## ONLINE SUPPLEMENTAL INFORMATION

# Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C

Sara R. Selitsky[1-3], Jeanette Baran-Gale[1,2], Masao Honda[4], Daisuke Yamane[3,5], Takahiro Masaki[3,5], Emily E. Fannin[2], Bernadette Guerra[6], Takayoshi Shirasaki[4], Tetsuro Shimakami[4], Shuichi Kaneko[4], Robert E. Lanford[6], Stanley M. Lemon[3,5+], and Praveen Sethupathy[1,2,5+*]

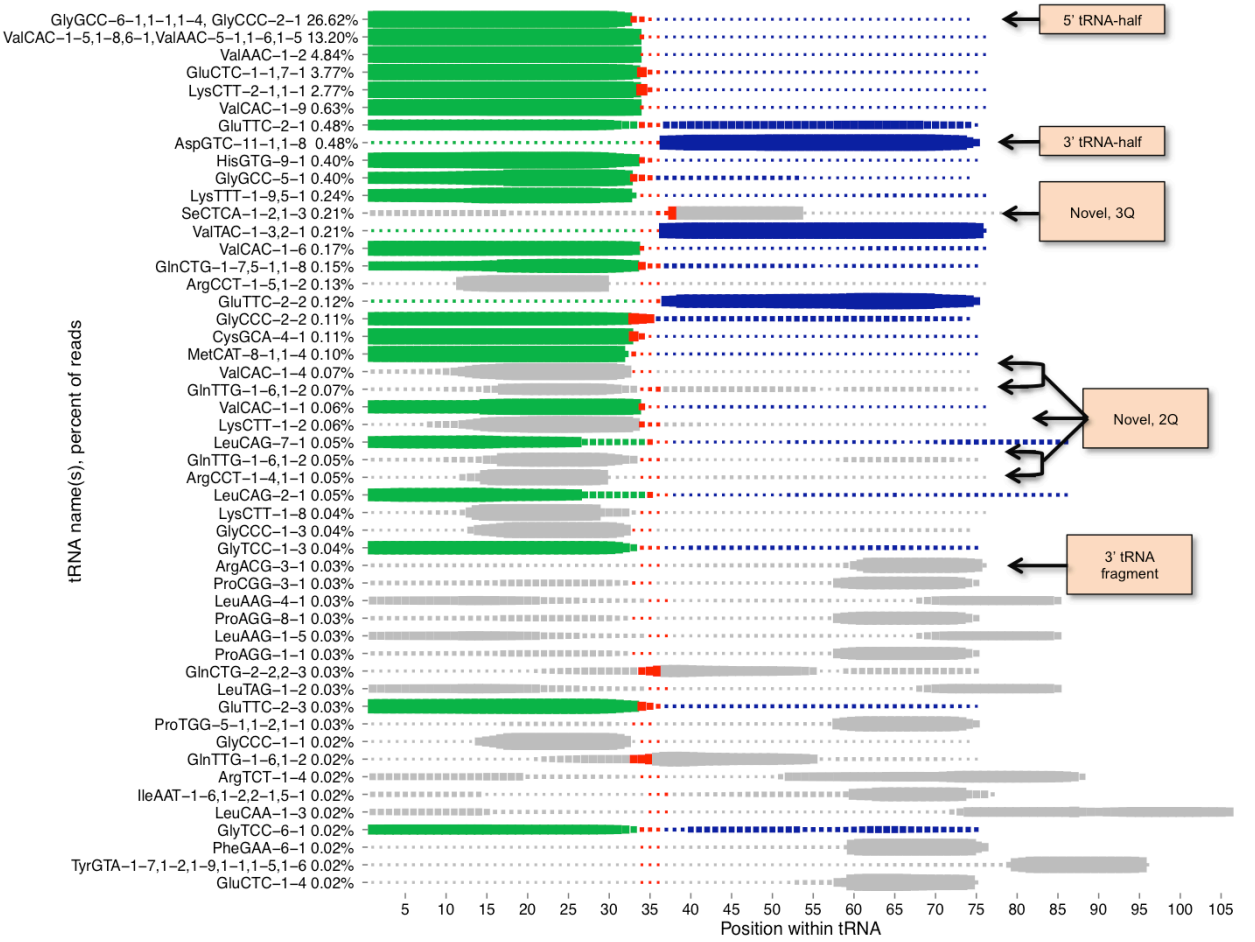
*[1b]ioinformatics and Computational Biology Curriculum, Departments of [2]Genetics, [3]Medicine and [3]Microbiology & Immunology, [5]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, United States of America*

