OMTO, Volume *26*

Supplemental information

Comprehensive landscape of tRNA-derived fragments in lung cancer

Zitong Gao, Mayumi Jijiwa, Masaki Nasu, Heather Borgard, Ting Gong, Jinwen Xu, Shaoqiu Chen, Yuanyuan Fu, Yu Chen, Xiamin Hu, Gang Huang, and Youping Deng

Supplementary Material:

Table S1: Clinical characteristics of all patients in five datasets.

Table S2: Information of 3’-tRDFs sequences including 3P/downstream sequences, 3a-tRF and 3b-tRF. Data resources are from tRFdb and Pekarsky’s study[1].

Table S3: Information of 5’-tRDFs sequences including 5P/upstream sequences, 5a-tRF, 5b-tRF and 5c-tRF. Data resources are from tRFdb[2,3] and Gtrnadb[4].

Table S4: Length of four subclass of tRDFs. This table excludes all tRDFs with zero expression from GEO and TCGA, 362 tRDFs are collected in total.

Table S5: 58 common tRDFs from four datasets. Collection of 58 tRDFs common in four datasets (GEO and TCGA-LUAD) includes sequence information, labels in former databases and unified name based on its origin tRNA.

Table S6: Correlation of 17 tRDFs that were paired as derived from the same tRNA.

Table S7: DE tRDFs in GSE83527 and GSE62182 (S7-1, 7-2); common DE from GSE83527 and GSE62182 (S7-3); independent validation prediction results from random forest including sensitivity, specificity, accuracy, and AUC (S7-4). DE tRDFs in TCGA-LUAD and TCGA-LUSC (S7-5, S7-6); tRDFs signatures in TCGA-LUAD, TCGA-LUSC (S7-7, S7-8).

Table S8: Univariate Cox regression in each stage group of TCGA-LUAD and TCGA-LUSC.

Table S9: 52 tRDFs that highly expressed in TCGA-LUAD cohort. Table contains sequence information, labels in former databases and unified name based on its origin tRNA.

Table S10: miRNA-tRDFs correlation analysis results by Pearson Correlation.

Table S11: mRNA-tRDFs correlation analysis results by Pearson Correlation.

Table S12: Correlation between differentially expressed genes with tRDFs signatures.

Table S13: Mutation frequency of tRDFs correlated gene which target in six signaling pathways.

Table S14: Correlation between genes that are enriched in KEGG pathways and tRDFs signatures.

Table S15: The correlation of tRDFs with TME infiltration.

Table S16: The correlation of tRDFs with immune-related genes.

Table S17: Primer information of RT-qPCR validation.

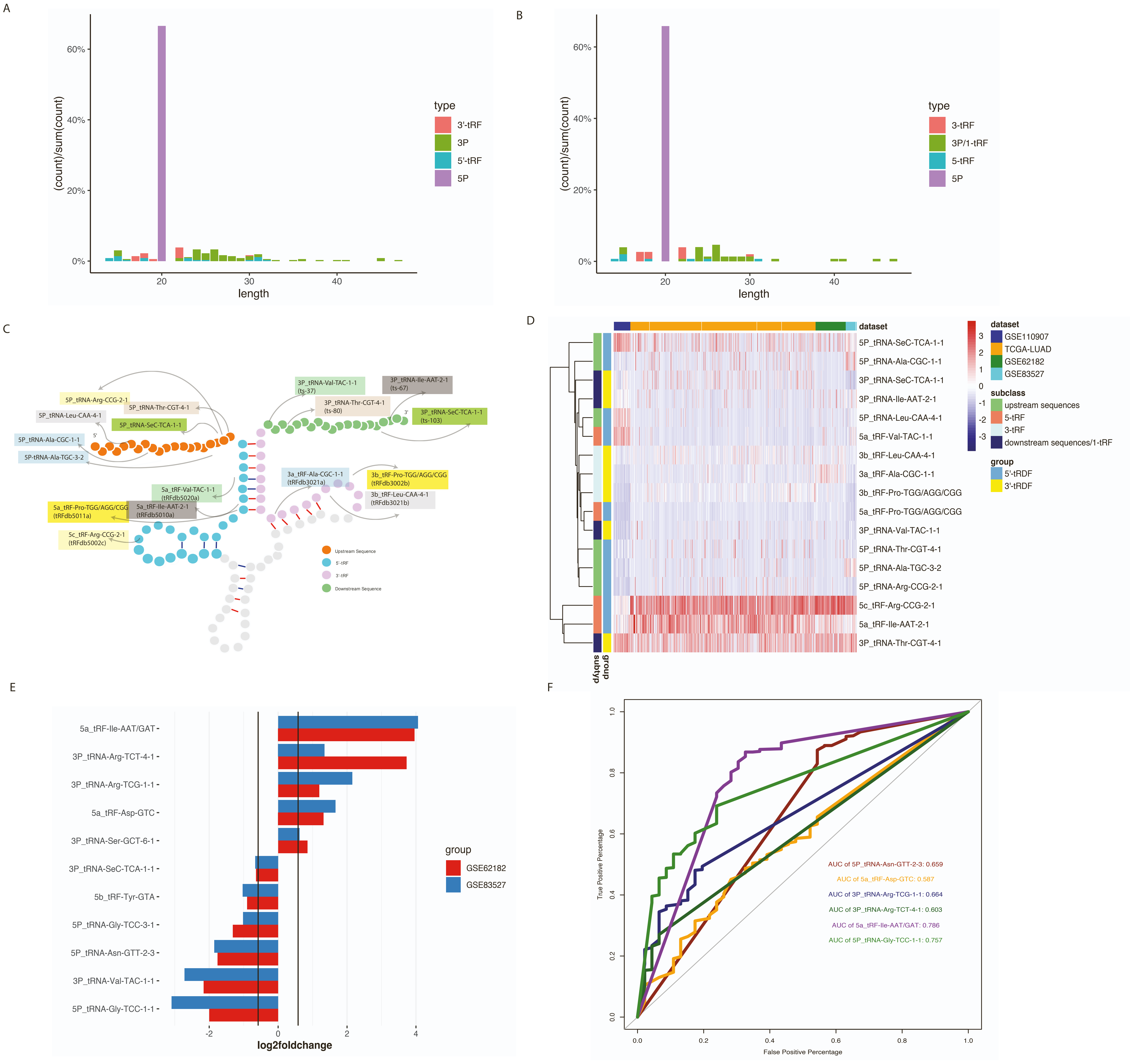
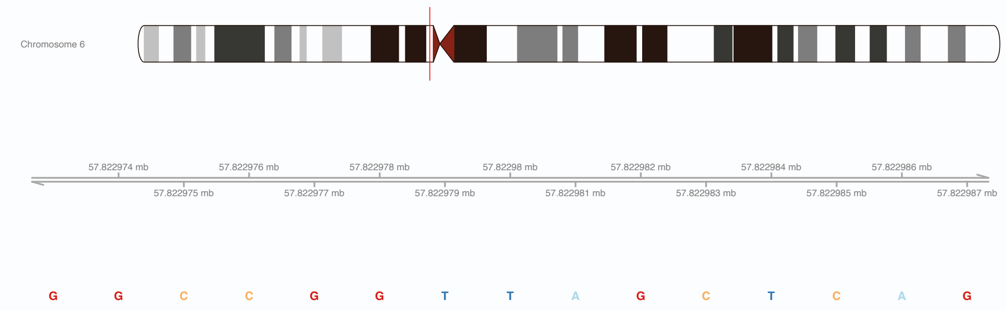


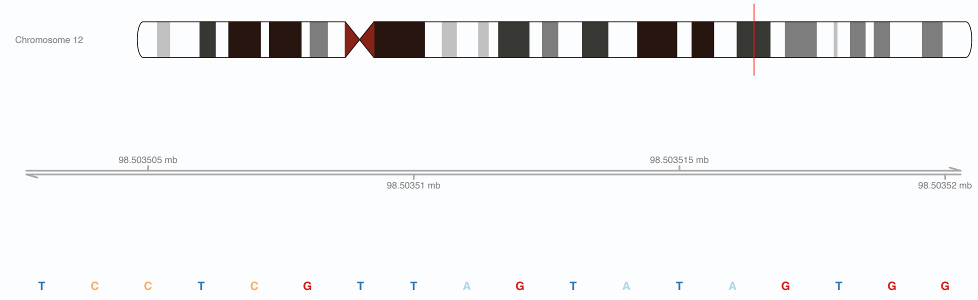
Figure legend:

A.B.Two bar plots showed length distribution of each type of tRDFs. (A) shows length distribution of 362 tRDFs from five datasets and (B) shows the 152 tRDFs found in common from only LUAD cohorts. Length of column indicates the percentage of tRDFs. C. tRNA labels the location of eight pairs, 17 tRDFs from 58 common tRDFs that were derived from the same tRNA. D. Heatmap and hierarchical clustering shows expression of the eight pairs of tRDFs originated from the same tRNA in four datasets. E. Fold change of 11 DE tRDFs in two GEO datasets, the length of column represents the value of fold change. F. Independent validation presents the diagnostic value of each signature as a single biomarker.

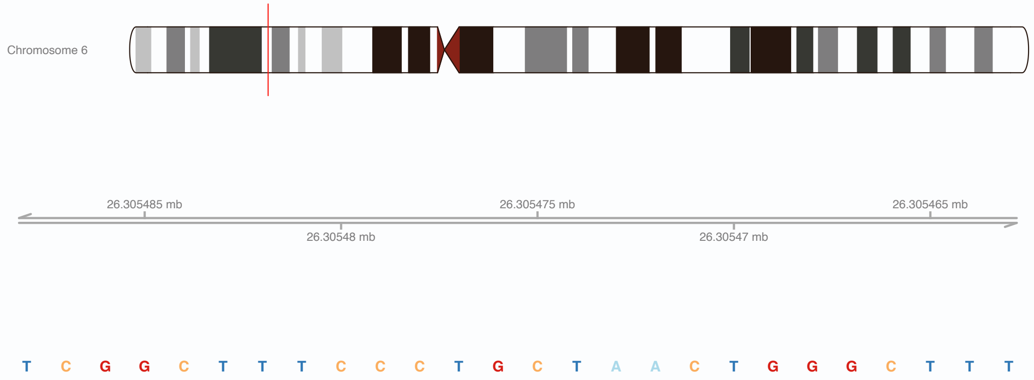
5a_tRF-Ile-AAT/GAT GRCh38/hg38 57822973-57822988 +



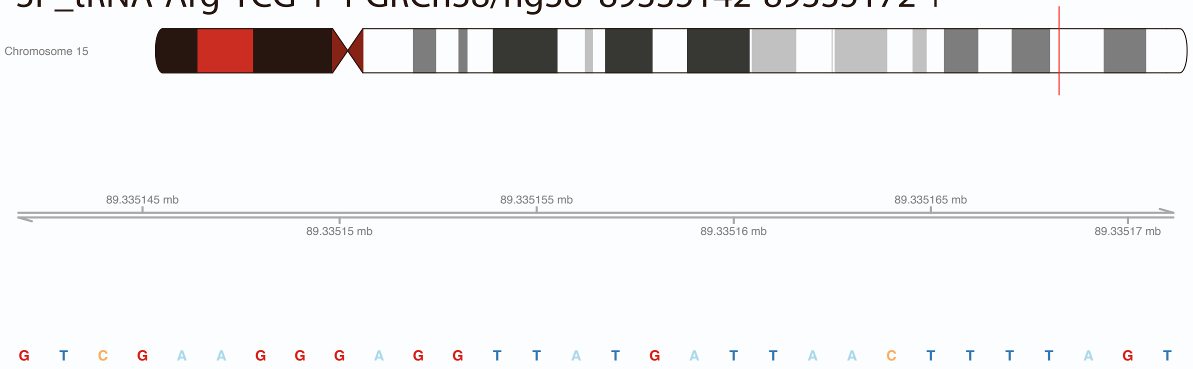
5a_tRF-Asp-GTC GRCh38/hg38 98503503-98503521 +



