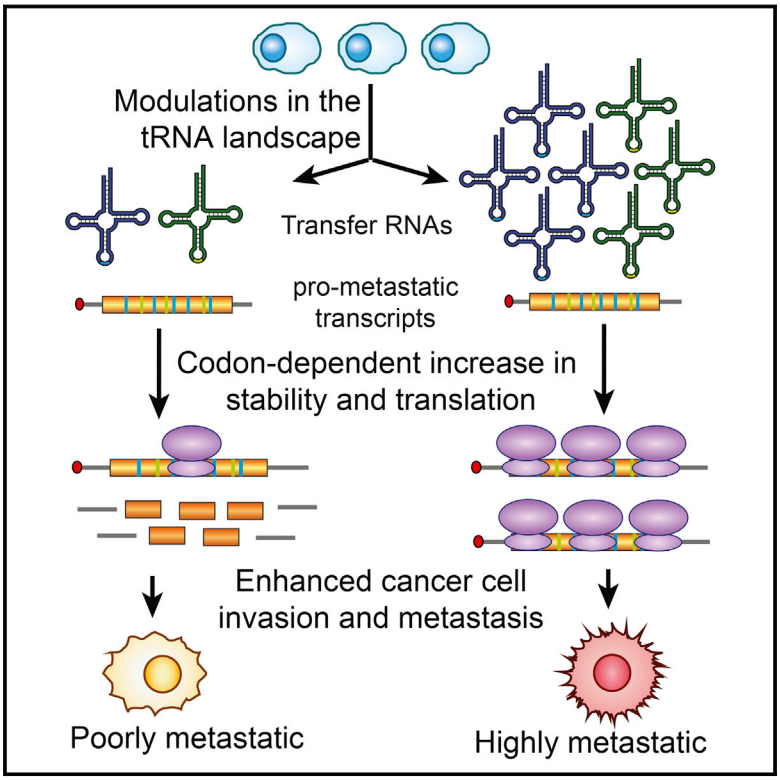
Article

Cell

Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression

Graphical Abstract



Authors

Hani Goodarzi, Hoang C.B. Nguyen, Steven Zhang, Brian D. Dill, Henrik Molina, Sohail F. Tavazoie

Correspondence

hgoodarzi@mail.rockefeller.edu (H.G.), stavazoie@mail.rockefeller.edu (S.F.T.)

In Brief

A new tRNA profiling method reveals that specific tRNAs are upregulated in metastatic breast cancer cells and drive metastasis by enhancing stability and translation of transcripts enriched for their cognate codons.

Highlights

- Specific tRNAs are upregulated in highly metastatic breast cancer cells
- TRNAs promote stability and translation of transcripts enriched for their codons
- $TRNA^{Glu}_{UUC}$ drives metastasis by directly upregulating *EXOSC2* and enhancing *GRIPAP1*

Accession Numbers

GSE77401

Goodarzi et al., 2016, Cell *165*, 1416–1427
June 2, 2016
http://dx.doi.org/10.1016/j.cell.2016.05.046

CellPress

Cell

Article

Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression

Hani Goodarzi,[1,3,*] Hoang C.B. Nguyen,[1,3] Steven Zhang,[1] Brian D. Dill,[2] Henrik Molina,[2] and Sohail F. Tavazoie[1,*]
[1]Laboratory of Systems Cancer Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[2]Proteome Resource Center, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[3]Co-first author
*Correspondence: hgoodarzi@mail.rockefeller.edu (H.G.), stavazoie@mail.rockefeller.edu (S.F.T.)
http://dx.doi.org/10.1016/j.cell.2016.05.046

SUMMARY

Transfer RNAs (tRNAs) are primarily viewed as static contributors to gene expression. By developing a high-throughput tRNA profiling method, we find that specific tRNAs are upregulated in human breast cancer cells as they gain metastatic activity. Through loss-of-function, gain-of-function, and clinical-association studies, we implicate tRNAGluUUC and tRNAArgCCG as promoters of breast cancer metastasis. Upregulation of these tRNAs enhances stability and ribosome occupancy of transcripts enriched for their cognate codons. Specifically, tRNAGluUUC promotes metastatic progression by directly enhancing *EXOSC2* expression and enhancing *GRIPAP1*—constituting an "inducible" pathway driven by a tRNA. The cellular proteomic shift toward a pro-metastatic state mirrors global tRNA shifts, allowing for cell-state and cell-type transgene expression optimization through codon content quantification. TRNA modulation represents a mechanism by which cells achieve altered expression of specific transcripts and proteins. TRNAs are thus dynamic regulators of gene expression and the tRNA codon landscape can causally and specifically impact disease progression.

INTRODUCTION

Transfer RNAs and the genetic code underlying protein synthesis are universal to all domains of life (Dever and Green, 2012). Despite this universality, genomes exhibit substantial variations in their preference for specific codons across their coding sequences. The source of this bias, though still debated, likely reflects selection for translational efficiency and accuracy (Drummond and Wilke, 2008; Plotkin and Kudla, 2011; Shah and Gilchrist, 2011). Importantly, even the genes within the same genome show high levels of variation in their codon preferences and synonymous codon usage bias. While rigorous proof remains lacking, there is substantial evidence linking these observed variations to different aspects of cellular biology. Given the link between protein synthesis rates, protein concentration, and optimized growth and function (Han et al., 2014; Li et al., 2014), it is conceivable that the components of translation machinery may affect protein expression levels in a concerted fashion. In *Saccharomyces cerevisiae*, the estimated translational speeds determined across all genes showed a significant correlation between codon usage bias and tRNA abundances, highlighting codon usage as an optimizing factor in overall cellular efficiency (Qian et al., 2012). At the same time, the role of tRNAs as direct modulator of translation efficiency in yeast has been challenged (Pop et al., 2014).

On the other hand, microarray-based analysis of tRNA abundances in various tissues has shown a significant correlation between tRNA content and codon usage bias of highly expressed tissue-specific proteins (Dittmar et al., 2006). Given that protein synthesis rate is correlated with tRNA abundance in transgene overexpression experiments (Zouridis and Hatzimanikatis, 2008), it is further hypothesized that tRNA content may effectively regulate the rate of translation for a subset of endogenous proteins (Gustafsson et al., 2004). In contrast, measurements of in vivo translational speeds in *S. cerevisiae* have shown that preferentially used codons may not necessarily be translated faster (Qian et al., 2012) and may instead stem from optimization for efficiency at a cellular level rather than for translational speed. The significant role of tRNAs in regulating gene expression is further supported by recent experiments revealing that variations in active tRNA content of bacteria play an adaptive role in response to environmental cues (Subramaniam et al., 2013). Importantly, a recent study based on microarray profiling of tRNA contents in various human cell lines and samples characterized distinct tRNA signatures to correlate with proliferation and differentiation, two distinct cellular programs (Gingold et al., 2014).

It has been hypothesized that modulations in tRNA content may impact the protein expression landscape of the cell (Begley et al., 2007; Chan et al., 2010; Dittmar et al., 2006; Pavon-Eternod et al., 2009). Moreover, rare codon bias has been shown to impact Kras-driven tumorigenesis (Pershing et al., 2015). More recently, codon optimality was also reported as a major determinant of mRNA stability (Presnyak et al., 2015). Additionally, mutation of a tissue-specific tRNA expressed in the mouse nervous system was implicated as the cause of neurodegeneration (Ishimura et al., 2014). However, the regulatory consequences of tRNA modulations and their potential direct roles in gene expression control and human disease remain poorly characterized. Here, we have performed an unbiased study of tRNA abundances in malignant and non-malignant human cell lines.

We find that highly metastatic sublines derived from distinct parental cancer cell populations exhibit similar modulations in their tRNA content relative to their isogenic poorly metastatic parental lines. Through loss-of-function and gain-of-function experiments, we establish a causal role for two specific tRNA species as promoters of breast cancer metastasis. Increased expression of these specific tRNAs reshapes protein expression through the direct modulation of ribosome occupancy and/or transcript stability of specific transcripts enriched for codon complementary to these tRNAs. We reveal that increased expression of a specific tRNA enhances the expression of a direct target gene downstream of the tRNA in a codon-specific manner. The downstream target(s) of this tRNA constitute novel promoters of metastasis and, in combination with their upstream regulatory tRNA, form a tRNA-activated pathway that drives cancer progression.

RESULTS

Metastatic Progression and Modulations in the tRNA Expression Landscape

To create a snapshot of the tRNA landscape in different cellular contexts, we measured the cellular content of various tRNA species in five different human cell lines (Minn et al., 2005; Tavazoie et al., 2008): a non-tumorigenic epithelial cell line (MCF10a), poorly metastatic breast cancer lines MDA-231 and CN34, and their respective highly metastatic sub-lines MDA-LM2 and CN-LM1a (Figure 1A). To do this, we developed an approach based on hybridization and subsequent ligation of complementary DNA probes. Due to their strong secondary structures and extensive base modifications, tRNAs are not suitable substrates for reverse transcription. Thus, their quantification using common cDNA-based approaches results in unpredictable biases and spurious measurements (Dittmar et al., 2006). While enzymatic removal of certain tRNA modifications and the application of highly processive reverse transcriptase enzymes provide a promising avenue for tRNA-sequencing and quantification (Zheng et al., 2015), high-quality tRNA profiling is still a need, and tRNA content and its regulatory roles remain poorly studied. Here, by relying on the hybridization and quantification of tRNA-specific probes, we have bypassed the first-strand cDNA synthesis step (Nilsson et al., 2000). Briefly, for each tRNA species, a pair of probes were designed that upon hybridization to their cognate tRNAs would provide a nick at the site of the anticodon. The nick was then repaired in a ligation reaction, giving rise to a first-strand cDNA. Biotinylation of the tRNA population and streptavidin-mediated co-precipitation steps were included to achieve a higher signal-to-noise ratio (Figures S1A–S1E; see Experimental Procedures for details). Following the successful splinted ligation of probes on their cognate tRNAs, high-throughput sequencing or qPCR can be used for relative quantification of tRNA levels (Figure 1A).

TRNA profiling revealed that tRNA expression levels in breast cancer lines were different from those of non-tumorigenic cells (Figure 1A). And more importantly, while cells from the MDA-231 and CN34 populations showed distinct tRNA profiles, we observed substantial similarities between the two cell lines with respect to their differential tRNA content when comparing each parental cell line and its metastatic derivative sub-line (Figures 1A and 1B). In other words, in vivo selection of MDA-231 and CN34 parental cancer cell populations for higher metastatic capacity selected for similar modulations in tRNA abundances. Such concerted modulation of tRNA levels suggested a potentially direct role for tRNAs in promoting cancer progression.

Among the tRNAs deregulated in the metastatic cells, $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ were consistently upregulated in both MDA-LM2 and CN-LM1a highly metastatic cells relative to their parental lines (Figure 1B). We further validated this upregulation using both splinted-ligation followed by qPCR and Northern blot analysis (Figures 1C and S1F). Importantly, qRT-PCR assays for pre- $tRNA^{Arg}_{CCG}$ and pre- $tRNA^{Glu}_{UUC}$ species revealed a significant upregulation in levels of the pre-cursors of these tRNAs as well (Figure 1D). Consistent with the increased pre-tRNA and mature tRNA levels expressed in highly metastatic cells, increased genomic copy number of $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ loci were observed in highly metastatic cells (Figure S1G).

To measure the phenotypic and molecular consequences of modulated levels of these tRNAs, we first tested whether stable cell lines overexpressing or knocked down for these tRNAs could be generated. Endogenous levels of $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ in the highly metastatic MDA-LM2 background could be effectively reduced by expressing short hairpins targeting these tRNAs respectively (Figure 1E). Similarly, expression levels of these tRNAs in MDA-parental cells were enhanced upon stable integration of additional copies of these tRNAs driven by a U6 promoter (Figure 1E). It should be noted that these observed overexpression and knockdowns (~2-fold as measured by quantitative PCR-based tRNA quantification) in the levels of these tRNAs were within physiological boundaries of their levels of endogenous modulation between the parental and the in vivo-selected highly metastatic cells (Figures 1C and 1E).

$tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ Promote Metastatic Progression

To test whether the increased expression levels of $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ play direct roles in conferring higher metastatic capacity, we employed short hairpins targeting $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ in metastatic MDA-LM2 cells. Reducing the levels of $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ by 40% and 70%, respectively, in metastatic MDA-LM2 cells (Figure 1E) to physiologically similar levels observed in poorly metastatic parental cells significantly reduced their lung colonization capacity (Figure 2A). Gross histology of extracted lungs also revealed significantly fewer metastatic nodules relative to MDA-LM2 control cells (Figure S2A). Consistent with these observations, overexpression of $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ in poorly metastatic MDA-parental cells significantly enhanced metastatic progression (Figures 2B and S2B). More strikingly, tRNA overexpressing cells exhibited enhanced orthotopic metastasis compared to control cells despite their significantly reduced primary tumor growth rates (Figures 2C and 2D). These results reveal that increased abundance of specific tRNAs can promote the metastatic phenotype of breast cancer cells.

Given that cancer cell invasion is a key phenotypic attribute required for metastatic progression from the mammary gland

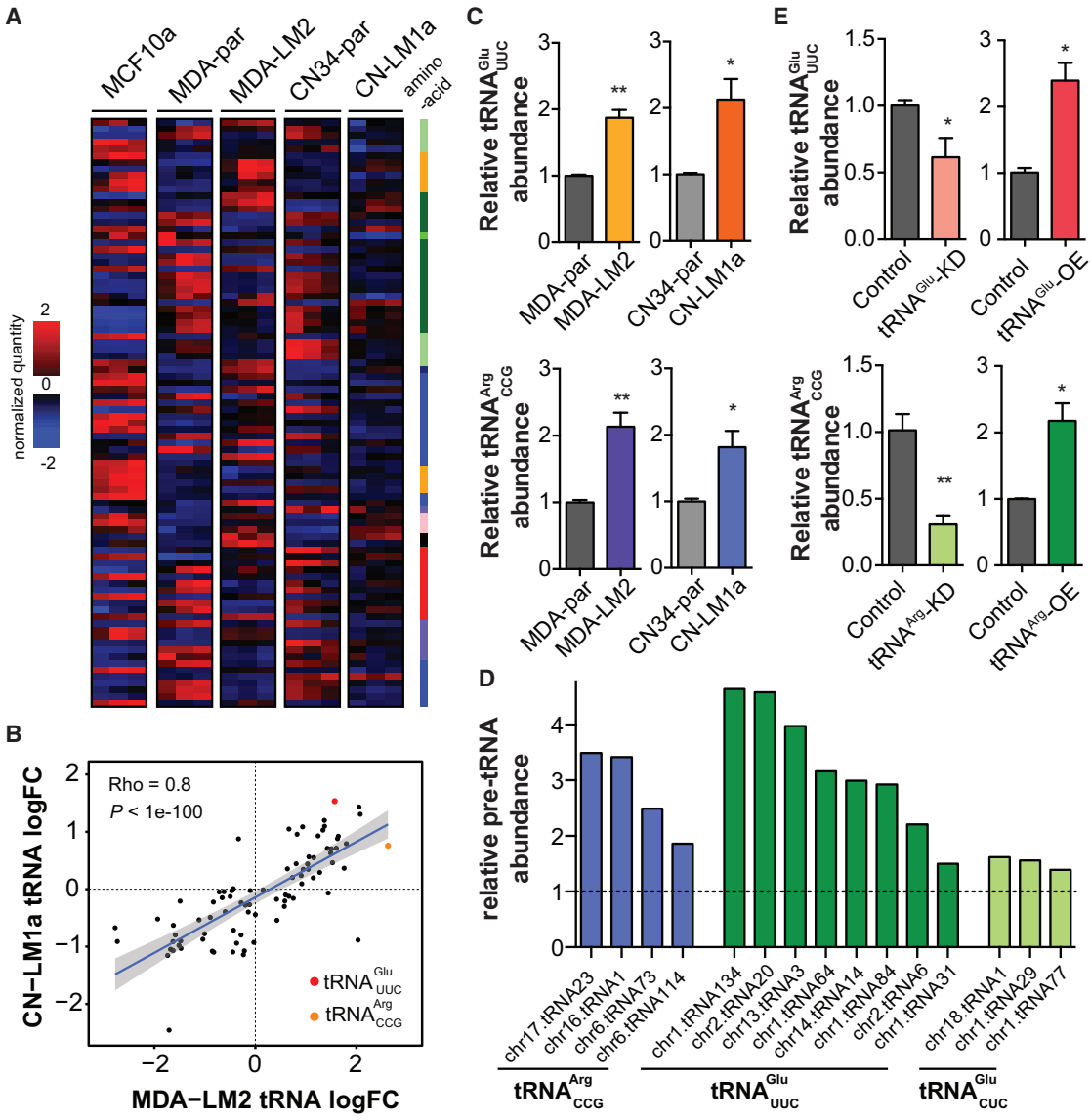


Figure 1. Transfer RNA Profiling of Metastatic and Non-metastatic Breast Cancer Lines

(A) Whole-genome tRNA profiling was performed for MCF10a, MDA-par, MDA-LM2, CN34-par, and CN-LM1a cell lines. Hierarchical clustering was used to cluster the resulting profiles. The tRNAs are labeled based on their cognate amino acid: A, G: light green; C: green; D, E, N, Q: dark green; I, L, M, V: blue; F, W, Y: lilac; H: dark blue; K, R: orange; P: pink; S, T: red.

