

AD _____

Award Number: W81XWH-07-1-0595

TITLE: tRNAs as Biomarkers and Regulators for Breast Cancer

PRINCIPAL INVESTIGATOR: Tao Pan, Ph.D.

CONTRACTING ORGANIZATION: University of Chicago
Chicago, IL 60637

REPORT DATE: August 2010

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

X Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 01-08-2010		2. REPORT TYPE Final		3. DATES COVERED (From - To) July 30, 2007-July 29, 2010	
4. TITLE AND SUBTITLE tRNAs as Biomarkers and Regulators for Breast Cancer				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-07-1-0595	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Tao Pan * taopan@uchicago.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Chicago Chicago, Illinois 60637				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Material Command Fort Detrick, Maryland, 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT tRNA reads mRNA codons in translation and is essential for protein synthesis. Like mRNA, tRNAs are also under stringent cellular control. Our proposal aims to determine the feasibility of using tRNA expression as biomarkers for breast cancer type and progression, and how tRNAs are used to regulate gene expression in breast cancer cells. We found that the expression of both nuclear and mitochondrial-encoded tRNAs can be useful as possible biomarkers for breast cancer. Furthermore, tRNA expression also correlates in some way to certain groups of breast cancer related genes. These results demonstrate that studies of tRNA and breast cancer biology will be useful in understanding breast cancer type and progression and may lead to new drug targets for breast cancer treatment.					
15. SUBJECT TERMS tRNA, biomarker, regulation of gene expression, breast cancer.					
16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 30	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
Introduction.....	4
BODY.....	4
Key Research Accomplishments.....	8
Reportable Outcomes.....	9
Conclusion.....	9
Appendices.....	11

Final Report

Proposal : W81XWH-07-1-0595, tRNAs as Biomarkers and Regulators for Breast Cancer.

Period: 07/30/07-07/29/10.

PIs: Tao Pan, Marsha Rosner.

INTRODUCTION:

tRNAs read mRNA codons in translation and are essential for protein synthesis. Like mRNA, the concentration and the identity of tRNAs are also under stringent cellular control. Our project aims to determine whether tRNAs serve as biomarkers for breast cancer tumor cells and whether tRNAs in normal and breast cancer cell lines control the expression of selected genes that are important for the tumorigenic process. Finding tRNA as biomarkers could open up a new and so far under-appreciated avenue for detecting the type and stage of breast cancer progression. Identification of tRNAs that regulate tumor development and progression could provide targets for a new class of cancer drugs.

BODY:

All figures and tables cited are in the appended publication: Pavon-Eternod et al.: tRNA over-expression in breast cancer and functional consequences. *Nucl. Acids Res.* 37, 7268-7280 (2009).

Task 1: Development of a “second-generation” tRNA microarray capable of studying single-nucleotide differences in human tRNA. → **Completed successfully.**

Our “first generation” microarray identifies tRNA based on hybridization alone and could not distinguish tRNAs with less than 8 nucleotide differences in their sequences (Fig. 1). To enable detection of single-base difference between tRNAs, we developed a “second-generation” method based on chemical ligation (Fig. 5A). Enzymatic ligation methods are commonly used to probe single-base differences in DNA sequences, but are inefficient on RNA templates and can result in poorly reproducible data. An appealing alternative to enzymatic ligation is non-enzymatic autoligation, in particular the 3’phosphorothioate-5’iodide autoligation chemistry (Fig. 4A). This reaction proceeds efficiently at moderate temperatures, yields high discrimination against point mutations, and works well on DNA templates.

The development of the second-generation method took place in two stages. In the first stage, a model 30-mer RNA was used to determine whether 5’iodide and 3’thiophosphate DNA oligonucleotide probes prefer one RNA base over the other in ligation (Fig. 4B). Ligation yield was highly sensitive to the presence of a mismatch at the ligation junction, so that ligation proceeded efficiently only when the DNA probes were fully complementary to the RNA template at the ligation junction (Fig. 4C). This method was verified on full-length purified yeast tRNA^{Phe} and selected tRNAs in total cellular RNA mixture (Fig. 4D). In the second stage, this autoligation chemistry was applied to the detection and quantitation of single-base differences at the first position of the tRNA anticodon in a microarray format (Fig. 5B, 5C, 5D).