*[4]Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan*

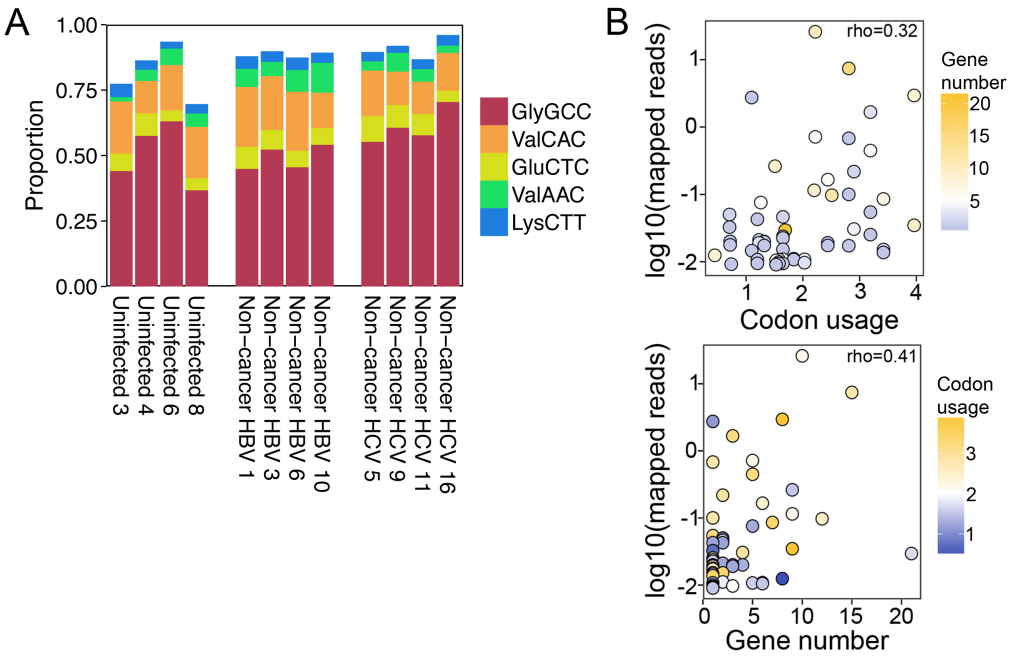
*[6]Department of Virology and Immunology, Texas Biomedical Research Institute and Southwest National Primate Research Center, San Antonio, Texas, United States of America*

[+]These authors contributed equally to this work.

* To whom correspondence should be addressed. Tel: 919-966-6387; Fax: 919-843-4682; Email: praveen_sethupathy@med.unc.edu




**Supplemental Figure 1.** Example of a tRNA coverage plot for small RNAs detected in non-malignant tissue from HBV-infected subject 6. Each line in the coverage plot represents tRNA families from which the tRNA-derived RNA could originate. The tRNAs are listed using a nomenclature that groups tRNA gene paralogs by sequence: "X-Y-Z", where X is the tRNA amino acid and anticodon, Y is the number of tRNA genes with the identical sequence, and Z is a unique identifier for each tRNA family. For example, GlyGCC-6-1 is a tRNA that carries Gly with a GCC in the anticodon, and which could be derived from 6 different tRNA genes and its unique ID is 1. Supplemental Table 5 provides the new tRNA sequence ID, the position of the anticodon for each tRNA sequence, the tRNA sequence, and all the tRNA genes to which each sequence ID corresponds. If a tRNA-derived RNA could have originated from multiple sequences, each is listed separately in the label for each row. If multiple sequences have the same anticodon, then the number identifiers are each listed, comma separated after the amino acid and anticodon. Superimposed on the coverage plot are boxes (peach) that show the name of a tRNA-derived RNA type and arrows pointing to examples of each type. Different tRNA-derived RNAs are named as suggested by Garcia-Silva *et al.*[1], differentiating between tRNA-fragments (tRFs) and tRNA-halves (tRHs). Also shown are novel tRNA-derived RNAs that we have termed "second quarter" (2Q) or "third quarter" (3Q) tRNA fragments, depending on where they originate from within the tRNA.