(B) Correlation plot for changes in tRNA levels between MDA-LM2 and CN-LM1a cells. Strong positive correlation suggests these two distinct metastatic derivatives employ similar approach in modulating tRNA levels to attain metastatic phenotypes. $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ are among the most highly upregulated in both MDA-LM2 and CN-LM1a.

(C) Quantitative PCR-based tRNA quantification validated the changes in the abundance of $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ in metastatic MDA-LM2 and CN-LM1a cells relative to their respective parental lines.

(D) Relative pre-tRNA abundances for $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ across multiple genetic loci as determined by quantitative RT-PCR. $tRNA^{Glu}_{CUC}$, pre-tRNAs for which deregulated expression was not observed, were also included for comparison.

(E) $TRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ were successfully overexpressed and knocked down as revealed by quantitative PCR. Note that manipulation of the levels of these two tRNAs occurs within the physiological boundaries of the parental or metastatic backgrounds. One-tailed Student's t test was used to measure statistical significance between the two samples in each experiment. Error bars indicate SEM. *p < 0.05 and **p < 0.01.

in vivo, we asked whether modulations in the levels of $tRNA^{Arg}_{CCG}$ or $tRNA^{Glu}_{UUC}$ could affect the invasiveness of these cells. We performed in vitro cancer cell invasion assays for $tRNA^{Arg}_{CCG}$ or $tRNA^{Glu}_{UUC}$ overexpressing cells (Figure 3A) and observed significant increases in invasion capacity by cells overexpressing either of these tRNAs. Conversely, highly metastatic MDA-LM2 cells expressing shRNAs targeting either of these tRNAs exhibited significantly reduced invasive capacity

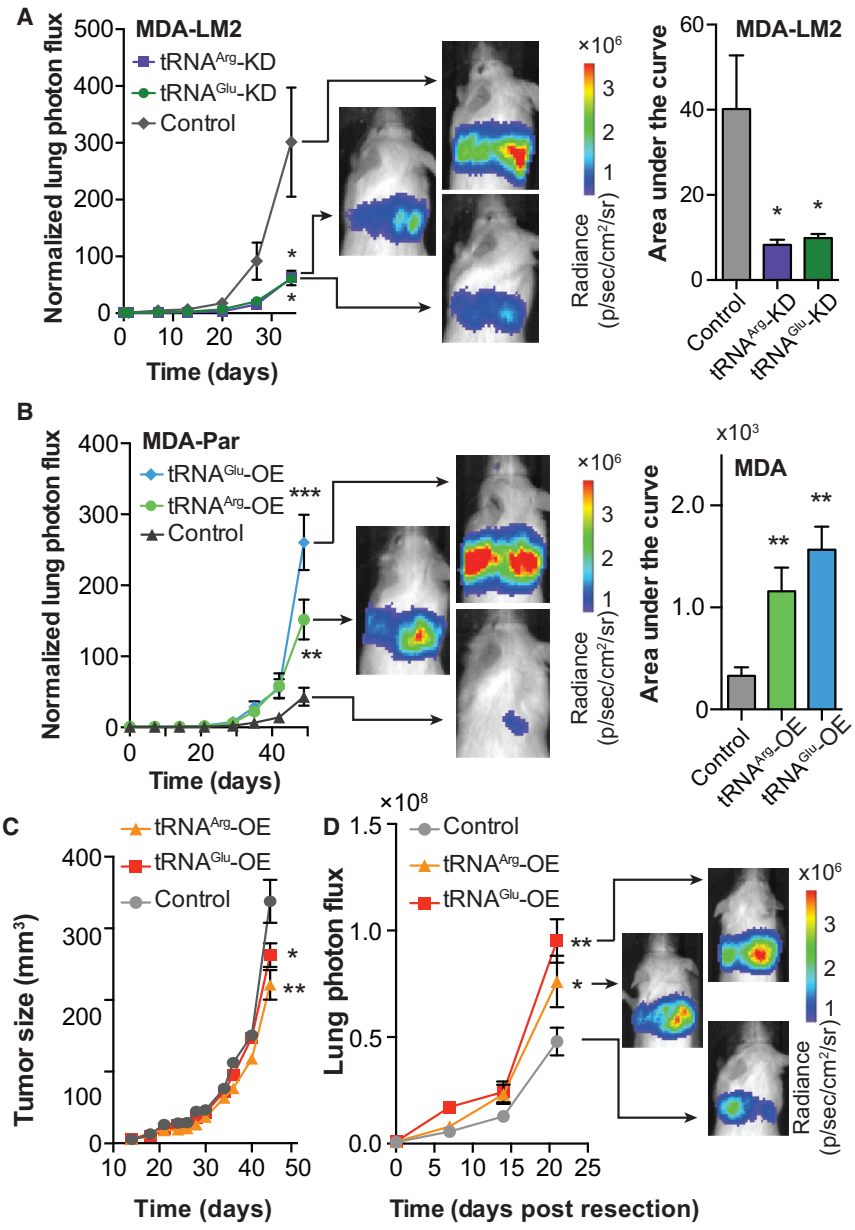


Figure 2. tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ Promote Metastatic Breast Cancer

(A) Bioluminescence imaging plot of lung colonization by MDA-LM2 cells expressing short hairpins targeting tRNA$^{Glu}_{UUC}$, tRNA$^{Arg}_{CCG}$, or a control hairpin (shControl); n = 5 in each cohort. Area-under-the-curve was also calculated for each mouse.

(B) Bioluminescence imaging plot of lung colonization by tRNA$^{Glu}_{UUC}$ or tRNA$^{Arg}_{CCG}$ overexpressing lines, as compared to control in MDA-parental cells; n = 5 in each cohort. Area-under-the-curve was also calculated for each mouse.

(C) Primary tumor growth measurement after orthotopic injection of Control, tRNA$^{Glu}_{UUC}$, or tRNA$^{Arg}_{CCG}$ overexpressing cells into the mammary fat pads of mice; n = 5 in each cohort.

(D) Orthotopic metastasis bioluminescence imaging plot of mice after primary tumor resection; n = 5 in each cohort. For comparing lung colonization, primary tumor growth, and orthotopic metastasis assays, two-way ANOVA was used to measure statistical significance. One-tailed Mann-Whitney test was used to measure statistical significance between the areas under the curves. Error bars indicate SEM. *p < 0.05, **p < 0.01, and ***p < 0.001.

(Figure 3B). The observed increase in in vitro invasion upon upregulation of these tRNAs was not due to a general increase in proliferative capacity, as overexpression of these tRNAs actually slightly decreased in vitro proliferation rates (Figure S2C). These in vitro findings are consistent with the tumor growth and metastatic phenotypes observed in vivo and provide further support to the notion that individual tRNAs can have specific, pro-metastatic phenotypic consequences.

To ensure the broad biological relevance of these findings and to rule out off-target effects due to shRNA expression, we performed the following experiments. First, we functionally tested the phenotypic effects of tRNA knockdown in an independent cell line—the metastatic CN-LM1a sub-line. Knockdown of each of these two tRNAs also reduced metastatic capacity of this breast cancer cell population as well (Figures S2D and S2E). We then overexpressed and knocked down each tRNA simultaneously and measured their metastatic capacity in vivo relative to control cells. Consistent with on-target effect of shRNAs, overexpressing each tRNA prevented the reduced metastatic phenotype observed upon depletion of each tRNA (Figure S3A). Lastly, to ensure that the observed phenotypes were specific to tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ and that varying the levels of another tRNA, which was not observed to be modulated in cancer cells upon selection for enhanced metastatic activity, would not manifest similar effects, we also varied the levels of tRNA$^{Tyr}_{GUA}$. Modulating tRNA$^{Tyr}_{GUA}$ in both knockdown and overexpression experiments did not significantly affect the metastatic activity of MDA-LM2 cells (Figure S3B).

To further ascertain whether these associations are clinically relevant to human breast cancer progression, we quantified the levels of these tRNAs in a cohort of primary tumors that had not metastasized as well as in a set that had exhibited clinical metastasis (n = 23). We observed significant upregulations in tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ levels across the metastatic primary tumors relative to non-metastatic primary tumors (Figure 3C). These findings reveal that the expression levels of tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ in human breast cancers correlate with and predict metastatic propensity.

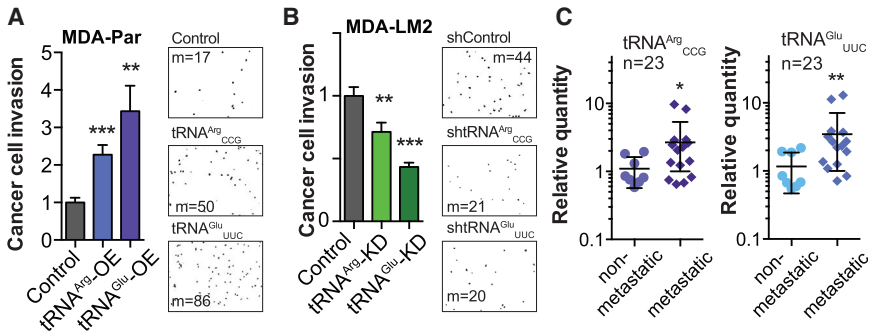


Figure 3. In Vitro Characterization of tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ and Their Clinical Associations with Breast Cancer Progression

(A) Overexpression of tRNA$^{Glu}_{UUC}$ or tRNA$^{Arg}_{CCG}$ in MDA-parental cells significantly increased cancer cell invasion. Also included are representative fields from the invasion inserts along with the median number of cells observed in each cohort.

(B) Conversely, tRNA$^{Glu}_{UUC}$ or tRNA$^{Arg}_{CCG}$ knockdown in MDA-LM2 cells significantly decreased cancer cell invasion. Also included are representative fields from the invasion inserts along with the median number of cells observed in each cohort.

(C) The relative abundance of tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ in primary breast tumor samples from patients who either developed metastatic relapse (n = 15) or remained disease-free (n = 8), measured using quantitative PCR. One-tailed Mann-Whitney test was used to establish statistical significance between the two cohorts. Error bars indicate SEM. *p < 0.05, **p < 0.01, and ***p < 0.001.

tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ Overexpression Impacts Transcript Stability and Translation

As previously mentioned, modulations in tRNA levels may impact the gene expression landscape of the cell. To assess the regulatory consequences of upregulating these tRNAs, we systematically measured their impact on post-transcriptional regulatory processes. First, we performed ribosome profiling in control as well as tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ overexpressing cells to provide a snapshot of changes in active translation (Ingolia et al., 2014). We observed a substantial positive correlation between relative ribosome occupancy of transcripts across biological replicates (Figure S4A). Importantly, the periodicity and length generally noted for ribosome protected fragments by other investigators were observed in this dataset as well (Figures 4A, 4B, S4B, and S4C). For each gene, we calculated a corrected ribosome-occupancy score to compare active translation in tRNA overexpressing cells relative to control cells (MDA-parental background; see Experimental Procedures). We subsequently asked whether the frequency of codons cognate to each overexpressed tRNA was informative of the observed changes in active translation. We observed that, in tRNA$^{Glu}_{UUC}$ overexpressing cells, genes with high GAR content in their coding sequence (GAA and GAG codons, since tRNA$^{Glu}_{YUC}$ can Wobble-base pair at the third nucleotide) were significantly enriched among those with higher relative ribosome occupancy (Figures 4C and S4D). Similarly, higher CGG content was associated with higher active translation in tRNA$^{Arg}_{CCG}$ overexpressing cells (Figure 4C).

To directly assess the impact of tRNA modulations on proteomic output, we measured the expression levels of roughly 4,000 proteins using mass-spectrometry-based quantification in tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$, as well as control cells (MDA-parental background). In tRNA overexpressing cells, we identified hundreds of proteins that were significantly altered in their expression levels (Figures S4E and S4F). To correct for changes in transcript levels, we performed transcriptomic measurements in tRNA$^{Arg}_{CCG}$ and tRNA$^{Glu}_{UUC}$ overexpressing as well as control cells. We then corrected the fold-changes in protein expression levels with those of their corresponding transcripts. As shown in Figure 4D, similar to ribosome profiling results, proteins with high GAR content in their genes (GAA and GAG codons) were significantly enriched among those upregulated in the tRNA$^{Glu}_{UUC}$ overexpressing cells. Moreover, proteins with high CGG content showed a significant enrichment among the upregulated genes in the tRNA$^{Arg}_{CCG}$ overexpressing line (Figure 4D). These observations are consistent with the hypothesis that protein expression levels can be modulated as a function of their codon usage and cellular tRNA content.

In addition to a direct impact on active translation, differential ribosome occupancy may also affect other aspects of RNA life cycle, specifically RNA stability (Huch and Nissan, 2014). To test whether tRNA abundance can likewise impact transcript stability, we performed α-amanitin-mediated whole-genome mRNA stability measurements in tRNA overexpressing and control cell lines. Consistent with a positive association between translation and mRNA stability (Coller and Parker, 2005; Huch and Nissan, 2014; Muhlrad et al., 1995), we observed a general stabilization of transcripts with higher GAR and CGG content in tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ overexpressing cells, respectively (Figure 4E). We also validated these observations by measuring transcript decay rates using qRT-PCR for a set of mRNAs showing differential stability in tRNA overexpressing cells (Figure 4F). Taken together, modulations in tRNA levels can have broad regulatory consequences across the proteome of cancer cells, which can partly be explained by variations in codon usage of the target proteins.