3P_tRNA-Ser-GCT-6-1 GRCh38/hg38 26305463-26305489 -



3P_tRNA-Arg-TCG-1-1 GRCh38/hg38 89335142-89335172 +



3P_tRNA-Arg-TCT-4-1 GRCh38/hg38 159141585-159141615 -

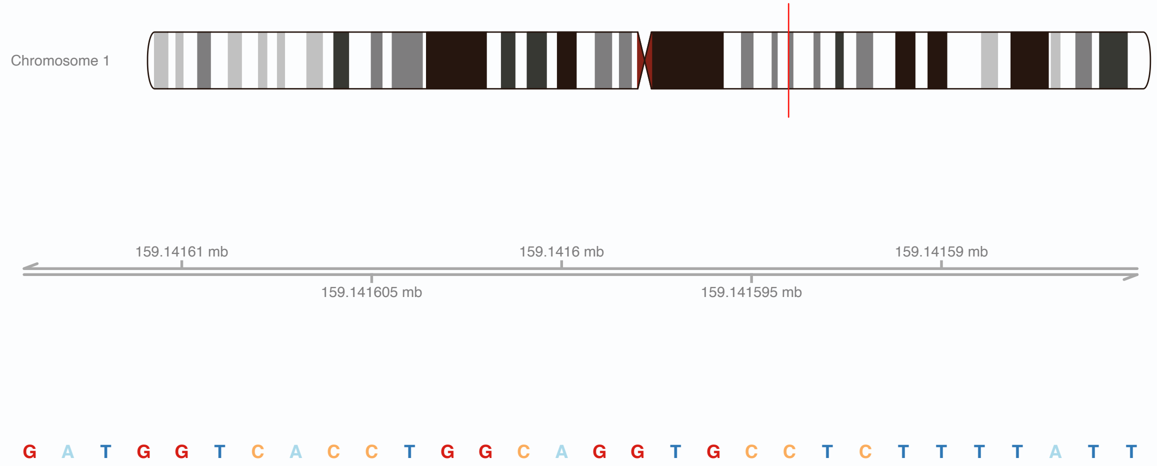


Figure legend:
Fig S2. Genome tracks of five upregulated tRDFs.

Figure S2 shows the genome tracks of five upregulated tRDFs: 3P_tRNA-Ser-GCT-6-1, 3P_tRNA-Arg-TCG-1-1, 5a_tRF-Asp-GTC, 3P_tRNA-Arg-TCT-4-1, 5a_tRF-Ile-AAT/GAT.

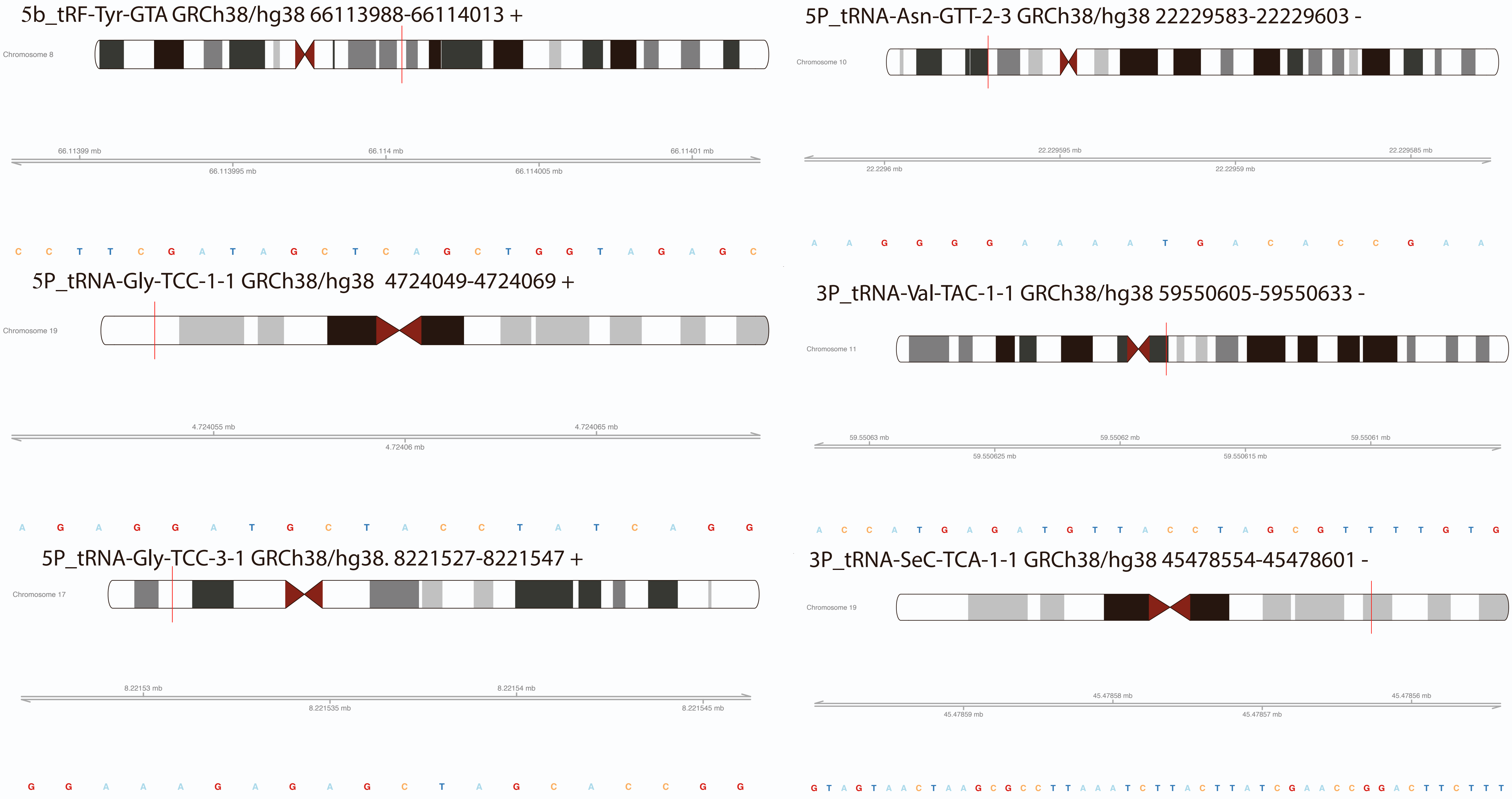
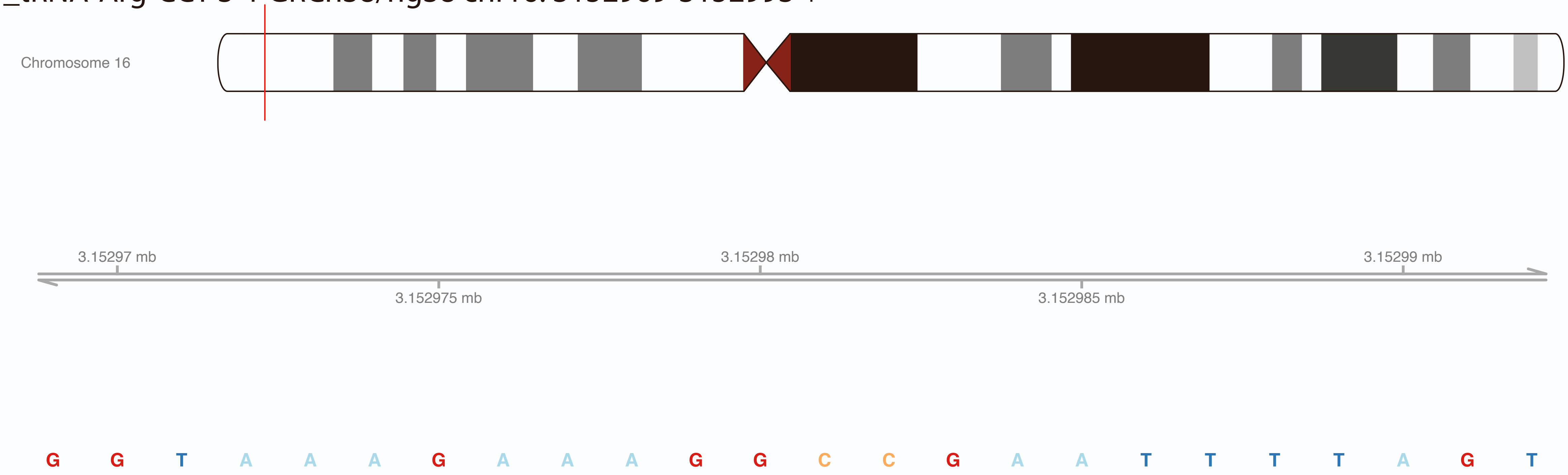


Figure legend:

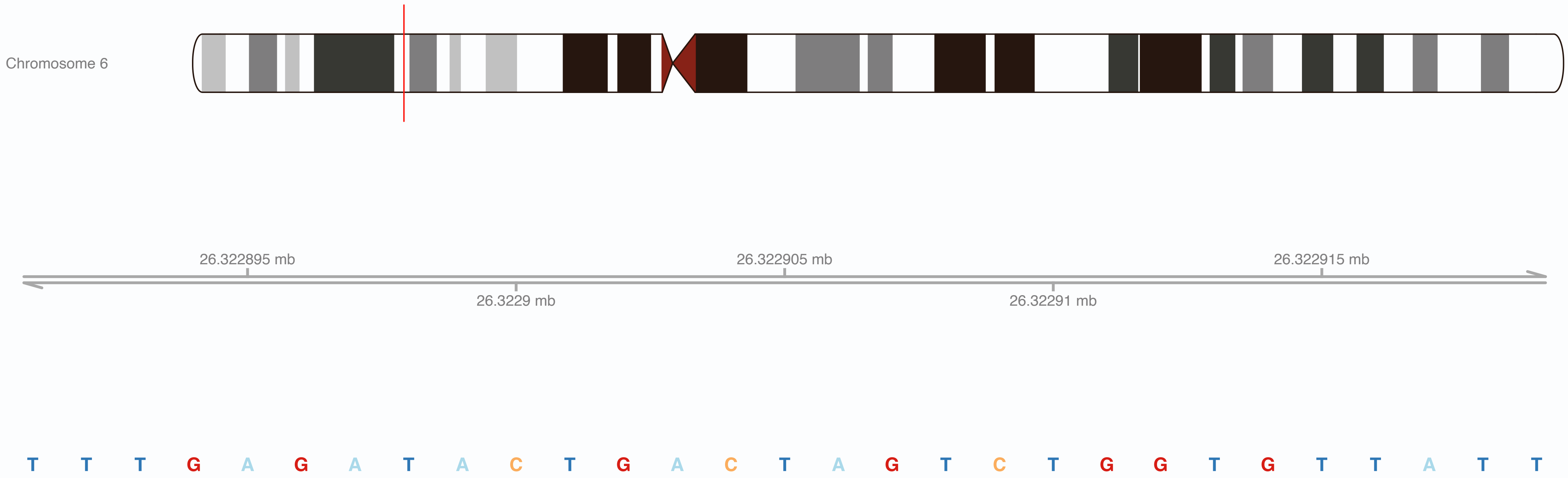
Fig S3. Genome tracks of Six downregulated tRDFs.