**Supplemental Figure 2.** (**a**) Proportional abundance of the top five tRNA-derived RNAs. Proportion of all tRNA-derived RNAs represented by each of the five most abundant tRNA-derived RNAs (rank based from Fig. 1c) in uninfected and non-malignant chronic viral hepatitis liver tissue. (**b**) Correlation between the $\log_{10}$ (relative abundance) of the 60 most abundant tRNA-derived RNAs (averaged across all 12 samples) and (top) codon usage (from the Genomic tRNA Database) and (bottom) tRNA gene number. Symbols are shaded according to the percent codon usage, or number of full-length tRNAs that each tRNA-derived RNA could originate from.

A

```
GlyGCC-6-1 GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCAATGCACCA
GlyCCC-2-1 GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTCCCACGCGGGAGACCCGGGTTCAATTCCCGGCCAATGCACCA
GlyGCC-1-1 GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTTGATTCCCGGCCAGTGCACCA
GlyGCC-1-4 GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCATGCGGGCGGCCGGGCTTCGATTCCTGGCCAATGCACCA
           ******************************** *** ***** * ** ** **  ***** ***** *******
```

B

```
ValCAC-1-5 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCGTAAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValCAC-1-8 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCG-AAAGGTCCCCGGTTCGAAACTGGGCGGAAACACCA
ValCAC-6-1 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCG-AAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValAAC-1-6 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCGCGGTTCGAAACCGGGCGGAAACACCA
ValAAC-5-1 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValAAC-1-5 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCCTGGATCAAAACCAGGCGGAAACACCA
           ********************************* ******* ********  ** ** ****  *************
```

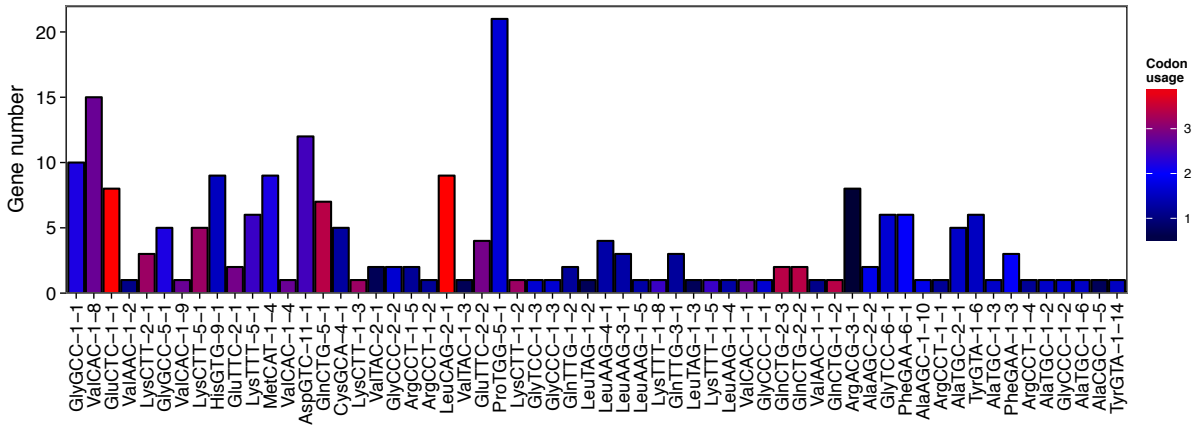
C

```
ValCAC-1-5 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCGTAAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValCAC-6-1 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCG-AAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValCAC-1-8 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCG-AAAGGTCCCCGGTTCGAAACTGGGCGGAAACACCA
ValAAC-1-5 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCCTGGATCAAAACCAGGCGGAAACACCA
ValAAC-5-1 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValAAC-1-6 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCG-AAAGGTCCGCGGTTCGAAACCGGGCGGAAACACCA
GlyGCC-1-1 GCATTGGTGGTTCAGTGGT-AGAATTCTCGCCTGCCACGCG-GGAGG-CCCGGGTTTGATTCCCGGCCAGTGCACCA
GlyGCC-6-1 GCATTGGTGGTTCAGTGGT-AGAATTCTCGCCTGCCACGCG-GGAGG-CCCGGGTTCGATTCCCGGCCAATGCACCA
GlyGCC-1-4 GCATTGGTGGTTCAGTGGT-AGAATTCTCGCCTGCCATGCG-GGCGG-CCGGGCTTCGATTCCTGGCCAATGCACCA
GlyCCC-2-1 GCATTGGTGGTTCAGTGGT-AGAATTCTCGCCTCCCACGCG-GGAGA-CCCGGGTTCAATTCCCGGCCAATGCACCA
           *  *  ** **  ****** *  *   ******  ** ***    *  **  *  *  *  *  ***     *****
```