Codon-Specific Modulation of tRNA$^{Glu}_{UUC}$ Downstream Targets and Clinical Associations

Given the regulatory consequences of tRNA modulations in cancer cells and their impact on metastatic capacity, we hypothesized that a set of core target transcripts may be the major drivers of the metastatic phenotype downstream of tRNA$^{Glu}_{UUC}$ as a global regulator. To identify these targets, we systematically combined the various whole-genome datasets we had compiled as part of this study. We specifically focused on the ribosome profiling data in tRNA$^{Glu}_{UUC}$ overexpressing cell line to identify potential targets that exhibited a higher rate of active translation when tRNA$^{Glu}_{UUC}$ was more abundant. We identified *EXOSC2* and *GRIPAP1* as such targets. Consistent with their higher ribosome occupancy in tRNA$^{Glu}_{UUC}$ overexpressing cell, these genes also displayed higher protein levels in highly metastatic

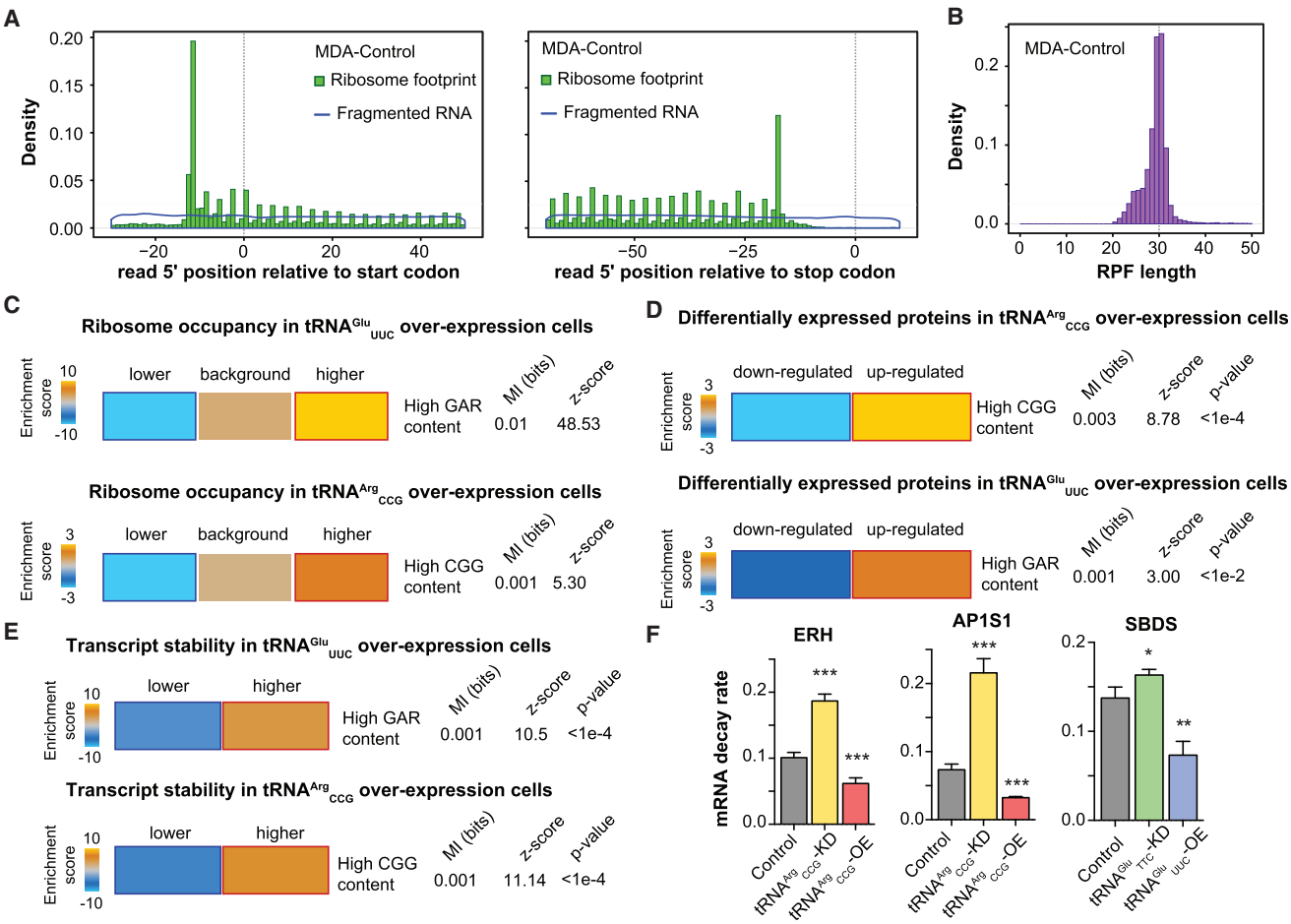


Figure 4. Post-Transcriptional Consequences of tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ Upregulation

(A) A hallmark of ribosome profiling libraries is a 3-nt periodicity. As an example, we have included the coverage of the 5′-end of reads along the coding sequence with respect to the start (left) or stop codon (right). In comparison, the total RNA library (fragmented RNA) did not exhibit this periodicity.

(B) Given the footprint of ribosomes on mRNAs, ribosome protected fragments (RPF) of ~30-nt are expected. Here, as an example, we have shown the RPF length distribution for our control samples.

(C) Genes with higher GAR and CGG contents exhibited significant enrichment among transcripts with increased ribosomal occupancy in tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ overexpressing cells, respectively.

(D) Genes with a high abundance of GAG and GAA (GAR) codons were significantly enriched among proteins significantly upregulated (corrected for their transcript changes) in tRNA$^{Glu}_{UUC}$ overexpression cells. Similarly, genes with higher CGG content exhibited a significant enrichment among the proteins up-regulated (after correction for transcript changes) upon overexpression of tRNA$^{Arg}_{CCG}$. The statistical significance of these enrichments was assessed using mutual-information calculations and associated *Z* score (based on randomized input vectors). Also included is the χ^2 p value for the associated contingency table. The heatmap was generated using the –log of the hypergeometric p value for enrichment and log of p value for depletion (collectively termed the enrichment score). The red and dark-blue borders indicate the statistical significance of the calculated hypergeometric p values (for details, see Goodarzi et al., 2009).

(E) Whole-genome transcript stability measurements reveal significant enrichment for genes with higher GAR content among those strongly stablized in tRNA$^{Glu}_{UUC}$ overexpressing line. Similarly, stability of transcripts with higher CGG content is also significantly enhanced in the context of tRNA$^{Arg}_{CCG}$ overexpression.

(F) ERH, AP1S1, and SBDS were chosen to validate by qRT-PCR the impact overexpressing or knocking down corresponding tRNAs has on mRNA stability as a function of decay rate.

cells relative to their poorly metastatic parental cells (both MDA and CN34 cell line backgrounds, data not shown). Quantitative western blots further validate the increased expression of these genes in tRNA$^{Glu}_{UUC}$ overexpressing background (Figure 5A). It should be emphasized that neither of these genes exhibited a significant upregulation at the transcript level (log fold-change of < 0.1 for both genes in both MDA- and CN34- backgrounds), highlighting that the observed upregulation for these genes was primarily post-transcriptionally mediated.

To test whether these genes impact metastatic progression downstream of tRNA$^{Glu}_{UUC}$, we conducted epistasis experiments in xenograft mouse models. *EXOSC2* and *GRIPAP1* were stably knocked down in the context of tRNA$^{Glu}_{UUC}$ overexpression, as well as in the control line, and these cells were injected into the tail-veins of NSG mice. Reduced EXOSC2 and GRIPAP1 expression levels substantially abrogated the enhanced metastatic colonization outcome caused by tRNA$^{Glu}_{UUC}$ overexpression (Figure 5B and data also shown as separated plots in Figure S5).

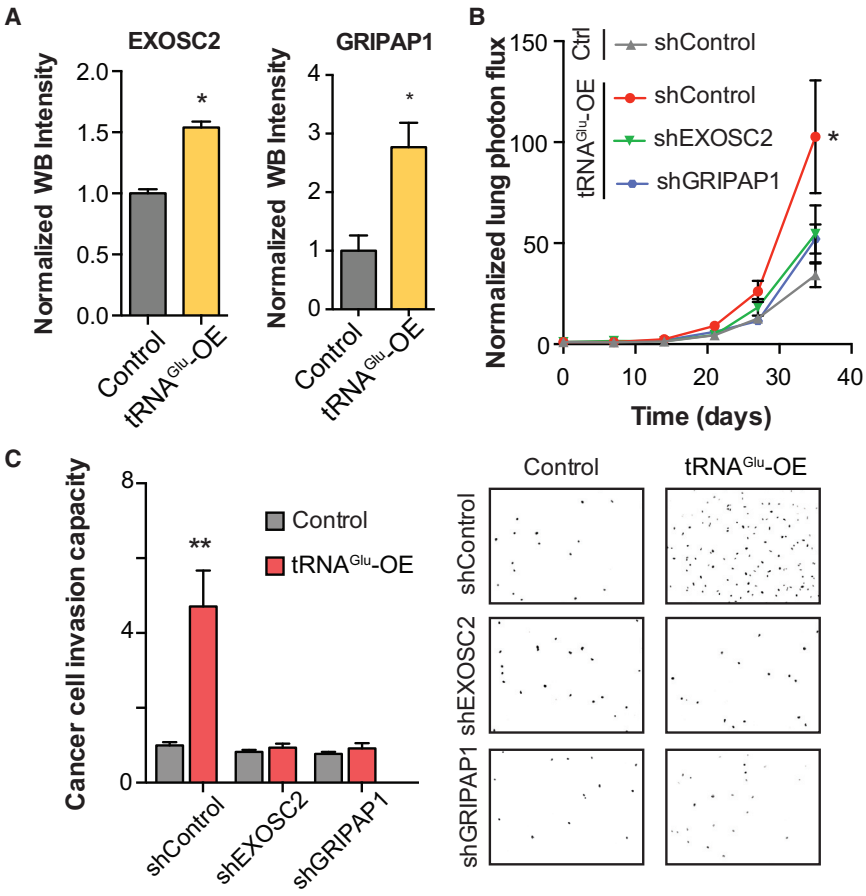


Figure 5. Variations in $tRNA^{Glu}_{UUC}$ Levels Post-Transcriptionally Modulate Expression of Breast Cancer Metastasis Promoters *EXOSC2* and *GRIPAP1*

(A) Endogenous EXSOC2 and GRIPAP1 protein levels as measured by quantitative western blotting (see Experimental Procedures) in $tRNA^{Glu}_{UUC}$ overexpressing or control lines.

(B) Bioluminescence imaging plot of lung colonization by *EXOSC2* and *GRIPAP1* knocked-down cells in MDA-parental overexpressing $tRNA^{Glu}_{UUC}$ relative to control cells expressing a control hairpin, shControl; n = 5 in each cohort. Two-way ANOVA was used to measure statistical significance.

(C) Knockdown of *EXOSC2* or *GRIPAP1* abrogated the enhanced invasion capacity of $tRNA^{Glu}_{UUC}$ overexpressing line. Also included are representative fields from the invasion inserts along with the median number of cells observed in each cohort. Error bars indicate SEM. *$p < 0.05$ and **$p < 0.01$ by one-tailed Student's t test.

Consistent with this, the in vitro invasiveness of $tRNA^{Glu}_{UUC}$ overexpressing line was also dramatically decreased upon silencing of *EXOSC2* and *GRIPAP1* (Figure 5C). It should be noted that reduced lung colonization and invasiveness were not observed in the control MDA-parental background upon *EXOSC2* or *GRIPAP1* depletion. This suggests that *EXOSC2* and *GRIPAP1* are downstream targets of $tRNA^{Glu}_{UUC}$ overexpression and that in the context of elevated $tRNA^{Glu}_{UUC}$, cancer cells can exploit these proteins to elicit a pro-metastatic phenotype. Furthermore, silencing *EXOSC2* and *GRIPAP1* did not completely abolish the metastatic phenotype of $tRNA^{Glu}_{UUC}$ overexpression, supporting the likely possibility that additional genes may operate downstream of this tRNA in driving metastasis.

To elucidate the importance of codon specificity on the effect of $tRNA^{Glu}_{UUC}$ on its targets, we performed a codon mutagenesis experiment in which every preferred cognate GAA codon in the *EXOSC2* coding sequence was switched to the synonymous GAG codon. As a negative control, we instead mutated every instance of a non-deregulated codon, Gly-GGG to Gly-GGC (see Figure S6A and Methods for details). Transcript stability assays for exogenously transfected wild-type, Gly-mutated, and Glu-mutated versions of *EXOSC2* revealed significant loss of the stabilizing effect of $tRNA^{Glu}_{UUC}$ overexpression in the GAA-to-GAG mutant (Figure 6A). Reduction in protein expression of the GAA-to-GAG (Glu) codon mutated transcript, and not of the GGG-to-GGC (Gly) codon mutated transcript (data not shown), further validates the codon-specific mechanism through which $tRNA^{Glu}_{UUC}$ confers its effect (Figures 6B, S6B, and S6C). These mutagenesis studies establish the direct interaction of a specific tRNA with its downstream target of regulation.

Because EXOSC2 and GRIPAP1 have not been previously implicated in breast cancer metastasis, we sought clinical association evidence for our experimental findings. Although we could not establish specificity for multiple commercially available GRIPAP1 antibodies, we identified an EXOSC2 antibody, which exhibited specificity in fixed tissue immunohistochemistry (IHC). Immunohistochemical staining of breast cancer progression tissue microarrays (TMA) for EXOSC2 revealed a positive association between its expression and clinical breast cancer progression. EXOSC2 protein expression was significantly higher in invasive breast cancer relative to normal breast tissues (Figures 6C, 6D, and S6D). More importantly, EXOSC2 protein levels were also significantly higher in primary tumors of patients with distant metastases compared to earlier stage tumors (Figures 6C, 6D, and S6D). These clinical association results not only support our in vitro and in vivo findings regarding the role of tRNA modulation in promoting metastasis but also support this tRNA-based pathway discovery approach as a means for identifying post-transcriptionally regulated targets that might have been otherwise missed by traditional transcriptomic profiling methods.

