Research Paper

The tsRNAs (tRFdb-3013a/b) serve as novel biomarkers for colon adenocarcinomas

Lihong Tan[1,*], Xiaoling Wu[1,*], Zhurong Tang[3], Huan Chen[2,3], Weiguo Cao[3], Chunjie Wen[3], Guojun Zou[2], Hecun Zou[3]

[1]Chongqing Medical and Pharmaceutical College, Chongqing 401331, China
[2]Yueyang Hospital of Traditional Chinese Medicine, Yueyang 414000, China
[3]Chongqing Medical University, Chongqing 400016, China
*Co-first authors

Correspondence to: Guojun Zou, Hecun Zou; **email:** yyzgj5589@163.com, https://orcid.org/0009-0007-9945-7199; zouhecun@cqmu.edu.cn
Keywords: tsRNA, tRFdb-3013a, tRFdb-3013b, colon adenocarcinomas, ST3GAL1
Received: July 13, 2023 **Accepted:** January 24, 2024 **Published:** March 6, 2024

ABSTRACT

The tsRNAs (tRNA-derived small RNAs) are a novel class of small non-coding RNAs derived from transfer-RNAs. Colon adenocarcinoma (COAD) is the most malignant intestinal tumor. This study focused on the identification and characterization of tsRNA biomarkers in colon adenocarcinomas. Data processing and bioinformatic analyses were performed with the packages of R and Python software. The cell proliferation, migration and invasion abilities were determined by CCK-8 and transwell assays. Luciferase reporter assay was used to test the binding of tsRNA with its target genes. With computational methods, we identified the tRNA fragments profiles within COAD datasets, and discriminated forty-two differentially expressed tsRNAs between paired colon adenocarcinomas and non-tumor controls. Among the fragments derived from the 3′ end of tRNA-His-GUG (a histidyl-transfer-RNA), tRFdb-3013a and tRFdb-3013b (tRFdb-3013a/b) were notably decreased in colon and rectum adenocarcinomas, especially, tRFdb-3013a/b might tend to be down-regulated in patients with lymphatic or vascular invasion present. The clinical survival of colorectal adenocarcinoma patients with low tRFdb-3013a/b expression was significantly worse than that of high expression patients. In colon adenocarcinoma cells, tRFdb-3013a could have inhibited cell proliferations, and reduced cell migration and invasion abilities. The enrichment analyses showed that most of tRFdb-3013a correlated-genes were enriched in the extracellular matrix associated GO terms, phagosome pathway, and a GSEA molecular signature pathway. Additionally, the 3′UTR of ST3GAL1 mRNA was predicted to contain the binding site of tRFdb-3013a/b, tRFdb-3013a/b might directly target and regulate ST3GAL1 expression in colon adenocarcinomas. These results suggested that tRFdb-3013a/b might serve as novel biomarkers for diagnosis and prognosis of colon adenocarcinomas, and act a key player in the progression of colon adenocarcinomas.

INTRODUCTION

The tsRNAs (tRNA-derived small RNAs) are a novel class of small non-coding RNA derived from transfer-RNA (tRNA), and also called as tRNA-derived fragments (tRFs) [1]. The tRNA fragments are produced through cleavage of mature or pre-mature tRNAs at various positions, the fragments are short and range in length from 16 to 35 nucleotides [2]. A general classification of tsRNAs is based on their position in relation to the parental tRNA sequence, which can be classified into some distinct categories, including 5′-tRFs, 3′-tRFs, and internal tRFs, these tsRNAs originate precisely from the span of mature tRNAs [2, 3]. Two remaining tsRNAs: 5′U-tRFs, which derived from the 5′-leader sequence of primary tRNA genes; and tRF-1

that comprise part of the 3′-trailer sequence [2]. Numerous researches suggested that the generation of tRNAs into specific small fragments was a regimented process, did not result from degradation [4]. These cleavages of tRNAs are context-dependent, including sex, ethnicity, disease, and tissue type of individuals. These dependencies increase the urgency of understanding the regulatory roles of tsRNAs, such efforts are gaining momentum, and comprise experimental and computational approaches [5, 6].

These efforts have discriminated more than thousands of microRNA isoforms via analysis of datasets in The Cancer Genome Atlas (TCGA) [7, 8]. It is worth mentioning that tsRNA is sequenced at an abundance comparable to that of the abundant miRNA [9]. Several studies have linked the tsRNAs to different human cancers, including glioma [10], liver [11] and breast cancer, etc. [12]. Growing evidences demonstrated that many tsRNAs were differentially expressed in the tissues of multiple tumor patients, these dysregulated tRNA fragments might contribute to various biology processes of cancer development and progression [13]. Colon adenocarcinoma (COAD) is the most common type of primary tumor originating in the intestinal tract [14], colon adenocarcinomas may be the higher mortality among gastrointestinal tumors based on epidemiological statistical reports [15]. The pathogenesis of colon adenocarcinomas is extremely complicated, as it involves the aberrant activation of proto-oncogenes and inactivation of anti-oncogenes [14, 16]. However, it is not clear whether the tsRNAs would serve as novel molecular biomarkers for diagnosis, prognosis and target therapy of colon adenocarcinomas.

In the present study, we aimed to identify the tsRNAs profiles within the sncRNA-seq data of human colon adenocarcinomas samples via a deterministic and exhaustive tsRNAs mining pipeline, then investigate the expression patterns and biology function of tRFdb-3013a and tRFdb-3013b in colon adenocarcinomas. Additionally, the relationships between tRFdb-3013a/b and its target genes are also studied in order to reveal the underlying mechanisms.

RESULTS

The identifications of tsRNAs within colon adenocarcinomas data

Firstly, we collected more than thousands of tsRNA signatures from tRFdb [9] and tRFexplorer databases, the tsRNA sequences have mapped within the specific regions of tRNA (transfer-RNA) as described previously [17]. As statistical analysis, these tsRNA fragments are mainly derived from 5′ end (5’-tRFs) and 3′ end of mature tRNA, as well as the 5′ leader (5′U-tRF) and 3′ trailer regions of primary tRNA-gene. With computational approaches, we determined two hundreds of tsRNAs and assessed their expression profiles within the sncRNA-seq data of TCGA-COAD, the summary of our identification pipeline was showed in Supplementary Figure 1A. After filtering, a total of 192 tsRNAs with available expression abundances were identified within COAD samples (Supplementary Table 3), of which about thirty-two percents were derived from the 3′ trailer regions (tRF-1s) of primary tRNA genes, and thirty percent tsRNAs from the 3′ end (3’-tRFs) of mature tRNAs (Supplementary Figure 1C). As for the chromosomes, more than thirty-five percent of tsRNAs are originated from the human chr1 region, and nearly twenty percent originated from the chr6 region (Supplementary Figure 1B).

Based on the differential expression analysis between paired colon adenocarcinoma samples and non-tumor controls, we discriminated a total of forty-two differentially expressed tsRNA-genes (DEGs), of which fifteen were down-regulated and twenty-seven were up-regulated (Figure 1A and Supplementary Table 4) in paired colon adenocarcinomas compared to non-tumor controls. Hierarchical cluster analysis and heatmap of differentially expressed tsRNA genes were performed and exhibited in Figure 1B. The expression patterns of tsRNAs were significantly distinguished between paired colon adenocarcinomas and non-tumor groups. As shown in the blue dotted line frame of Figure 1B, several tsRNAs (including ts-43, ts-44, tRFdb-3013a and tRFdb-3013b) caught our attention that derived from tRNA-His-GUG, which is a histidyl-transfer-RNA [18].

Significance of tRFdb-3013a and tRFdb-3013b within colon adenocarcinomas

As shown in Figure 2A, we have identified seven tsRNA fragments derived from tRNA-His-GUG or *tRNA-His-GTG* gene in the sncRNA-seq datasets of colon adenocarcinomas, there are two 3’-tRFs fragments (tRFdb-3013a and tRFdb-3013b) that can re-map to the cleavages of 3′ end of several different mature tRNA-His-GUG isoforms (Supplementary Figure 2A), and five tRF-1 fragments (ts-30, ts-43, ts-44, ts-1 and ts-88) that can re-map to the down-stream regions (35 bases) of primary *tRNA-His-GTG* genes. To reveal the implication of these tsRNAs in colon adenocarcinomas, the expression patterns of above tsRNAs were analyzed in the samples of TCGA-COAD dataset. It showed that the expressions of four tsRNAs were down-regulated in colon adenocarcinoma samples compared to paired controls (Supplementary Figure 2B, all P <0.05, Supplementary Table 5), and compared with non-tumor controls, the expressions of tRFdb-3013a and tRFdb-3013b

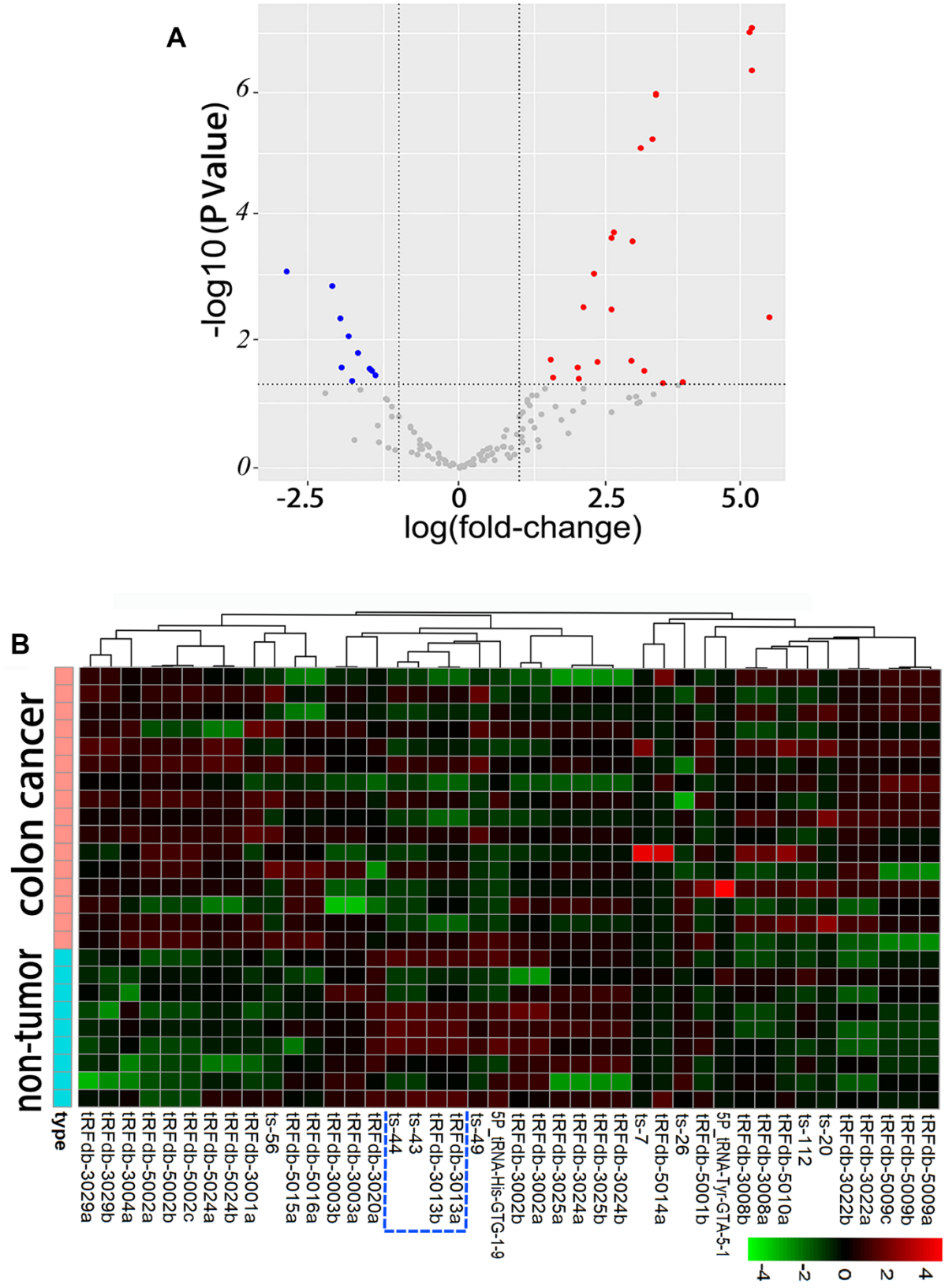


Figure 1. Differential expression analysis between paired colon adenocarcinomas (n=16) and non-tumor controls (n=9) of TCGA-COAD dataset. (**A**) Volcano plots of the differentially expressed tsRNAs, including 15 up-regulated and 27 down-regulated tsRNAs. (**B**) Hierarchical clustering and heatmap of differentially expressed tsRNAs, the up-regulated tsRNAs are clustered in red-shaded areas, and the green-shaded areas indicate down-regulation.