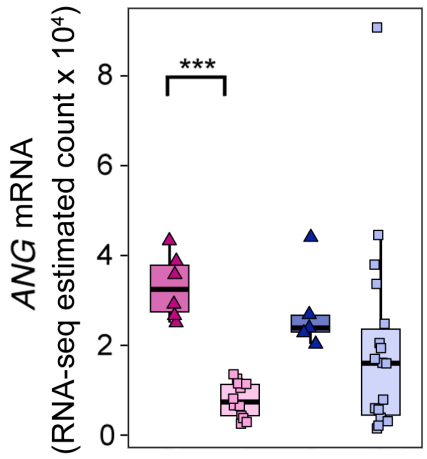
D

```
GluCTC-7-1 TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCTCTCACCGCCGCGGCC-CGGGTTCGATTCCCGGTCAGGGAACCA
GluTTC-2-1 TCCCACATGGTCTAGCGGTTAGGATTCCTGGTTTTCACCCAGGCGGCC-CGGGTTCGACTCCCGGTGTGGGAACCA
LysCTT-2-1 GCCCGGCTAGCTCAGTCGGTAGAGCATGGGACTCTTAATCCCAGGGTCGTGGGTTCGAGCCCCACGTTGGGCGCCA
LysCTT-5-1 GCCCGGCTAGCTCAGTCGGTAGAGCATGAGACTCTTAATCTCAGGGTCGTGGGTTCGAGCCCCACGTTGGGCGCCA
GlyGCC-1-1 GCATTGGTGGTTCAGTGGT-AGAATTCTCGCCTGCCACGCGGGAGG-CCCGGGTTTGATTCCCGGCCAGTGCACCA
GlyGCC-5-1 GCATGGGTGGTTCAGTGGT-AGAATTCTCGCCTGCCACGCGGGAGG-CCCGGGTTCGATTCCCGGCCCATGCACCA
ValCAC-6-1 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCGAAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValAAC-1-2 GTTTCCGTAGTGTAGTGGTCATCACGTTCGCCTAACACGCGAAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
ValCAC-1-9 GCTTCTGTAGTGTAGTGGTTATCACGTTCGCCTCACACGCGAAAGGTCCCCGGTTCGAAACCGGGCAGAAGCACCA
HisGTG-9-1 GCCGTGATCGTATAGTGGTTAGTAC-TCTGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAGTCACGGCACCA
                 * *   **  *  *          *              *   **** **  **          ***
```