Association among tRNA Preference, Ribosomal Occupancy, and Protein Expression

Given their crucial roles in translation, deregulations in tRNA abundance could strongly impact the protein expression landscape of the cell. While $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ were

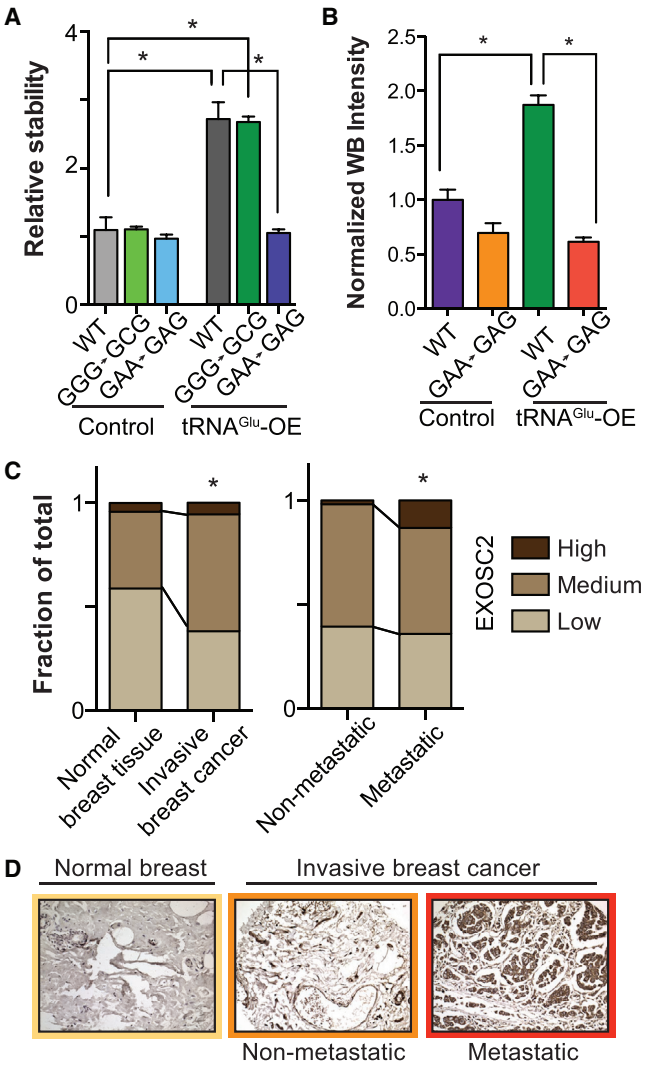


Figure 6. Codon-Specific Modulation of *EXOSC2* and Its Clinical Association

(A) Relative transcript stability measured by qRT-PCR (see Experimental Procedures) of exogenous wild-type, GGG-to-GCG (Gly) codon mutated, and GAA-to-GAG (Glu) codon mutated transcript in control and $tRNA^{Glu}_{UUC}$ overexpressing backgrounds. While overexpression of $tRNA^{Glu}_{UUC}$ significantly stabilized wild-type and GGG-to-GCG (negative control) transcripts, such an effect was absent when the specific cognate Glu codons were mutated.

(B) Quantitative western blot demonstrated similar loss of translational enhancement brought about by overexpressing $tRNA^{Glu}_{UUC}$ when its cognate codons were mutated. Error bars indicate SEM. *$p < 0.05$ by one-tailed Student's t test.

(C) Stacked bars representing the fraction of tissue samples from TMA with respectively low, medium, and high intensity of EXOSC2 in normal breast tissues and invasive breast cancer tissues (n = 46 and 160, respectively). Also shown are fractions of tissues of different EXOSC2 intensity in breast cancer from patients without metastasis and those with detected metastasis in distant organs (n = 107 and 53, respectively). Hypergeometric p values were calculated to assess the significance of the increase in the frequency of samples with higher intensities; *$p < 0.05$.

(D) Shown are representative tissue-microarray immunohistochemical images of stained tissues of median score from normal breast, non-metastatic invasive breast cancer, and metastatic breast cancer tissues. Higher EXOSC2 intensity positively correlated with disease progression stage.

significantly upregulated in highly metastatic cells, the modifications in the tRNA profiles of highly metastatic cells relative to poorly metastatic cells were not limited to these two tRNAs. To test the consequences of the broader tRNA modulations, we first asked whether the observed changes in tRNA content, across all measured tRNA species, were informative of expression levels of proteins based on their coding sequences alone. As mentioned before, differential tRNA expression levels can impact translational outcome by affecting the abundance of ribosome-bound transcripts. We thus sought to quantify the likely impact of variations in tRNA content on active translation and protein expression given the frequency of each codon for every gene. For this, we defined tRNA preference scores as the sum of changes in the tRNA content in each background (log-ratio in MDA-231 and CN34 backgrounds) across all codons of a given gene (MDA- and CN- preference scores; see Experimental Procedures). In this scoring scheme, genes whose codons are favored, based on the upregulation of their cognate tRNAs, are assigned positive scores, and those whose codons on average are associated with downregulated tRNAs are assigned negative scores.

We performed whole-genome ribosome profiling on poorly metastatic parental lines, MDA-par and CN34-par, as well as their highly metastatic derivatives, MDA-LM2 and CN34-LM1a. The ribosome-protected fragments (RPF) were normalized to total RNA (TT) for every cell line to account for differences in gene expression. As expected, the RPF/TT ratio between parental and derivative lines, as well as between biological duplicates, demonstrated a strong positive correlation (Figures 7A and S7A). TRNA preference scores of every gene were then overlapped with its respective corrected ribosome footprints. Interestingly, transcripts with higher tRNA preference scores were strongly enriched among those bound more by ribosomes (Figure 7B). We subsequently measured differential protein levels between the metastatic and parental lines using stable-isotope labeling by amino acids in cell culture (SILAC, Ong et al., 2003; also see Figure S7B). To correct for protein expression changes due to variations in transcript abundances, we normalized the change in protein expression for each gene to its transcript level in each background. In accordance with the increased ribosomal occupancy, we observed a significant enrichment of genes with high tRNA preference scores among those translationally upregulated in the highly metastatic sublines in both MDA-231 and CN34 backgrounds (Figures 7C). Importantly, consistent with the observed correlation between the changes in tRNA abundance in the MDA and CN34 backgrounds, we also observed a highly significant correlation between tRNA preference scores calculated across all coding sequences (Figure S7C).

Our findings reveal that differential tRNA expression is informative of changes in translational landscapes. However, to show that there in fact may be a causal link between global tRNA content and protein expression, we took advantage of two synthetic constructs based on the coding sequence of Renilla luciferase. We designed a comparative luciferase coding sequence by scanning the gene and inserting the codon variant whose cognate tRNA had the highest relative expression in CN34 versus MDA-parental cells (CN-optimized luciferase). We should emphasize that the chosen codons were not the ones

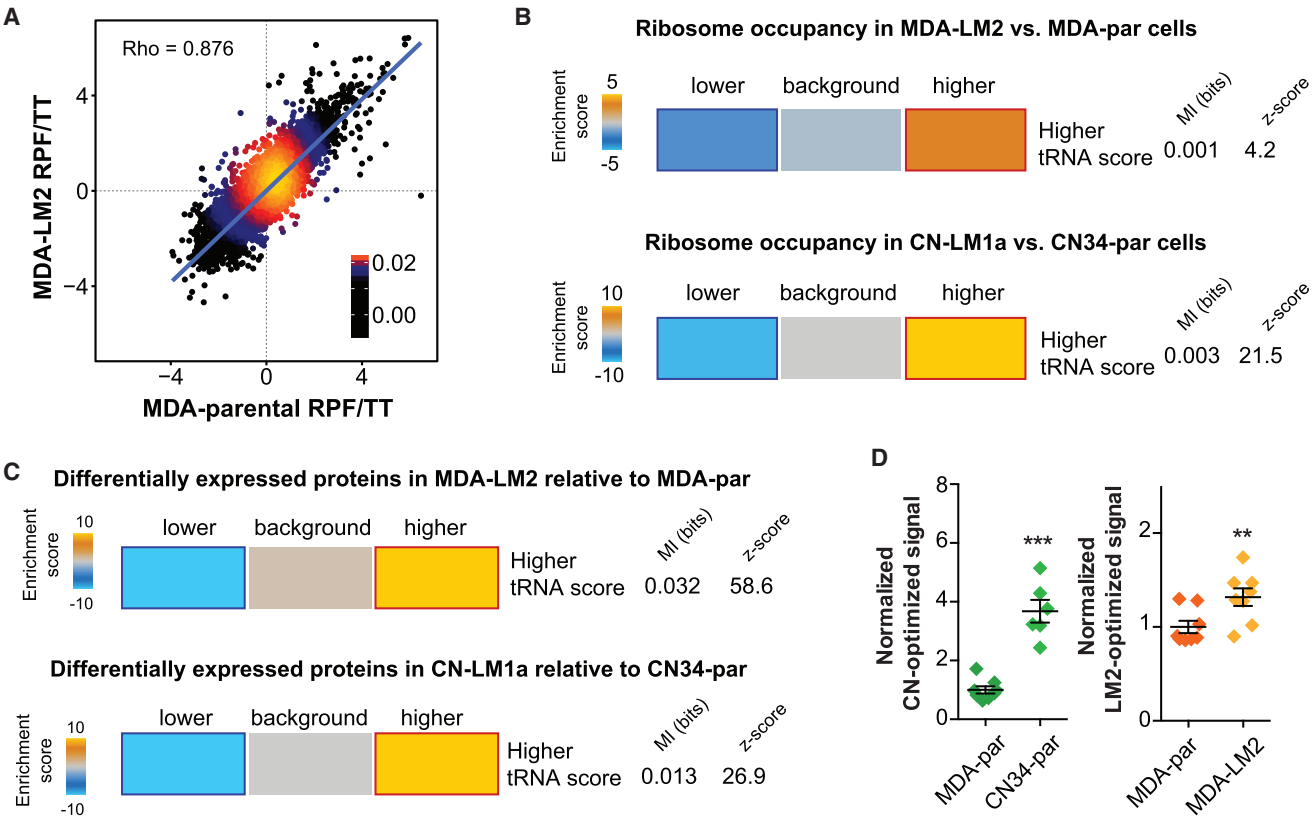


Figure 7. tRNA Preference Scores Were Informative of Differential Ribosome Occupancy and Protein Expression

(A) As an example, we have shown the linear regression of MDA-parental and MDA-LM2 ribosome protected fragment to total RNA ratio (RPF/TT). RPF reads were normalized to TT reads to correct for variation in transcript expression in each cell line.

(B) Genes with positive tRNA preference score (based on derivative versus parental tRNA profiling results) were significantly enriched among transcripts with higher corrected ribosome occupancy values in both MDA-LM2 and CN-LM1a relative to MDA-par and CN34-par, respectively (see Experimental Procedures). In other words, coding sequences with more favorable codon content, based on changes in tRNA abundance between parental cells and their highly metastatic derivatives, exhibited a more active translation.

(C) Genes with positive tRNA preference score were significantly enriched among the proteins upregulated in MDA-LM2 cells and CN-LM1a compared to MDA-par and CN34-par, respectively. The significance of these enrichments was determined by calculating mutual-information values and their associated *Z* scores (based on randomized input values). Also included is the χ^2 p value for the associated contingency table. The enrichment score, based on which the heatmap was generated, is defined as the –log of hypergeometric p value for enrichment (gold) and log of p value for depletion (blue). The red and dark-blue borders indicate the statistical significance of the calculated hypergeometric p values (Goodarzi et al., 2009).

(D) Normalized relative luciferase activity for CN-optimized and LM2-optimized luciferase constructs.

with highest tRNA levels but rather the ones with highest CN34 to MDA-231 ratios. We similarly constructed an LM2-optimized luciferase coding sequence comparing MDA-LM2 tRNA levels to those of the parental MDA-231. We then measured luciferase activity of the CN-optimized construct in both MDA- and CN34-parental cells. Interestingly, we observed a substantially higher luciferase signal in CN34-parental cells (Figure 7D). We next focused on cancer cell populations originating from the same patient (MDA and MDA-LM2 cells). Expression of the LM2-optimized construct similarly showed a higher signal in MDA-LM2 line relative to MDA-parental cells (Figure 7D). The magnitude of the change also supports the level of variation we had initially observed with the tRNA profile in CN34 being more distinct from the MDA lines and MDA-231 and MDA-LM2 showing more similar tRNA profiles. These findings reveal that, even within the same cancer type, the proteomic output of cells obtained from two distinct patients is constrained by the cells' tRNA contents. Moreover, even within a given patient, cancer subpopulations' expression output can be significantly impacted by the tRNA codon landscape within cell sub-populations.

DISCUSSION

Collectively, our findings demonstrate that changes in tRNA abundance can modulate protein expression in the cell and that cancer cells can evolve to fine-tune the expression of multiple promoters of cancer progression through modulations in tRNA levels. Consistent with previous reports (Pavon-Eternod et al., 2009), we observed that the tRNA profiles of breast cancer lines can markedly differ from non-cancerous epithelial cells. More importantly, the tRNA profile is further modified en route to higher metastatic capacity. $TRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$

were significantly upregulated in breast cancer lines relative to epithelial cells, but this increase was further augmented in metastatic lines relative to their parental population. Moreover, we find that the levels of these tRNAs are higher in metastatic human tumors relative to non-metastatic ones. We have shown that the increased expression of these tRNAs contributes to the metastatic phenotype. We propose that modulations in specific tRNA levels enhance the translational efficiency of genes that are promoters of metastasis in a codon-dependent manner. A detailed analysis of differential protein expression and its relationship to codon preference revealed a modest albeit highly significant correlation between these parameters.

Variations in the abundance of tRNA isoacceptors impact the rate of translation in vivo (Fredrick and Ibba, 2010). Transient pauses at codons with rare cognate tRNAs can also affect protein folding and translocation (Zhang and Ignatova, 2011). As a result, even synonymous substitutions can alter the rate of translation and impact protein dynamics (Kirchner and Ignatova, 2015). In humans, tissue-specific tRNA expression mirrors the codon usage of tissue-specific proteins, indicating an association between the active tRNA pool and translational demand (Dittmar et al., 2006). Recently, it was also hypothesized that the translational machinery can be locked into a "proliferation" versus a "differentiation" program (Gingold et al., 2014). Our results, however, indicate that changes in tRNA expression states in the context of pathophysiology are more dynamic and their modulations can have specific phenotypic consequences such as enhanced invasiveness. As previously mentioned, the role of global tRNA content in translation efficiency and the role of codon usage in concerted regulation of protein expression have been debated (dos Reis et al., 2004). Based on our results, our model sides with those reporting a significant selection on codon usage adaptation with respect to tRNA content (Comeron, 2004; Gingold et al., 2014). It was based on this tRNA program that we successfully identified and functionally validated novel promoters of metastatic colonization.

In many organisms, such as bacteria, yeast, and even human, tRNA gene copy number is correlated with codon usage of highly expressed genes (Novoa and Ribas de Pouplana, 2012; Novoa et al., 2012). However, in yeast, modulating the levels of a rare $tRNA^{Arg}_{CCU}$ did not impact elongation rate or translational efficiency, while $tRNA^{Thr}_{UGU}$ knockdown, achieved by deleting three of the four copies of the heavily used $tRNA^{Thr}_{UGU}$, had a modest effect on efficiency (Pop et al., 2014). Thus, the observed correlation between codon bias and efficiency may have arisen from the selection on highly expressed genes to utilize translation machinery efficiently (Pop et al., 2014). In our data, the association between tRNA abundance and codon usage were more pronounced among highly expressed genes. For example, limiting our analysis to roughly 300 genes with one SD above average expression (based on whole-transcriptomic measurement), a significant correlation between codon CGG content and protein levels in $tRNA^{Arg}_{CCG}$ overexpressing cells (Rho = 0.12, p = 0.03) was observed—a correlation that is otherwise less pronounced across all the detected proteins (Rho = 0.06). Another point to consider is that higher levels of these specific tRNAs were specifically selected for in the context of pathological disease progression. It is possible that modulating other tRNAs, as was the case for $tRNA^{Tyr}$, will not result in significant changes in protein expression levels or may not be tolerated by the cells under study. Furthermore, it is plausible that the link between tRNA levels and translational efficiency is more pronounced in mammalian cells selected in vivo than it is in yeast cells grown in culture. More importantly, in our data, codon usage explains a fraction of protein expression changes, implying that (1) tRNA levels may affect protein expression in ways other than direct translation (Lee et al., 2006) and stability, and/or (2) positive and negative feedback loops may amplify and propagate the regulatory consequences of modulations in tRNA levels.