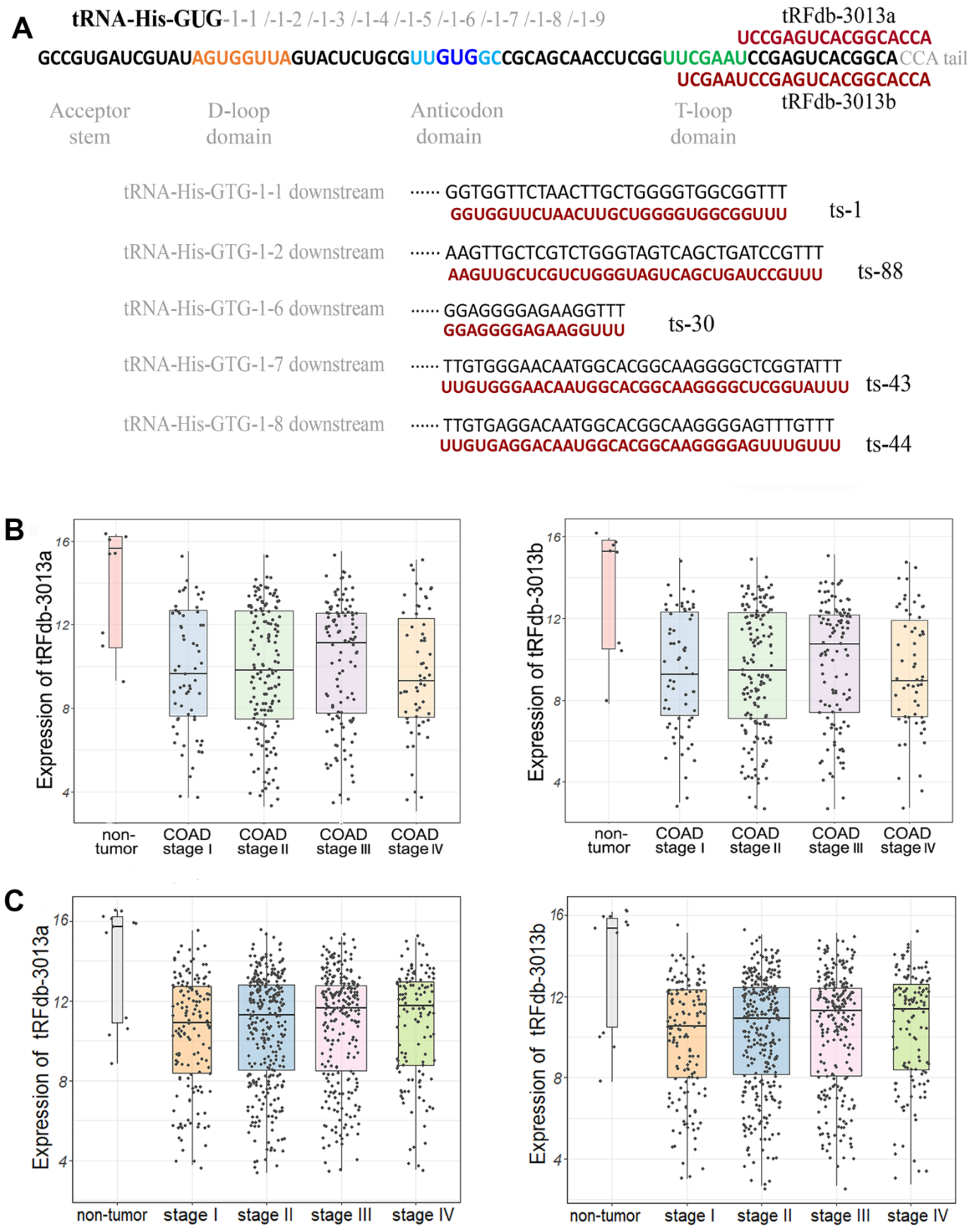


Figure 2. (**A**) Representative images of the sequence alignments between tsRNAs and its tRNA sources. (**B**) The scatter plus box plots of tRFdb-3013a and tRFdb-3013b expression (log_2TPM) between non-tumor controls and colon adenocarcinomas (COAD) with four stages (both P <0.01). (**C**) The scatter plus box plots of tRFdb-3013a and tRFdb-3013b expression (log_2TPM) between non-tumor controls and colon plus rectal adenocarcinomas (COAD-READ) with four stages (both P <0.01).

(known simply as tRFdb-3013a/b) were significantly down-regulated in colon adenocarcinoma and/or rectum adenocarcinoma samples with four stages (Figure 2B, 2C, all P <0.01). In our clinical tissue samples detections, it also demonstrated that the tRFdb-3013a/b expressions were down-regulated in colon adenocarcinoma compared to non-tumor controls (Supplementary Figure 3A, 3B).

Subsequently, we investigated the relationship between tsRNA expressions and overall survival of colon and/or rectum adenocarcinoma samples using KM (Kaplan–Meier) curve analysis [10]. Based on the median value of tsRNAs expression, primary tumor samples were divided into low-expressions and high-expressions group [19]. As shown in Figure 3B, the clinical survival of colon adenocarcinoma patients with low-expressions of tRFdb-3013a or tRFdb-3013b was worse than that of the high-expression patients (P =0.029 and P =0.026); As for colon adenocarcinoma plus rectum adenocarcinoma (COAD-READ), the clinical survival of patients with low-expressions of tRFdb-3013a or tRFdb-3013b were also significantly worse than that of the high-expression patients (P =0.018 and P =0.011, Figure 3A). What's more, the Cox regression analysis of overall survival within

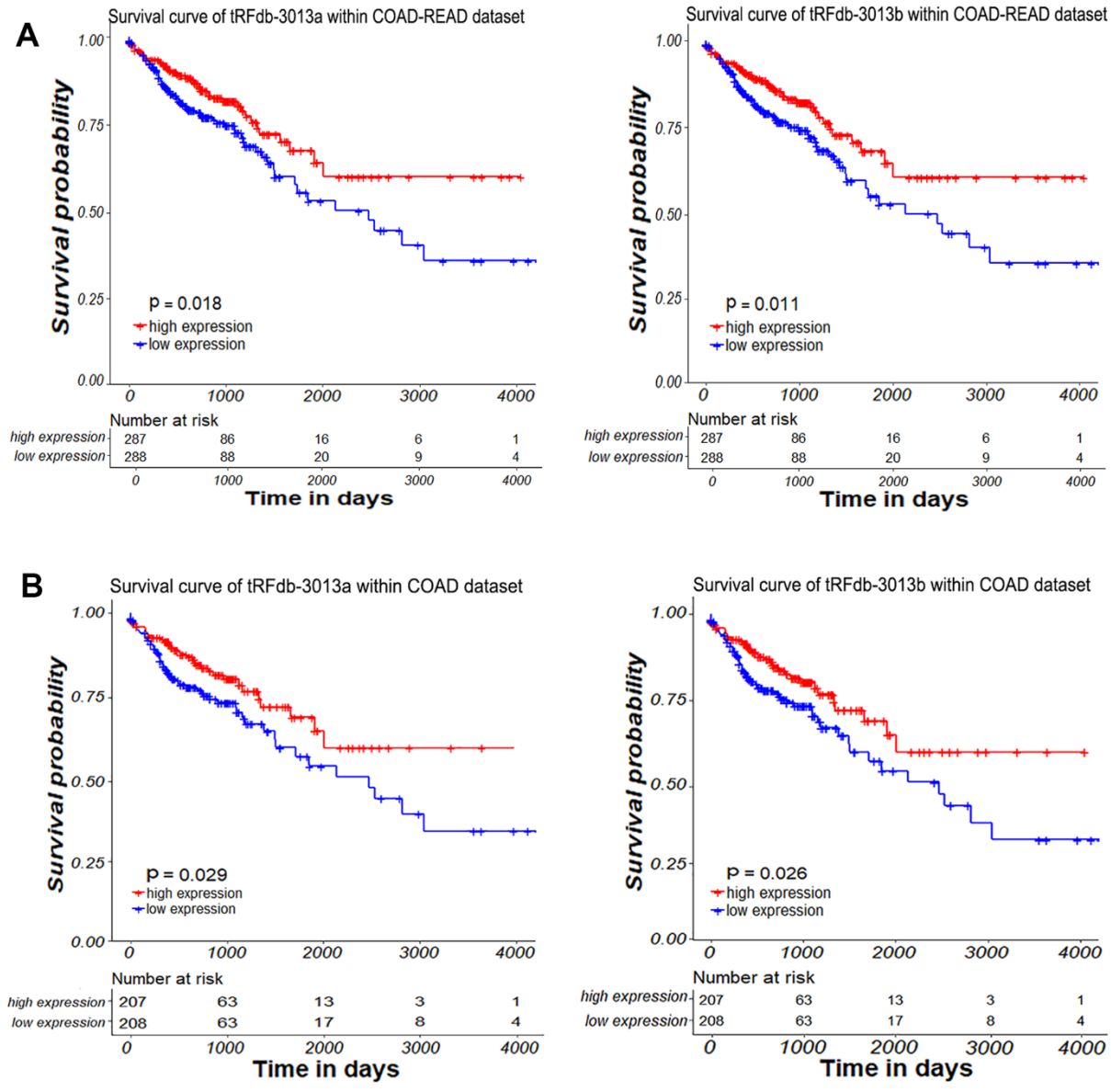


Figure 3. The survival analysis of tRFdb-3013a and tRFdb-3013b. (**A**) KM curve survival analysis with log-rank tests based on tRFdb-3013a and tRFdb-3013b expression within colon adenocarcinoma plus rectum adenocarcinoma (COAD-READ) samples (n=805) of TCGA datasets. (**B**) KM survival analysis with log-rank tests based on tRFdb-3013a and tRFdb-3013b expression within colon adenocarcinoma samples (n=587) of TCGA-COAD dataset.

COAD-READ patients were performed [20]. As shown in Figure 4A, the forest plot described that univariate Cox regression analysis of patients showed that low tRFdb-3013a expression (P =0.012), increased age (P <0.001), high AJCC pathologic stage (P <0.01 for stage III and P <0.001 for stage IV), advanced metastasis pathologic spread (P <0.001 for M1) and lymph node pathologic spread (P <0.001 for N2) were the risk factors associated with prognosis. Subsequent multivariate Cox regression analysis results revealed that low tRFdb-3013a expression (HR=1.684, P = 0.006) were an independent prognosis factor for survival of patients (Figure 4B), in addition to the increased age, advanced metastasis and lymph node pathologic spread. Similar results were also obtained from the Cox regression analysis of tRFdb-3013b. These results suggested that down-regulated tRFdb-3013a/b was associated with poor survival prognosis of colon adenocarcinoma and rectum adenocarcinoma patients.

Nevertheless, does the survival prognosis role of tRFdb-3013a/b specific to colorectal adenocarcinomas, and how about other types of tumors? In order to understand the clinical significances of tRFdb-3013a and tRFdb-3013b within multiple types of tumors, Cox regression analysis with hazard ratio test based on tRFdb-3013a/b

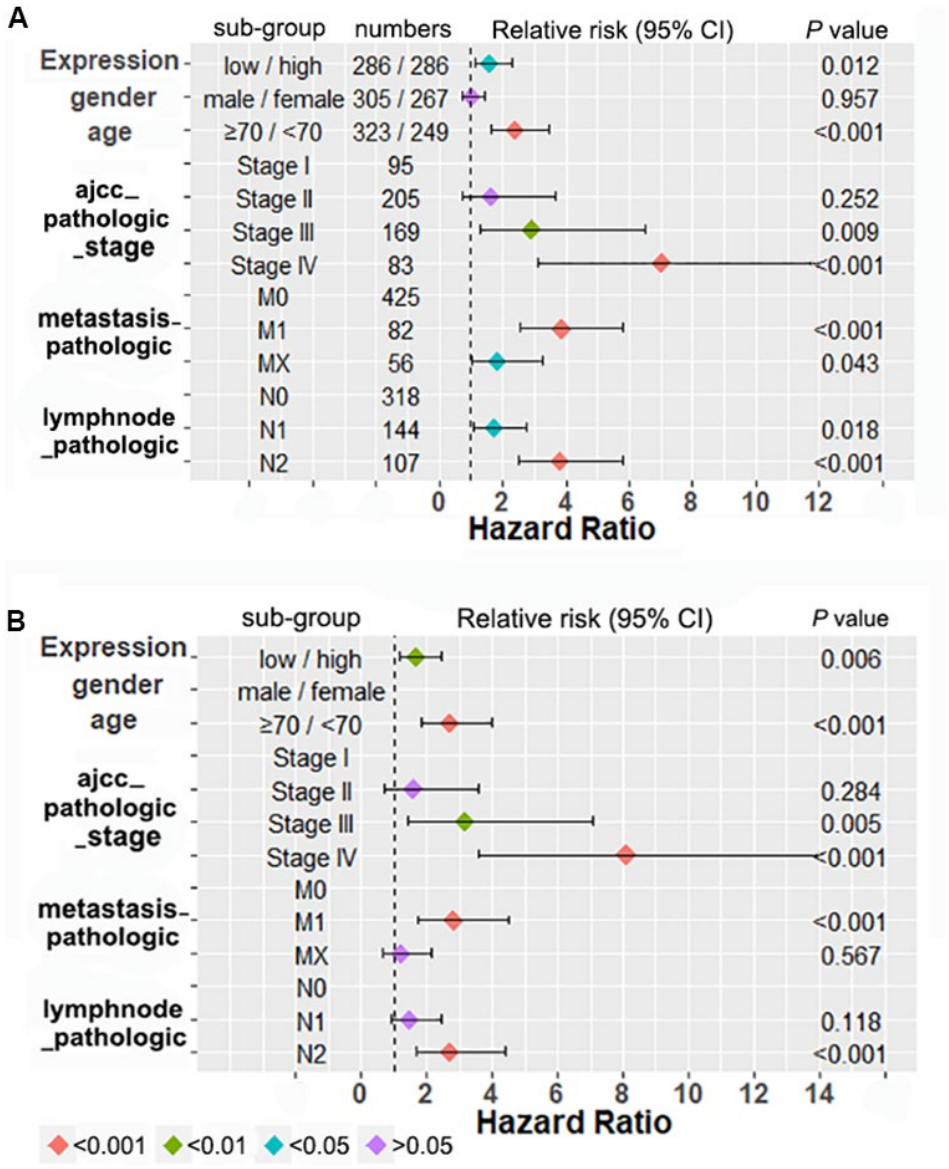


Figure 4. Cox regression analysis of overall survival within colon adenocarcinoma plus rectum adenocarcinoma patients. (**A**) Univariate analysis of tRFdb-3013a expressions and other risk factors within patients of TCGA dataset. (**B**) Multivariate Cox regression analysis of tRFdb-3013a expressions and other risk factors within patients.

expressions were performed within many types of tumors on TCGA datasets. As shown in the left of Figure 5, the forest plot described that tRFdb-3013a and tRFdb-3013b down-regulation were the risk factors associated with prognosis for five types of tumors in TCGA datasets, including COAD, READ, HNSC, KIRC, and LGG (all $P \leq 0.03$). Additionally, the differential expression analysis with log2 fold-change test of tRFdb-3013a and tRFdb-3013b were performed between primary tumor samples and paired non-tumor controls. As expected, the bar graphs showed that tRFdb-3013a/b expressions were markedly down-regulated in primary tumor samples compared to non-tumor controls for other six types of tumors ($P<0.01$, $|\log(FC)| >1$, Figure 5-right). Taken together, the results implicated that only within colon and rectum adeno-carcinoma, the expressions of tRFdb-3013a and tRFdb-3013b were both down-regulated, and could be served as the independent risk biomarkers for prognosis.

The correlations of tRFdb-3013a/b expression with primary clinical pathological characteristics were analyzed using *Chi*-square test, and visualized via hierarchical-clustering heatmaps [21]. As shown in Supplementary Figure 4, tRFdb-3013a/b might tend to be decreased in the patients with lymphatic or vascular invasion present. However, no significant relationship was found between tRFdb-3013a/b expressions and other pathological characteristics, such as grade, gender, MLH1 silencing, methylation subtype, colon polyps and MSI status, etc. (all $P > 0.05$). Above all, findings demonstrated that tRFdb-3013a and tRFdb-3013b may serve as novel biomarkers for diagnosis and prognosis of colon adenocarcinomas.