Task 2: Development of a “third-generation” method capable of studying the expression of >90% of human tRNA species. → **Did not complete due to technical difficulties.**

After the successful development of the “second-generation” microarray method (task 1), we realized that the most effective way for completing task 2 is to apply the newly available, high throughput DNA sequencing method (e.g. Solexa technology by Illumina). This new technology offers a much higher data output at a reduced cost.

To our dismay, we found out that the cDNA synthesis of cellular tRNAs could not be performed reproducibly in a quantitative fashion. This inadequacy is most likely derived from the presence of a large number of post-transcriptional modifications in cellular tRNAs (~13 modifications for each tRNA on average). Since the cDNA synthesis is the first step for high throughput DNA sequencing and quantitative comparison is our ultimate goal, this technical difficulty represents a high enough hurdle that further experiments on high throughput DNA sequencing of total cellular tRNA were not attempted.

Task 3: Evaluate the usefulness of tRNA expression pattern as biomarkers. → **Completed successfully.**

To explore the potential of tRNAs to define breast cancer signatures, we used tRNA microarrays to generate comparative tRNA profiles for three non-tumorigenic breast epithelial cell lines (MCF10A, 184 A1, 184 B5) and six breast cancer cell lines (BT-474, HCC70, MCF7, MDA-MB-231, MDA-MB-436, ZR-75-1). The breast cancer cell lines cover a range of physiological and molecular properties (Table S2). Importantly, genome-wide mRNA expression data are available for these cell lines to allow tRNA expression and codon usage correlation analysis.

Significant differences are observed in the expression levels of tRNA among non-tumorigenic and tumorigenic breast cell lines (Fig. 2). The global level of nuclear- and mitochondrial-encoded tRNAs can be approximated separately by the median and mean sample-to-MCF10A ratio. For nuclear-encoded tRNAs, this ratio is 0.7-0.8 for the other non-tumorigenic cell lines and 2-3 for the tumorigenic cell lines. These differences are even more pronounced for mitochondrial-encoded tRNA levels: the sample-to-MCF10A ratio for non-tumorigenic cell lines is approximately 0.75, but as much as 5 for the tumorigenic cells.

The over-expression of tRNA in tumorigenic relative to non-tumorigenic cell lines is also selective: certain individual tRNAs are more strongly over-expressed than others. Variations in the relative expression of nuclear-encoded tRNAs carrying certain amino acid types are readily observed across tumorigenic cell lines. For example, nuclear-encoded tRNAs carrying polar amino acids (e.g. Ser, Thr, and Tyr) are up-regulated 3 to 4-fold in breast cancer cell lines relative to MCF10A, while nuclear-encoded tRNAs carrying small amino acids (Ala, Cys, Gly, Pro) are up-regulated only 1.5 to 2-fold. These differences can be observed more clearly when the tRNA expression is normalized to the median value for either nuclear- or mitochondrial-encoded tRNAs within each cell line (Fig. 2D, 2E). Differences in the relative expression of tRNA isoacceptors also become more apparent after normalization. Selective up-regulation of tRNA levels is also observed for mitochondrial-encoded tRNAs: certain tRNAs are expressed up to 2-fold above median, and others are expressed 4-fold below median. While nuclear-encoded tRNA expression patterns are remarkably similar across cell lines, mitochondrial-encoded tRNA expression patterns exhibit greater variations.

We also determined that tRNA over-expression in breast cancer cell lines does not simply reflect an increase in proliferation rate (Table S3). Under the culture conditions used in this study, doubling times range from 21 to 47 hours for the non-tumorigenic cell lines, versus 15 to 35 hours for the tumorigenic cell lines. Thus, the proliferation of all three non-tumorigenic cell lines with low tRNA content is comparable to that of tumorigenic breast cancer cell lines with much higher tRNA content. Doubling times do not correlate with either nuclear or mitochondrial global tRNA levels.

tRNA over-expression in patient-derived breast tumor samples: In all breast cancer cell lines examined, we observe global tRNA over-expression and differential tRNA isoacceptor expression. To generalize our results in cell lines to breast cancer *in vivo*, we measured tRNA expression patterns in 9 patient-derived breast tumor samples and 3 normal breast tissue samples (Table S1). The 9 breast tumor samples were selected from the three major subtypes of breast cancer: luminal (ER+, HER2-), basal (ER-, HER2-), and HER2+. For consistency of data analysis, all samples were run using MCF10A as a reference in array experiments.