Taken together, our study puts forth the notion that cancer cells, in addition to many known regulatory mechanisms, exhibit tRNA landscape modulations that modify the expression of promoters of cancer progression. Our work reveals that specific tRNAs can form "inducible" pathways with their direct target transcripts, which are enriched for their cognate codons. Such target transcripts can become stabilized in the context of their favored tRNA content or can be more effectively translated—ultimately yielding greater protein output. It should be noted that, while we have focused on breast cancer metastasis, this approach based on tRNA profiling is general in concept and can be employed to study other diseases and can be extended to other models and species, as well as to developmental processes where similar tRNA modulations may govern developmental gene expression programs and phenotypic outcomes.

EXPERIMENTAL PROCEDURES

Transfer RNA Profiling

For each family of mature tRNAs (with introns removed and containing terminal CCA) with a similar consensus sequence and a common anticodon, a pair of probes is designed so that, upon annealing to the complementary tRNA, the resulting nick in the DNA-RNA hybrid is located at the site of the anticodon. We designed a total of 67 probe-pairs to cover the majority of cytosolic tRNAs. The downstream probes were 5′-phosphorylated to enable enzymatic ligation. Small-RNAs extracted from cells were subjected to deacylation and biotinylation, followed by hybridization with the probe library. The DNA/RNA hybrids were then subjected to overnight ligation with T4 DNA ligase. MyOne-C1 Streptavidin Dynabeads (Invitrogen) were then used to purify biotinylated DNA/RNA hybrids, and the ligated probes were then eluted after incubation with RNase H and RNase A followed by incubation with an elution buffer (50 mM Tris [pH 8], 10 mM EDTA, 1% SDS; incubate at 65°C for 30 min with intermittent vortexing). Eluted probes were subsequently cleaned up and minimally PCR amplified (12–15 cycles) for high-throughput sequencing.

Low-Throughput tRNA Quantification

The preparation of samples for low-throughput tRNA quantification was identical to the high-throughput tRNA profiling protocol described above, with the following exception: (1) instead of using a library of probes, the RNA samples were hybridized to a single probe-pair (matching the tRNA of interest); (2) the amplification step is replaced with quantitative PCR using SYBR Green (Life Technologies) per manufacturer's instructions.

Whole-Genome Ribosomal Occupancy Profiling

The procedure was performed with Truseq Ribo Profile for mammalian cells (Illumina) per manufacturer's instructions. An input of 50×10^6 cells were harvested for each replicate (biological duplicates for every cell line profiled), and libraries were sequenced using Illumina Nextseq 500 at the Rockefeller Genomics Center.

Codon-Specific Mutational Assays

Wild-type and mutated versions of EXOSC2 coding sequences were synthesized by IDT. Every Glu-GAA codon was mutated to Glu-GAG. As a negative control, another construct in which every Gly-GGG codon was mutated to Gly-GGC was used (Figure S6A). Gly-GGG and Gly-GGC codons were chosen because there was no significant difference in their respective tRNA levels based on our high-throughput tRNA profiling analysis. Each synthetic coding sequence also contained a 3′-FLAG-tag. The constructs were then cloned into the psiCHECK2 backbone (replacing synthetic Renilla Luciferase gene) together with an upstream Tetracycline-response element.

Animal Studies

All mouse studies were conducted according to a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the Rockefeller University.

ACCESSION NUMBERS

The data for high-throughput sequencing and microarray profiling experiments are deposited at GEO under the accession number GSE77401.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.05.046.

AUTHOR CONTRIBUTIONS

H.G. and S.F.T. conceived the project. S.F.T. supervised all research. S.F.T., H.G., and H.C.B.N. wrote the manuscript. H.G., H.C.B.N., S.Z., B.D.D., and H.M. designed, performed, and analyzed the experiments.

ACKNOWLEDGMENTS

We thank C. Alarcon for technical help in performing Northern blot experiments. We also thank P. Freddolino, C. Alarcon, H. Najafabadi, and A. Nguyen for comments on previous versions of this manuscript. We are grateful to C. Zhao, C. Lai, and N. Nnatubeugo of the Rockefeller Genomics Resource Center for assistance with next-generation RNA sequencing and microarray profiling. H.G. was previously supported by a Ruth L. Kirschstein National Research Service Award (T32CA009673) and is currently supported by an NIH Pathway to Independence Award (K99CA194077). S.F.T. is a Department of Defense Era of Hope Scholar and a Department of Defense Breast Cancer Collaborative Scholars and Innovators Award recipient.

Received: January 20, 2016
Revised: April 4, 2016
Accepted: May 13, 2016
Published: June 2, 2016

REFERENCES

Begley, U., Dyavaiah, M., Patil, A., Rooney, J.P., DiRenzo, D., Young, C.M., Conklin, D.S., Zitomer, R.S., and Begley, T.J. (2007). Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol. Cell *28*, 860–870.

Chan, C.T., Dyavaiah, M., DeMott, M.S., Taghizadeh, K., Dedon, P.C., and Begley, T.J. (2010). A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet. *6*, e1001247.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. Cell *122*, 875–886.

Comeron, J.M. (2004). Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. Genetics *167*, 1293–1304.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb. Perspect. Biol. *4*, a013706.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet. *2*, e221.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. *32*, 5036–5044.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341–352.

Fredrick, K., and Ibba, M. (2010). How the sequence of a gene can tune its translation. Cell *141*, 227–229.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. Cell *158*, 1281–1292.

Goodarzi, H., Elemento, O., and Tavazoie, S. (2009). Revealing global regulatory perturbations across human cancers. Mol. Cell *36*, 900–911.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Han, K., Jaimovich, A., Dey, G., Ruggero, D., Meyuhas, O., Sonenberg, N., and Meyer, T. (2014). Parallel measurement of dynamic changes in translation rates in single cells. Nat. Methods *11*, 86–93.

Huch, S., and Nissan, T. (2014). Interrelations between translation and general mRNA degradation in yeast. Wiley Interdiscip. Rev. RNA *5*, 747–763.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep. *8*, 1365–1379.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Kirchner, S., and Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nat. Rev. Genet. *16*, 98–112.

Lee, J.W., Beebe, K., Nangle, L.A., Jang, J., Longo-Guess, C.M., Cook, S.A., Davisson, M.T., Sundberg, J.P., Schimmel, P., and Ackerman, S.L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature *443*, 50–55.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. Nature *436*, 518–524.

Muhlrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. Mol. Cell. Biol. *15*, 2145–2156.

Nilsson, M., Barbany, G., Antson, D.O., Gertow, K., and Landegren, U. (2000). Enhanced detection and distinction of RNA by enzymatic probe ligation. Nat. Biotechnol. *18*, 791–793.

Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet. *28*, 574–581.

Novoa, E.M., Pavon-Eternod, M., Pan, T., and Ribas de Pouplana, L. (2012). A role for tRNA modifications in genome structure and codon usage. Cell *149*, 202–213.

Ong, S.E., Kratchmarova, I., and Mann, M. (2003). Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J. Proteome Res. *2*, 173–181.

Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M.R., and Pan, T. (2009). tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res. *37*, 7268–7280.

Pershing, N.L., Lampson, B.L., Belsky, J.A., Kaltenbrun, E., MacAlpine, D.M., and Counter, C.M. (2015). Rare codons capacitate Kras-driven de novo tumorigenesis. J. Clin. Invest. *125*, 222–233.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol. Syst. Biol. *10*, 770.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., and Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124.

Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. *8*, e1002603.

Shah, P., and Gilchrist, M.A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. Proc. Natl. Acad. Sci. USA *108*, 10231–10236.

Subramaniam, A.R., Pan, T., and Cluzel, P. (2013). Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. Proc. Natl. Acad. Sci. USA *110*, 2419–2424.

Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massagué, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. Nature *451*, 147–152.

Zhang, G., and Ignatova, Z. (2011). Folding at the birth of the nascent chain: coordinating translation with co-translational folding. Curr. Opin. Struct. Biol. *21*, 25–31.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nat. Methods *12*, 835–837.

Zouridis, H., and Hatzimanikatis, V. (2008). Effects of codon distributions and tRNA competition on protein translation. Biophys. J. *95*, 1018–1033.

Cell

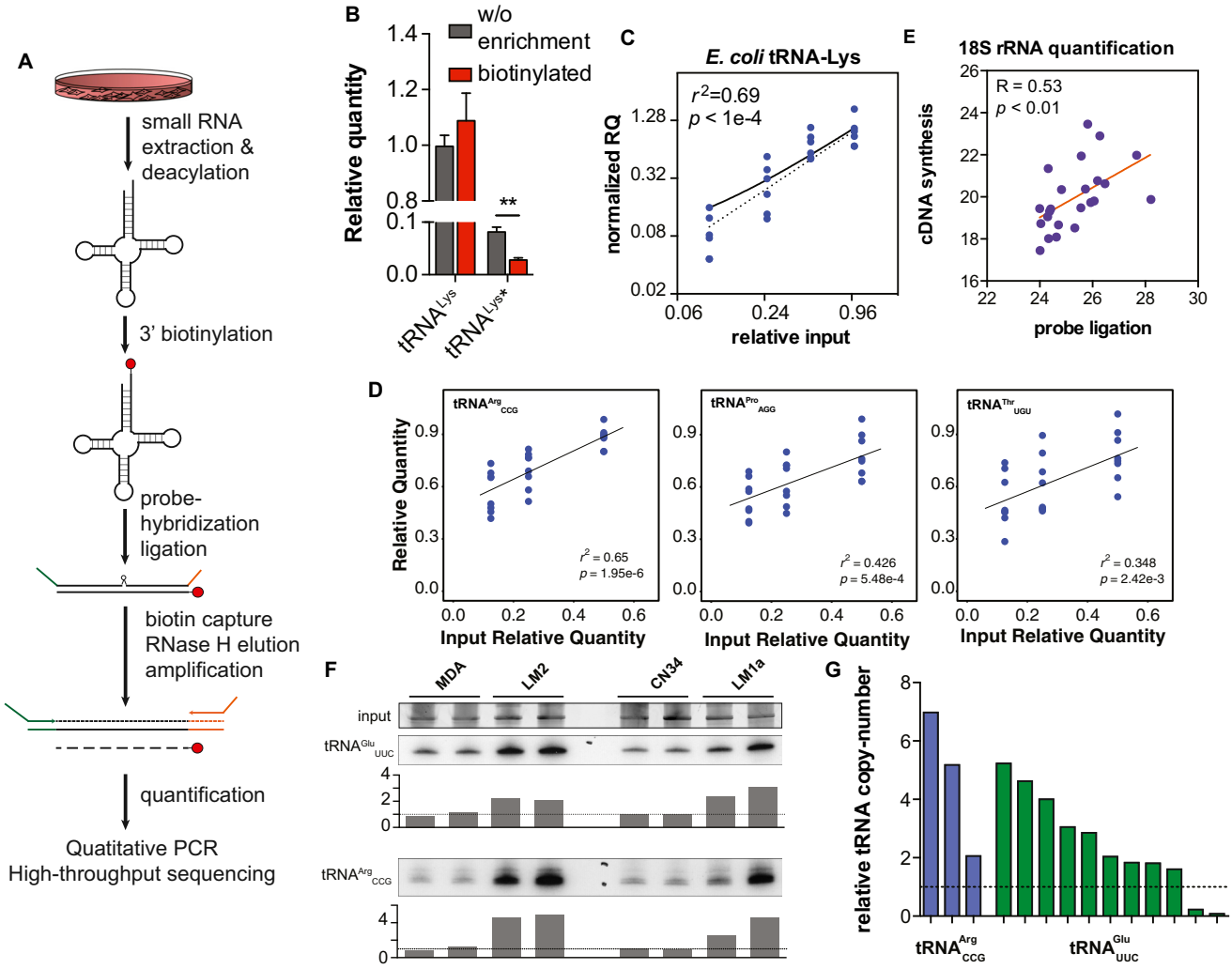


Figure S1. tRNA Profiling Approaches and Controls, Related to Figure 1

(A) For tRNA profiling, total small RNA from growing cells was extracted and subjected to deacylation and 3′-biotinylation. Transfer RNA-specific probe-pairs were hybridized to tRNAs and ligated using T4 DNA ligase. Hybridized and ligated probes were then extracted by streptavidin-based biotin precipitation followed by RNase-mediated elution. Then depending on the assay, either high-throughput sequencing or quantitative PCR assays were used to quantify the levels of each probe in the population.

(B) tRNA Lysine of *E. coli* was used for benchmarking our approach. Here, we have compared the quantification of probes perfectly designed against $tRNA^{Lys}$ relative to a probe with a mismatch at the ligation site ($tRNA^{Lys*}$). Without 3′ biotinylation and enrichment for tRNA hybridized probes, $tRNA^{Lys*}$ was measured at ~10% of $tRNA^{Lys}$; the addition of the Streptavidin-mediated precipitation step resulted in another 10-fold increase in signal-to-noise ratio.

(C) A standard curve created using $tRNA^{Lys}$ at different concentrations.

(D) Quantification of three different human tRNAs at different concentrations of input RNA from four separate cell-lines.

(E) We quantified 18S rRNA levels in 23 samples from tumors using first-strand cDNA synthesis followed by quantitative PCR. We then used the probe hybridization/ligation method to quantify 18S rRNA in the same samples. As shown here, we observed a significant agreement between the two distinct measurements.

(F) [32P]-labeled antisense RNA probes (IDT) were used to perform Northern blotting for $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ in MDA-Par, MDA-LM2, CN34- Par, and CN-LM1a samples (in biological replicates). Band intensities in each sample were then quantified and normalized to input RNA. In all cases, consistent with our tRNA profiling results, the levels of these tRNAs showed a significant upregulation.

(G) Relative difference in genomic copy numbers across multiple loci of $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ as measured by quantitative PCR.