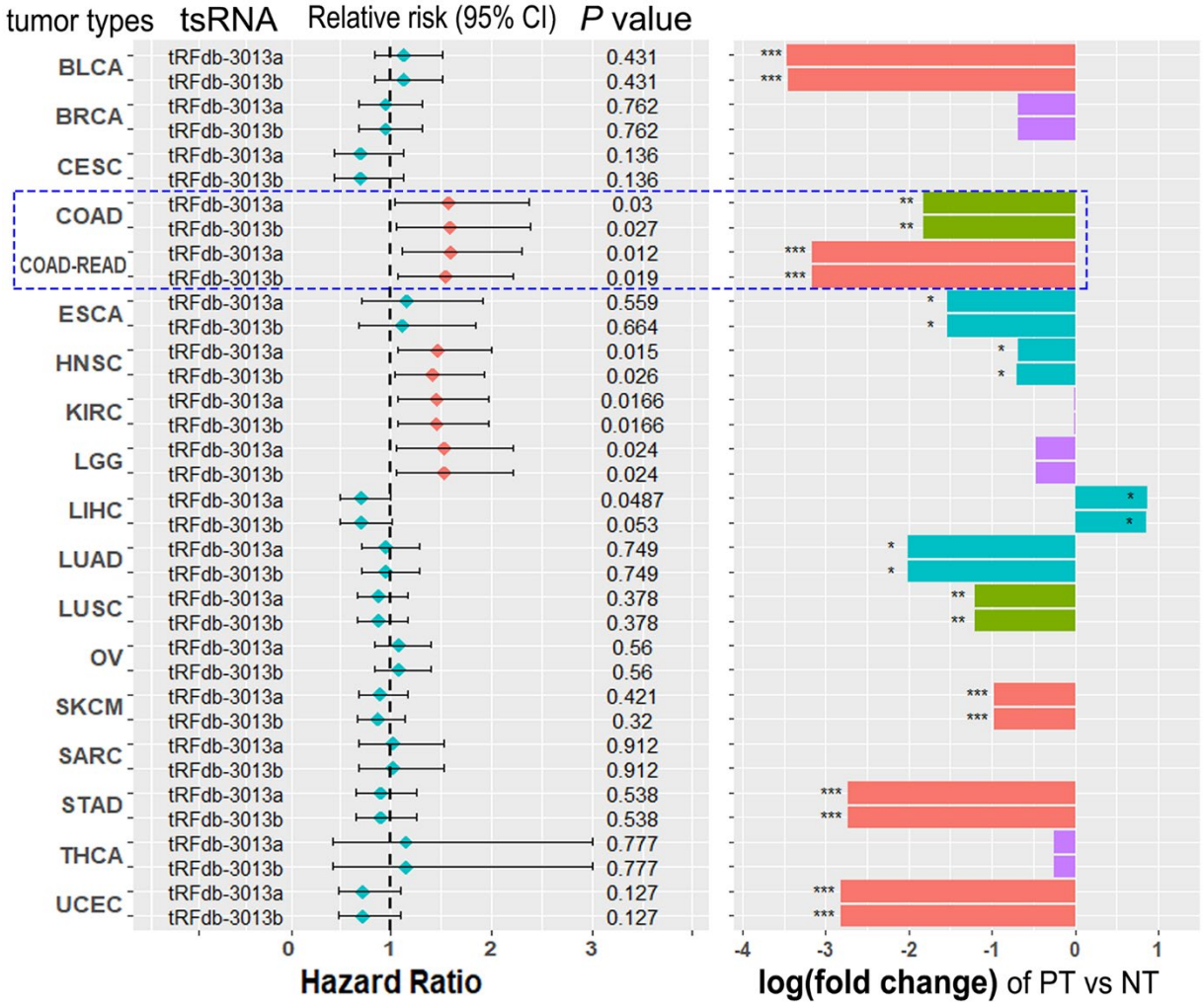


Figure 5. The clinical significances of tRFdb-3013a and tRFdb-3013b within multiple different types of tumors. (Left) The forest plots of Cox regression analysis with hazard ratio test based on tRFdb-3013a and tRFdb-3013b expressions within many types of tumors on TCGA dataset. (Right) The bar graph of differential expression analysis with a log2 fold-change test of tRFdb-3013a and tRFdb-3013b between primary tumor (TP) samples and non-tumor (NT) controls in multiple types of tumors.

Biological insights of tRFdb-3013a/b on colon adenocarcinoma cells

To further explore the roles of tRFdb-3013a/b on colon adenocarcinomas. The mimic or inhibitor was used to over-expressed or inhibit tRFdb-3013a in cancer cells (Supplementary Figure 5A). Cell proliferations of colon adenocarcinomas were detected using CCK8 assay, it showed that the proliferations were significantly suppressed in cells transfected tRFdb-3013a mimic compared to negative control (NC; both $P<0.05$; Figure 6A). Transwell and scratch assays were performed to test the effect of tRFdb-3013a on invasion and migration abilities of colon adenocarcinoma cells. The results displayed that migratory rates of cells transfected with tRFdb-3013a mimic were reduced in comparison with negative controls (Figure 6B and Supplementary Figure 5B); the numbers of invasive cells with tRFdb-3013a mimic were significantly reduced, and that cell numbers with tRFdb-3013a inhibitor were increased than those of negative controls (Figure 6C). These suggested that tRFdb-3013a might be effective to regulate the cell proliferations, cell migration and invasion abilities of colon adenocarcinomas.

Moreover, the enrichment analyses were performed to provide a biological insight concerning the roles of tRFdb-3013a/b, mRNA expression profiling of TCGA-COAD data was used to find the 500-top related-genes of tRFdb-3013a/b via *Spearman*'s correlation method. As shown in Figure 7A, for gene ontology (GO) analysis [22], some tRFdb-3013a related-genes were enriched in the specific GO terms, such as extracellular matrix organization, collagen-containing extracellular matrix, and extracellular matrix structural constituent (EMILIN1, ADAT2, FBLN1). The dot-plots (Figure 7B) displayed that tRFdb-3013a related-genes were enriched in several KEGG pathway, including phagosome (C1R). In the GSEA analyses [23] (Figure 7C), a molecular signature (LMTK3 target genes) was found to be associated with tRFdb-3013a, such as FBLN1 and ILVBL-AS1 were both the tRFdb-3013a related-genes. The enrichment analysis of tRFdb-3013b were also conducted and presented in Supplementary

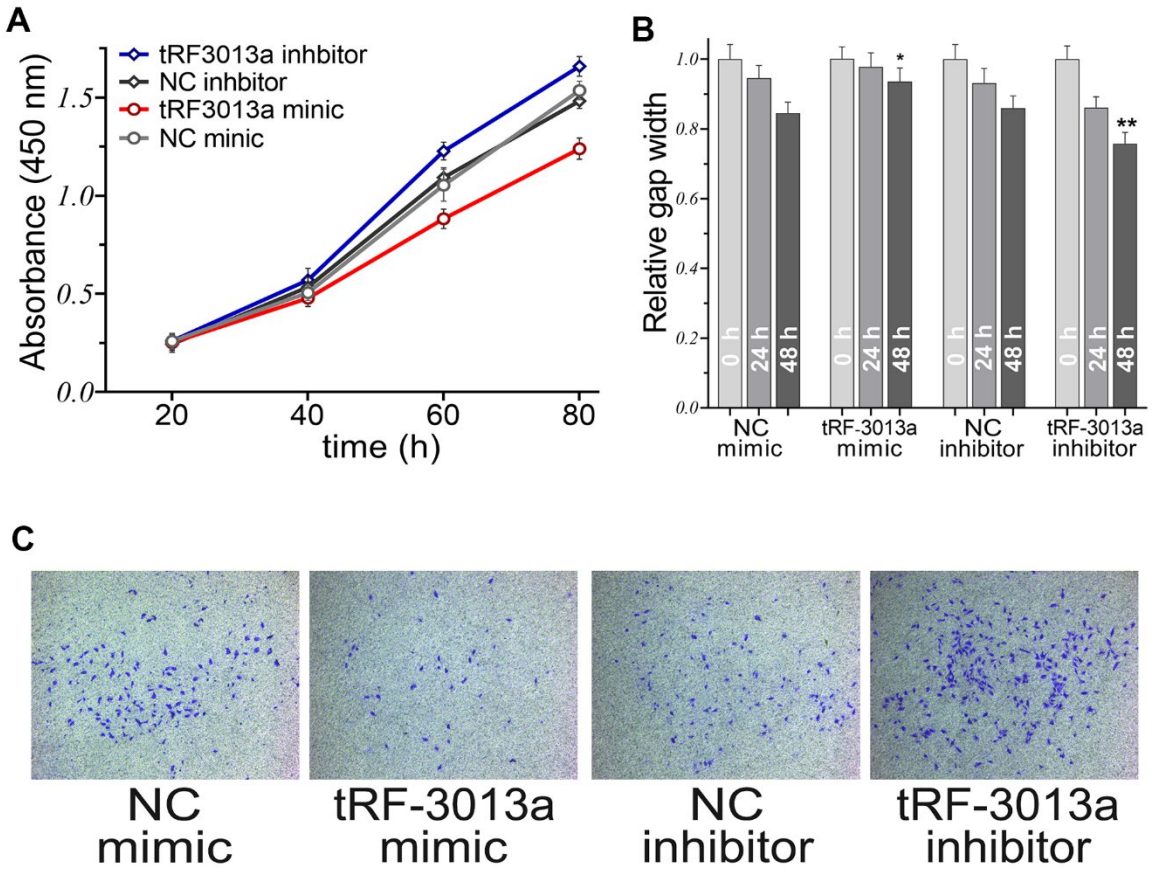


Figure 6. The roles of tRFdb-3013a on cell proliferations, cell migration and invasion abilities of colon adenocarcinomas. (**A**) CCK8 assay was performed to measure the cell proliferations of colon adenocarcinoma SW480 cell. (**B**) Relative gap analysis scratch assays were used to assess the migration ability of SW480 cell. (**C**) Transwell assays were performed to assess cell invasions of colon adenocarcinomas. ** $P < 0.01$, * $P < 0.05$. The "tRF-3013a" means as tRFdb-3013a, the "NC" means as negative control.

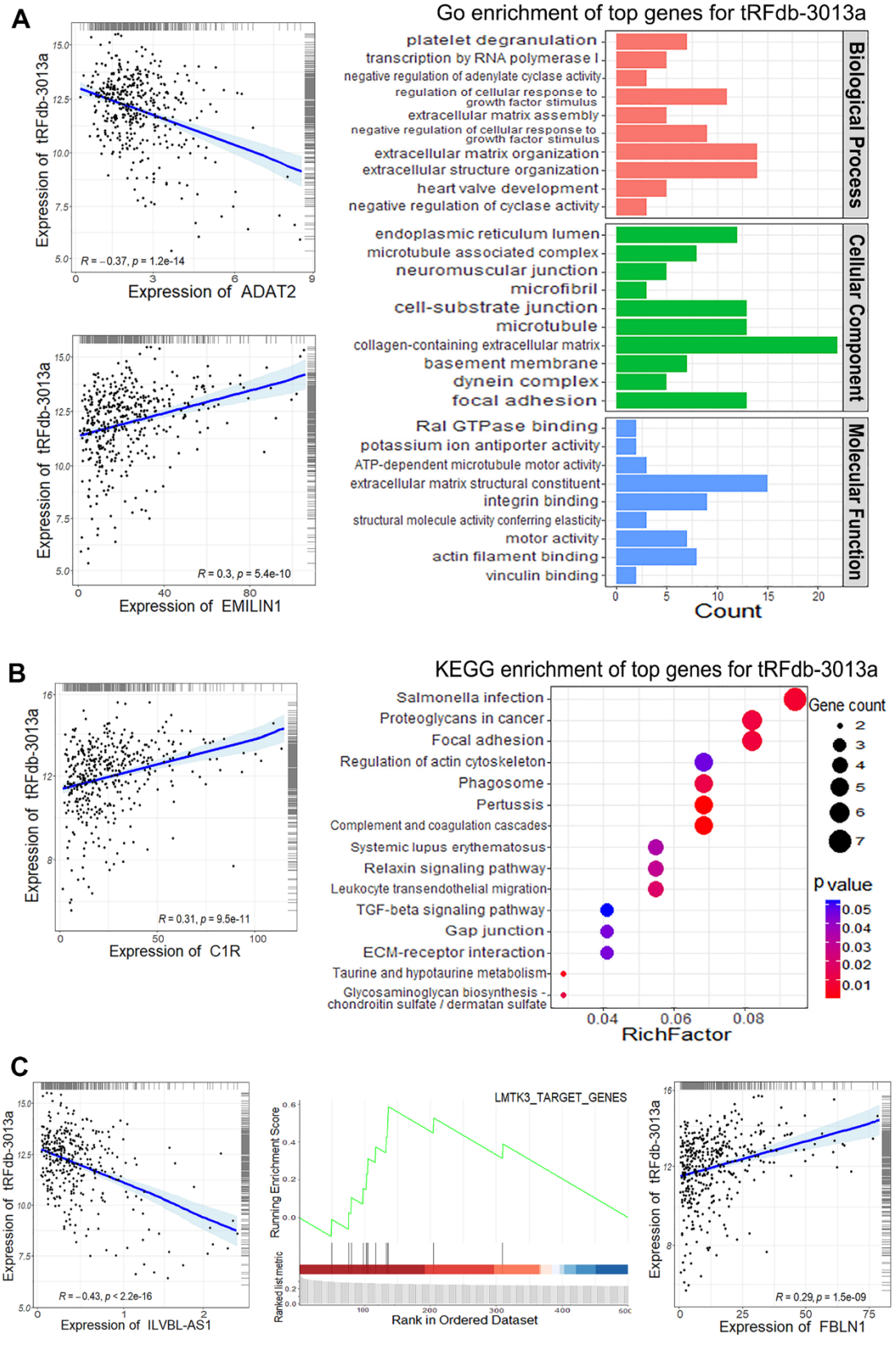


Figure 7. The enrichment analysis of tRFdb-3013a related-genes within TCGA-COAD data. (**A**) The top gene ontology (GO) terms, including biological process, cellular component and molecular function. (**B**) The top KEGG pathway for tRFdb-3013a related-genes. The correlation scatter-plots of tRFdb-3013a and its related-genes (ADAT2, EMILIN1, and C1R). (**C**) GSEA (gene set enrichment analysis) plots of the molecular signatures, the scatter plots for tRFdb-3013a and its related-genes (ILVBL-AS1, FBLN1).

Materials (Supplementary Figure 6). These data suggested that tRFdb-3013a/b might play a critical role in tumor progression of colon adenocarcinomas.

The tRFdb-3013a/b target and regulate ST3GAL1 expression in colon adenocarcinomas

The researches have revealed that tRNA fragments could bind to argonaute (AGO) protein and lead to gene silencing by targeting the 3' untranslated region (3'UTR) of some mRNAs in a manner similar to miRNAs [5, 11]. Via using bioinformatics database and tools [24], many potential targets of tRFdb-3013a/b were predicted and selected out, the interaction relationship of tRFdb-3013a/b and their target genes were presented in Figure 8A.