As in the breast cancer cell lines, we observe significant differences in global tRNA expression levels among breast tumor and normal breast samples (Fig. 3). For nuclear-encoded tRNAs, the mean sample-to-MCF10A ratios is 0.2 to 0.5 for normal breast tissue samples, compared to 2 to 4 for breast tumor samples. This translates to up to 20-fold up-regulation of global nuclear-encoded tRNA levels in breast tumors relative to normal breast. For mitochondrial-encoded tRNAs, the mean sample-to-MCF10A ratio is 0.4 to 1 for normal breast tissue samples, compared to 1.2 to 5 for breast tumor samples. This translates to up to 13-fold up-regulation of global mitochondrial-encoded tRNA levels in breast tumors relative to normal breast.

Consistent with our results on cell lines, tRNA over-expression in breast tumor samples is also selective: certain individual tRNAs are more strongly over-expressed than others (Fig. 3C). Among the top 10 over-expressed, nuclear-encoded tRNAs, six tRNA species overlap for cell lines and tumor tissues. We readily observe variations in the relative expression of both nuclear- and mitochondrial-encoded tRNAs. These variations correlate with both the cognate amino acid type and tRNA isoacceptor identity. We also observe variations in the relative expression of tRNA isoacceptors. These variations become more clear when tRNA expression levels are normalized to the median value within each sample (Fig. 3D). However, tRNA expression pattern in the three major sub-types of breast cancer is very similar, suggesting that tRNA over-expression is a general consequence for all breast tumors.

We conclude that elevated tRNA levels are a hallmark of breast cancer. Our results suggest that tRNAs can be used as biomarkers for breast cancer, although it is best suitable to determine tumorigenic versus normal breast cells. Furthermore, tRNA expression measured by microarrays more accurately reflects the “functional abundance” of individual tRNA species.

Task 4: Identify correlations between tRNA and mRNA expression. → **Completed successfully.**

The differential expression of tRNA isoacceptors can be used to regulate translational efficiency via the codon usage of specific genes. In prokaryotes and fungi, differences in the abundance of tRNA isoacceptors are correlated with codon preferences of genes encoding highly expressed proteins and impact the synthesis of these proteins. In humans, tissue-specific differences

in the expression of individual tRNA species can correlate to the codon usage of highly-expressed, tissue-specific genes, although this correlation was seen only for a small number of tissues.

tRNA isoacceptor levels in breast cancer cell lines may also correlate with the codon usage of certain genes that are important for cancer. Finding such correlations would suggest an additional level of translational regulation for breast cancer cell lines, and by extension for breast cancer *in vivo*. At least three groups of genes are relevant in seeking codon usage-tRNA correlation for cancer cell lines: cell-line specific genes, cancer-related genes and house-keeping genes (Fig. 6A). Cell-line specific genes can be important for distinct tumorigenic properties across cell lines; cancer-related genes can be important for general tumor initiation and progression; and house-keeping genes are important for cell growth and architecture.

To identify cell line-specific genes, we used the publicly available mRNA expression data for the breast cancer lines and MCF10A (<http://www.ebi.ac.uk/arrayexpress/>, accession number E-TABM-157). To identify cell line-specific genes, we used publicly available mRNA expression data (<http://www.ebi.ac.uk/arrayexpress/>, accession number E-TABM-157). Cell line-specific genes were selected based on mRNA expression level (7- to 15-fold above the median expression level determined for all genes) and high cell line/MCF10A expression ratios (top 20 to 30 genes). A cell line-specific gene set was determined for each breast cancer cell line. To identify cancer-related genes, we selected from a comprehensive list of potential breast cancer diagnostic markers (http://www.sabiosciences.com/gene_array_product/HTML/OHS-402.html). Genes in this group are highly associated with breast cancer. Functional groupings used in our study include: cell cycle, cell growth and proliferation, ECM molecules, protein kinases, and transcription factors/regulators. To identify house-keeping genes, we selected the 30 most highly expressed house-keeping genes as defined in a previous report. This list includes ribosomal proteins, actin, ubiquitin, and others.