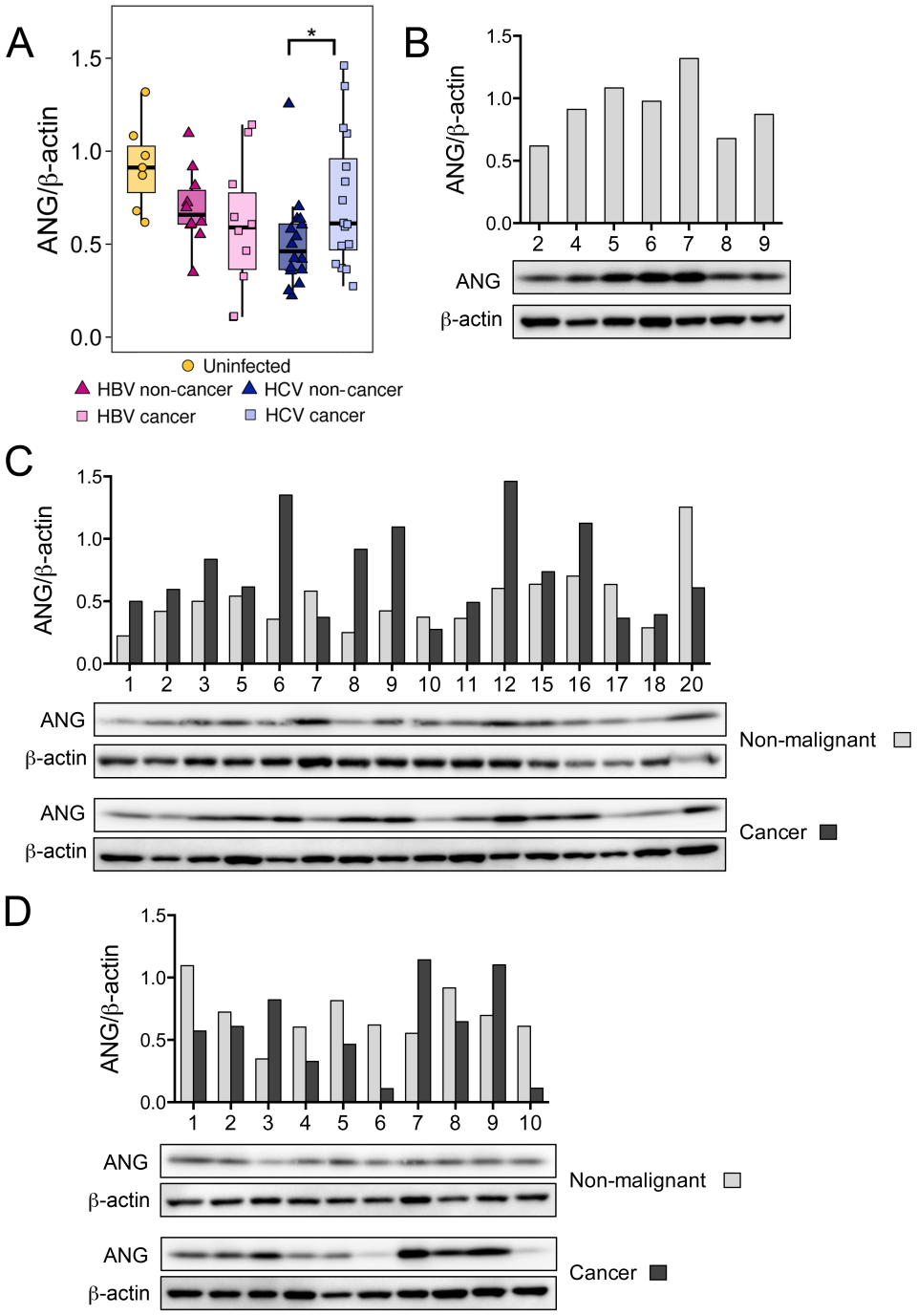
**Supplemental Figure 3** Multiple sequence alignment of tRNA genes from which 5' tRHs could have originated. (**a**) Nucleotide sequences of tRNA genes from which 5' $tRH^{Gly}$ ("Gly[C/G]CC") could originate. (**b**) Nucleotide sequences of tRNA genes from which 5' $tRH^{Val}$ ("Val[A/C]AC") could originate. (**c**) Nucleotide sequences of tRNA genes from (a) and (b). (**d**) Nucleotide sequences from representative tRNA genes from which the 10 most abundant tRNA-derived RNAs in non-malignant tissue could originate. Multiple sequence alignments were generated by ClustalW2[2] of tRNA sequences[3]. tRNA nomenclature is described in the legend to Supplemental Figure 1. The anticodon is highlighted in yellow.