Among these targets, we picked out a molecular with strong binding score, *ST3GAL1* gene, which encodes a type II membrane protein with sialyl-transferase activity [25]. The 3′UTR position of *ST3GAL1* mRNA contained a putative binding site of tRFdb-3013a/b (Figure 8B and

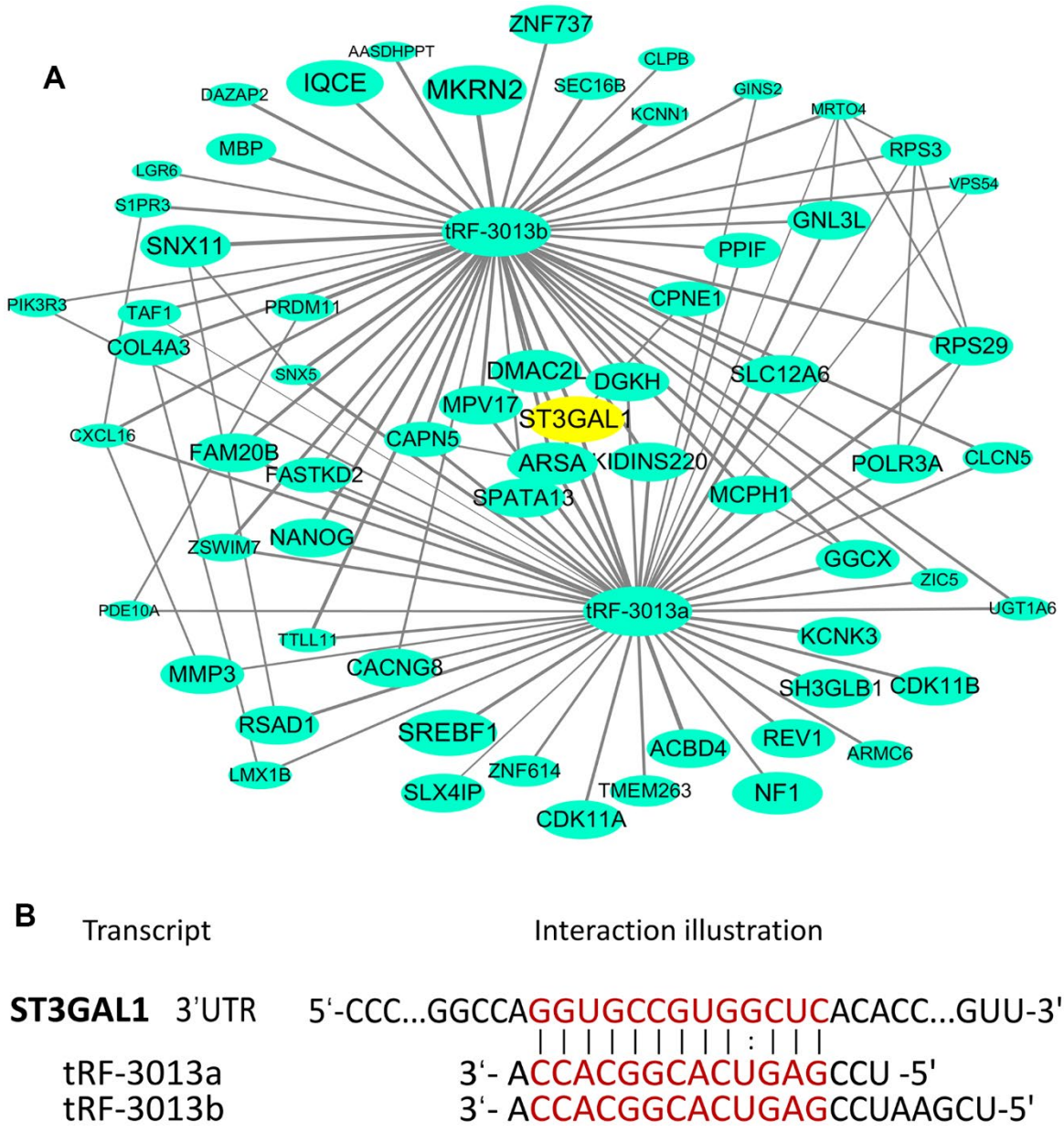


Figure 8. (**A**) The interactional network that involved tRFdb-3013a/b and their target genes based on tRFtarget database, the nodes represent target genes, and the lines represent the interaction relationships. (**B**) The binding sites of tRFdb-3013a/b in the 3' UTR region of *ST3GAL1* based on bioinformatics prediction.

Supplementary Figure 7). Dual-luciferase reporter assays were used to validate their direct targeting affinity, it showed that relative luciferase activity was decreased in the co-transfection of pmirRB-ST3GAL1 wild-type reporter and tRFdb-3013a/b mimic (both $P < 0.05$, Figure 9B-left), but no significant difference in the mutant reporter. Moreover, the analysis of TCGA-COAD and GSE39582 data showed that ST3GAL1 was highly expressed in colon adenocarcinomas compare to non-tumor controls (Figure 9A, both $P < 0.01$), as mentioned above, tRFdb-3013a and tRFdb-3013b were conversely decreased in colon adenocarcinomas (Figure 2B). In addition, the experiments demonstrated that expression levels of ST3GAL1 mRNA and protein were down-regulated in colon adenocarcinoma cells transfected with tRFdb-3013a/b mimic (both $P < 0.01$, Figure 9B-right, 9C). These findings indicated tRFdb-3013a and tRFdb-3013b might directly target ST3GAL1 and regulate ST3GAL1 expressions in colon adenocarcinoma cells.

DISCUSSION

Identifications and characterizations of tsRNAs are accelerated sharply in the last few years, due in large part to the emergence of next generation sequencing system-level approaches to molecular genetics [5]. In the present study, we firstly identified some tsRNAs within the sncRNA-seq datasets of colon adenocarcinoma samples, most of tsRNAs were derived from the 3′ end of mature tRNA or 3′ trailer regions of primary tRNA genes. Furthermore, a total of forty-two differentially expressed tsRNA genes were discriminated by bioinformatics analysis, of which fifteen were down-regulated and twenty-seven were up-regulated between colon adenocarcinomas and non-tumor controls. In our previous cytobiology study for microRNA, we have noted that it is easy to get the over-expression of microRNA with miRNA mimic, but hard to inhibit or decrease microRNA expression via miRNA inhibitor *in vitro*. Similarly, for the down-regulated tsRNA, we could make it over-expressed in cells via tsRNA mimic. Therefore, we selected several downregulated tsRNAs for further research. Among them, we picked out seven tsRNAs derived from tRNA-His-GUG, which was a histidyl-transfer-RNA [18]. There were two 3'-tRFs fragments (tRFdb-3013a, whose tDRnamer ID is tDR-60:76-His-GTG-1-M2; and tRFdb-3013b, whose tDRnamer ID is tDR-55:76-His-GTG-1-M2 [26]) can re-map to the cleavage of 3′ end of mature tRNA-His-GUG, the expressions of tRFdb-3013a and tRFdb-3013b were markedly down-regulated in colorectal adenocarcinomas compared to non-tumor controls. KM curve analysis displayed that clinical survival of colon adenocarcinoma patients with low tRFdb-3013a/b expressions were significantly worse than that of the high-expression patients, it suggested that down-regulated tRFdb-3013a and tRFdb-3013b were associated with poor survival prognosis of patients with colon adenocarcinoma and/or rectum adenocarcinoma. In the further analyses of other tumor types from TCGA datasets, however, only within colon adeno-carcinoma (COAD) and/or rectum adenocarcinoma (READ), the expressions of tRFdb-3013a and tRFdb-3013b were both down-regulated, and could be served as risk biomarkers for prognosis, not for other type tumors. Further analyses revealed that tRFdb-3013a/b expression might tend to decrease in the patients with lymphatic or vascular invasion present. These findings implied that tRFdb-3013a and tRFdb-3013b might serve as novel biomarkers for diagnosis and prognosis of colorectal adenocarcinomas.

Growing evidences have demonstrated that dysregulated tRNA fragments may contribute to various biological processes of cancer development and progression [13]. For example, Tao et al. have revealed the expression profiles of tsRNAs in human colorectal cancer (CRC) specimens via sequencing analyses, and identified a specific tsRNA derived from the internal of mature histidyl-tRNA, 5'tiRNA-His-GTG (its base sequence is GCCGTGATCGTATAGTGGTTAGTACTCTGCGTTGT), which is up-regulated in CRC tissues [11]. Then, some experiments have revealed that an inhibition of 5'tiRNA-His-GTG suppressed cell proliferations and colony formations, induced cell apoptosis, and affected tumor growth in nude mice, these indicate that 5'tiRNA-His-GTG may act an oncogenic role in the progression of colorectal cancers [11]. In glioma, Ren et al. have determined the differential expressed tsRNAs between tumor tissues with matched non-tumor controls, and found some tRNA-Cys-GCA derived tsRNAs were significantly decreased in tissues from gliomas. Moreover, lower expression of tRNA-Cys-GCA derived tsRNA is associated with poor survival outcome of glioma patients [10]. In present study, certain cyto-biology assays were performed to provide a biological insight concerning the roles of tRFdb-3013a on colon adenocarcinoma cells. It demonstrated that tRFdb-3013a over-expression suppressed cell proliferations, migration and invasion abilities of colon adenocarcinomas *in vitro*. Moreover, the enrichment analysis showed tRFdb-3013a was associated with specific GO terms, such as extracellular matrix organization, collagen-containing extracellular matrix, and extracellular matrix structural constituent (EMILIN1, ADAT2, FBLN1); the tRFdb-3013a related-genes might refer to phagosome pathway (C1R), as well as the GSEA molecular signature genes (FBLN1 and ILVBL-AS1). These results indicated that down-regulated tRFdb-3013a might act a suppressive role in cell migration and invasion of colon adeno-carcinomas via specific signaling pathway.

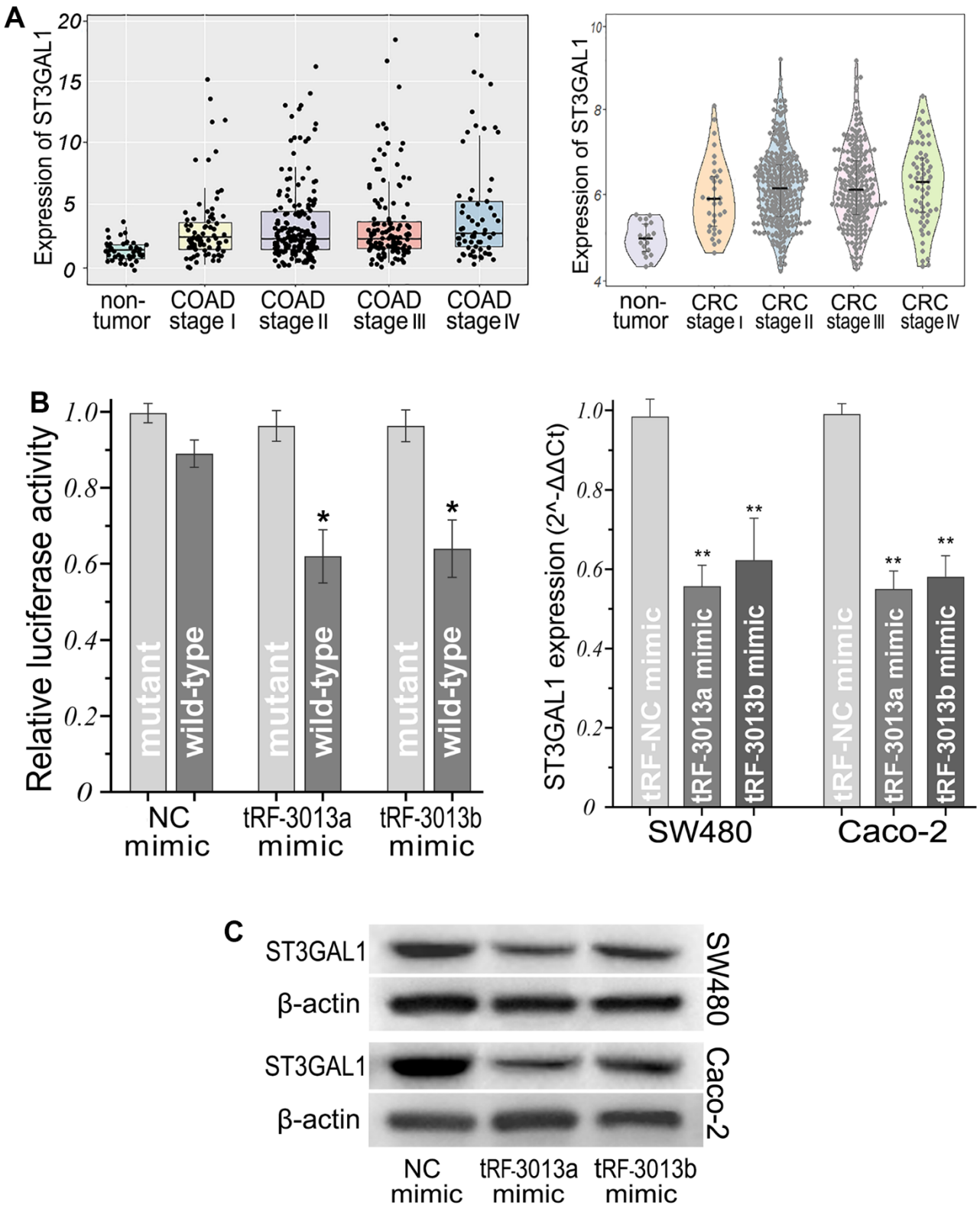


Figure 9. ST3GAL1 was directly regulated by tRFdb-3013a and tRFdb-3013b in colon adenocarcinomas. (**A**) ST3GAL1 was highly expressed in colon adenocarcinomas compared to non-tumor controls of TCGA-COAD and GSE39582 data (log_2TPM, both P <0.01), COAD: colon adenocarcinomas, CRC: colorectal cancers. (**B**-left) The luciferase reporter assay of colon adenocarcinoma cell co-transfected with pmiR-RB-ST3GAL1 wild-type or mutant report vector, together with tRFdb-3013a/b or NC mimic. (**C**) Expression levels of ST3GAL1 mRNA (**B**-right) and protein in colon adenocarcinomas cells transfected with tRFdb-3013a/b mimic or control. ** P < 0.01, * P < 0.05. The "tRF-3013a/b" means as tRFdb-3013a/b, "NC" means as negative control.