Figure S3 shows the genome tracks of six down regulated tRDFs: 5b_tRF-Tyr-GTA, 5P_tRNA-Gly-TCC-1-1, 5P_tRNA-Gly-TCC-3-1, 5P_tRNA-Asn-GTT-2-3, 3P_tRNA-Val-TAC-1-1, 3P_tRNA-SeC-TCA-1-1.

3P_tRNA-Arg-CCT-3-1 GRCh38/hg38 chr16: 3152969-3152993 +



3P_tRNA-Arg-TCG-2-1 GRCh38/hg38 chr6: 26322891-26322920 +



5P_tRNA-SeC-TCA-1-1 GRCh38/hg38 chr19: 45478688-45478708 -

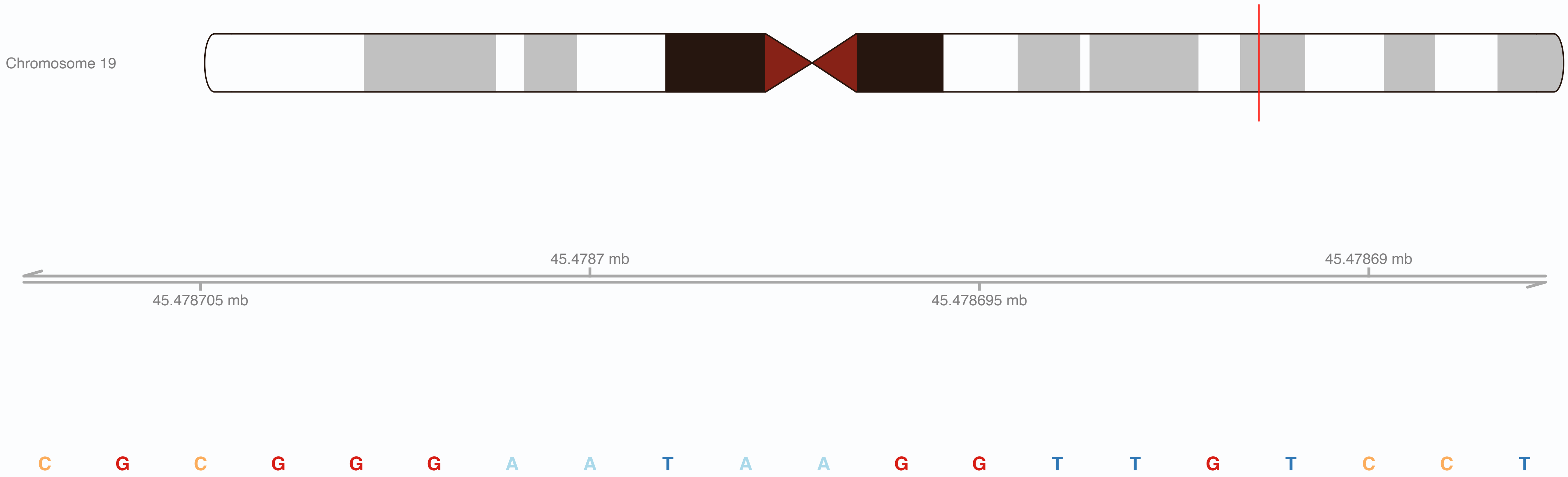
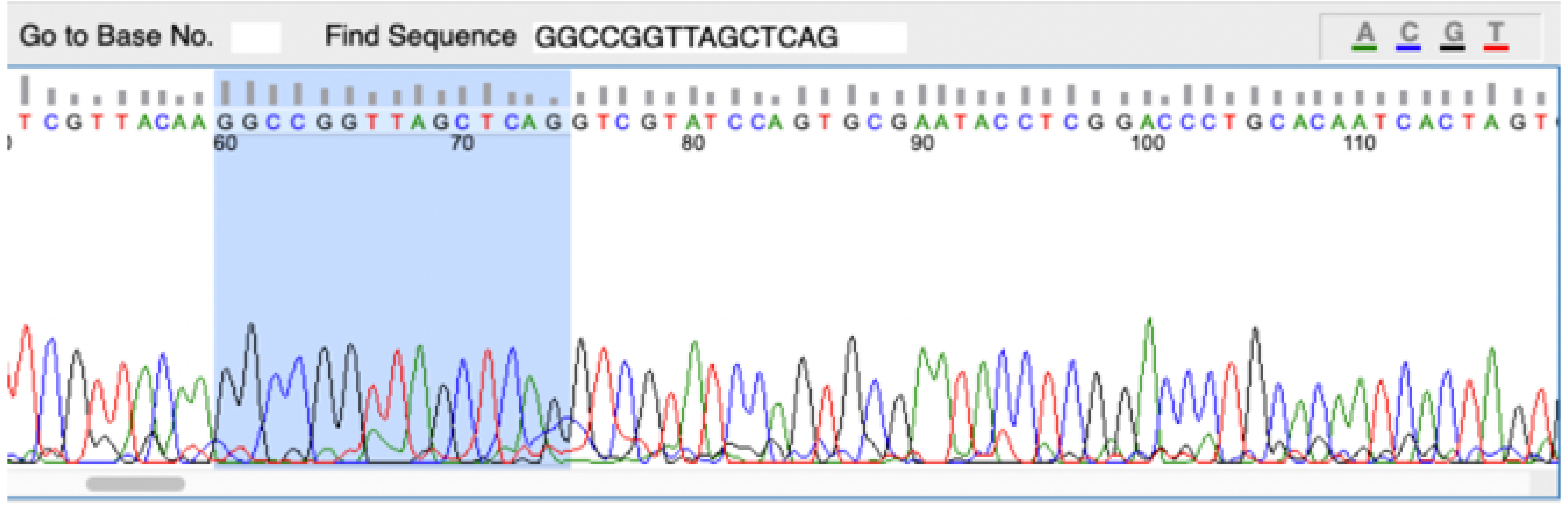


Figure legend:

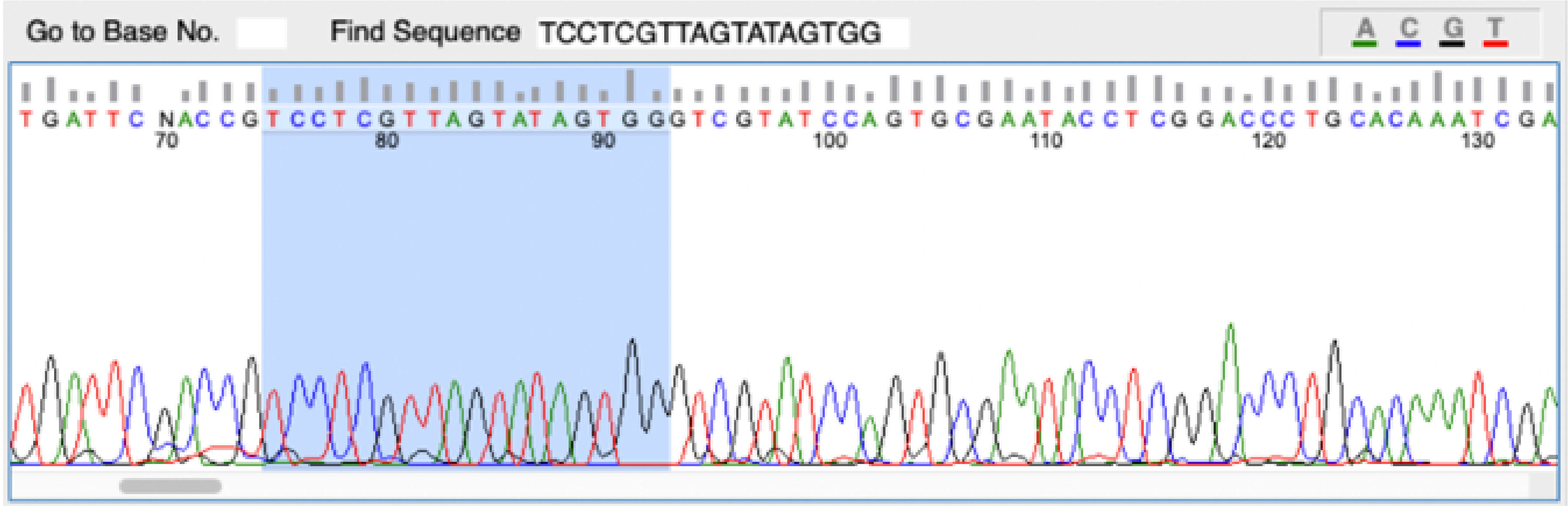
Fig S4. Genome tracks of three unchanged tRDFs.

Figure S4 shows the genome tracks of three unchanged tRDFs, 3P_tRNA-Arg-CCT-3-1, 3P_tRNA-Arg-TCG-2-1, and 5P_tRNA-SeC-TCA-1-1.

