship of features for potential target site with the probabilistic model or conservative model. The color of boxes represents the coefficient of correlations and * represents the significance of

correlations (Pearson's test). C. The intersection of target genes (tRFs in CLASH) by three methods of tRFTars and miRNA target predicting models (TargetScan and miRanda). D. Ternary plot of the number of targets of each tRFs. The value to each axis represents the proportion of targets predicted by corresponding models relative to all potential targets. The node color represents the number of potential targets by intersection of three models. As the number of targets increased, the node color changes from red to blue. The node size indicates the number of seed pairings in whole 3'-UTR. The larger the node is, the greater number of seed matches the 3'-UTRs have.

A

The screenshot shows the tRFTars web application interface. On the left, there are search filters including 'Select species' (Human/SRCH37), 'Select tRF type' (tRF-3), 'Select tRF annotation' (tRFdb ID), 'Group by sites' and 'Group by transcript', 'search by tRF (tRNA anticodon)', 'Please type the name' (3015a), 'Sorting algorithm' (SVM-GA scores (Recommended)), and 'confidence' (high confidence). On the right, there is a table with columns: tRNA, Alias, gene, ENTREZID, transcript, seed, length, start, match_3, match_1. The table shows 10 entries for tRNA 3015a.

B

Sorting algorithm

- tRFdb ID ▲
- tRFdb ID
- tRNA anticodon

C

search

- search by tRF(ID/tRNA anticodon) ▲
- search by tRF(ID/tRNA anticodon)
- search by gene (Official gene symbol)
- search by transcripts (RefSeq)

Sorting algorithm

- SVM-GA scores (Recommended) ▲
- SVM-GA scores (Recommended)
- Conservative scores
- Probabilistic model

D

Download the CSV file

Show 10 entries

tRNA	Alias	gene	ENTREZID	transcript	seed	length	start	match_3	match_1	position_1_A	positive_probability	seed_conservative_score	FDR	MFE
1	3001a	[LeuTAG]_LeuAAG	RPL	126432	NM_108445.3	TCCCAC	1274	55	1	0	0.964905142	0	0.03316373	-9.61
2	3001a	[LeuTAG]_LeuAAG	RPL	126432	NM_01195633.1	TCCCAC	1274	55	1	0	0.964849007	0	0.03316373	-9.61
3	3001a	[LeuTAG]_LeuAAG	NGO	98157	NM_011257.3	TCCCAC	1045	299	1	1	0.961387251	0	0.01290878	-21.58
4	3001a	[LeuTAG]_LeuAAG	SOC37	30837	NM_014596.3	TCCCAC	6111	3864	1	0	0.960705309	0	0.1450131	-18.92
5	3001a	[LeuTAG]_LeuAAG	CCNYL1	151195	NM_152823.2	TCCCAC	2476	1586	1	0	0.959378993	0.03333333	0.1220933	-14.52
6	3001a	[LeuTAG]_LeuAAG	SH2D1B	117157	NM_013282.4	TCCCAC	2014	462	1	1	0.958423014	0	0.1178436	-17.31
7	3001a	[LeuTAG]_LeuAAG	KMT5B	51111	NM_01300906.1	TCCCAC	2825	651	0	0	0.956814303	1	0.04314784	-20.91
8	3001a	[LeuTAG]_LeuAAG	KMT5B	51111	NM_01300907.1	TCCCAC	2825	651	0	1	0.956568905	1	0.04314784	-20.91
9	3001a	[LeuTAG]_LeuAAG	BCKDK	10295	NM_005881.3	TCCCAC	526	476	0	1	0.956240311	0	0.01641594	-17.76
10	3001a	[LeuTAG]_LeuAAG	CCNYL1	151195	NM_01142300.1	TCCCAC	2476	1586	1	0	0.955742095	0.03333333	0.122093288	-14.52

Showing 1 to 10 of 2,269 entries

Figure 5

Search the tRF targets in tRFTars with an example of 3001a. A. Overview of the Search page interface. B. Users can search the target according to the tRFs in tRFdb or anticodon/ amino acid of the source tRNA. C. Users can input the corresponding name and rank the target candidates according to positive possibility assigned by the SVM-GA model, the conservative score or the sites overrepresented by the probabilistic model. D. The result panel of the target site of tRF 3001a.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarydata.pdf](#)
- [Supplementarytable7.xlsx](#)
- [Supplementarytable13.xlsx](#)