Though the biological function of tsRNAs is complex and require further elucidation, current knowledge of their function has been a key focus on cancer researches. Previous researches have revealed that tsRNAs could bind to AGO proteins and play a critical role by directly sponging the 3'UTRs of some mRNAs in a manner similar to miRNAs [5]. For example, Goodarzi et al. have identified the new class of tsRNAs derived from several tRNA through hypoxic stress induction, these tsRNAs could bind to the 3'UTRs of *YBX1* gene, which encoded an RNA-binding protein, then suppressed the stability of multiple oncogenic transcripts, including *EIF4EBP1* and *AKT1*, results in the tumor-suppressive and metastasis-suppressive activities in breast cancers [12]. Our bioinformatics analyses displayed the interaction relationships of tRFdb-3013a/b and their predicted target genes, and found that tRFdb-3013a/b could bind to the 3'UTR of *ST3GAL1*, which was highly expressed in colon adenocarcinomas. ST3GAL1 is a membrane protein with sialyl-transferase activity, which has been reported to promote tumor metastasis and invasion by previous study [25]. The luciferase reporter assay showed that tRFdb-3013a/b could directly target 3′UTR of ST3GAL1 and decrease its relative luciferase activity. In addition, tRFdb-3013a over-expressions could down-regulate the expression levels of ST3GAL1 protein and mRNA in colon adenocarcinoma cells. These results suggest tRFdb-3013a and tRFdb-3013b might directly target the 3′UTR of ST3GAL1 and regulate ST3GAL1 expressions in colon adenocarcinomas.

CONCLUSIONS

Taken together, our investigation identified the tsRNAs profiles, particularly several tsRNAs derived from tRNA-His-GUG, in the sncRNA-seq data of colon adenocarcinoma samples. Among these, tRFdb-3013a and tRFdb-3013b were significantly down-regulated in colon and rectum adenocarcinomas, the expressions of tRFdb-3013a/b were associated with clinical survival prognosis of colon and rectum adenocarcinoma patients. Over-expression of tRFdb-3013a inhibited cell proliferations, cell migrations and invasions of colon adenocarcinomas. The down-regulated tRFdb-3013a might act as an anti-oncogene role in colon adenocarcinomas via specific signaling pathway. Additionally, tRFdb-3013a/b could directly target the 3′UTR of ST3GAL1 and regulate ST3GAL1 expressions in colon adenocarcinomas. These results implied that tRFdb-3013a/b may play an important role in the progression of colon adenocarcinomas, and could be explored as novel biomarkers for diagnosis and prognosis of colon and rectum adenocarcinomas.

MATERIALS AND METHODS

Bioinformatic analysis

The RNA expression profiles of TCGA datasets were downloaded and processed using TCGAbiolinks package [8]. The differential expression analysis between paired colon adenocarcinoma samples and non-tumor controls was performed with the limma package [33]. An absolute log2 fold-change ($|\log_2 FC|$) more than one and *P*-value less than 0.05 were set as cut-off point. Hierarchical clustering of differentially expressed tsRNA genes was performed based on expression values to verify the difference between paired colon adenocarcinomas and non-tumors [19]. Visualization of the tsRNAs including volcano plots and heatmaps was performed with the ggplot2 [34] and pheatmap packages of R, respectively. Survival analysis with Kaplan-Meier (KM) curves plus the log-rank test was performed via the R survival and survminer packages. Hierarchical cluster analysis was performed on tsRNAs expression and primary pathology characteristics, its corresponding heatmap was visualized by pheatmap package (https://cran.r-project.org/web/packages/pheatmap/). *Spearman's* correlation was calculated for each pair of tsRNA and mRNA, its scatter diagrams were plotted by ggplot2 and ggpubr package in R software [34]. Some enrichment analyses were performed to provide biology views that regard tRFdb-3013a/b-related genes, the molecular signatures database (MSigDB) was used as gene set for gene set enrichment analysis (GSEA) [23]. Gene ontology, KEGG pathway, and GSEA analysis were visualized by clusterProfiler and enrichplot packages [22, 35]. The computational binding interactions between tsRNA and targets were predicted by tRFtarget databases [24]. The interaction networks of top-50 targets and tRFdb-3013a/b were performed and visualized via Cytoscape (v3.7.2).

Cell culture and transfections

Human colon adenocarcinoma cell lines SW480 and Caco-2 were purchased from the Cell Bank of Chinese Academy of Sciences (Shanghai, China). The cells were cultured in Leibovitz's L-15 medium or Eagle's Minimum Essential Medium (MEM; Servicebio, Wuhan, China) supplemented with 10% or 20% fetal bovine serum (FBS; Servicebio) at 37° C in a humidified incubator with 5 % CO_2. The cell lines at 50–70% confluences were transfected with micromolar either mimic for tRFdb-3013a/b (tRF-3013a/b mimic), or inhibitor for tRFdb-3013a (tRF-3013a inhibitor), or negative controls (NC mimic, NC inhibitor) using the Lipofectamine RNAimax reagent (Invitrogen, Thermo Fisher Scientific, USA). The mimic and inhibitor for tRFdb-3013a/b and negative controls were purchased from RiboBio BioTech (Guangzhou, China).

RNA extraction and qRT-PCR analysis

Total RNAs were extracted from cultured cells using the RNAex reagent (AGbio, Changsha, China). One microgram total RNAs of each sample were reversely transcribed into cDNA under standard conditions by using PrimeScript RT reagent Kit with gDNA Eraser (Takara, Shiga, Japan). Quantitative real-time polymerase chain reaction (qRT-PCR) was performed with the SYBR® Premix DimerEraser™ (Takara) on the CFX connect real-time system (Bio-Rad, USA). The Bulge-Loop miRNA qRT-PCR Stater Kit (RiboBio, China) with specific stem-loop RT primers was used to quantify the expression of tsRNAs, following the manufacturer's protocol. RNU6 (U6) was used as the internal control for tsRNAs template normalization, and ACTB (β-actin) for mRNA template normalization. Relative quantification of gene expression was calculated by the comparative cycle-threshold (CT) method [19]. Stem-loop primers for tRFdb-3013a/b, and RNU6 were designed and synthesized by RiboBio BioTech (Guangzhou, China). The general primers for ST3GAL1 and ACTB were designed and synthesized by Sangon Biotech (Shanghai, China), and the primers sequences were listed in Supplementary Table 2.

Protein extraction and Western blot

Total proteins were extracted from cultured cells using RIPA buffer (Beyotime, Shanghai, China) and the protein concentrations were determined using a BCA protein assay kit (Servicebio). The specific primary antibody (anti-ST3GAL1 antibody diluted at 1:1000, Abcam, USA; anti-β-actin antibody diluted at 1:1500, Sigma-Aldrich, USA). The β-actin served as an endogenous protein for normalization. The protein bands were visualized using ChemiScope S6 imaging system and processing software (Clinx, Shanghai, China).

Cell proliferation, migration and invasion assays

Cell proliferation assays were performed with Cell Counting Kit-8 (CCK-8; Selleck, USA), the absorbance (A) in each well was measured at 450 nm with multi-mode Microplate Reader (Varioskan LUX, Thermo Fisher Scientific). Cell migration and invasion abilities were tested via the scratch and transwell assays, the images of scratch gap and invasion cells were recorded and photographed using Nikon eclipse Ti-S microscope (Tokyo, Japan).

Luciferase reporter assay

The wild-type or mutant sequences of ST3GAL1 3'UTR containing putative tRFdb-3013a and tRFdb-3013b binding sites were subcloned into pmir-RB-Report vector (RiboBio, China). SW480 cells were co-transfected with pmir-RB-Report vector plus or without tRFdb-3013a/b. The luciferase activity was measured using the Dual-luciferase Reporter Assay System (Promega, USA).

Statistical analysis

The statistical analyses were performed with the R software (https://www.r-project.org/), GraphPad Prism version 8.0 (GraphPad Software, Inc., USA) was used for graphing and analysis. Data were exhibited as means ± standard deviation (SD). The *Chi*-square test was used to compare the categorical variables. Regarding the numerical variables, statistical significance of differences between two groups was assessed using two-sided Student's t test; and comparisons of multiple groups were made by one-way analysis of variance. All experiments were performed in triplicate and *P* values less than 0.05 were considered a statistically significant difference.

Data availability statement

The present study used secondary data which are available in the public domain. The dataset has no identifiable information about the survey participants. All data and materials used during the current study are available from the corresponding author on reasonable request. Some limitations about tsRNA identifications were elaborated and described in Supplementary Materials.

AUTHOR CONTRIBUTIONS

Conception and design: L. Tan and H. Zou; Experiments, data acquisition and analysis: X. Wu, L. Tan, Z. Tang, and H. Zou; Writing, review, and revision of the manuscript: X. Wu, H. Chen, G. Zou, and H. Zou; Study supervision: G. Zou, C. Wen and W. Cao. All authors have contributed significantly and agreed with the content of the final manuscript.

CONFLICTS OF INTEREST

All authors declare no conflicting interests.

ETHICAL STATEMENT AND CONSENT

This study was approved by the Ethics Committees of Chongqing Medical University (KJQN202200430). All methods are carried out in accordance with relevant guidelines and regulations. The patients provided written informed consent.

FUNDING

This work was sponsored by Natural Science Foundation of Chongqing, China (cstc2021jcyj-msxmX0302), and supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202200430). This work was also supported by grants from the National Natural Science Foundation of China (No. 82003854), and the Program for Youth Innovation in Future Medicine from Chongqing Medical University (W0066). The computing work was partly supported by the Supercomputing Center of Chongqing Medical University.

REFERENCES

1. Kim HK, Fuchs G, Wang S, Wei W, Zhang Y, Park H, Roy-Chaudhuri B, Li P, Xu J, Chu K, Zhang F, Chua MS, So S, et al. A transfer-RNA-derived small RNA regulates ribosome biogenesis. Nature. 2017; 552:57–62. https://doi.org/10.1038/nature25005 PMID:29186115

2. Shi J, Zhang Y, Zhou T, Chen Q. tsRNAs: The Swiss Army Knife for Translational Regulation. Trends Biochem Sci. 2019; 44:185–9. https://doi.org/10.1016/j.tibs.2018.09.007 PMID:30297206

3. Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell. 2017; 170:61–71.e11. https://doi.org/10.1016/j.cell.2017.06.013 PMID:28666125

4. Magee R, Rigoutsos I. On the expanding roles of tRNA fragments in modulating cell behavior. Nucleic Acids Res. 2020; 48:9433–48. https://doi.org/10.1093/nar/gkaa657 PMID:32890397

5. Xie Y, Yao L, Yu X, Ruan Y, Li Z, Guo J. Action mechanisms and research methods of tRNA-derived small RNAs. Signal Transduct Target Ther. 2020; 5:109. https://doi.org/10.1038/s41392-020-00217-4 PMID:32606362

6. Zou H, Wu LX, Tan L, Shang FF, Zhou HH. Significance of Single-Nucleotide Variants in Long Intergenic Non-protein Coding RNAs. Front Cell Dev Biol. 2020; 8:347. https://doi.org/10.3389/fcell.2020.00347 PMID:32523949

7. Telonis AG, Magee R, Loher P, Chervoneva I, Londin E, Rigoutsos I. Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. Nucleic Acids Res. 2017; 45:2973–85. https://doi.org/10.1093/nar/gkx082 PMID:28206648

8. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44:e71. https://doi.org/10.1093/nar/gkv1507 PMID:26704973

9. Kumar P, Mudunuri SB, Anaya J, Dutta A. tRFdb: a database for transfer RNA fragments. Nucleic Acids Res. 2015; 43:D141–5. https://doi.org/10.1093/nar/gku1138 PMID:25392422

10. Ren J, Wu X, Shang FF, Qi Y, Tang Z, Wen C, Cao W, Cheng Q, Tan L, Chen H, Zhou HH, Zou H. The tRNA-Cys-GCA Derived tsRNAs Suppress Tumor Progression of Gliomas via Regulating VAV2. Dis Markers. 2022; 2022:8708312. https://doi.org/10.1155/2022/8708312 PMID:36426134

11. Tao EW, Wang HL, Cheng WY, Liu QQ, Chen YX, Gao QY. A specific tRNA half, 5'tiRNA-His-GTG, responds to hypoxia via the HIF1α/ANG axis and promotes colorectal cancer progression by regulating LATS2. J Exp Clin Cancer Res. 2021; 40:67. https://doi.org/10.1186/s13046-021-01836-7 PMID:33588913

12. Goodarzi H, Liu X, Nguyen HC, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell. 2015; 161:790–802. https://doi.org/10.1016/j.cell.2015.02.053 PMID:25957686

13. Yu M, Lu B, Zhang J, Ding J, Liu P, Lu Y. tRNA-derived RNA fragments in cancer: current status and future perspectives. J Hematol Oncol. 2020; 13:121. https://doi.org/10.1186/s13045-020-00955-6 PMID:32887641

14. Bien J, Lin A. A Review of the Diagnosis and Treatment of Metastatic Colorectal Cancer. JAMA. 2021; 325:2404–5. https://doi.org/10.1001/jama.2021.6021 PMID:34129003

15. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020; 70:145–64. https://doi.org/10.3322/caac.21601 PMID:32133645

16. Zou H, Wen C, Peng Z, Shao YY, Hu L, Li S, Li C, Zhou HH. P4HB and PDIA3 are associated with tumor

progression and therapeutic outcome of diffuse gliomas. Oncol Rep. 2018; 39:501–10. https://doi.org/10.3892/or.2017.6134 PMID:29207176

17. La Ferlita A, Alaimo S, Veneziano D, Nigita G, Balatti V, Croce CM, Ferro A, Pulvirenti A. Identification of tRNA-derived ncRNAs in TCGA and NCI-60 panel cell lines and development of the public database tRFexplorer. Database (Oxford). 2019; 2019:baz115. https://doi.org/10.1093/database/baz115 PMID:31735953

18. Veneziano D, Tomasello L, Balatti V, Palamarchuk A, Rassenti LZ, Kipps TJ, Pekarsky Y, Croce CM. Dysregulation of different classes of tRNA fragments in chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 2019; 116:24252–8. https://doi.org/10.1073/pnas.1913695116 PMID:31723042

19. Xu B, Liang J, Zou H, Wang J, Xiong Y, Pei J. Identification of Novel tRNA-Leu-CAA-Derived tsRNAs for the Diagnosis and Prognosis of Diffuse Gliomas. Cancer Manag Res. 2022; 14:2609–23. https://doi.org/10.2147/CMAR.S367020 PMID:36072386

20. Zou H, Wu LX, Yang Y, Li S, Mei Y, Liu YB, Zhang L, Cheng Y, Zhou HH. lncRNAs PVT1 and HAR1A are prognosis biomarkers and indicate therapy outcome for diffuse glioma patients. Oncotarget. 2017; 8:78767–80. https://doi.org/10.18632/oncotarget.20226 PMID:29108264

21. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer International Publishing : Imprint: Springer, pp. 1 online resource (XVI, 260 pages 32 illustrations, 140 illustrations in color. 2016. https://doi.org/10.1007/978-3-319-24277-4

22. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021; 2:100141. https://doi.org/10.1016/j.xinn.2021.100141 PMID:34557778

23. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25. https://doi.org/10.1016/j.cels.2015.12.004 PMID:26771021

24. Li N, Shan N, Lu L, Wang Z. tRFtarget: a database for transfer RNA-derived fragment targets. Nucleic Acids Res. 2021; 49:D254–60. https://doi.org/10.1093/nar/gkaa831 PMID:33035346

25. Wu X, Zhao J, Ruan Y, Sun L, Xu C, Jiang H. Sialyltransferase ST3GAL1 promotes cell migration, invasion, and TGF-β1-induced EMT and confers paclitaxel resistance in ovarian cancer. Cell Death Dis. 2018; 9:1102. https://doi.org/10.1038/s41419-018-1101-0 PMID:30375371

26. Holmes AD, Chan PP, Chen Q, Ivanov P, Drouard L, Polacek N, Kay MA, Lowe TM. A standardized ontology for naming tRNA-derived RNAs based on molecular origin. Nat Methods. 2023; 20:627–8. https://doi.org/10.1038/s41592-023-01813-2 PMID:36869120

27. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 2016; 44:D184–9. https://doi.org/10.1093/nar/gkv1309 PMID:26673694

28. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell. 2018; 173:283–5. https://doi.org/10.1016/j.cell.2018.03.042 PMID:29625045

29. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41:D991–5. https://doi.org/10.1093/nar/gks1193 PMID:23193258

30. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID:19261174

31. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–9. https://doi.org/10.1093/bioinformatics/btu638 PMID:25260700

32. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–30. https://doi.org/10.1093/bioinformatics/btt656 PMID:24227677

33. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/Bioconductor package for

integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44:e71. https://doi.org/10.1093/nar/gkv1507 PMID:26704973

34. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47. https://doi.org/10.1093/nar/gkv007 PMID:25605792

35. Xu B, Liang J, Zou H, Wang J, Xiong Y, Pei J. Identification of Novel tRNA-Leu-CAA-Derived tsRNAs for the Diagnosis and Prognosis of Diffuse Gliomas. Cancer Manag Res. 2022; 14:2609–23. https://doi.org/10.2147/CMAR.S367020 PMID:36072386

36. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Use R!,. (Cham: Springer International Publishing: Imprint: Springer). 2016:1. https://doi.org/10.1007/978-3-319-24277-4

37. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25. https://doi.org/10.1016/j.cels.2015.12.004 PMID:26771021

38. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45:D353–61. https://doi.org/10.1093/nar/gkw1092 PMID:27899662

39. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021; 2:100141. https://doi.org/10.1016/j.xinn.2021.100141 PMID:34557778

40. Li N, Shan N, Lu L, Wang Z. tRFtarget: a database for transfer RNA-derived fragment targets. Nucleic Acids Res. 2021; 49:D254–60. https://doi.org/10.1093/nar/gkaa831 PMID:33035346

41. Martinho C, Lopez-Gomollon S. Detection of MicroRNAs by Northern Blot. Methods Mol Biol. 2023; 2630:47–66. https://doi.org/10.1007/978-1-0716-2982-6_4 PMID:36689175

SUPPLEMENTARY MATERIALS

Supplementary Methods

Data collection and processing

The sequences of human tsRNAs were taken from tRFdb database (http://genome.bioch.virginia.edu/trfdb/) and tRFexplorer database (https://trfexplorer.cloud/), including different types of tsRNA fragments: 5′-tRFs, 3′-tRFs, 5′U-tRFs and tRF-1 [1, 2]. The hg19 reference genome annotations and corresponding sequences of tRNAs in humans were taken from GtRNAdb (genomic tRNA database, http://gtrnadb.ucsc.edu/) [3]. Then, a custom annotation of the reference human gene containing only known tsRNAs was assembled. The small non-coding RNA sequencing (sncRNA-seq) data of colon adenocarcinoma (COAD) samples and rectum adenocarcinoma (READ) samples on TCGA, the RNA sequencing datasets of TCGA-COAD and TCGA-READ samples, and the corresponding patient clinical information were retrieved from GDC data portal [4]. The primary clinical and molecular pathology characteristics of COAD patients were listed in Supplementary Table 1. The gene expression profiles of GSE39582 were downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database, the gene expression microarrays are based on Affymetrix Human Genome U133 Plus 2.0 Array platform (Affymetrix, USA) [5].

A conservative pipeline was implemented to get an accurate estimation of tsRNAs expression as demonstrated previously [2], the data processing and flow chart for tsRNAs identification are showed in Supplementary Figure 1A. The quality controls and data filtering of raw files on sncRNA-seq data were pre-processed via using Cutadapt and FastQC method. Then, clean reads of small RNA sequencing were re-mapped to the reference human genome (GRCh37 /hg19) and the sequences of tsRNAs annotation file via applying Bowtie software [6]. After alignment, only the mapped reads could be quantified to count the number of reads belonging to each of the candidate tsRNAs with HTSeq software [7]. Finally, the expression value of tsRNAs was calculated and normalized as transcripts per million reads (TPM) of total raw counts [8], and the average expression values less than one log2TPM were filtered to eliminate random degradation sequences.

The clinical specimens

Fifteen carcinoma tissues and five para-carcinoma tissues were obtained from patients diagnosed with colon adenocarcinomas undergoing surgical resections at the Department of Gastrointestinal Surgery, the First Affiliated Hospital of Chongqing Medical University from February 2022 to October 2023. After excisions, tissues were immediately frozen in liquid nitrogen for subsequent use. This study was approved by the Ethics Committees of Chongqing Medical University and the patients provided written informed consent.

Northern blotting

Total RNA was isolated from the cells using TRIzol reagent (Invitrogen, USA). Before loading, the RNA samples were denatured at 65° C for 5 min and chilled on ice immediately. The samples were separated on a 15% denaturing polyacrylamide gel and electrophoretically transferred to a charged nylon membrane (Labselect, China). Following cross-linking and pre-hybridization, the RNA was hybridized with the corresponding 3′-digoxigenin (DIG)-labeled DNA probe at 42° C overnight [17]. After washing and blocking, the membrane was incubated with anti-DIG antibody solution (Servicebio, China). The chemiluminescence signal was captured and analyzed via a ChemiScopeS6 imaging system (Clinx, China). The DIG-labeled tRFdb-3013a/b probe (5’-TGGTGCCGTGACTCGGA-3′), the DIG-labeled tRNA-His-GUG probe (5’-CGGCCACAACGCAGAGTACT-3′) and the 5S rRNA probe (as an internal control) were synthesized by Sangon Biotech (Shanghai, China).

Limitations

In this study, tsRNAs refer to that fragments derived from the tRNAs based on the GtRNAdb, tRFdb and tRFexplorer databases by a conservative identification pipeline. With computational approaches, we determined the tsRNAs expression profile within the TCGA-COAD dataset. Particularly, our bioinformatic analysis has focused on the role of tRFdb-3013a and tRFdb-3013b, which are two of the tsRNA fragments derived from tRNA-His-GTG gene. Furthermore, there should be many ways, through diversification of methods, addressing the tRNA-derived fragments mining with a conservative pipeline. Recently, a tRNA-derived fragments (tRFs) repository has been released, MINTbase v2.0, in which more than ten thousand tRFs were mined from the TCGA datasets with their own deterministic and exhaustive pipeline.

There were some differences or limitations between some naming system of tsRNA, while compared to the profiles of tsRNAs that identified through tDRnamer and MINTbase method and our identification pipeline within COAD dataset. Take the example of tRFdb-3013a and tRFdb-3013b, two of the fragments

derived from tRNA-His-GTG gene; as shown in the Supplementary Figure 8A, 8B, these are a total of forty-eight fragments derived from the 3′ end of mature tRNA-His-GTG in tDRnamer, among them, twenty fragment isoforms could be precisely aligned to the sequence of tRFdb-3013a, and twenty-eight fragment isoforms can be aligned to the sequence of tRFdb-3013b. Actually, two tRF isoforms (tDR-60:76-His-GTG-1-M2, whose MINTbase ID is tRF-17-8US5652; and tDR-55:76-His-GTG-1-M2, whose MINTbase ID is tRF-22-WB8US5652) were two of the most abundant fragments that identified from the TCGA datasets, and may be seen as a typical representative of tRFdb-3013a and tRFdb-3013b. Moreover, the primers for tRFdb-3013a/b detection with stem-loop qRT-PCR assay, were designed to match the sequence (TCCGAGTCACGGCA or TCGAATCCGAGTCACGGCA), hence the expression levels should refer to the fold-change of tRFdb-3013a/b, which may be detected, not only one fragment isoform, in the qRT-PCR experiments. In this case, tsRNAs generally refer to the small RNAs derived from the tRNAs, instead of the tRFs or tRNA isoforms with highly similar sequences, to avoid overestimation and artifacts that are presented in tDRnamer and MINTbase.

As mentioned in the GtRNAdb, there are about 64 classes of tRNAs which correspond to twenty-two kinds of amino acids in the human genome, since each amino acid has many different anticodons such as histidine, which has two specific anticodons and that makes up two isotypes of tRNAs (tRNA-His-ATG and tRNA-His-GTG). Hence it is conservatively estimated that not more than four tsRNAs derived from the 3′ end of two mature tRNA isotypes for histidine. Undoubtedly, it will be more complex in the human genome transcriptional system, of the tRNA-His-GTG isotypes, twenty tRNA isoforms (including tRNA-His-GTG-1-1, tRNA-His-GTG-1-2, tRNA-His-GTG-1-6, tRNA-His-GTG-1-7, and tRNA-His-GTG-1-8, etc.) were identified with almost similar sequences, and sometimes only individual bases differ between these twenty sequences (As shown in Figure 2A). Even though with respect to tDRnamer and MINTbase which contain thousands of tRFs in TCGA datasets, it is not surprising that only two hundred tsRNAs were identified in COAD samples based on the GtRNAdb, tRFdb and tRFexplorer databases with our conservative pipeline.

Supplementary References

1. Kumar P, Mudunuri SB, Anaya J, Dutta A. tRFdb: a database for transfer RNA fragments. Nucleic Acids Res. 2015; 43:D141–5. https://doi.org/10.1093/nar/gku1138 PMID:25392422

2. La Ferlita A, Alaimo S, Veneziano D, Nigita G, Balatti V, Croce CM, Ferro A, Pulvirenti A. Identification of tRNA-derived ncRNAs in TCGA and NCI-60 panel cell lines and development of the public database tRFexplorer. Database (Oxford). 2019; 2019:baz115. https://doi.org/10.1093/database/baz115 PMID:31735953

3. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 2016; 44:D184–9. https://doi.org/10.1093/nar/gkv1309 PMID:26673694

4. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell. 2018; 173:283–5. https://doi.org/10.1016/j.cell.2018.03.042 PMID:29625045

5. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41:D991–5. https://doi.org/10.1093/nar/gks1193 PMID:23193258

6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID:19261174

7. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–9. https://doi.org/10.1093/bioinformatics/btu638 PMID:25260700

8. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–30. https://doi.org/10.1093/bioinformatics/btt656 PMID:24227677

9. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44:e71. https://doi.org/10.1093/nar/gkv1507 PMID:26704973

10. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47. https://doi.org/10.1093/nar/gkv007 PMID:25605792

11. Xu B, Liang J, Zou H, Wang J, Xiong Y, Pei J.

Identification of Novel tRNA-Leu-CAA-Derived tsRNAs for the Diagnosis and Prognosis of Diffuse Gliomas. Cancer Manag Res. 2022; 14:2609–23. https://doi.org/10.2147/CMAR.S367020 PMID:36072386

12. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Use R!,. (Cham: Springer International Publishing: Imprint: Springer). 2016:1. https://doi.org/10.1007/978-3-319-24277-4

13. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25. https://doi.org/10.1016/j.cels.2015.12.004 PMID:26771021

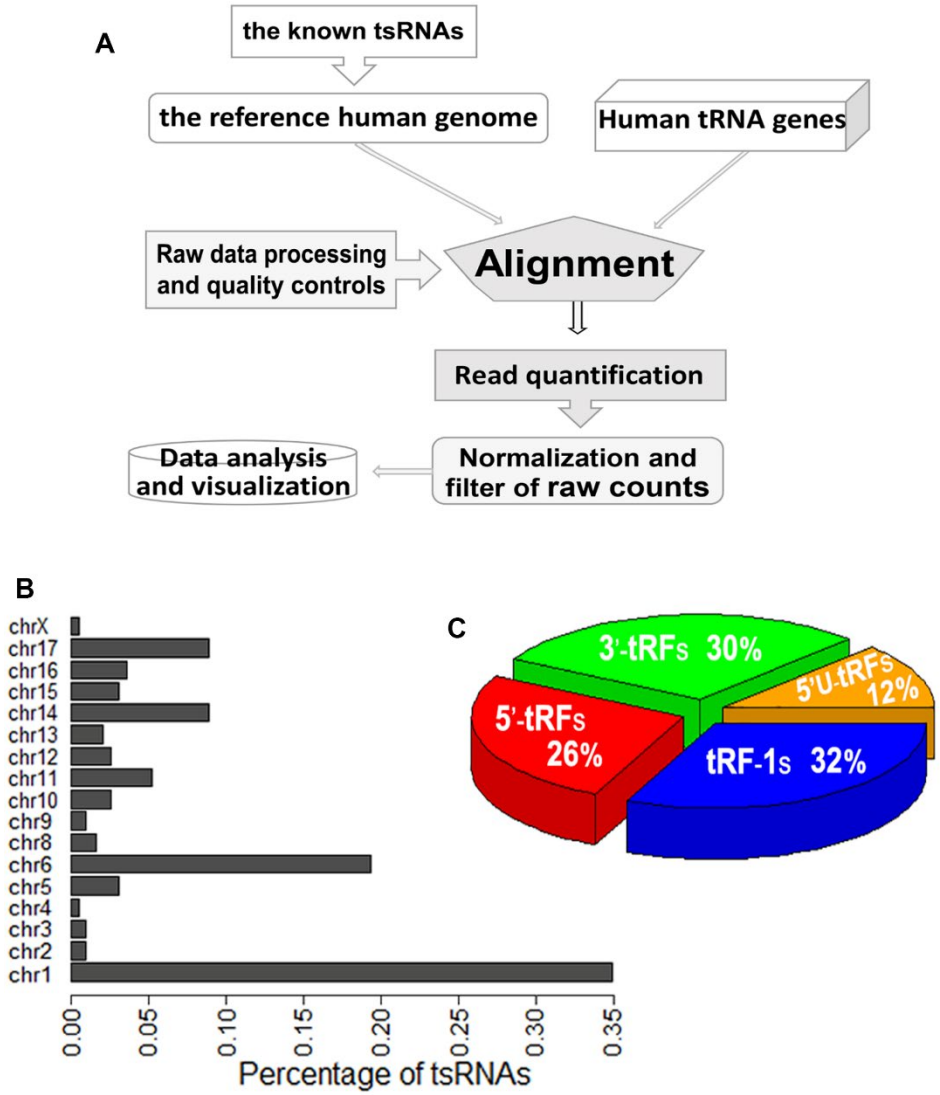
14. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45:D353–61. https://doi.org/10.1093/nar/gkw1092 PMID:27899662

15. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021; 2:100141. https://doi.org/10.1016/j.xinn.2021.100141 PMID:34557778

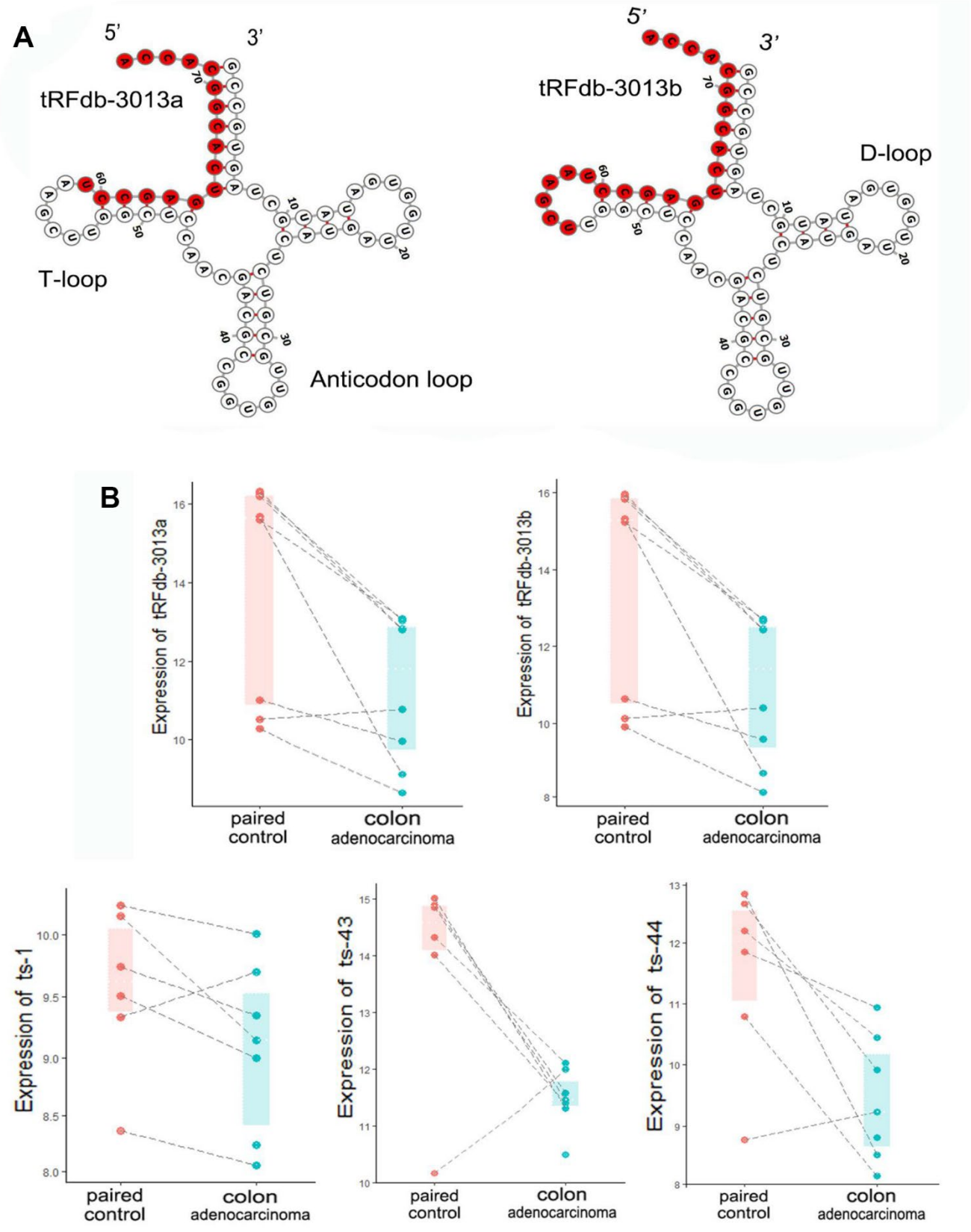
16. Li N, Shan N, Lu L, Wang Z. tRFtarget: a database for transfer RNA-derived fragment targets. Nucleic Acids Res. 2021; 49:D254–60. https://doi.org/10.1093/nar/gkaa831 PMID:33035346

17. Martinho C, Lopez-Gomollon S. Detection of MicroRNAs by Northern Blot. Methods Mol Biol. 2023; 2630:47–66. https://doi.org/10.1007/978-1-0716-2982-6_4 PMID:36689175

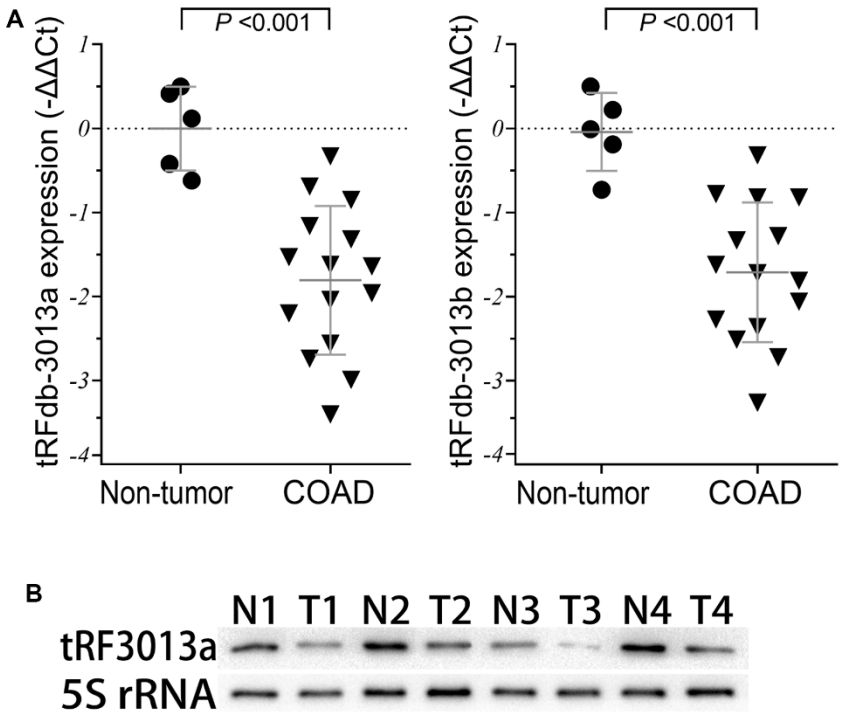
Supplementary Figures



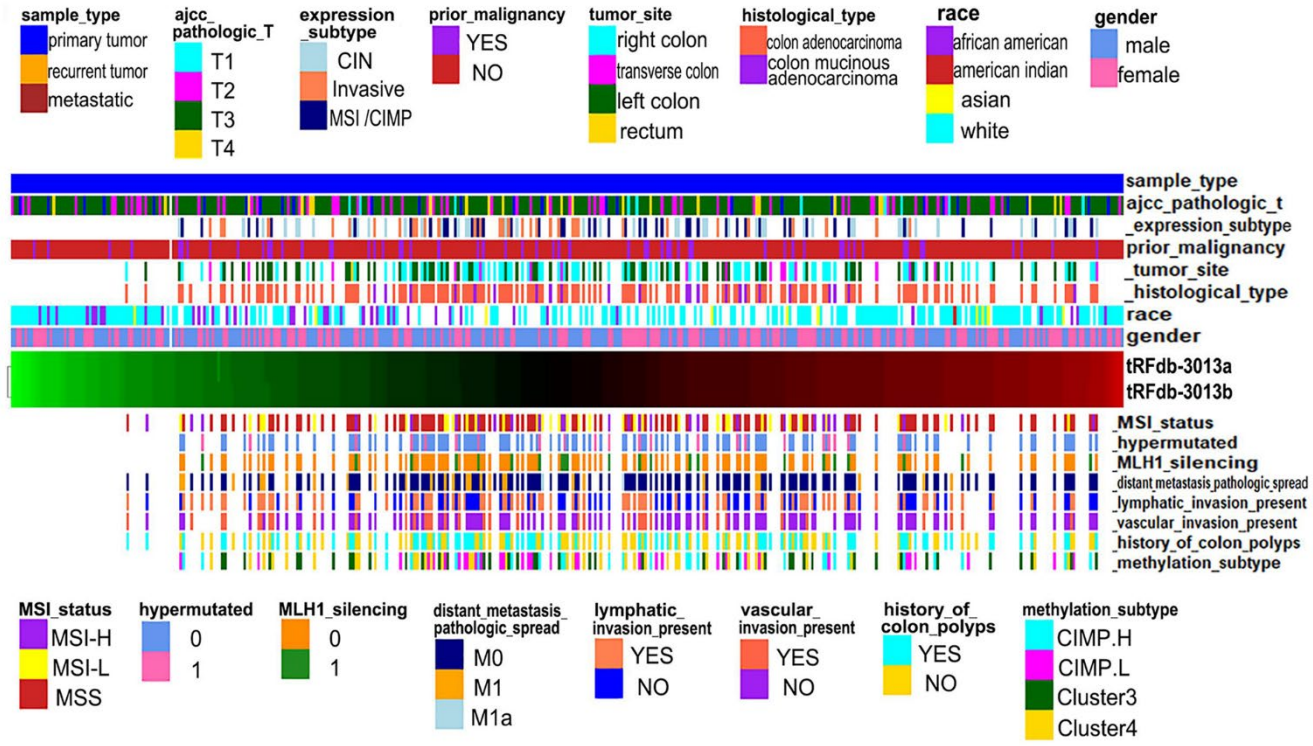
Supplementary Figure 1. A summary of tsRNAs identification and characterization. (**A**) The flow chart of data processing and tsRNAs identification pipeline. (**C**) The characterization of tsRNA corresponding derived tRNA gene sources and (**B**) the chromosome locations statistical analysis.



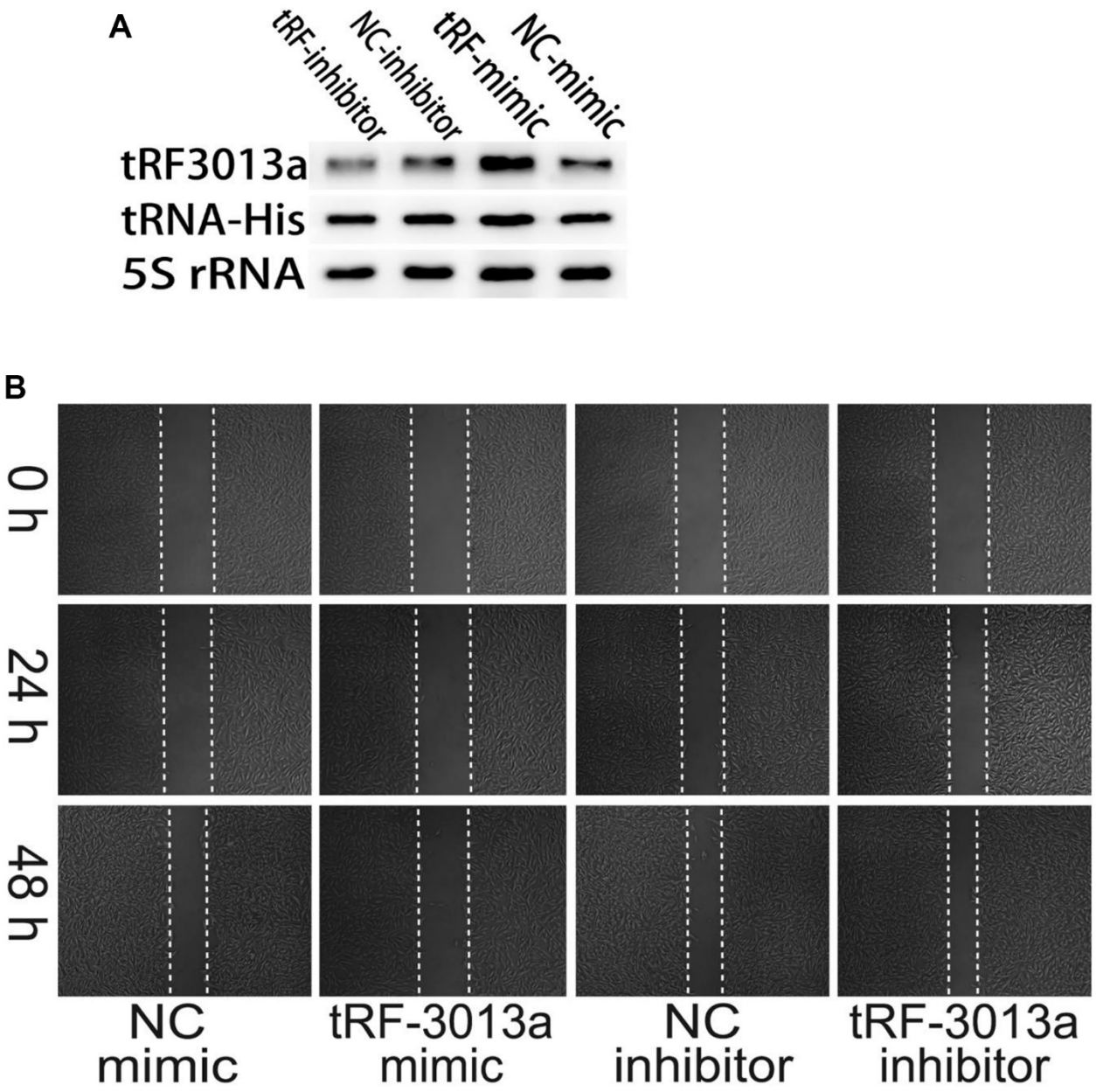
Supplementary Figure 2. (**A**) Representative secondary structure of tRNA-His-GUG and its derived fragments (tRFdb-3013a and tRFdb-3013b). (**B**) The expression patterns of several tRNA-His-GTG derived tsRNA fragments (including tRFdb-3013a, tRFdb-3013b, ts-1, ts-43 and ts-44) in paired colon adenocarcinomas and non-tumor controls of TCGA-COAD datasets.