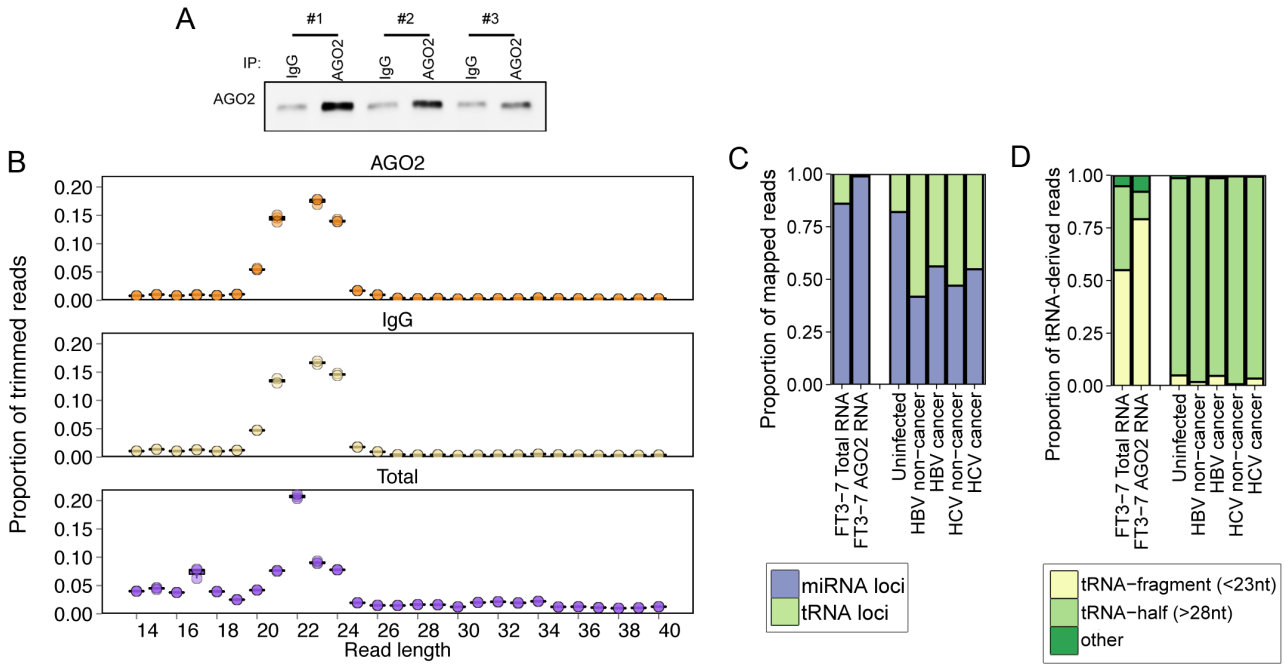

**Supplemental Figure 4.** tRNA gene number and codon usage. The height of each bar depicts the number of tRNA genes from which each of tRNA-derived RNAs could be derived, while the color represents the percent codon usage for that tRNA (downloaded from Genomic tRNA database). Representative tRNA names for tRNA-derived RNAs are ordered from left to right based on the average tRNA-derived RNA abundance in non-malignant tissues.




**Supplemental Figure 5.** *ANG* RNA-seq estimated count (RSEM) from The Cancer Genome Atlas (TCGA) RNA-sequence data: non-cancer (*n*=6) and cancer (*n*=12) liver tissue from chronic hepatitis B subjects, and non-cancer (*n*=5) and cancer (*n*=18) from chronic hepatitis C subjects. ***P < 0.005, calculated by Mann-Whitney *U*-test.