For each gene set, we compiled gene sequences and analyzed them for codon usage (expressed as number per one thousand codons, www.bioinformatics.org/sms2) (Tables S8, S9). Because certain tRNAs read more than one codon, we converted the obtained codon usage into tRNA-based codon usage. For example, tRNA^{Arg(ICG)} reads both CGU and CGC. Its tRNA-based codon usage is therefore equal to the sum of the CGU and CGC codon usages. For simplicity, we refer to tRNA-based codon usage as codon usage throughout our analysis.

We first compared the codon usages of each gene set in the three gene groups (cell line-specific, cancer-related, and housekeeping) against each other (Fig. 6B). Though there is limited overlap across cell line-specific gene sets, their codon usage was remarkably similar (average $r_s = 0.92 \pm 0.04$). The codon usage of cell-line specific gene sets also correlates with the codon usage of the housekeeping genes (average $r_s = 0.80 \pm 0.08$), but to a significantly lesser extent (p-value <0.01). Similarly, the codon usages of cancer-related gene sets correlate remarkably well with each other (average $r_s = 0.92 \pm 0.05$) but to a lesser extent with the housekeeping genes (average $r_s = 0.83 \pm 0.06$, p-value <0.01). These results suggest functionally different gene groups have significantly different codon usages.

Is the codon usage of cell-line specific genes related to the over-expression patterns of nuclear-encoded tRNAs in breast cancer cells? We plotted relative isoacceptor levels (derived from comparative tRNA measurements) versus the codon frequency of cell-line specific genes (derived

from comparative mRNA analysis) to determine whether the changes in tRNA levels favor the codon usage of these genes. No obvious correlations were observed. The absence of correlations may be explained by the low mRNA level of these genes at the global scale (Fig. 6A). The cell line-specific genes identified for this study are expressed only 7- to 15-fold above the median expression level of all genes, compared to approximately 80-fold above the median for the housekeeping genes. Among human tissues examined, a significant tRNA abundance-codon usage correlation was found only in liver. The mRNA levels of liver-specific genes approach those of house-keeping genes (200-fold above median), sufficiently high for tRNAs to adjust to their respective codon usages. Because the codon usage of line-specific genes is different from that of house-keeping genes, adjusting tRNA levels to favor expression of cell line-specific genes would be unfavorable for translation of house-keeping genes.

The codon usage of cancer-related genes, however, appears to have some correlations to relative tRNA over-expression in cancer cell lines (Fig. 6C-E). Since the tRNA over-expression pattern is similar for the nuclear-encoded tRNAs across all lines examined, we used the average tRNA over-expression for all 6 cancer lines for this analysis. Bearing in mind that tRNA over-expression cannot favor all codons of cancer-related genes because such tRNA adjustment would diminish the translational efficiency of house-keeping genes, we reasoned that a positive correlation should reveal itself only for codons that are strongly over-represented in cancer-related relative to house-keeping genes. A positive correlation between tRNA over-expression and codon usage is indeed observed for codons over-represented by 2-fold or more for the cell cycle, extracellular matrix, and transcription factor groups (Fig. 6C). Among the group of 28 cell cycle genes ranked by their average mRNA expression levels in all lines, a similar correlation can be seen when comparing the codons that are over-represented in the 9 genes with the highest mRNA expression versus the 9 genes with the lowest mRNA expression (Fig. 6D). Finally, tRNA^{Arg} isoacceptors seem to be particularly tuned to increase the translational efficiency of Arg-codons of the cell cycle, extracellular matrix and transcription factor genes (Fig. 6E).

Task 5: Identify correlations between tRNA and active protein synthesis. → **Continuing. Future studies pending.**

Based on the results from Task 4, we have selected a range of gene targets for protein expression studies. These genes are all known to be important in breast cancer development and progression. To determine the tRNA dependences of the expression levels for these genes, we have made mammalian expression plasmid constructs with these genes fused to easily assayable reporter proteins. We are still in the process of obtaining specific, isogenic cell lines that selectively overexpress specific tRNA species to better pinpoint the mechanism of altered translation for these genes.