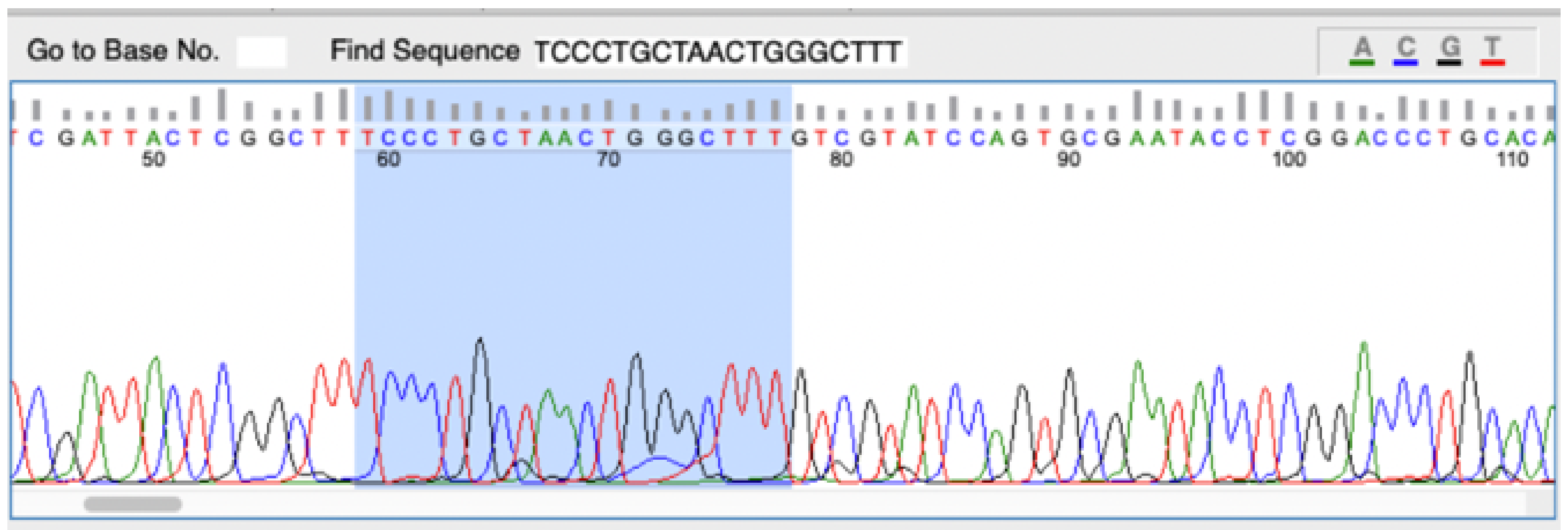
5a_tRF-Ile-AAT/GAT


Go to Base No.
Find Sequence GGCCGGTTAGCTCAG

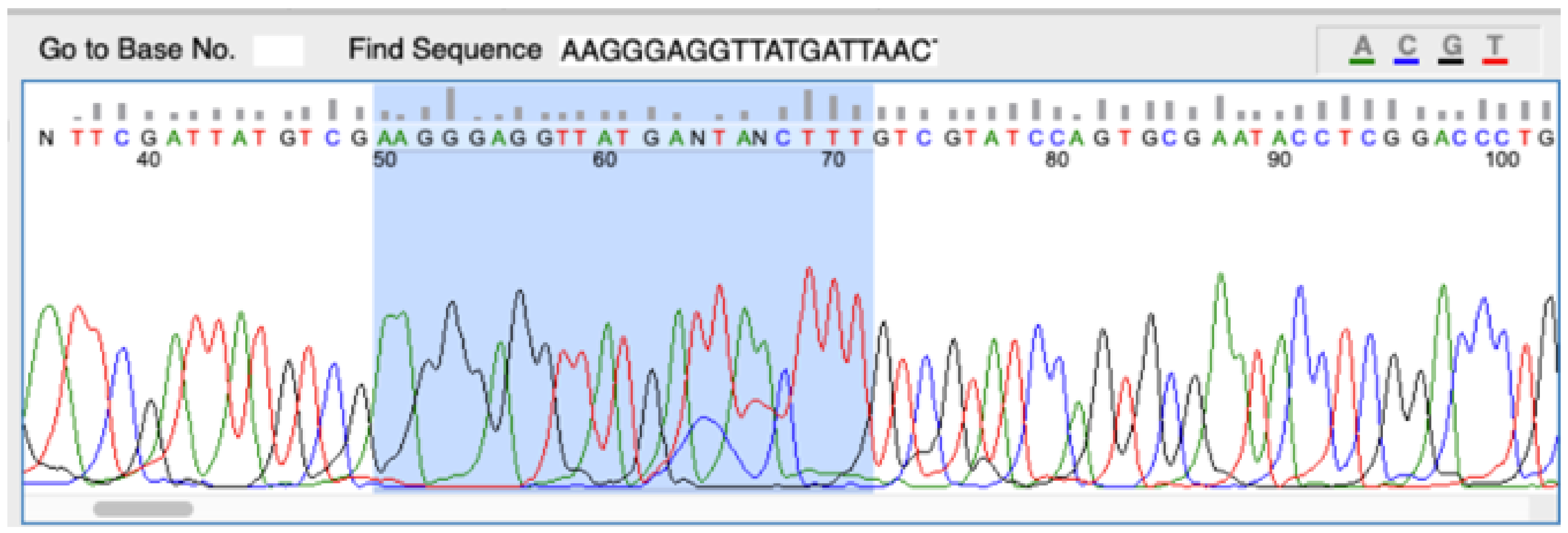
5a_tRF-Asp-GTC


Go to Base No.
Find Sequence TCCTCGTTAGTATAGTGG

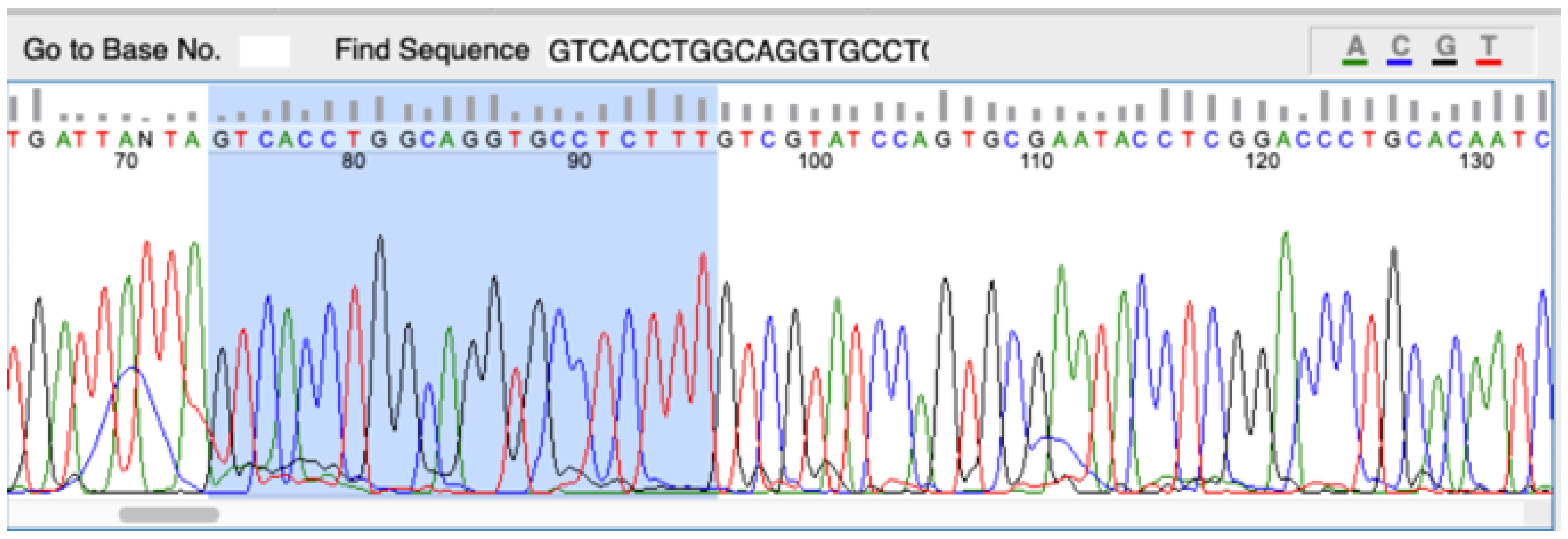
3P_tRNA-Ser-GCT-6-1


Go to Base No.
Find Sequence TCCCTGCTAACTGGGCTTT

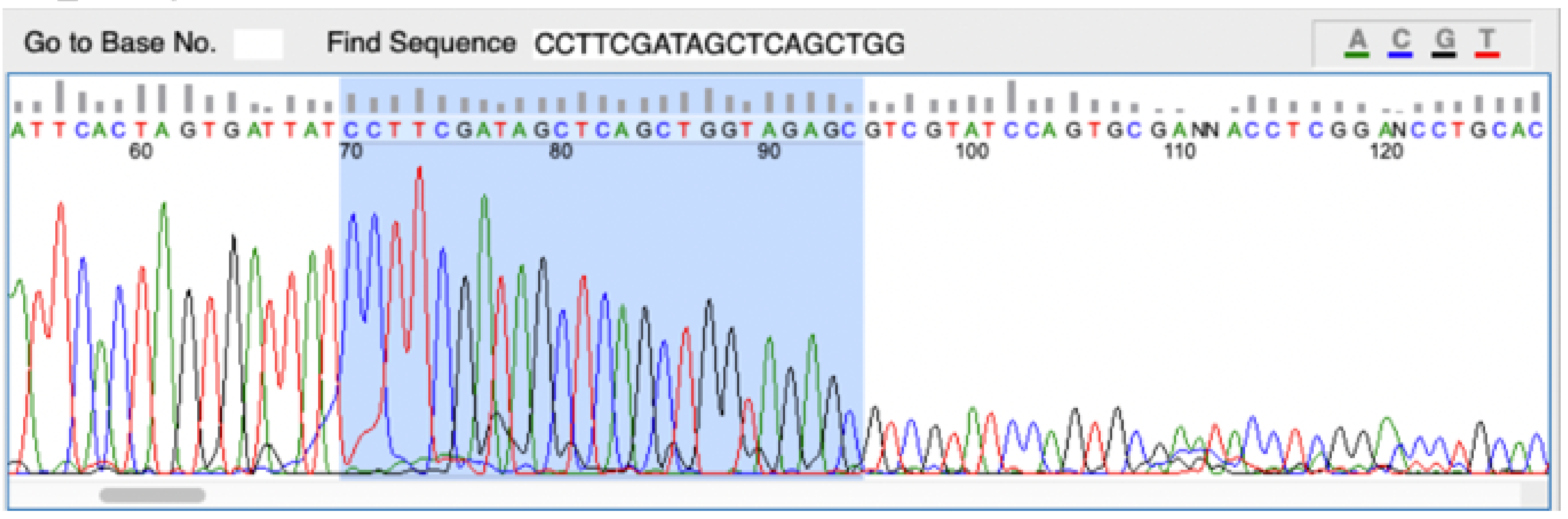
3P_tRNA-Arg-TCG-1-1


Go to Base No.
Find Sequence AAGGGAGGTTATGATTAAC

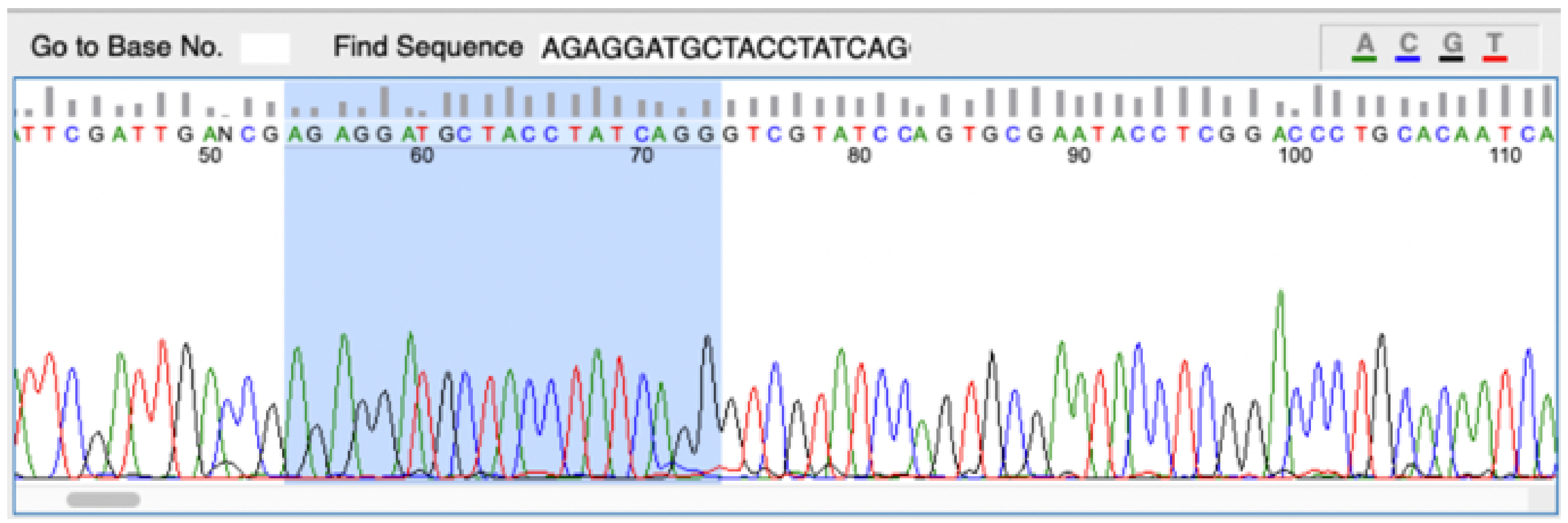
3P_tRNA-Arg-TCT-4-1


Go to Base No.
Find Sequence GTCACCTGGCAGGTGCCT

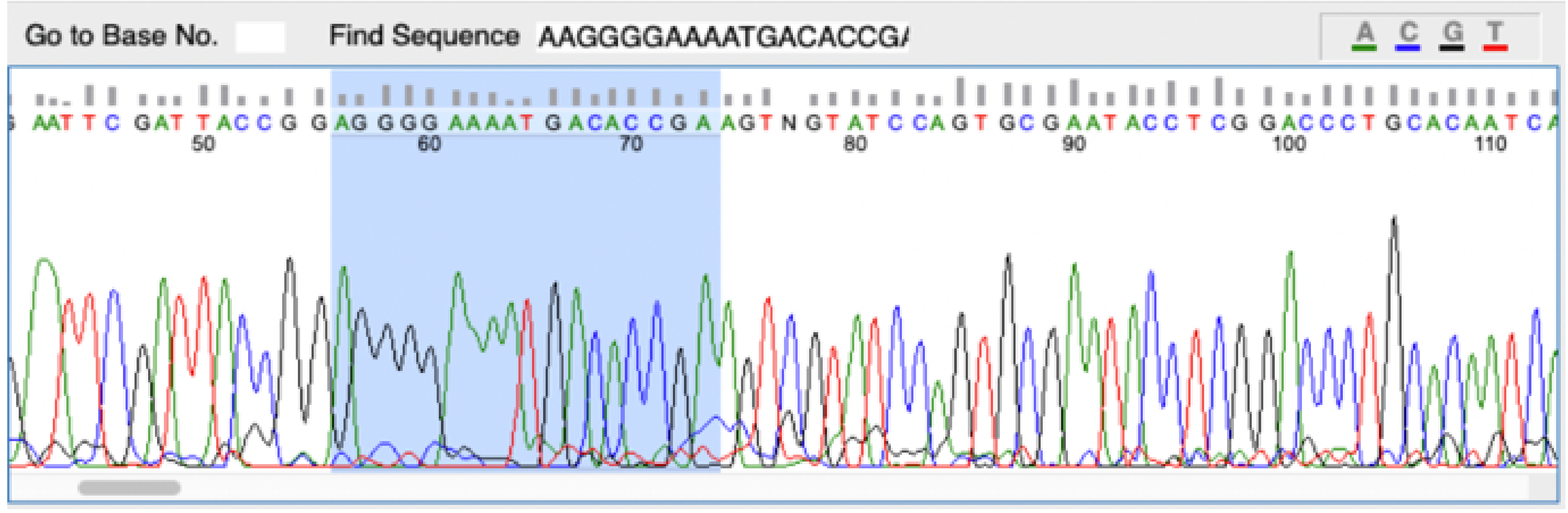
5b_tRF-Tyr-GTA


Go to Base No.
Find Sequence CCTTCGATAGCTCAGCTGG

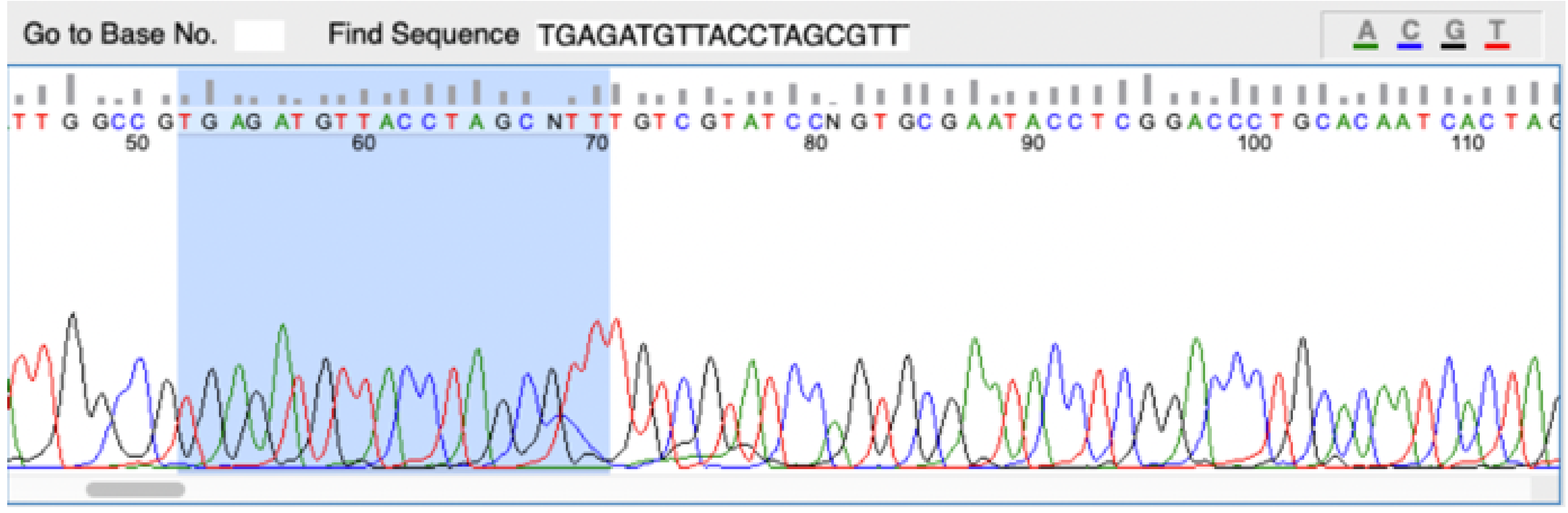
5P_tRNA-Gly-TCC-1-1


Go to Base No.
Find Sequence AGAGGATGCTACCTATCAG

5P_tRNA-Asn-GTT-2-3


