tRFtar: Identifying 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

Zhenning Wang (josieon826@sina.cn)
The First Affiliated Hospital of China Medical University

Research

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

DOI: https://doi.org/10.21203/rs.3.rs-52933/v1

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

Background: The tRNA-derived fragments (tRFs) are 14–40 nucleotides, small non-coding RNAs from specific tRNA cleavages, and they have key regulatory functions in many biological processes. Many studies showed that tRFs are associated with Argonaute complexes and inhibit gene expression in the same manner as miRNAs. However, there are currently no tools to accurately predict tRF target genes.

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

Results: We first identified features that globally influenced tRF targeting. Among them, the most significant ones were minimum free folding energy (MFE), position 8 match, number of bases paired in tRF-mRNA duplex, and length of tRF, which were consistent with previous findings. We built the model with the area under the receiver operating characteristic (ROC) curve (AUC) = 0.980 (0.977-0.983) in the training process and 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’-UTR of human genome. Seven of nine target / non-target genes of tRFs confirmed by reporter assay were predicted.

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

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

Figures
Figure 1

The workflow of the target prediction pipeline of tRFTar.
**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 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 tRFTar 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.
Figure 5

Search the tRF targets in tRFTar 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.

- Supplementarymaterial.pdf