Future studies of identifying the underlying cause and effect of tRNA overexpression may be continued using the funding from the Idea Expansion Award derived from this study (BC097967, funding recommended).

KEY RESEARCH ACCOMPLISHMENTS:

- Determined the feasibility of using tRNA expression as possible biomarkers for breast cancer cell lines.

- Determined the feasibility of using tRNA expression as possible biomarkers for tumor versus normal breast tissues.
- Determined that tRNA expression has some correlation to optimal translation of breast cancer related genes.
- Derived specific hypotheses on linking tRNA over-expression to altered translation of specific genes that are important for breast cancer development and progression. These hypotheses are still being tested.

REPORTABLE OUTCOMES:

- A full paper that includes results described under Tasks 1, 3 and 4.

Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha R. Rosner*, Tao Pan*: tRNA Over-expression in Breast Cancer and Functional Consequences. *Nucl. Acids Res.* 37, 7268-7280 (2009).

- An Idea Expansion Award (BC097967, tRNA and its activation targets as biomarkers and regulators of breast cancer) based on these studies has been recommended for funding starting Oct. 1, 2010.

CONCLUSION:

Our results show for the first time that tRNAs are good candidates as molecular biomarkers for breast cancer cells. This conclusion, initially established in breast cancer cell lines, has been firmly established in tumor and normal breast tissues. Both nuclear and mitochondrial-encoded tRNAs show distinct patterns in different breast cancer cells with varying tumorigenic characteristics.

Another important aspect of our tRNA study is to discover the possibility that altering tRNA expression in breast cancer may lead to changes in cellular behavior. We have found correlations between tRNA overexpression and the codon usage of breast cancer related genes. Altering this relationship may lead to the identification of certain tRNAs or their associated protein enzymes as potential new drug targets for breast cancer.

Biography of all publications and meeting abstracts:

Publication:

1. Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha R. Rosner*, Tao Pan*: tRNA Over-expression in Breast Cancer and Functional Consequences. *Nucl. Acids Res.* 37, 7268-7280 (2009).

Meeting abstracts:

2. Era of Hope 2008: Mariana Pavon Eternod, Susanna Gomes, Marsha Rosner, Tao Pan: "Transfer RNAs as biomarkers and regulators of gene expression in breast cancer."
3. Midwest Breast Cancer Symposium 2009: Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha Rosner, Tao Pan: "Transfer RNA Over-Expression and Functional Consequences".

4. 14th annual RNA society meeting, 2009: Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha Rosner, Tao Pan: “tRNA over-expression in breast cancer and functional consequences”.
5. 23rd International tRNA workshop, 2010: Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha Rosner, Tao Pan: “tRNA over-expression in breast cancer and functional consequences”.

List of personnel receiving pay from the research effort:

1. Tao Pan (PI).
2. Marsha Rosner (PI).
3. Mariana Pavon Eternod (Graduate student, worked on Tasks 1-5).
4. Susanna Gomes (Senior Research Tech, worked on Tasks 3 and 5).
5. Qing Dai (Research Associate, worked on Task 1).
6. Renaud Geslain (Research Associate, worked on Task 1).
7. Jeffrey Goodenbour (Graduate student, worked on Task 2).
8. Eva Eves (Senior Research Associate, worked on Tasks 3 and 5).

tRNA over-expression in breast cancer and functional consequences

Mariana Pavon-Eternod¹, Suzanna Gomes², Renaud Geslain¹, Qing Dai¹,
Marsha Rich Rosner^{2,*} and Tao Pan^{1,*}

¹Department of Biochemistry and Molecular Biology and ²Ben May Department of Cancer Research, University of Chicago, Chicago, IL 60637, USA