Go to Base No.
Find Sequence AAGGGGAAAATGACACCG

3P_tRNA-Val-TAC-1-1


Go to Base No.
Find Sequence TGAGATGTTACCTAGCGTT

3P_tRNA-Arg-CCT-3-1

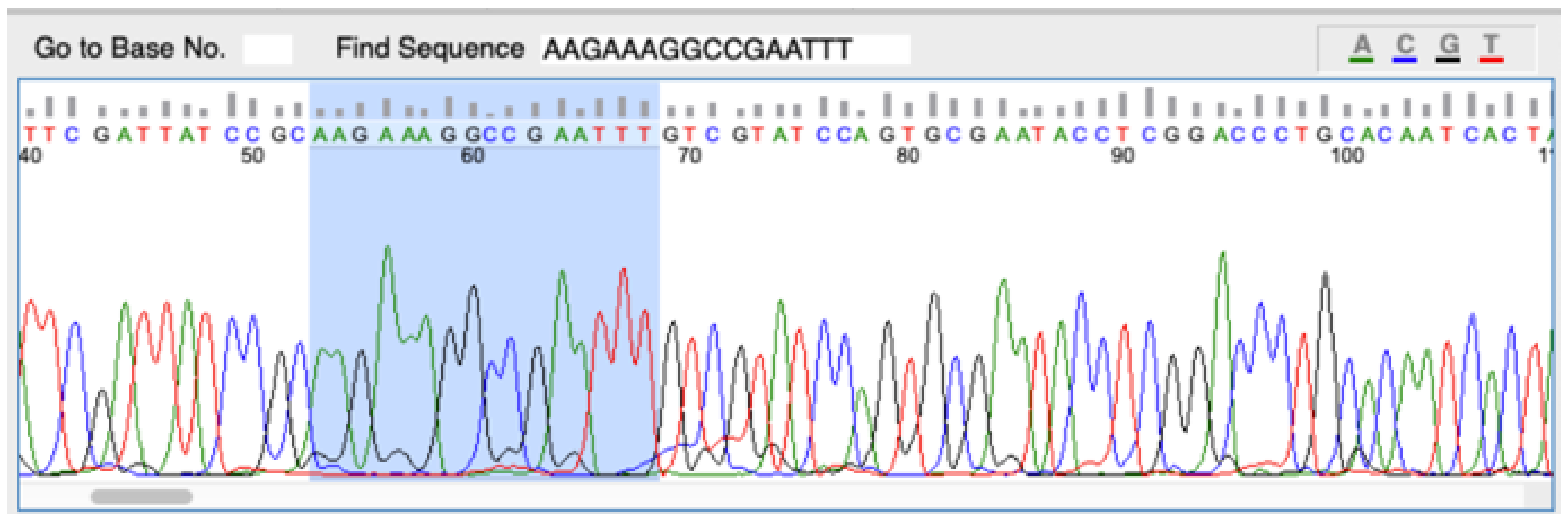

Go to Base No.
Find Sequence AAGAAAGGCCGAATTT

Figure legend:
Fig S5. Sanger sequencing results of tRDFs.
Figure S5 shows the Sanger sequencing results of all the tRDFs we could amplify through TOPO TA clone experiments. It includes 5a_tRF-Ile-AAT/GAT, 5a_tRF-Asp-GTC, 3P_tRNA-Ser-GCT-6-1, 3P_tRNA-Arg-TCG-1-1, 3P_tRNA-Arg-TCT-4-1 (up regulated); 5b_tRF-Tyr-GTA, 5P_tRNA-Gly-TCC-1-1, 5P_tRNA-Asn-GTT-2-3, 3P_tRNA-Val-TAC-1-1 (down regulated); and 3P_tRNA-Arg-CCT-3-1 (unchanged tRDF).

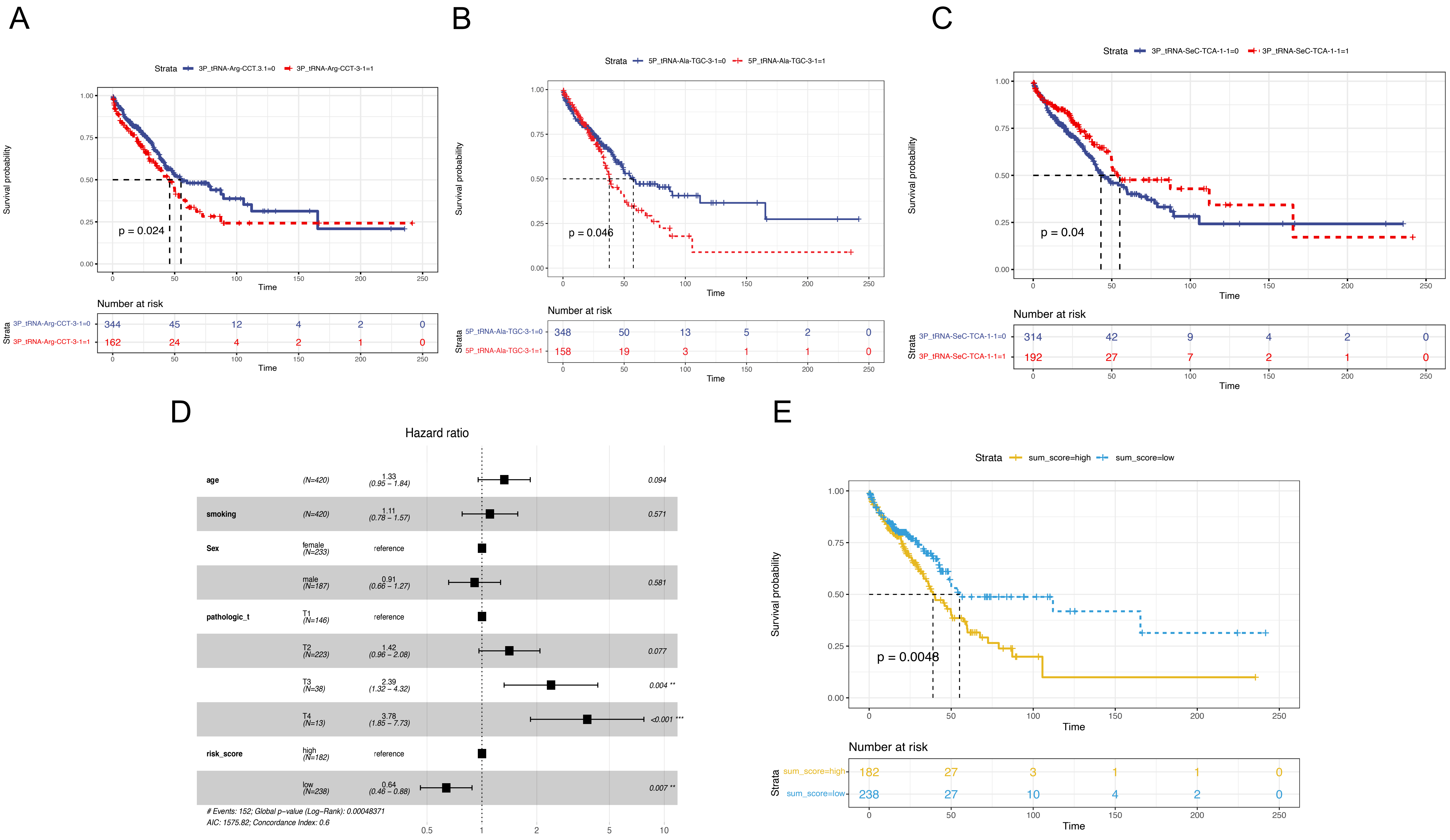


Figure legend:

Fig S6. Prognostic value of tRDFs in TCGA-LUAD among all patients.