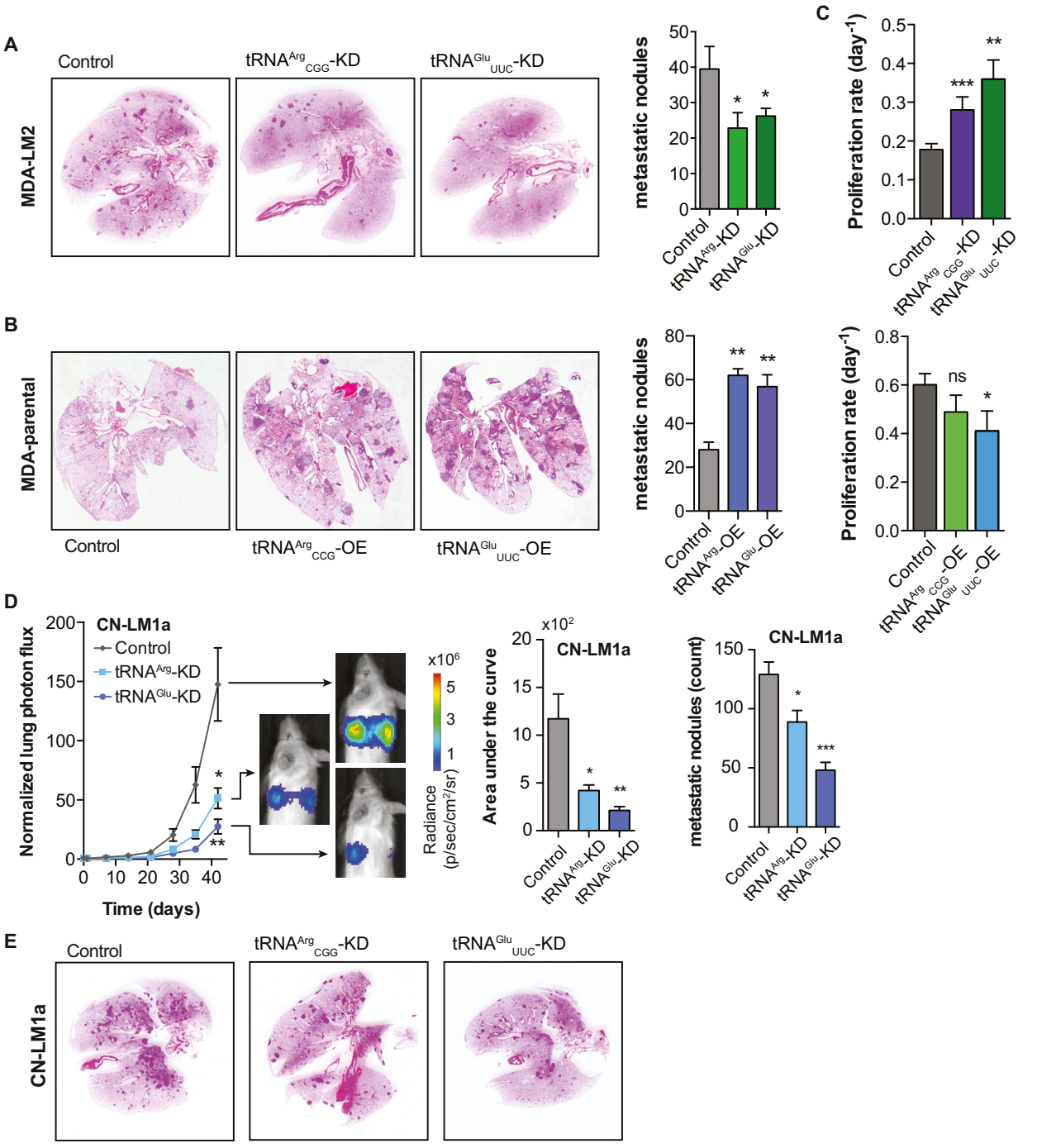


Figure S2. Gross Histology of Lungs Extracted from Mice Injected with MDA-LM2 and MDA-par Cells in which Specific tRNAs Are Knocked Down or Overexpressed, Related to Figure 2

(A and B) Lungs were extracted, fixed, sectioned, and subjected to hematoxylin and eosin staining. Visible macro-metastatic nodules were then counted for each cohort. Shown are representative lung slices from each cohort along with comparison of the number of visible nodules in each cohort.

(C) The proliferation of cells in vitro was performed for $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ overexpressing cells relative to control MDA-parental cells.

(D) Bioluminescence imaging plot of lung metastasis by CN-LM1a cells expressing short hairpins targeting $tRNA^{Glu}_{UUC}$, $tRNA^{Arg}_{CCG,}$ or a control hairpin (shControl); n = 4-5 in each cohort. Area-under-the-curve was also calculated for each mouse.

(E) Lungs were extracted, fixed, sectioned, and subjected to hematoxylin and eosin staining. Shown are representative lung slices from each cohort. Error bars indicate SEM. One-sample t test was used to calculate the associated *p*-values; *p < 0.05, **p < 0.01, and ***p < 0.001.

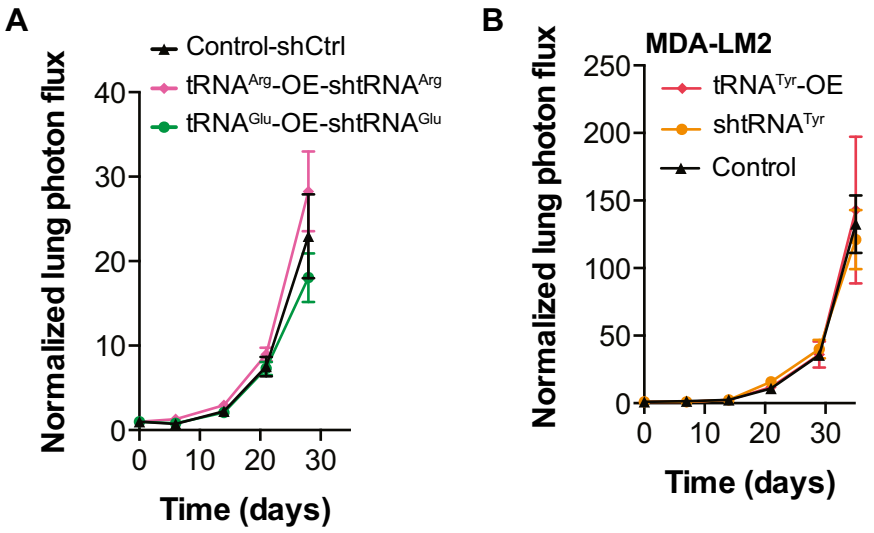


Figure S3. Modulations of Specific tRNAs Affect Metastatic Progression in Breast Cancer, Related to Figure 3

(A) Bioluminescence imaging plot of lung metastasis by MDA-parental cells simultaneously silencing and overexpressing $tRNA^{Glu}_{UUC}$ and $tRNA^{Arg}_{CCG}$ respectively (in comparison to shControl; n = 5 in each cohort).

(B) Bioluminescence imaging plot of lung metastasis by MDA-LM2 cells silencing or overexpressing $tRNA^{Tyr}$ relative to a control hairpin (shControl); n = 5 in each cohort.

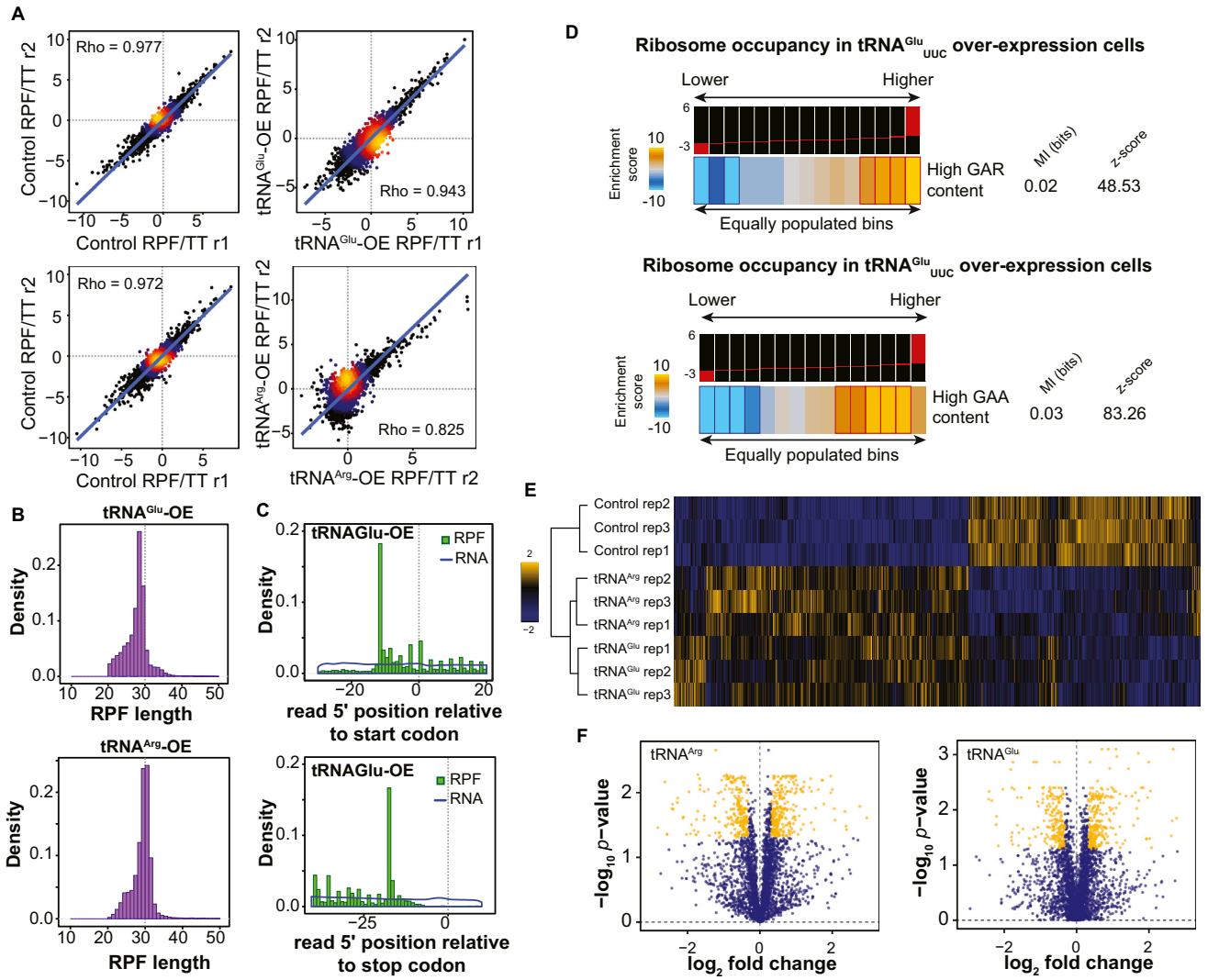


Figure S4. tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ Overexpression Alters Ribosomal Occupancy and Translational Landscapes, Related to Figure 4

(A) Ribosome protected fragment (RPF) reads were normalized to total RNA reads (TT) to correct for variations in transcript levels. Linear regression plots show expected positive Spearman's correlation of RPF/TT ratio between biological replicates in control, tRNA$^{Glu}_{UUC}$, and tRNA$^{Arg}_{CCG}$ overexpressing lines. It should be noted that since tRNA$^{Arg}_{CCG}$-OE and tRNA$^{Glu}_{UUC}$ -OE samples were prepared independently, they each have their corresponding control samples.

(B) The distribution of ribosome protected fragment (RPF) lengths in each sample.

(C) Observed periodicity of ribosome footprints in tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ overexpressing lines, which is the hallmark of the ribosome profiling approach, along with coverage of total RNA along the transcripts as control.

(D) Enrichment/depletion patterns of transcripts with higher GAR and GAA contents in the ribosome profiling measurements. Coding sequences with either high GAR (matching YUC) or GAA (matching UUC) content exhibited a highly significant enrichment among those with increased ribosome occupancy.

(E) Overexpression of tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ in MDA-parental cells resulted in pervasive protein expression modulations relative to the control cells.

(F) Volcano plots depicting the protein expression modulation in MDA-Par cells in which tRNA$^{Glu}_{UUC}$ and tRNA$^{Arg}_{CCG}$ were overexpressed.

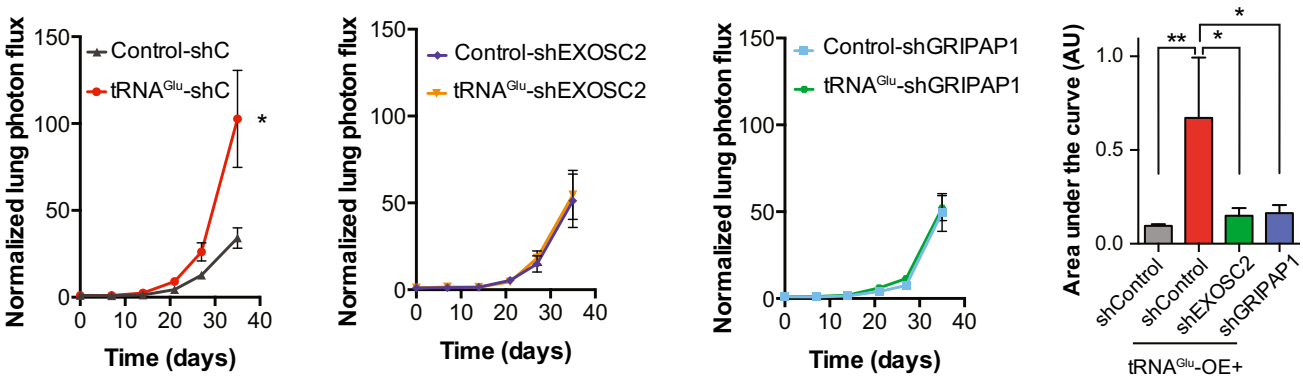


Figure S5. *EXOSC2* and *GRIPAP1* Act as Promoters of Breast Cancer Metastasis, Related to Figure 5

Bioluminescence imaging plot of lung colonization by MDA-231 cells expressing short-hairpins against *EXOSC2* or *GRIPAP1* in $tRNA^{Glu}_{UUC}$ overexpressing and control backgrounds; n = 5 in each cohort. Area-under-the-curve was also calculated for each group. While knock-down of *GRIPAP1* and *EXOSC2* reversed the lung colonization capacity of $tRNA^{Glu}_{UUC}$ overexpressing cells, it had little to no effect on control cells with expression level of $tRNA^{Glu}_{UUC}$ observed in parental cells. The *p*-values were calculated using one-sample one-tailed Mann-Whitney tests; *p < 0.05, and **p < 0.01.

