

Supporting Information

tsRFun: a comprehensive platform for decoding human tsRNA expression, functions and prognostic value by high-throughput small RNA-Seq and CLIP-Seq data

Authors:

Jun-Hao Wang^{1,3}, Wen-Xin Chen¹, Shi-Qiang Mei¹, Yue-Dong Yang², Jian-Hua Yang^{1,*}, Liang-Hu Qu^{1,*}, Ling-Ling Zheng^{1,*}

Affiliation:

¹Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, Sun Yat-sen University, Guangzhou 510275, P. R. China

²National Supercomputer Center in Guangzhou, Sun Yat-sen University, Guangzhou, China

³School of Medicine, Sun Yat-sen University, Shenzhen, Guangdong, China.

***Corresponding author:**

Mailing address: Biotechnology Research Center, Sun Yat-sen University, Guangzhou 510275, P. R. China

Phone: 86-20-84112399

Fax: 86-20-84036551

Jian-Hua Yang, Email: yangjh7@mail.sysu.edu.cn

Liang-Hu Qu, Email: lssqlh@mail.sysu.edu.cn

Ling-Ling Zheng, Email: zhengl33@mail.sysu.edu.cn

Supplementary Figures

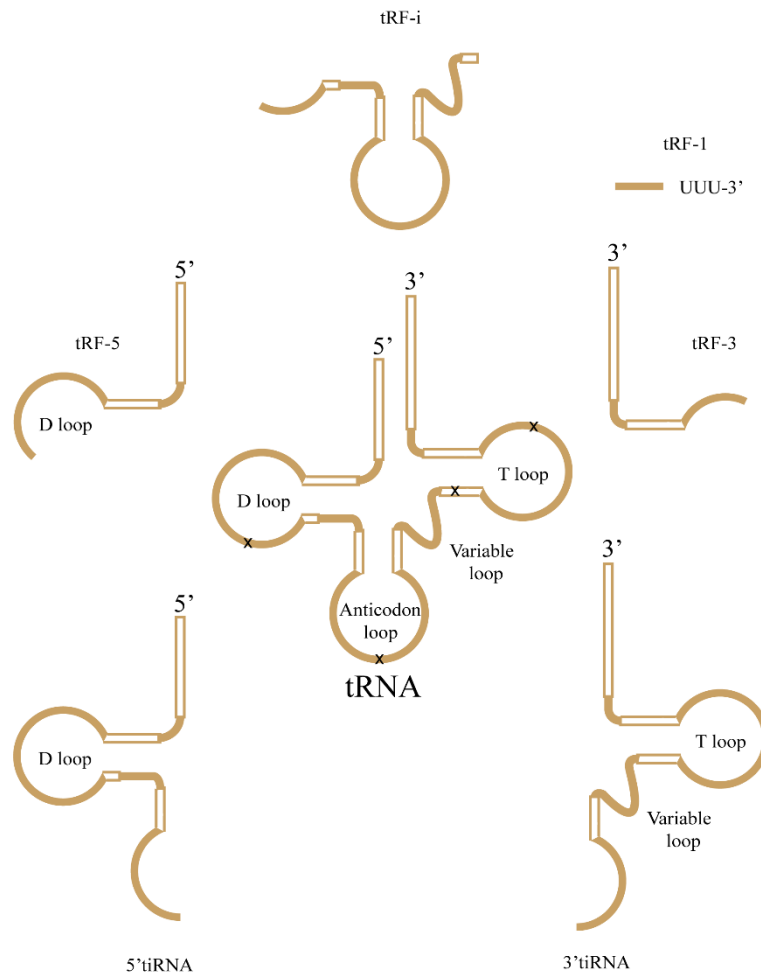


Figure S1. Biogenesis and classification of tRFs and tiRNAs. tRF-1 is generated from precursor tRNA and cleaved by RNase Z or ELAC2 at 3' trailer. tRF-3 and tRF-5 are generated from the 3' and 5' ends of the mature tRNAs, respectively. tRF-i is mainly from the internal region of mature tRNA. tiRNAs are generated by specific cleavage by angiogenin in the anticodon loops of mature tRNAs.