A-C. Kaplan-Meier curves show overall survival between high (red) and low expression (blue) in 3P_tRNA-Arg-CCT-3-1, 5P_tRNA-Ala-TGC-3-1, and 3P_tRNA-SeC-TCA-1-1 from the TCGA-LUAD cohort. The group of TCGA-LUAD samples is shown at the bottom of the chart. $P < 0.05$ in the two-sided log-rank test were considered statistically significant. D. Kaplan-Meier Curve shows risk score related survival in high score (yellow) and low score (blue) groups. $P < 0.05$ in the two-sided log-rank test were considered statistically significant. E. Multivariate Cox regression model analysis, including age, smoking status, pathologic T and risk score in TCGA-LUAD cohort.

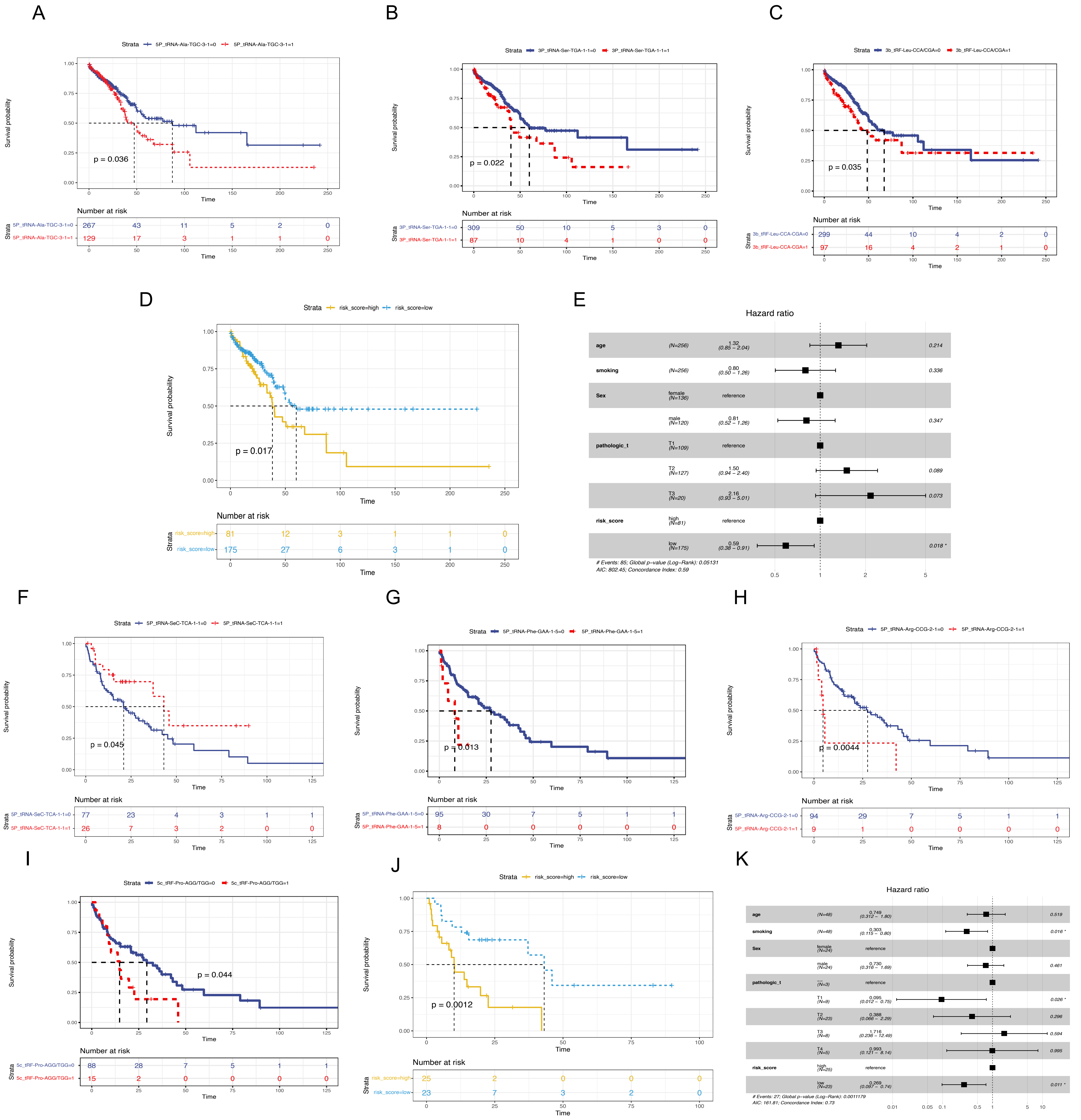


Figure legend:

Fig S7. Prognostic value of tRDFs in TCGA-LUAD among early-stage patients and later stages of patients.

A-C. Kaplan-Meier curves show survival between high (red) and low expression (blue) in 5P_tRNA-Ala-TGC-3-1, 3P_tRNA-Ser-TGA-1-1, and 3b_tRF-Leu-CCA/CGA from LUAD patients in early stages (stage I-II). The group of TCGA-LUAD samples is shown at the bottom of the chart. P < 0.05 in the two sides log-rank test were considered statistically significant.
D. Kaplan-Meier Curve shows risk score related survival in the high score (yellow) and low score (blue) groups. P < 0.05 in the two sides log-rank test were considered statistically significant.
E. Multivariate Cox regression model analysis, including age, smoking status, pathologic T, and risk score in TCGA-LUAD cohort. F-I. Kaplan-Meier curves show survival between high (red) and low expression (blue) in 5P_tRNA-SeC-TCA-1-1, 5P_tRNA-Phe-GAA-1-5, 5P_tRNA-Arg-CCG-2-1, and 5c_tRF-Pro-AGG/TGGc from LUAD patients in later stage (stage III-IV).
J. Kaplan-Meier Curve shows risk score related survival in high score (yellow) and low score (blue) groups. K. Multivariate Cox regression model analysis, including age, smoking status, pathologic T, and risk score in TCGA-LUAD cohort.

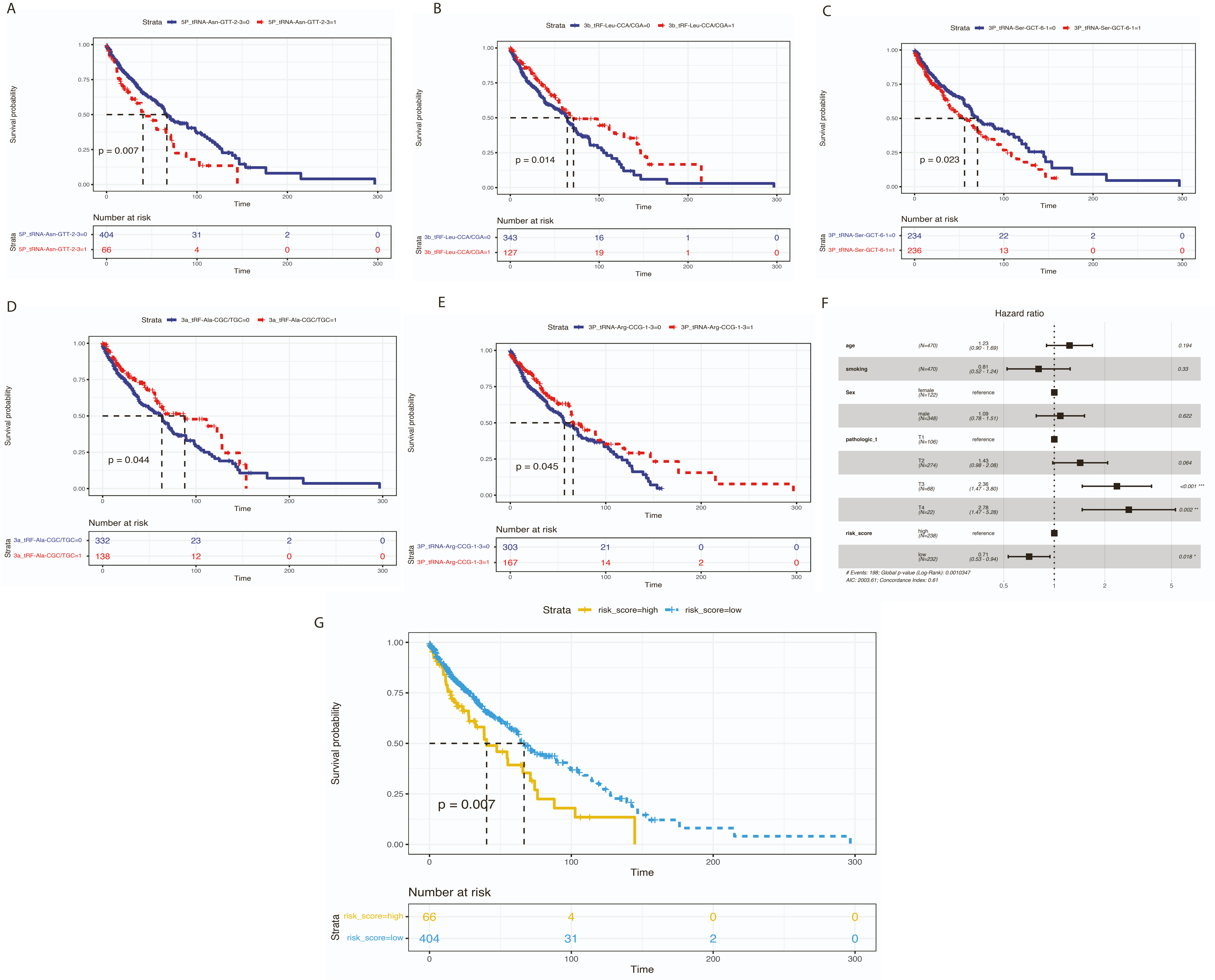


Figure legend:
Fig S8. Prognostic value of tRDFs in TCGA-LUSC among all patients.
A-E. Kaplan-Meier curves show overall survival between high (red) and low expression (blue) in 5P_tRNA-Asn-GTT-2-3, 3b_tRF-Leu-CCA/CGA, 3P_tRNA-Ser-GCT-6-1, 3a_tRF-Ala-CGC/TGC and 3P_tRNA-Arg-CCG-1-3 from the TCGA-LUSC cohort. The group of TCGA-LUSC samples is shown at the bottom of the chart. P < 0.05 in the two sides log-rank test were considered statistically significant. F. Multivariate Cox regression model analysis, including age, smoking status, pathologic T, and risk score in TCGA-LUSC cohort. G. Kaplan-Meier Curve shows risk score related survival in the high score (yellow) and low score (blue) groups. P < 0.05 in the two-sided log-rank test was considered statistically significant.