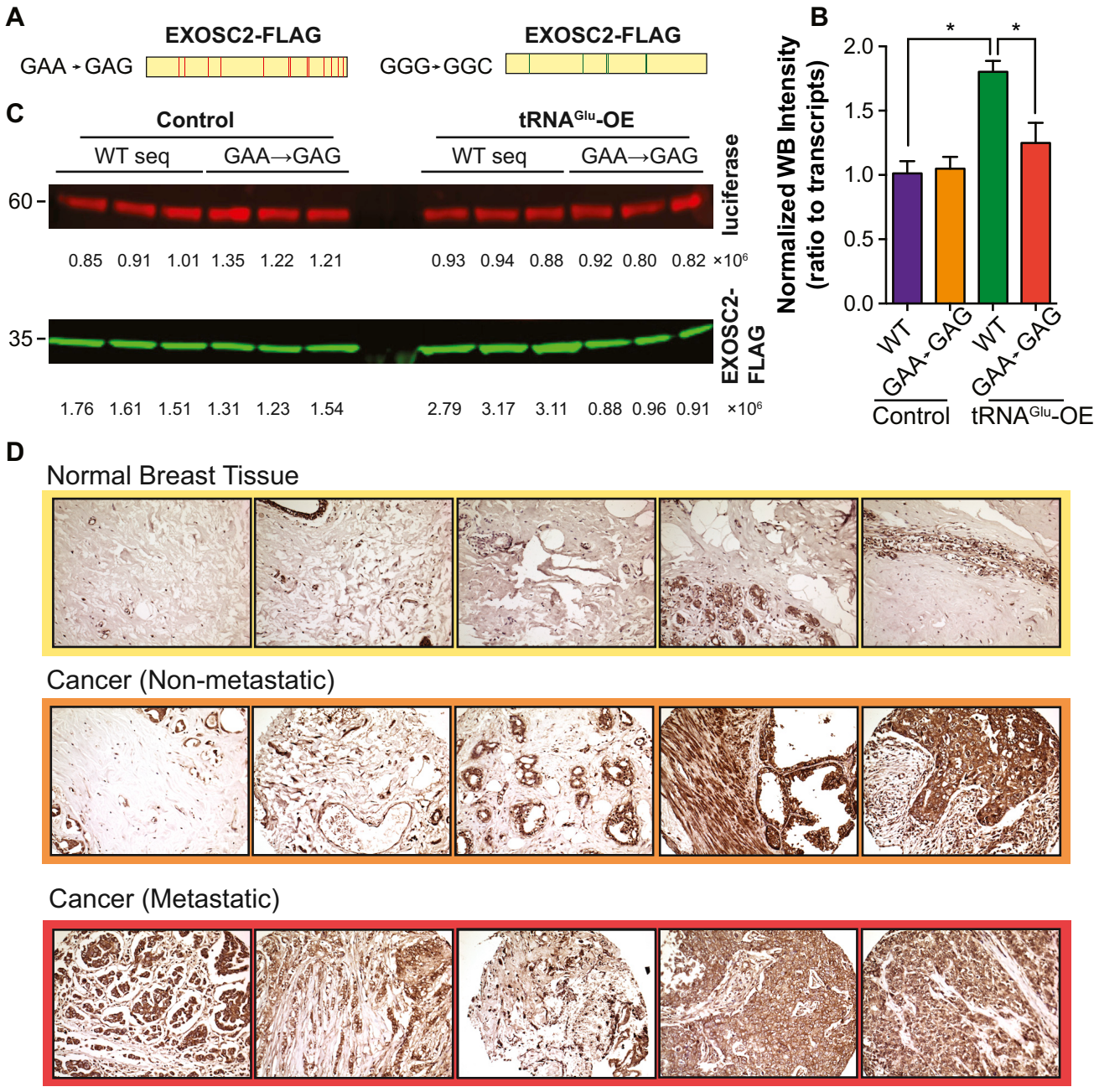


Figure S6. Codon-Specific Effect of $tRNA^{Glu}_{UUC}$ Modulations and Downstream Clinical Significance, Related to Figure 6

(A) Schematic map of mutational positions on synthetic *EXOSC2* constructs.

(B) Protein expression level (measured by quantitative WB) corrected for transcript level to demonstrate the translational-specific effect of $tRNA^{Glu}_{UUC}$ overexpression on transcript with or without its cognate codons. The p-values were calculated using one-sample one-tailed Mann-Whitney tests; *$p < 0.05$, and **$p < 0.01$.

(C) Shown are quantitative Western blots of control and $tRNA^{Glu}_{UUC}$ overexpressing cells transfected with either WT or GAA-to-GAG mutated codons in biological triplicates. Anti-FLAG antibody was used to measure different variants of exogenous EXOSC2 and anti-Luciferase antibody was used as a loading and transfection control (see Methods). Band intensities were quantified using LICOR quantification software.

(D) Shown are representative tissue-microarray immunohistochemical images of normal breast, non-metastatic invasive cancer, and metastatic cancer tissues (5 per cohort). Higher EXOSC2 intensity correlates with later disease progression stage.

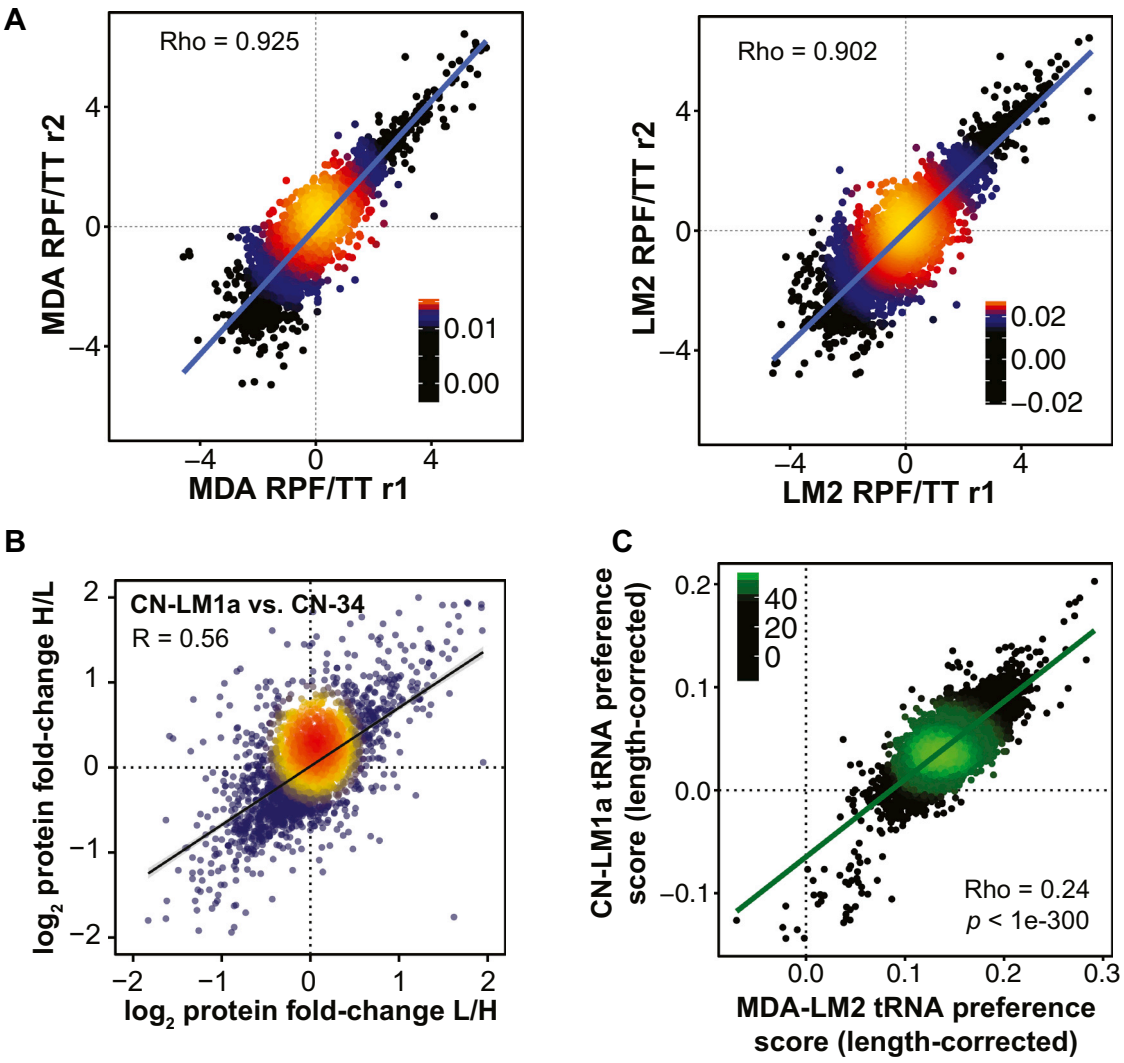


Figure S7. Differential Protein Expression and tRNA Preference Scores, Related to Figure 7

(A) Correlations between RPF/TT ratios in biological replicates in each background.

(B) As expected, label-swapped biological replicates quantified in our SILAC measurements exhibited a highly significant correlation. About 3,000 proteins were successfully quantified in both replicates.

(C) Consistent with tRNA modulations being similar between metastatic and parental lines in MDA-231 and CN-34 backgrounds, the tRNA preference scores (corrected for length of coding sequences) calculated for every protein based on differential tRNA expression in each background were highly correlated (Spearman's Rho = 0.24, p < 1e-300).

Cell, Volume *165*

Supplemental Information

Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression

Hani Goodarzi, Hoang C.B. Nguyen, Steven Zhang, Brian D. Dill, Henrik Molina, and Sohail F. Tavazoie

Extended Experimental Procedures

Cell culture

CN34 and MDA-MB-231 cells and their respective sub-lines, CN-LM1a and MDA-LM2, were propagated in DMEM-based media supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin and fungizone (Life Technolgies).

Stable cell line generation

For generation of lentivirus, 293T cells were seeded onto 10cm plates such that cell confluency will be approximately 70% the next day. 3 µg each of pRSV-Rev, pCMV-VSVG-G and pCgpV packaging vectors were co-transfected with 9 µg the appropriate pLKO-backboned targeting plasmid using 45 µl of Lipofectamine 2000 in antibiotic-free media. After 24 hours, the media was replaced with fresh antibiotic-free media. After 48 hours, virus-containing supernatant was collected and centrifuged for 10min at 2000rpm before being filtered through a 0.45 µm filter. For transduction of cells, 2mL of the appropriate virus was used to transduce 100k cells in the presence of 8 µg/mL polybrene. Media was replaced 24 hours later. 48 hours after transduction, antibiotic selection was performed with either blasticidin (10 µg/mL) or puromycin (2 µg/mL) for 2–7 days alongside a population of untransduced control cells. For each stable knockdown line, four independent shRNAs were tested and the best two shRNAs (>90% knockdown) for each cell-line were used (except for hairpins against tRNAs). For tRNA knock-downs, CGCTTTCACCGCCGCGGCCCGGGTT and AAGGCGTCTGATTCCGGATCAGAA were targeted in tRNA-Glu and tRNA-Arg respectively. For tRNA over-expression constructs, chr1.trna5-tRNA-Glu and chr16.trna1-tRNA-Arg sequences were used.

Northern blotting

10 µg of RNA was separated on a Urea-PAGE gel at 300V. After SYBR Gold-based staining and imaging, RNA was transferred to a nylon membrane in 0.5X TBE and UV crosslinked (240mJ/cm^2). The membrane was then pre-hybridized using UltraHyb-Oligo (Ambion) at 40C. RNA oligos were radiolabeled with [γ-^{32}P]ATP (3uL) and T4 PNK (NEB) at 37C for 30min and purified using G-25/G-50 columns. After hybridization overnight, the blot was washed twice and developed.

Western blotting

Cellular lysates were prepared by lysing cells (5-10 million) in ice-cold RIPA buffer containing protease and phosphatase inhibitors (Roche). Cellular debris was removed by centrifugation (12,000 rpm) for 20 min at 4 °C. Samples were denatured in loading buffer separated using SDS–PAGE, transferred to a PVDF membrane (Pierce), blocked and probed using target-specific antibody: EXOSC2 (Proteintech, 14805-1-AP; 1:500), GRIPAP1 (Proteintech, 15806-1-AP; 1:500). Bound antibodies were chemiluminescently detected using horseradish peroxidase–conjugated secondary antibodies (1:10,000), ECL Western Blotting Substrate (Pierce) and the SRX-101A (Konica Minolta) developer, according to the manufacturer's instructions. The membrane was then stripped (Restore Western Blot Stripping Buffer, Pierce) and re-probed (at RT for 1 hr) and re-developed (similar to the previous step) using an alpha-tubulin antibody (1:5,000; Sigma) or GAPDH (1:5,000, Cell signaling) as an internal control.

Quantitative Western Blot

Odyssey™ Quantitative Western Blot (LICOR) was used to quantify protein expression levels. Briefly, the procedure is identical to western blotting, except that instead of using horseradish peroxidase–conjugated secondary antibodies, species-specific fluorescent IRDye® secondary

antibodies were used. The membrane was then imaged with the Odyssey® Sa Infrared Imaging System at the Rockefeller University Center for High Throughput Screening. Quantification was done using Image Studio™ Lite (LICOR) and statistical significance was determined with unpaired *t*-test.

Transfer RNA profiling

For each family of mature tRNAs (with introns removed and containing terminal CCA) with a similar consensus sequence and a common anticodon, a pair of probes is designed so that upon annealing to the complementary tRNA, the resulting nick in the DNA-RNA hybrid is located at the site of the anticodon. We designed a total of 67 probe-pairs to cover the majority of cytosolic tRNAs. The downstream probes were 5'-phosphorylated to enable enzymatic ligation. The 67 probe-pairs were combined into four separate batches based on their melting temperature. The batches were then aliquoted to avoid repeated freeze-thaw cycles and stored at -80C.

For tRNA profiling, cells were subjected to small-RNA extraction (microRNA Purification Kit, Norgen) and $tRNA^{Lys}$ of *E. coli* (Sigma) was added at set concentrations as spike-in (varied depending on the RNA concentration). To deacylate the tRNA population, the samples were incubated in 100mM Tris-HCl (pH 9.0) at 37C for 30min. To stop the reaction, an equal volume of acetate buffer (pH 4.8) and NaCl were added to the final concentration of 50mM. The samples were then subjected to RNA precipitation at -20C overnight. Following re-suspension, RNA samples were 3' biotinylated in 25% DMSO using T4 RNA ligase (RNA 3' End Biotinylation Kit, Thermo Scientific) at 16C overnight. The samples were then subjected to chloroform extraction and RNA precipitation. The RNA population was then divided into four batches and each batch was hybridized with one of the four pools of probe- pairs (in 10mM Tris-HCl, 5mM NaCl, 100µM EGTA) by incubating at 90C for 5min followed by slow cool down to 55-60C (based on Tm of the pool). The hybridized samples were then subjected to ligation at 16C overnight (given the lower efficiency of T4 DNA ligase for DNA/RNA hybrids, overnight ligation is essential). MyOne-C1Streptavidin Dynabeads (Invitrogen) were then used to purify biotinylated DNA/RNA hybrids per the manufacturer's instructions. Ligated probes were then eluted after incubation with RNase H and RNase A (30min at 37C) followed by incubation with an elution buffer (50 mM Tris pH 8, 10 mM EDTA, 1% SDS; incubate at 65C for 30min with intermittent vortexing). Following elution, the samples were purified with DNA Clean & Concentrator™-5 (Zymo Research) per manufacturer's instructions and minimally amplified (12-15 cycles) with primers against the universal sequences included at the two ends of the ligated probes. This amplification step is required to add the linkers required for high-throughput Ilumina sequencing and barcoding. For this study, we prepared libraries in biological triplicate from each cell-line. The samples were then sequences on two lanes of a HiSeq 2000 Illumina sequencer. The obtained reads were then aligned against the fasta file carrying all probe sequences. Sequences containing degenerate positions were multiplied so that all possible versions were present in the reference library. The relative quantity of each probe was calculated by normalizing the number of mapped reads to the total number of reads and statistical significance was assessed using the DESeq package in R.

Low-throughput tRNA quantification

The preparation of samples for low-throughput tRNA quantification was identical to the protocol described above, with the following exception: (i) instead of using a library of probes, the RNA samples were hybridized to a single probe-pair (matching the tRNA of interest); (ii) the amplification step is replaced with quantitative PCR using SYBR Green (Life Technologies) per manufacturer's instructions. The qPCR primers were CACGACGCTCTTCCGATCT and TTCTTTGCAGTGTCGTGG matching the 3' and 5' ends of the probe-pairs respectively. We used

$tRNA^{Lys}$ of *E. coli* or small-RNA populations extracted from cells to benchmark the protocol and create the standard curves and compare the efficiency of T4 DNA ligase in recognizing a mismatch at the location of the nick (Fig. S1A-C). This qPCR-based measurements of specific tRNA species was also used to validate tRNA expression changes between cell-lines and also cell-lines generated to test tRNA knock-down and over-expression of $tRNA^{Arg}$ and $tRNA^{Glu}$.