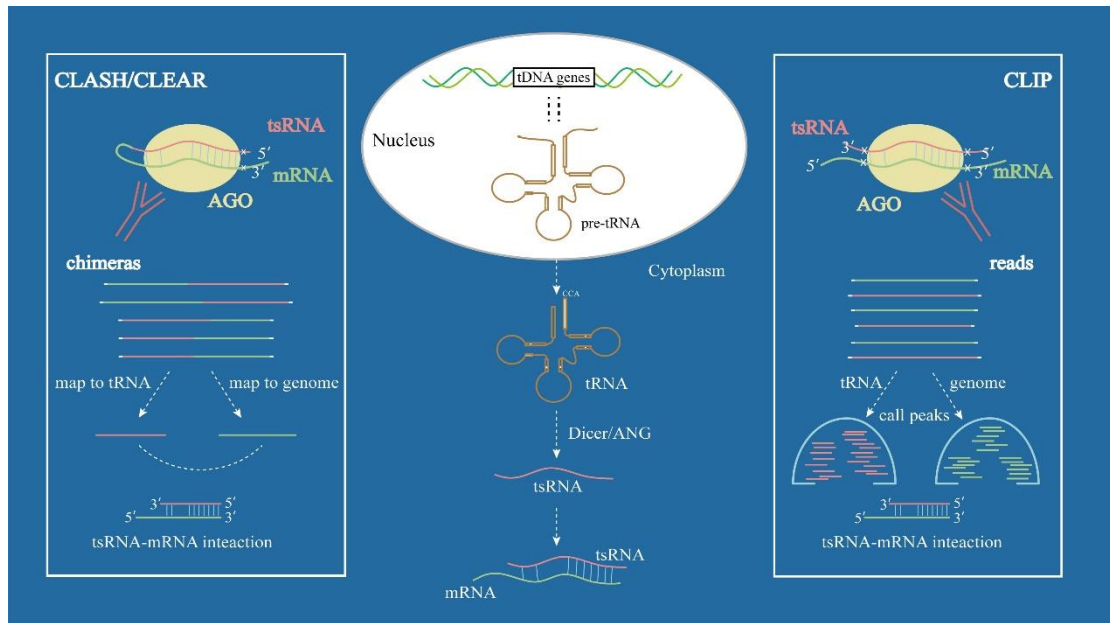


Figure S2. Schematic diagram of tsRNA target identification. For CLASH/CLEAR data (left), we match the chimeric reads to tRNA and genome reference respectively, and then detect their complementary pairing relationship. For CLIP data (right), we match the reads to tRNA and genome reference, then call peaks to find enriched clusters, and check the complementary pairing relationship between these clusters.

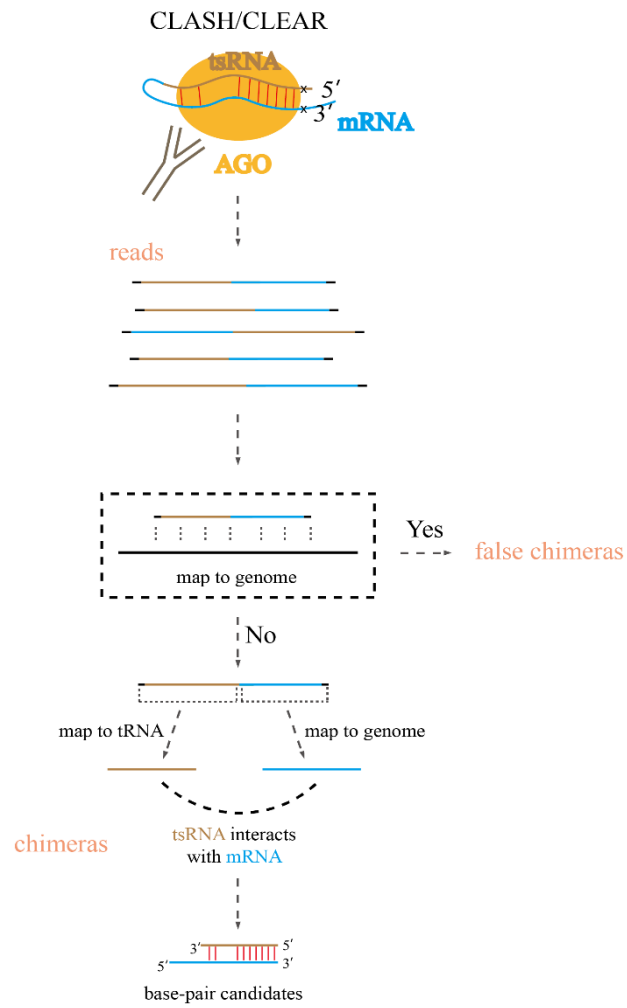


Figure S3. The analysis strategy to identifying tsRNA-target chimeras.

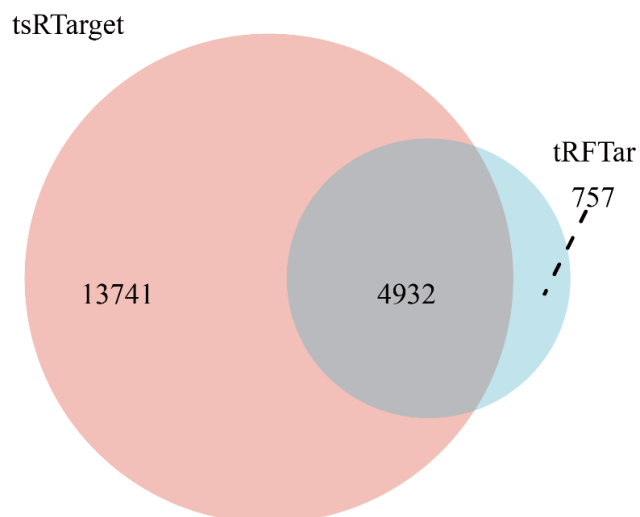


Figure S4. Venn diagram of tsRNA-mRNA chimera results identified by tsRTarget and tRFTar from CLASH/CLEAR and CLIP-Seq data.