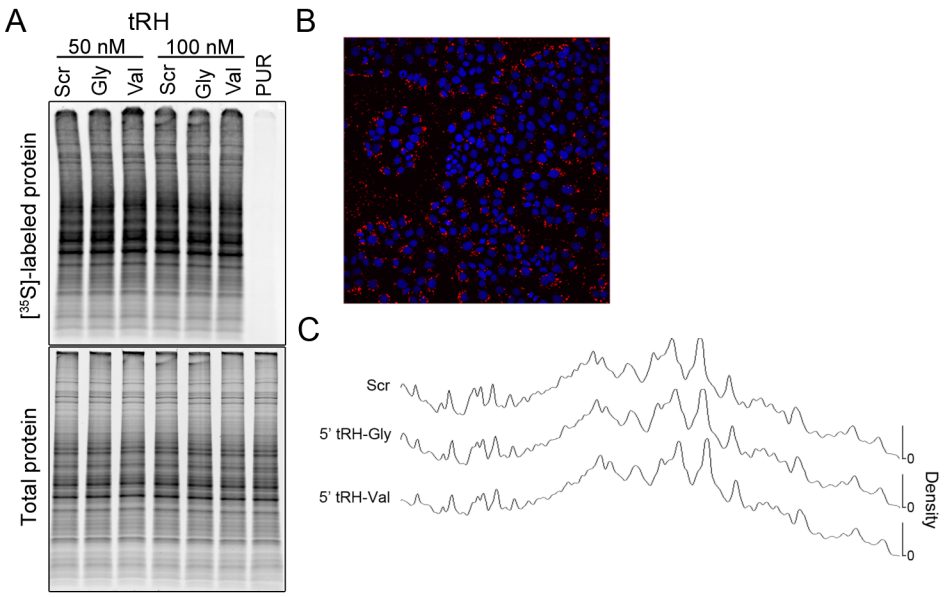



**Supplemental Figure 6.** Immunoblots showing angiogenin (ANG) expression in uninfected tissues and in paired non-malignant and cancer tissue samples. (**a**) ANG protein expression determined from immunoblots (see panels b-d). (**b**, **c**, **d**) ANG blot intensity measured by densitometry was normalized to b-actin signal (loading control) in blots of proteins extracted from uninfected (n=10) and paired non-malignant (top rows) and cancer (bottom rows) tissues from HBV- (n=10) and HCV-infected (n=16) subjects. Densitometry quantitation (ANG/b-actin) is shown above the blots. In panels (c) and (d), lightly shaded bars represent non-malignant tissue, and darkly shaded bars cancer tissue.




**Supplementary Figure 7.** To determine whether tRHs associate with argonaute 2 (AGO2), the main effector protein within RISC, we sequenced small RNAs associating with AGO2 in a pull-down assay and compared these to the total small RNA pool in FT3-7 hepatoma cells. (**a**) AGO2 immunoprecipitation. Immunoblots of products from immunoprecipitation experiments (n= 3 technical replicates) carried out with antibody against AGO2 and isotype-matched control (IgG). (**b**) Length distribution of reads from sequencing of small RNAs associating with AGO2 pull-down (*n*=3 replicates), isotype control (IgG) pull-down (*n*=3 replicates), and total RNA from FT3-7 cells. The first peak (17 nts) in total RNA is from ribosomal RNA. (**c**) Median proportion of reads mapping to miRNA versus tRNA loci in FT3-7 cells and AGO2 immunoprecipitates shown in comparison with small RNA from liver of uninfected subjects and malignant and non-malignant tissue from subjects with chronic hepatitis B and C (*n*=4 each). (**d**) Median proportion of reads mapping to tRNAs that correspond to tRFs (<23 nts), tRHs (>28 nts), or neither ("other").

Approximately 14% of small RNAs in FT3-7 cells mapped to tRNA genes, compared to <1% of those in the AGO2-associated fraction. Furthermore, the relative abundance of 5' tRHs was ~3-fold less in AGO2-associated RNA compared to the total small RNA pool. Thus, tRHs are not selectively associated with AGO2 and likely do not function generally as miRNAs.




**Supplemental Figure 8.** Metabolic labeling of newly synthesized proteins in human hepatoma (Huh7) cells transiently transfected with synthetic 5' tRHGly (Gly[C/G]CC), 5' tRHVal (Val[A/C]AC) (the two most abundant tRHs found in human liver), or a 31 nt scrambled sequence oligonucleotide (50 nM or 100 nM each). (**a**) The top panel shows [$^{35}$S]-labeled proteins synthesized in the 12 hours prior to cell lysis separated by SDS-PAGE (phosphoimager analysis); the bottom panel shows total protein abundance in each cell lysate (Sypro-Ruby staining). The lane labeled "PUR" was loaded with lysate from cells treated with puromycin. (**b**) Human hepatoma (RCF-26) cells were transfected with fluorescently labeled 5' tRH$^{Val}$ (20 nM) to gage transfection efficiency. Signal appears present in the majority of cells. (**c**) Densitometry traces from SDS-PAGE loaded with [$^{35}$S]-labeled protein products (100 nM results in panel a). Transfection of cells with high concentrations of tRHs appears to have no significant effects on cellular translation.

## SUPPLEMENTAL REFERENCES

1 Garcia-Silva, M. R., Cabrera-Cabrera, F. & Güida, M. C. Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms. *Genes*, (2012).

2 Larkin, M. A., Blackshields, G., Brown, N. P. & Chenna, R. Clustal W and Clustal X version 2.0. *Bioinformatics*, (2007).

3 Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research* **37**, D93-97, (2009).