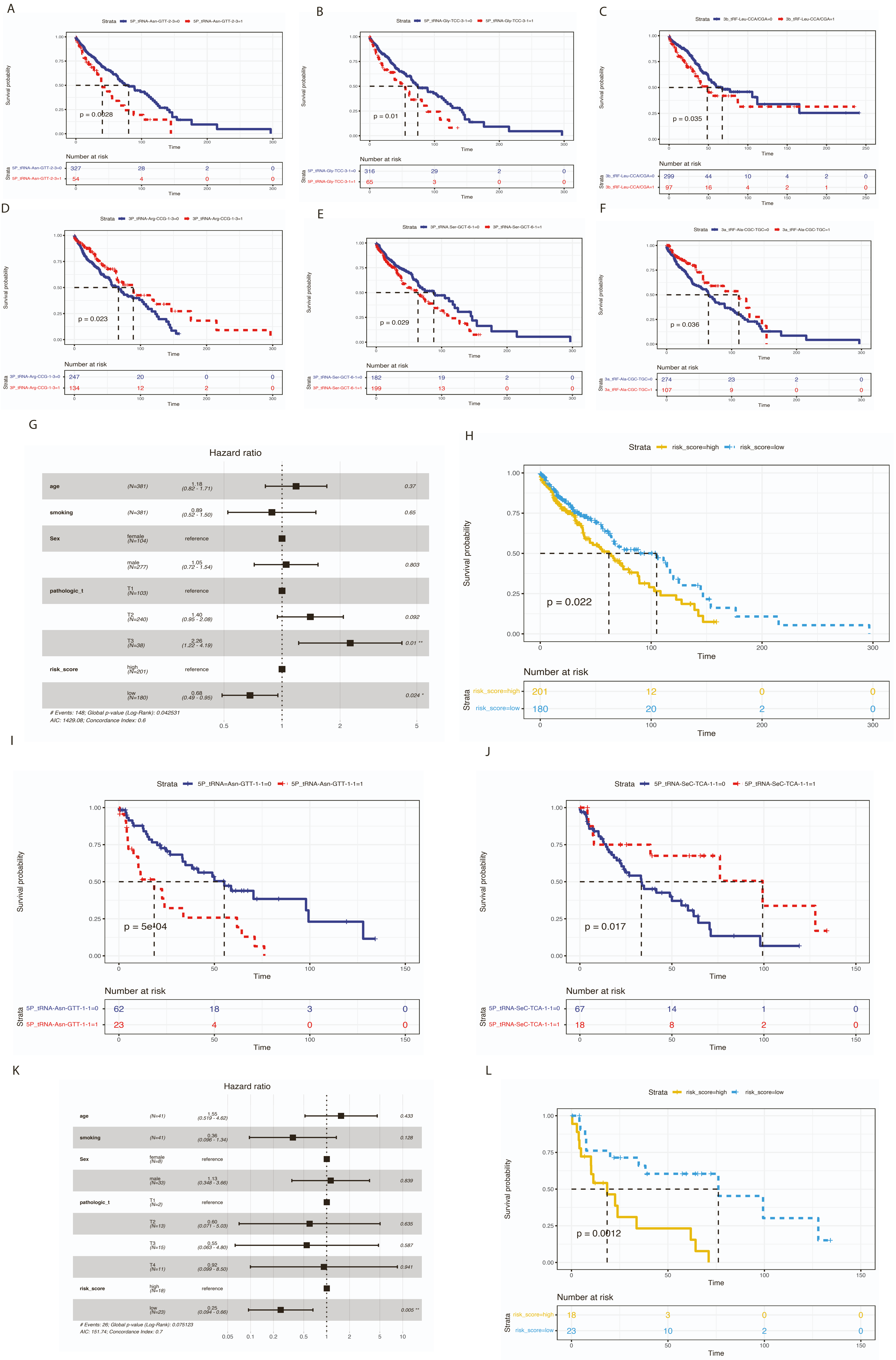


Figure legend:

Fig S9. Prognostic value of tRDFs in TCGA-LUSC among early-stage patients and later stages of patients.

A-F. Kaplan-Meier curves show early stages (stage I-II) survival between high (red) and low expression (blue) in 5P_tRNA-Asn-GTT-2-3, 5P_tRNA-Gly-TCC-3-1, 3b_tRF-Leu-CCA/CGA, 3P_tRNA-Arg-CCG-1-3, 3P_tRNA-Ser-GCT-6-1, and 3a_tRF-Ala-CGC/TGC from TCGA-LUSC cohort. G. Multivariate Cox regression model analysis, including age, smoking status, pathologic t and risk score in TCGA-LUSC cohort. H. Kaplan-Meier Curve shows risk score related survival in the high score (yellow) and low score (blue) groups. P < 0.05 in the two-sided log-rank test was considered statistically significant. I.J. Kaplan-Meier curves show survival between high (red) and low expression (blue) in 5P_tRNA-Asn-GTT-1-1 and 5P_tRNA-SeC-TCA-1-1 from the later stage (stage III-IV) TCGA-LUSC cohort. K. Multivariate Cox regression model analysis, including age, smoking status, pathologic T, and risk score in TCGA-LUSC cohort. L. Kaplan-Meier Curve shows risk score related survival in the high score (yellow) and low score (blue) groups. P < 0.05 in the two sides log-rank test were considered statistically significant.

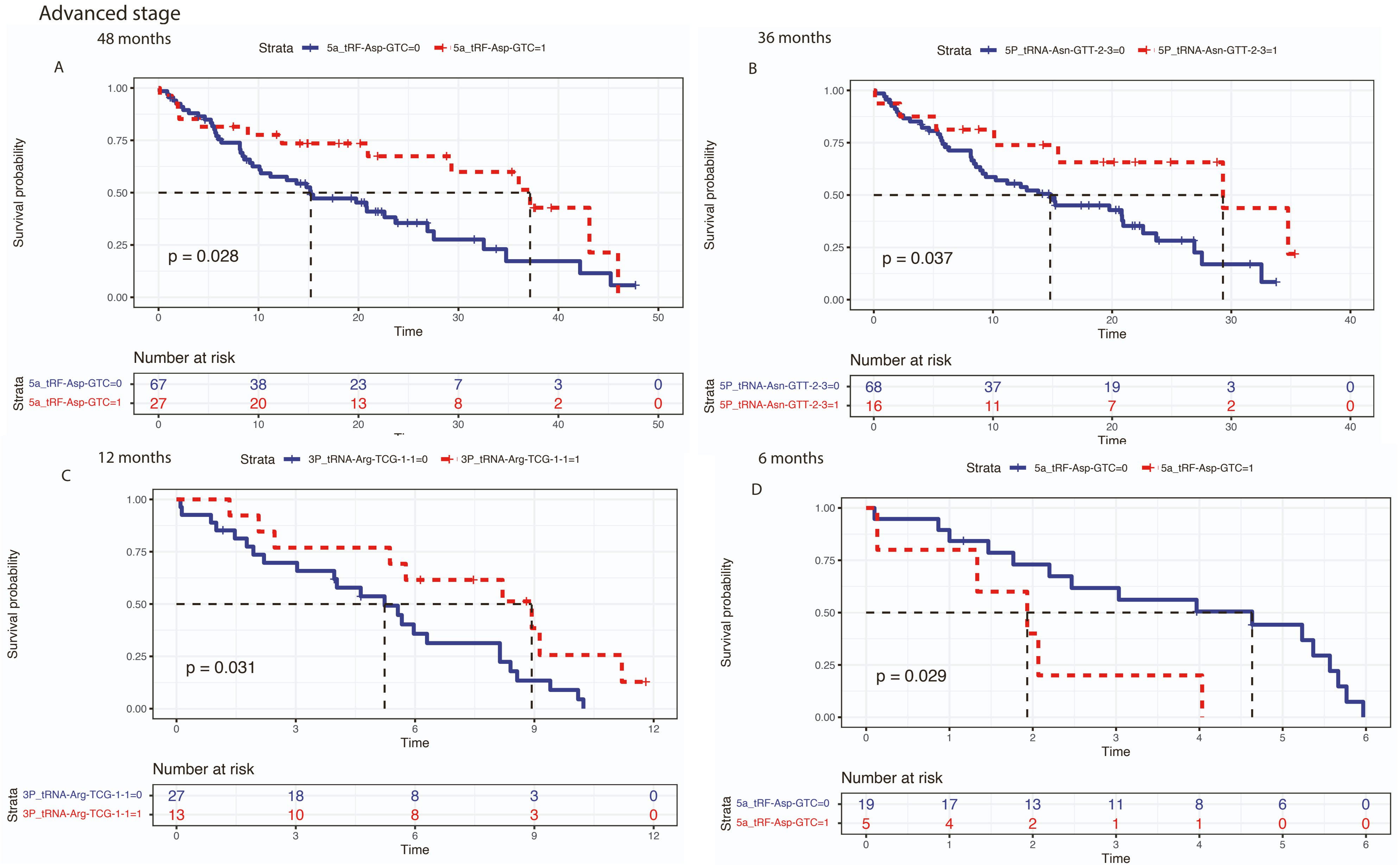


Figure legend:

Fig S10 Prognostic value of tRDF signatures in TCGA-LUAD among later stage patients with certain time endpoint.

A-D. Survival between 5a_tRF-Asp-GTC in 48 months, 5P_Asn-GTT-2-3 in 36 months, 3P_tRNA-Arg-TCG-1-1 in 12 months, 5a_tRF-Asp-GTC in 6 months among advanced stage patients.

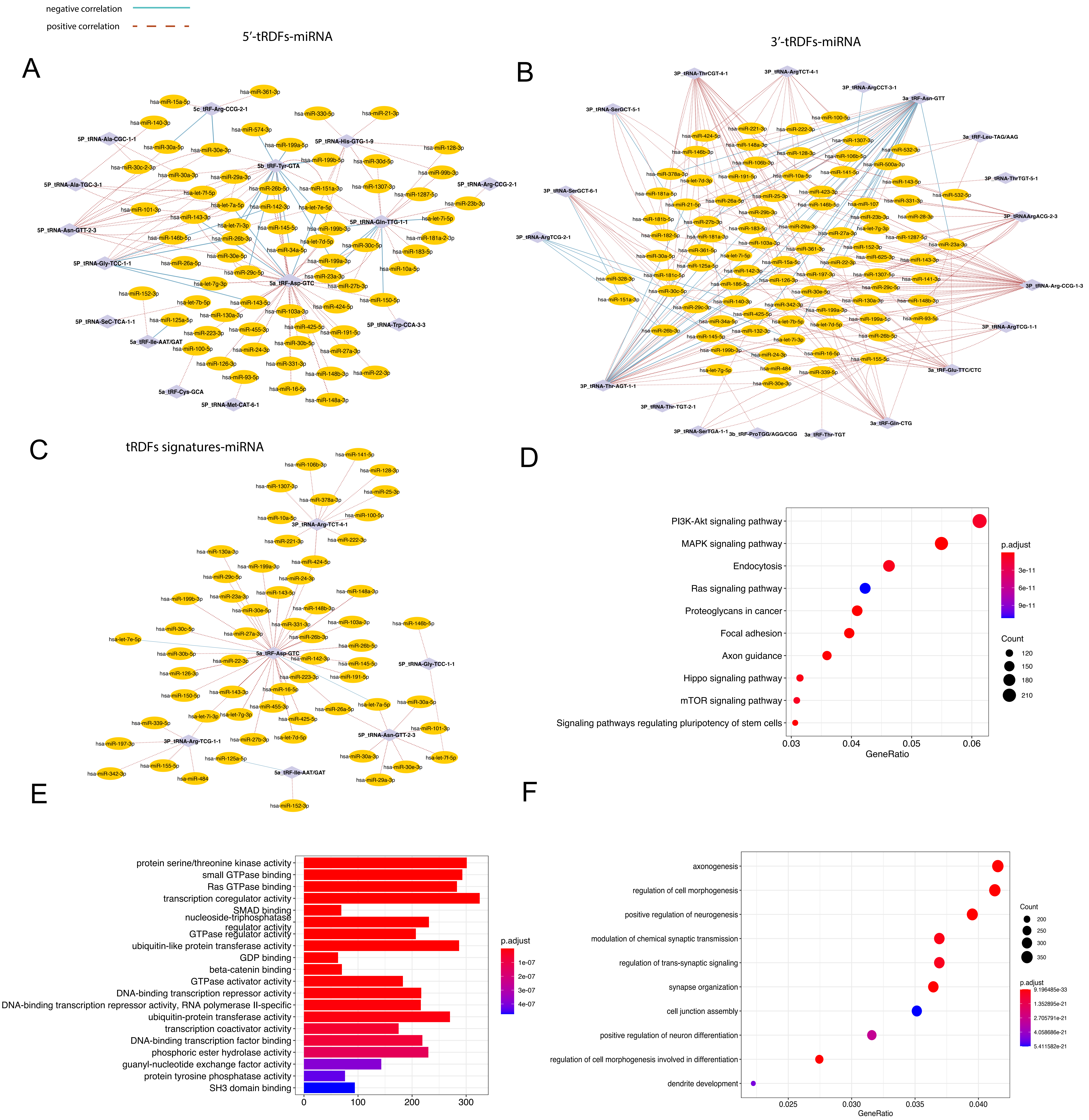


Figure legend:

Fig S11. Transcriptional and post-transcriptional regulation of tRDFs.