Supplementary Tables

Table S1. A list of gene sets used in gene enrichment analysis.

ID	Name
1	All Canonical Pathways
2	All Immunologic Signatures
3	All Oncogenic Signatures
4	BioCarta Pathways
5	Cancer Gene Neighborhoods
6	Chemical and Genetic Perturbations
7	Disease Ontology
8	GO Biological Processes
9	GO Cellular Components
10	GO Molecular Functions
11	Hallmark GeneSets
12	KEGG Pathways
13	PANTHER Pathways
14	Reactome Pathways
15	Transcription Factor Targets

Table S2. Alignment of sequencing reads to tRNA-Gly-GCC-1-3 transcript. The first two lines of the table indicate the sequence and the RNA secondary structure of tRNA. Each next row represents a sequencing read and its corresponding position on the tRNA. The asterisk indicates the read identified as tsRNA by the corresponding tool or article.

tRNA-Gly-GCC-1-3 sequence	Read	Abundance	Lee	tsRFinder	Other
	Length		et al.		Tools
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGGGAGCCCGGGTTGCGATCCCGGCCCATGCACCA					
(((((((C.(((C.....))))).(((C.....))))).(((C.....))))))))))....					
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGGGAG.....	45	3			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGGGAG.....	44	2			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGGGA.....	43	2			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGGG.....	42	7			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCGG.....	41	1			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACGCG.....	40	19			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCACG.....	38	9			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCAC.....	37	24			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCCA.....	36	106			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGCC.....	35	544			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTGC.....	34	37			*
GCATGGGTGGTTCAGTGGTAGAATTCGCGCTG.....	33	370			*
GCATGGGTGGTTCAGTGGTAGAATTCGCG.....	30	3122		*	*

GCATGGTGGTTCAGTGGTAGAATTCTCG.....	29	148	*
GCATGGTGGTTCAGTGGTAGAATTCTC.....	28	85	*
GCATGGTGGTTCAGTGGTAGAATTCT.....	27	30	*
GCATGGTGGTTCAGTGGTAGAATTC.....	26	8	*
GCATGGTGGTTCAGTGGTAGAATT.....	25	20	*
GCATGGTGGTTCAGTGGTAGAAT.....	24	8	*
GCATGGTGGTTCAGTGGTAGAA.....	23	4	*
GCATGGTGGTTCAGTGGTAGA.....	22	15	*
GCATGGTGGTTCAGTGGTAG.....	21	5	*
GCATGGTGGTTCAGTGGTA.....	20	9	*
GCATGGTGGTTCAGTGGT.....	19	9	*
GCATGGTGGTTCAGTGG.....	18	15	*
GCATGGTGGTTCAGTG.....	17	2	*
GCATGGTGGTTCAGT.....	16	7	*
GCATGGTGGTTCAG.....	15	6	*

Table S3. A list of tsRNAs-target interactions detected by tsRTarget and tRFtar tool.

	tsRTarget	tRFtar	overlap
Targeted gene number	18,673	5,689	4,932

Table S4. The positive and negative sets in the simulated dataset.

	Read number	Read abundance
Positive set	100	82,027
Negative set	29,727	75,880
Total	29,827	157,907

Table S5. The performance among MINTmap, SPORTS, and tsRFinder tools.

	Precision (%)	Sensitivity (%)	Specificity (%)	False positive rate (%)	False negative rate (%)	Accuracy (%)
tsRFinder	81.82	81	99.94	0.06	0.19	99.86
MINTmap	0.73	61	71.99	28.01	0.39	71.95
SPORTS	0.72	88	59.27	40.73	0.12	59.37

Table S6. The prediction result of tsRFinder tool.

	Actual_true	Actual_false	Total
Predicte_true	81	18	99
Predicte_false	19	29,709	29,728
Total	100	29,727	29,827

Table S7. The prediction result of MINTmap tool.

	Actual_true	Actual_false	Total
Predicte_true	61	8,326	8,387
Predicte_false	39	21,401	21,440
Total	100	29,727	29,827

Table S8. The prediction result of SPORTS tool.

	Actual_true	Actual_false	Total
Predicte_true	88	12,108	12,196
Predicte_false	12	17,619	17,631
Total	100	29,727	29,827