We also used probes against the 18S rRNA to compare quantifications based on probe-hybridization/ligation to those obtained from cDNA synthesis followed by qPCR (Fig S2B). The following probe sequences were synthesized (IDT) to detect 18S levels:

/5Phos/GGTAGTAGCGACGGGCGGTGTGTACAAAGGGC AGATCGGAAGAGCGTCGTG and TTCTTTGCAGTGTCGTGG CCGATCCGAGGGCCTCACTAAACCATCCAATC.

Whole-genome transcript stability measurements

Cells were seeded in triplicates at 80% confluence in 6-well plates in DMEM-based media enriched with 10% FBS and. 24 hours later, media was spiked with 10 μg/ml of α-amanitin (Sigma). Total RNA was extracted using the manufacturer's protocol (Norgen) at 0 and 8 hr time points after α-amanitin treatment. Samples were then labeled using TargetAmp-Nano Labeling Kit for Illumina (Epicenter). Labeled RNA was purified using RNeasy Minelute Kit (QIAGEN) and submitted for analysis to the Rockefeller University genomics core facility using Illumina HT-12 v4 Expression BeadChip microarrays. The Lumi package in R was used to transform and normalize signal intensities. The transformed signal at 0 and 8 hr time points was then normalized based on input RNA and used to estimate a decay rate for each transcript. For transcript-specific stability measurements, qRT-PCR was used to determine relative quantities of tRNA-overexpression dependent stabilized targets at different time points using 18S as endogenous control.

Quantitative RT-PCR

Transcript levels were measured using quantitative RT-PCR by first converting total RNA to cDNA (SuperScript III, Life Technologies) followed by SYBR Green quantification (Life Technologies) per the manufacturer's instructions. We used the following primers for qRT-PCR: EXOSC2: TGGCTCGCAAGCCTCTTAG, TGTGTCCGTAGTGATTGTATCCC; GRIPAP1: GGACAAACAACTACCAGCTTTCA, GCGACCTTCTGTCGAAGACT; ERH: AATGAATCCCAACAGTCCCTCT, CAGCTCGGTAAACCAGGCAG; AP1S: CGGTTCATGCTATTATTCAGCCG, CCGTTCCTTGTCCGAAGTGG; SBDS: ACCAACCAGATCCGCCTAAC, CGACGACCTTGTTTTTGTAGCA; EXOSC2-FLAG: AACTGCATCATCTCGCTGGTAAC, CTTGTCGTCATCGTCTTTGTAGT; Luciferase: CTGACCGGCAAGCTGGA, GACTCTAGAATTATTACACGGCGATCT. For pre-tRNA expression level and tRNA copy number, we designed 15-nt long qRT-PCR primers for each tRNA locus (hg19) by including 10-nt from the upstream (forward primer) or downstream (reverse primer) in each primer.

Whole-genome ribosomal occupancy profiling

The procedure was performed with Truseq Ribo Profile for mammalian cells (Illumina) per manufacturer's instructions. An input of 50×10^6 cells were harvested for each replicate (biological duplicates for every cell line profiled) and libraries were sequenced using Illumina Nextseq 500 at the Rockefeller Genomics Center. For analysis, we first performed quality trimming and linker removal (cutadapt v1.8). We then used bowtie2 to remove reads that map to contaminating RNAs (*e.g.* rRNA sequences). Tophat and cufflinks were consequently used to map the resulting reads to the human

transcriptome and to compare transcript abundances. We normalized the RPKM estimates from ribosome-protected fragments (RPF) to the total RNA from each sample (TT) and used the resulting logFC values to compare control, $tRNA^{Arg}_{CCG}$-OE, and $tRNA^{Glu}_{UUC}$-OE samples. The MDA and CN34 parental cells and their highly metastatic derivatives (MDA-LM2, and CN-LM1a) were similarly analyzed. To correct for GC bias, when present in the resulting datasets, we applied a lowess regression model based on GC content of coding sequences and subtracted the resulting values from the observed logFC values prior to performing enrichment analysis. This step is of especial importance for $tRNA^{Arg}_{CCG}$, whose anticodon has a GC content of 1.

Codon-specific mutational assays

Wild-type and mutated versions of EXOSC2 coding sequences were synthesized by IDT. Every Glu-GAA codon was mutated to Glu-GAG. As a negative control, another construct in which every Gly-GGG codon was mutated to Gly-GGC was used (Figure S6A). Gly-GGG and Gly-GGC codons were chosen because there was no significant difference in their respective tRNA levels based on our high-throughput tRNA profiling analysis. Each synthetic coding sequence also contained a 3'-FLAG-tag. The constructs were then cloned into the psiCHECK2 backbone (replacing synthetic Renilla Luciferase gene) together with an upstream Tetracycline-response element.

For transcript stability studies, 1×10^6 MDA-Ctrl and MDA-$tRNA^{Glu}_{UUC}$ overexpressing cells were seeded overnight in biological triplicates in 100mm tissue culture dishes. 1µg of psiCHECK2 (containing either wild-type, Gly-mutated, or Glu-mutated versions of EXOSC2) and 1 µg of pTet-OFF (Clontech) were co-transfected into cells using 10 µL of Lipofectamine 2000. 24 hours post-transfection, doxycycline was added to a final concentration of 200ng/mL to stop transcription. Untreated cells were harvested for the time 0 data point. 8 hours later, doxycycline-treated cells were harvested. RNA and protein was extracted as described before, with one exception: RNA was treated with DNase I (NEB) for 30 minutes at 37°C to remove any contaminating plasmid DNA.

Quantitative RT-PCR primers were designed such that the forward primer targets the EXOSC2 sequence that is identical among the wild-type and mutated versions of the transcript, while the reverse primer mapped to the FLAG-tag. As a control for transfection efficiency, primers were designed against the co-transcribed Firefly luciferase gene on psiCHECK2. To ensure the absence of contaminating plasmid DNA, a reverse transcriptase-free control reaction was run side-by-side, together with qPCR primers that target the ampicillin resistance gene on psiCHECK2. Stability was measured by comparing the relative changes in normalized transcript levels between 0h and 8h time point. For protein expression measurement, anti-FLAG antibody (Sigma) was used with anti-Firefly Luciferase (Novus, NB100-1677) as a control.

Lung colonization assays

All mouse studies were conducted according to a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the Rockefeller University. Seven- to eight-week-old age-matched female NOD/SCID gamma mice (Jackson Labs) were used for lung colonization assays. In all cases, 5×10^4 cells in 100 µL PBS were injected via tail-vein along with cells expressing a neutral hairpin as control. In every case, 4-5 mice were included in each cohort. The metastatic growth was tested using both two-way ANOVA as a function of time and sub-line identity and also *t*-test based comparison of area under the curve for each mouse.

Primary tumor growth and orthotopic metastasis

1×10^6 triple-reporter labeled cells were suspended in 100 µl of 1:1 PBS:Matrigel mixture and injected into subcutaneous flanks of mice at two sites. Tumor growth was measured using digital calipers starting 7 days after injection when palpable tumors can be measured accurately. Tumor volumes were calculated using the formula, Volume = $(\text{width})^2 \times (\text{length})/2$. Tumors were resected as they reached a growth plateau and lung metastasis was quantified through bioluminescent imaging. Statistical significance was determined as described above.

Histology

For gross macroscopic metastatic nodule visualization, mice lungs (from each cohort) were extracted at specific time-points post-injection and 5µm thick lung tissue sections were stained with hematoxylin and eosin (H&E). The number of macroscopic nodules was then recorded for each section. Unpaired *t*-test was used to test for significant variations.

Tissue Microarray and immunohistochemistry

Breast cancer progression tissue microarrays (TMA) were obtained from the Cancer Diagnosis Program. Immunohistochemistry was performed using an antibody targeting EXOSC2 (Proteintech, 14805-1-AP, 1:100 dilution) and VECTASTAIN ABC (Vectorlabs, PK-4000) per manufacturers' instructions with a citrate buffer (pH 6.0) antigen retrieval step. Staining was assigned a score of either 1, 2, or 3 corresponding to the low, medium, or high intensity of the signal (in a blinded fashion). A score of N/A was assigned to damaged tissues or to those sections for which staining was incomplete as judged by the scorers. The scores were then counted and compared among different clinical stages of interest (i.e., non-invasive vs. invasive, non-metastatic vs. metastatic). χ^2 statistics was used to initially test for the difference in the percentage of each intensity score in each cohort. Hypergeometric p-values were then calculated to assess the increase in the frequency of samples with higher intensity among the metastatic tumors (reported in the bargraphs). Three blinded researchers who were not authors were also asked to independently score the samples to ensure the unbiased nature of the findings. Representative images of stained tissues were taken at the Rockefeller Bio-Imaging Resource Center (Zeiss -field fluorescence/brightfield/DIC microscope).

Cancer cell invasion assays

Invasion assays were performed as previously described (Png et al., 2012). Briefly, cells were starved for 24 hours in DMEM-based media supplemented with 0.2% FBS. 5×10^5 cells were seeded in the same conditions in trans-well invasion chambers (BD Biosciences) and incubated for 16-18 hours. Inserts stained with DAPI and imaged using an inverted fluorescence microscope (Zeiss Axiovert 40 CFL), sampling five fields per insert. Invading cells were counted in a blinded fashion using ImageJ.

Cancer cell proliferation

Roughly 10,000 cells were seeded into three 6-well wells and subsequently were trypsinized and viable cells were counted using a hemocytometer at days 1, 3 and 5. An exponential model was then used to fit a growth rate for each sample ($\ln(N_{t-1}/N_1) = rt$ where t is measured in days). The experiment was performed blinded in biological quadruplicates and unpaired *t*-test was used to test for significant variations.

Codon-optimized luciferase assays

Approximately 200,000 cells were grown in 24-well plates in quadruplicate and were transfected as previously described with vector codon-optimized luciferase constructs. 48 hours post-transfection,

luciferase activity was measured according the manufacturer's protocol. Briefly, cells were washed using 1xPBS and lysed using 500ul of PLB for 15 min at room temperature. Lysate was then plated in technical quadruplicates onto a 96 well plate. Immediately prior to measuring firefly luciferase activity, 100ul of LARII solution was added to each sample. Subsequently, 100ul of Stop&Glo Reagent was added to measure Renilla luciferase activity. Half of the lysate was used to extract and measure luciferase transcript levels using qRT-PCR.

LC-MS/MS analysis of protein expression

For the label-free quantitation analysis of expression levels of $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ against control MDA cells, 100 µg protein from each sample was acetone precipitated overnight. Precipitates were dissolved in 8M urea (GE Healthcare) with 0.1 M ammonium bicarbonate (Sigma) and 10 mM DTT (Sigma), incubated with shaking at room temperature for 1h, followed by iodoacetamide (Sigma) alkylation of cysteines. Samples were diluted to below 4M urea before lysyl endopeptidase (Wako) digestion overnight. Samples were then diluted below 2 M urea before adding trypsin (Promega) digestion for 6 h. Peptides were desalted/concentrated using STAGE-type C18 tips (Ishihama et al., 2006) and resuspended in 2% acetonitrile /2% formic acid.

1 µg of each sample was analyzed by nano LC-MS/MS (Dionex 3000 HPLC coupled to a Q-Exactive mass spectrometer, ThermoFisher Scientific). Peptides were separated at 200 nL/min using a gradient increasing from 5% B to 45% B in 133 minutes (A: 0.1% formic acid, B: acetonitrile/0.1% formic acid). Peptides were loaded onto a trap column prior to separation on a packed-in-emitter C18 column (75 µm by 12 cm, 3 µm particles - Nikkyo Technos Co., Ltd. Japan). The mass spectrometer was operated in 'preferred mode' fragmentation up to 20 ions per cycle using an under fill ratio of 1%. MS spectra (m/z range: 300-1400) were recorded at a resolution of 70,000 (AGC: 5e5) and MS/MS spectra at 17,500 (AGC: 2e5) with a lowest m/z of 100. Generated LC-MS/MS data were queried against Uniprot's complete Human Proteome (July 2014) and quantitated using MaxQuant 1.5.0.9 (Cox et al., 2014). In short, Peptide-Spectrum Match false discovery rate and protein false discovery rate was set to 1%. Match between runs were used for label free quantitation (LFQ). A total of 5,096 proteins were matched. Proteins were filtered requiring that a protein be matched in a minimum 2 out of 3 biological replicates. 3,475 proteins fulfilled this criterion. Excellent correlation (r2 ~0.99) was observed between replicates. Missing LFQ values was replaced with imputated LFQ values (width: 0.3, down shift: 1.8). Differences between the three conditions were assessed by a multiple ANOVA test using a permutation based FDR cut-off of $p<0.05$.

For the correlation of tRNA preference scores to protein abundance changes between metastatic and parental cell lines using SILAC quantitation, each cell line was split into two populations of approximately 1×10^6 cells, each cultured for 10 days in DMEM-Flex based media supplemented with dialyzed FBS (Life Technologies) and either heavy arginine and lysine (^{10}R,^{6}K) or normal arginine and lysine (0.1mg/ml). Cells were passaged every 2 days to maintain appropriate confluency in 10 cm^2 plates. After 10 days, heavy-labeled parental cells were mixed with light-labeled derivative cells (and vice versa) in equal numbers and lysed according to the manufacturer's protocol. Briefly, cells were trypsinized and resuspended with cell counting performed twice using a manual cytometer to ensure consistency. 5×10^5 cells from each population were mixed together and spun down at 1000 rpm for 5 minutes. Cells were lysed using SILAC Phosphoprotein Lysis Buffer A (per manufacturer instructions). Fifty µg of each heavy/light mixture was prepared as above, followed by STAGE-type SAX microfractionation as described previously (Wisniewski et al., 2009). In brief, desalted peptides in a pH 11 buffer were loaded onto STAGE-type SAX tips, collecting the flow-through, and eluted with decreasing pH: 8, 6, 5, 4, and 3, for a total of 6 fractions. Fractions and an unfractionated sample

were analyzed by LC-MS/MS and MaxQuant/Perseus as above, yielding 4,601 protein identifications after filtering contaminants and 3,756 protein ratios based on proteins with at least 2 peptides. Additionally, analysis of unmixed, heavy-only samples demonstrated better than 98% SILAC incorporation for each cell line.

Differential gene expression analysis

Protein expression is a function of mRNA abundance. To correct differential protein expression levels for $tRNA^{Arg}_{CCG}$ and $tRNA^{Glu}_{UUC}$ over-expressing cells against control MDA cells, we performed microarray-based transcriptomic measurements in biological replicates. Sample preparation and data analysis was performed as described before (Goodarzi et al., 2014). The fold-changes in the levels of the detected proteins were then corrected by the fold-change in the transcript levels. We used a similar approach for correcting differential protein expression between parental and highly metastatic sub- lines.

References

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Molecular & cellular proteomics : MCP *13*, 2513-2526.

Goodarzi, H., Zhang, S., Buss, C.G., Fish, L., Tavazoie, S., and Tavazoie, S.F. (2014). Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. Nature *513*, 256-260.

Ishihama, Y., Rappsilber, J., and Mann, M. (2006). Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. Journal of proteome research *5*, 988-994.

Png, K.J., Halberg, N., Yoshida, M., and Tavazoie, S.F. (2012). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. Nature *481*, 190-194.

Wisniewski, J.R., Zougman, A., and Mann, M. (2009). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. Journal of proteome research *8*, 5674-5678.