A. Correlation network of 5′-tRDFs with miRNAs, red dashed line represents positive correlation and green solid negative correlation represents negative correlation. B. Correlation network of 3′-tRDFs with miRNAs, the red dashed line represents positive correlation and green solid negative correlation represents negative correlation. C. Correlation network of six tRDFs signatures with miRNAs. D. Dot plot shows six signature tRDFs correlated miRNA that targeted gene enriched in KEGG pathway. E, F. Bar plots and dot plots show Molecular Function and Biological Processes of six tRDFs correlated miRNA targeted gene enrichment analysis.

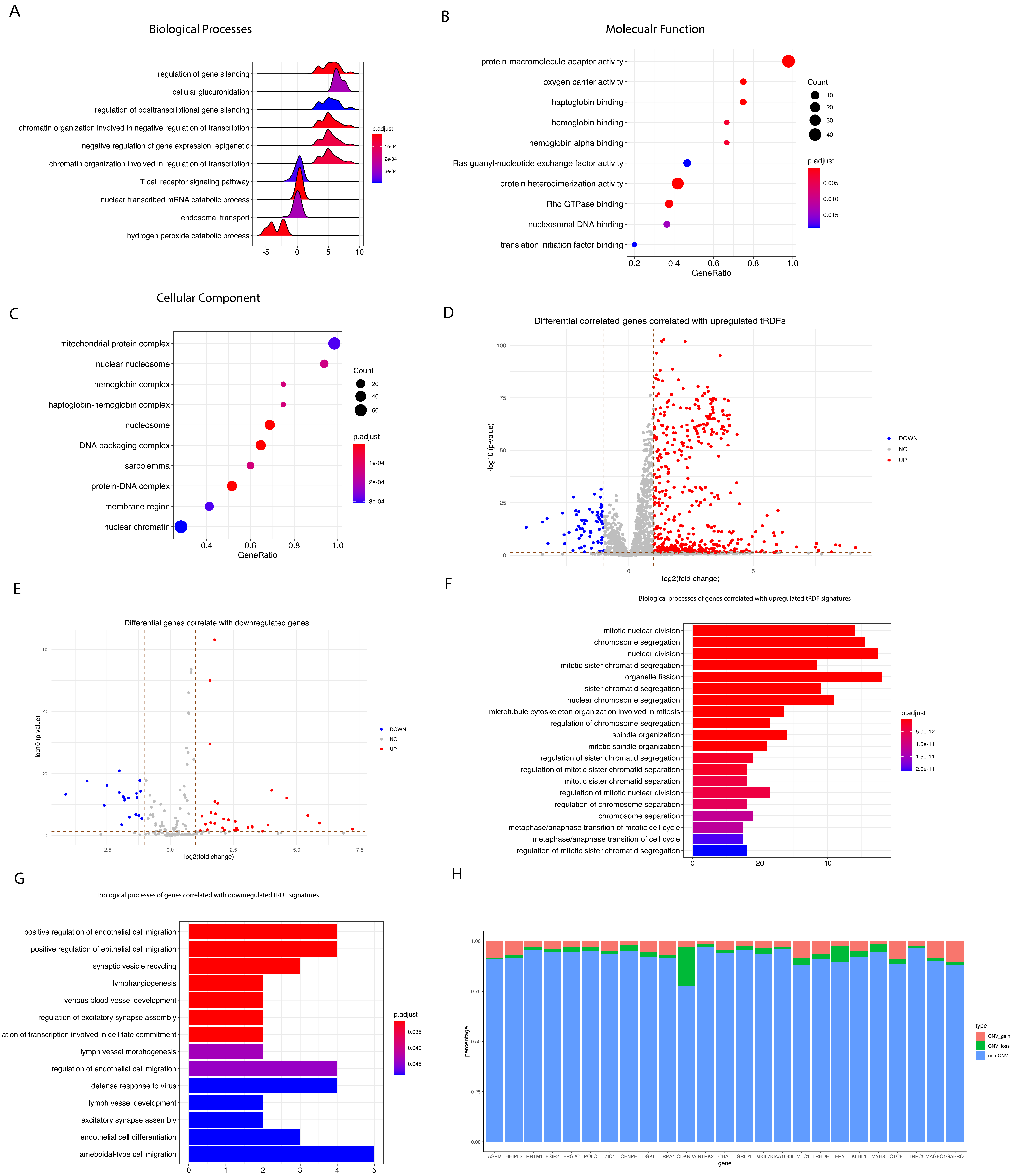


Figure legend:
Fig S12. GSEA and mutation results of tRDFs
A. B. C. Ridgeline plot and dot plot show the top 10 Biological Processes/Molecular Function/Cellular Component of 5782 tRDFs correlated gene from GSEA enrichment analysis(A). In ridge plot, the abscissa indicates the expression distribution of core enriched genes in biological processes(B), in dot plot, the abscissa indicates the gene ratio(C). D. Volcano plot of 2382 genes that correlated with four upregulated tRDFs signatures. E. Volcano plot of 164 genes that correlated with two downregulated tRDFs signatures. F. Biological processes of upregulated tRDF signatures correlated genes. G. Biological processes of downregulated tRDF signatures correlated genes. H. Bar graphs showing the frequency of CNV gain (red), loss (green), and non_CNV (blue) of 25 genes with high mutation frequency.

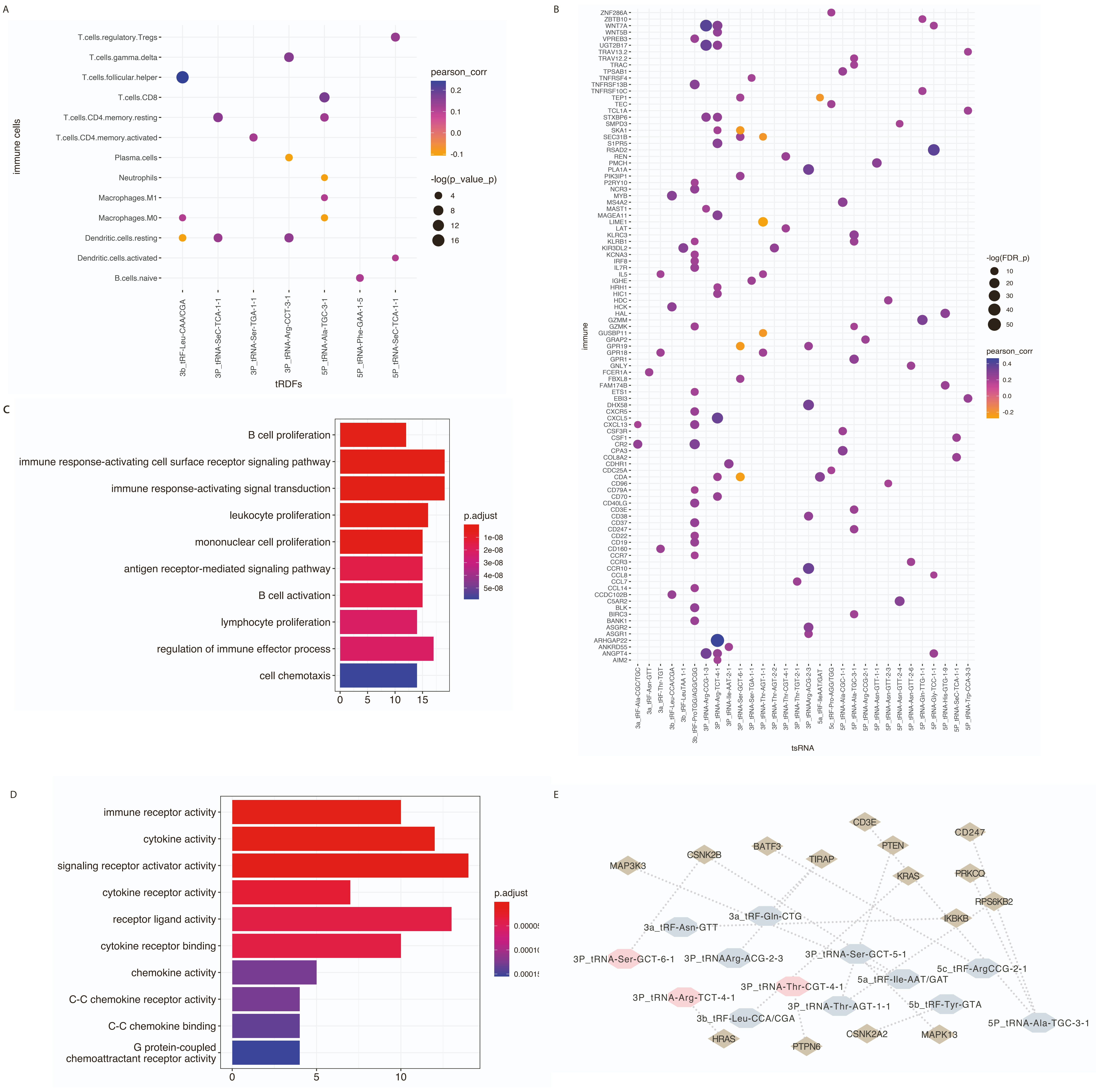


Figure legend:

Fig S13. Enrichment results of immune-related gene.

A. Seven prognostic related tRDFs we once identified through survival analysis were correlated with 13 types of immune cells. B. Dot plot showed correlation between 100 immune-related gene and tRDFs. C.D. Biological Processes (C) and Molecular Function (D) of 100 immune-related genes that correlated with 30 tRDFs.

E. Correlation network of tRDFs and correlated genes that were enriched in the signaling pathway of PD-L1 expression and PD-1 checkpoint pathways.

Reference

1. Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L.Z., Pass, H.I., Kipps, T.J., Liu, C.G., and Croce, C.M. (2016). Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. Proc Natl Acad Sci U S A *113*, 5071-5076. 10.1073/pnas.1604266113.
2. Kumar, P., Mudunuri, S.B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. Nucleic Acids Res *43*, D141-145. 10.1093/nar/gku1138.
3. tRFdb: a relational database of tranfer RNA related Fragments. (2012). http://genome.bioch.virginia.edu/trfdb/index.php.
4. GtRNAdb: tRNAscane-SE analysis of complete genomes. (2021). http://gtrnadb.ucsc.edu/.