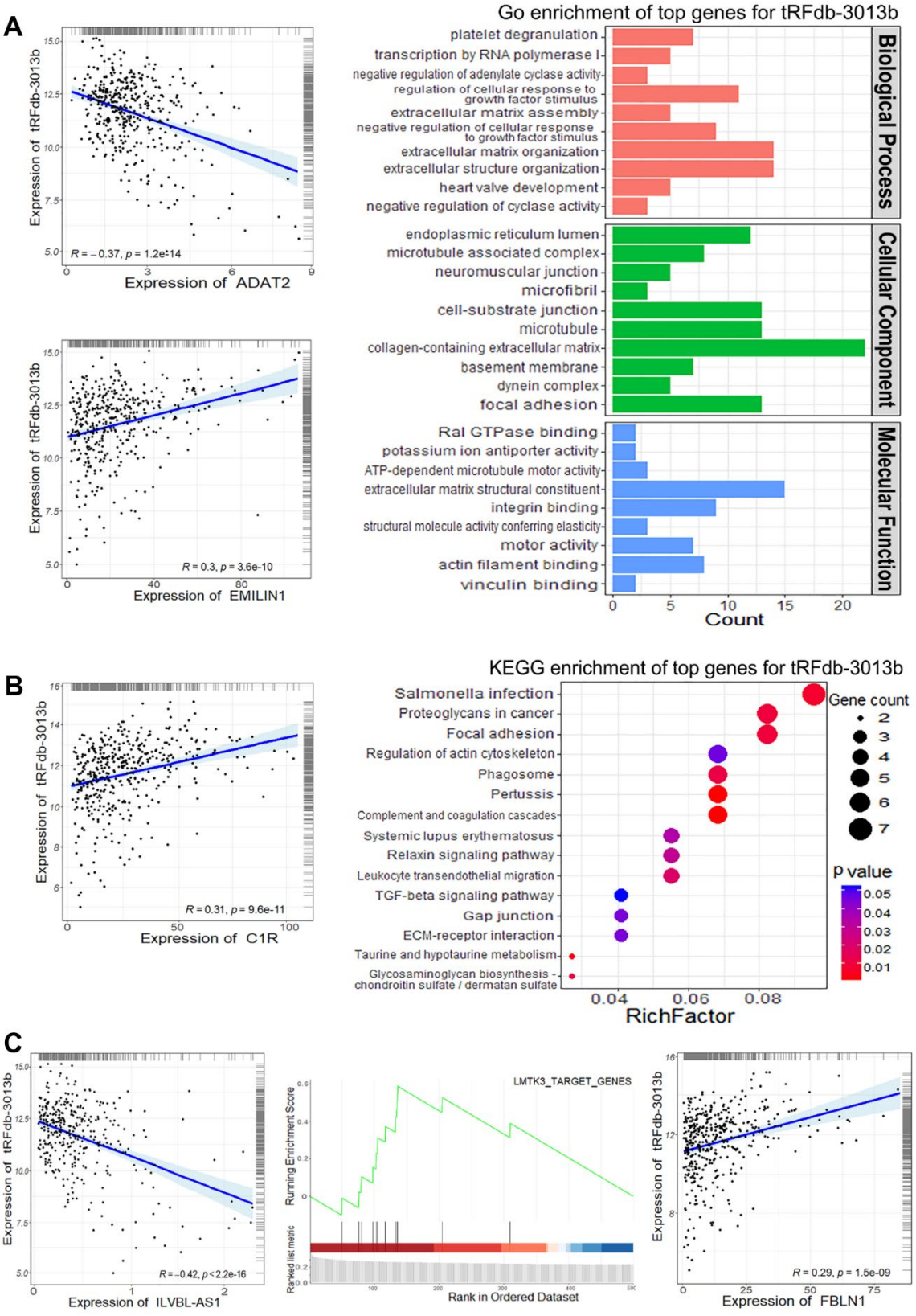
Supplementary Figure 3. (**A**) The boxplot of tRFdb-3013a/b expression between non-tumor controls and colon adenocarcinomas with the stem-loop tsRNA qRT-PCR analysis. (**B**) The Northern blotting analysis of tRFdb-3013a (tRF3013a) between non-tumor controls (N1-4) and colon adenocarcinoma (T1-4) samples (all P <0.01).



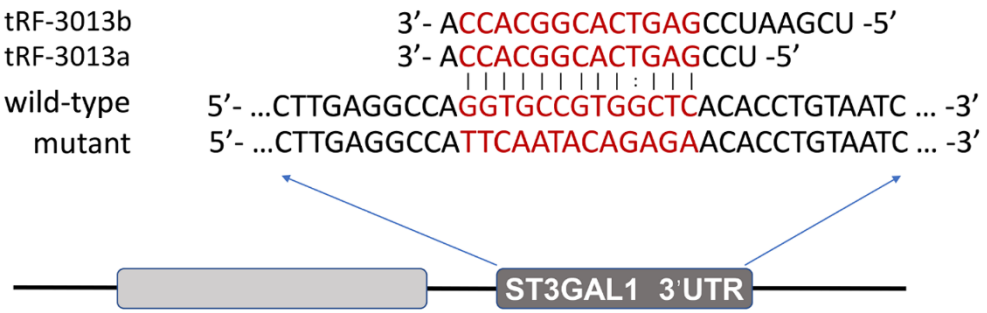
Supplementary Figure 4. The correlation analysis of tRFdb-3013a/b expression with the primary pathology characteristics. Hierarchical clustering heatmap of tRFdb-3013a/b expression and the pathology characteristics parameters.



Supplementary Figure 5. (**A**) The Northern blotting analysis of tRFdb-3013a (tRF3013a, both P <0.05) and tRNA-His-GUG (tRNA-His) in the cells transfected with tRFdb-3013a mimic (tRF-mimic) or inhibitor (tRF-inhibitor) or negative controls (NC-mimic, NC-inhibitor). (**B**) Relative gap analysis scratch assay was used to assess the migration ability of SW480 cell.



Supplementary Figure 6. The enrichment analyses of tRFdb-3013b correlated-genes within TCGA-COAD datasets. (**A**) The top gene ontology (GO) terms, including biological process, cellular component and molecular function. (**B**) The top KEGG pathway for tRFdb-3013b correlated-genes. The correlation scatter-plots of tRFdb-3013b and its correlated-genes (ADAT2, EMILIN1, and C1R). (**C**) GSEA (gene set enrichment analysis) plots of the molecular signatures, the scatter plots for tRFdb-3013b and its correlated-genes (ILVBL-AS1, FBLN1).



Supplementary Figure 7. Schematic diagram of the interaction position between tRFdb-3013a/b and the seed regions within the 3’UTR wide-type region and mutation region of ST3GAL1 gene in the luciferase reporter assay.

A

tDRnamer ID	Fragment sequence	length	MINTbase ID	reads	RPM
tDR-61:76-His-GTG-1-M2	**CCGAGTCACGGCACCA**	16	tRF-16-308HP1B	**541**	32.08
tDR-60:76-His-GTG-1-M2	**TCCGAGTCACGGCACCA**	17	tRF-17-8US5652	**947**	72.07
tDR-59:76-His-GTG-1-M2	A**TCCGAGTCACGGCACCA**	18	tRF-18-HS16MFD2	**416**	18.58
tDR-58:76-His-GTG-1-M2	AA**TCCGAGTCACGGCACCA**	19	tRF-19-DR5MEJE2	49	6.998
tDR-57:76-His-GTG-1-M2	GAA**TCCGAGTCACGGCACCA**	20	tRF-20-OML32826	61	17.99
tDR-56:76-His-GTG-1-M2	CGAA**TCCGAGTCACGGCACCA**	21	tRF-21-LE308HP1B	**321**	42.75
tDR-60:75-His-GTG-1-M2	**TCCGAGTCACGGCACC -**	16	tRF-16-8US565D	**162**	27.8
tDR-59:75-His-GTG-1-M2	A**TCCGAGTCACGGCACC -**	17	tRF-17-HS16MFJ	7	3.955
tDR-58:75-His-GTG-1-M2	AA**TCCGAGTCACGGCACC -**	18			
tDR-57:75-His-GTG-1-M2	GAA**TCCGAGTCACGGCACC -**	19	tRF-19-OML328FV	1	2.008
tDR-56:75-His-GTG-1-M2	CGAA**TCCGAGTCACGGCACC -**	20	tRF-20-LE308HP1	17	6.55
tDR-60:74-His-GTG-1-M2	**TCCGAGTCACGGCAC - -**	15			
tDR-59:74-His-GTG-1-M2	A**TCCGAGTCACGGCAC - -**	16	tRF-16-HS16MFD	5	4.889
tDR-58:74-His-GTG-1-M2	AA**TCCGAGTCACGGCAC - -**	17			
tDR-57:74-His-GTG-1-M2	GAA**TCCGAGTCACGGCAC - -**	18	tRF-18-OML328D1	1	1.176
tDR-56:74-His-GTG-1-M2	CGAA**TCCGAGTCACGGCAC - -**	19	tRF-19-LE308HI1	3	5.345
tDR-59:73-His-GTG-1-M2	**ATCCGAGTCACGGCA - - -**	15			
tDR-58:73-His-GTG-1-M2	AA**TCCGAGTCACGGCA- - -**	16			
tDR-57:73-His-GTG-1-M2	GAA**TCCGAGTCACGGCA- - -**	17	tRF-17-OML3282	1	1.225
tDR-56:73-His-GTG-1-M2	CGAA**TCCGAGTCACGGCA- - -**	18	tRF-18-LE308HDU	5	2.312

B

tDRnamer ID	Fragment sequence	length	MINTbase ID	reads	RPM
tDR-55:76-His-GTG-1-M2	**TCGAATCCGAGTCACGGCACCA**	22	tRF-22-WB8US5652	**6503**	62.09
tDR-54:76-His-GTG-1-M2	T**TCGAATCCGAGTCACGGCACCA**	23	tRF-23-YUHS16MFD2	**100**	35.37
tDR-53:76-His-GTG-1	GT**TCGAATCCGAGTCACGGCACCA**	24	tRF-24-7SDR5MEJE2	86	14.38
tDR-52:76-His-GTG-1	GGT**TCGAATCCGAGTCACGGCACCA**	25	tRF-25-R9OML32826	29	5.159
tDR-51:76-His-GTG-1	CGGT**TCGAATCCGAGTCACGGCACCA**	26	tRF-26-MNLE308HP1B	12	3.181
tDR-50:76-His-GTG-1	TCGGT**TCGAATCCGAGTCACGGCACCA**	27	tRF-28-48YUHS16MFD2	8	1.718
tDR-55:75-His-GTG-1-M2	**TCGAATCCGAGTCACGGCACC -**	21	tRF-21-WB8US565D	**1541**	16.82
tDR-54:75-His-GTG-1-M2	T**TCGAATCCGAGTCACGGCACC -**	22	tRF-22-YUHS16MFJ	15	2.22
tDR-53:75-His-GTG-1	GT**TCGAATCCGAGTCACGGCACC -**	23	tRF-23-7SDR5MEJV	4	1.443
tDR-52:75-His-GTG-1	GGT**TCGAATCCGAGTCACGGCACC -**	24			
tDR-51:75-His-GTG-1	CGGT**TCGAATCCGAGTCACGGCACC -**	25			
tDR-50:75-His-GTG-1	TCGGT**TCGAATCCGAGTCACGGCACC -**	26			
tDR-55:74-His-GTG-1-M2	**TCGAATCCGAGTCACGGCAC - -**	20	tRF-20-WB8US565	**113**	3.881
tDR-54:74-His-GTG-1-M2	T**TCGAATCCGAGTCACGGCAC - -**	21	tRF-21-YUHS16MFD	18	5.239
tDR-53:74-His-GTG-1	GT**TCGAATCCGAGTCACGGCAC - -**	22	tRF-22-7SDR5MEJ1	5	2.81
tDR-52:74-His-GTG-1	GGT**TCGAATCCGAGTCACGGCAC - -**	23			
tDR-51:74-His-GTG-1	CGGT**TCGAATCCGAGTCACGGCAC - -**	24	tRF-24-MNLE308HI1	1	1.298
tDR-50:74-His-GTG-1	TCGGT**TCGAATCCGAGTCACGGCAC - -**	25	tRF-26-48YUHS16MFD	5	18.6
tDR-55:73-His-GTG-1-M2	**TCGAATCCGAGTCACGGCA - - -**	19	tRF-19-WB8US5IU	44	5.828
tDR-54:73-His-GTG-1-M2	T**TCGAATCCGAGTCACGGCA - - -**	20	tRF-20-YUHS16MF	25	12.33
tDR-53:73-His-GTG-1	GT**TCGAATCCGAGTCACGGCA - - -**	21	tRF-21-7SDR5MEJB	14	10.81
tDR-52:73-His-GTG-1	GGT**TCGAATCCGAGTCACGGCA - - -**	22	tRF-22-R9OML3282	2	1.378
tDR-51:73-His-GTG-1	CGGT**TCGAATCCGAGTCACGGCA - - -**	23	tRF-23-MNLE308HDU	1	1.608
tDR-50:73-His-GTG-1	TCGGT**TCGAATCCGAGTCACGGCA - - -**	24	tRF-25-48YUHS16MF	2	1.075

Supplementary Figure 8. Representative images of the tRFdb-3013a (**A**) or tRFdb-3013b (**B**) sequence alignments of some fragments derived tRNA-His-GTG from tRF fragments within tDRnamer and MINTbase v2.0.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3–5.

Supplementary Table 1. The primary clinical and molecular pathology characteristics parameters in the samples of TCGA-COAD datasets.

Supplementary Table 2. Primers used in real-time PCR assay.

Primer names	Primer sequences (5′-3′)
ST3GAL1 forward	CACCCACACCCCTGTATTCTCC
ST3GAL1 reverse	TGTCCTGACCCAAGCTCAATGC
ACTB forward	CATGTACGTTGCTATCCAGGC
ACTB reverse	CTCCTTAATGTCACGCACGAT
tRFdb-3013a forward	TCCGAGTCACGGCACCA
tRFdb-3013b forward	TCGAATCCGAGTCACGGCACCA
U6 forward	GGAACGATACAGAGAAGATTAGC
U6 reverse	TGGAACGCTTCACGAATTTGCG

Supplementary Table 3. A summary of the identified tsRNAs with available expression abundance in the datasets of colon adenocarcinomas.

Supplementary Table 4. The differentially expressed tsRNA genes between colon adenocarcinomas and non-tumor controls.

Supplementary Table 5. The expression profiles of tRNA-His-GUG derived tsRNA fragments in the samples of TCGA-COAD datasets.