Received April 20, 2009; Revised September 3, 2009; Accepted September 6, 2009

ABSTRACT

Increased proliferation and elevated levels of protein synthesis are characteristics of transformed and tumor cells. Though components of the translation machinery are often misregulated in cancers, what role tRNA plays in cancer cells has not been explored. We compare genome-wide tRNA expression in cancer-derived versus non-cancer-derived breast cell lines, as well as tRNA expression in breast tumors versus normal breast tissues. In cancer-derived versus non-cancer-derived cell lines, nuclear-encoded tRNAs increase by up to 3-fold and mitochondrial-encoded tRNAs increase by up to 5-fold. In tumors versus normal breast tissues, both nuclear- and mitochondrial-encoded tRNAs increase up to 10-fold. This tRNA over-expression is selective and coordinates with the properties of cognate amino acids. Nuclear- and mitochondrial-encoded tRNAs exhibit distinct expression patterns, indicating that tRNAs can be used as biomarkers for breast cancer. We also performed association analysis for codon usage-tRNA expression for the cell lines. tRNA isoacceptor expression levels are not geared towards optimal translation of house-keeping or cell line specific genes. Instead, tRNA isoacceptor expression levels may favor the translation of cancer-related genes having regulatory roles. Our results suggest a functional consequence of tRNA over-expression in tumor cells. tRNA isoacceptor over-expression may increase the translational efficiency of genes relevant to cancer development and progression.

INTRODUCTION

Misregulation of components of the translation machinery can lead to malignant transformation (1–3). In particular,

deregulation of RNA polymerase III (polIII) and its products has been observed in a wide range of transformed cells, including ovarian and breast carcinomas (4–8). Over-expressed polIII RNAs include transfer RNAs (tRNA) which are directly involved in translation. A recent report demonstrates that such deregulation is not only a by-product but can drive transformation (9). In that study, modest over-expression of initiator tRNA drives cell proliferation and results in malignant transformation of immortalized mouse fibroblasts. As for other tRNAs, their expression or activity can control translational elongation because they read mRNA codons embedded in the coding sequence of each gene at various speeds (10). tRNA levels may therefore be relevant for the translation of key genes required for the tumorigenic process, and may play a central role in cancer development and progression.

A major challenge in cancer research is the identification of accurate molecular signatures for diagnosis and treatment (11–13). Though diverse genomic and proteomic approaches have been applied to this problem, no genome-wide studies of tRNA expression have been carried out in the context of breast cancer. Approximately 450 tRNA genes comprised of over 270 different sequences have been annotated in the human genome, with 22 additional genes present in human mitochondrial DNA (14,15). Human tRNAs are grouped into 49 isoacceptor families to decode all codons specified by the genetic code (16).

Here we describe the comparative analysis of tRNA levels for non-cancer-derived breast epithelial cell lines and breast cancer cell lines, as well as breast tumor and normal breast tissue samples. We apply a previously described tRNA microarray used to study tissue-specific expression of human tRNAs (17). We show that tRNA levels are significantly elevated for both tumor cell lines and tumor tissues. We also develop a chemical ligation method and apply it in a microarray format for the detection and comparison of all tRNA isoacceptors in the breast cell lines. We perform association analysis of

*To whom correspondence should be addressed. Tel: 773-702-0380; Fax: 773-702-4476; Email: m-rosner@uchicago.edu
Correspondence may also be addressed to Tao Pan. Tel: 773-702-4179; Fax: 773-702-0439; Email: taopan@uchicago.edu

tRNA isoacceptors to the codon usage of the house-keeping genes, the cell line-specific genes, and cancer-related genes. We find that the cell line-specific genes and cancer-related genes have similar codon usages that are distinct from house-keeping genes. Over-expressed tRNA isoacceptors do not favor the codon usage of the house-keeping genes or the cell line-specific genes. Instead, a subset of the over-expressed tRNAs favors over-represented codons among the cancer-related genes.

MATERIALS AND METHODS

Cell lines and RNA isolation

All cell lines were purchased from American Type Culture Collection (ATCC, Manassas, VA) and maintained according to ATCC recommendations. MCF7 cells were cultured in DMEM medium (Invitrogen, 11965-092) supplemented with 10% FBS and 1% penicillin/streptomycin (P/S). MCF10A cells were cultured in 1:1 (+I factor, Vectro)-DMEM/F12 (Invitrogen, 11330-032) supplemented with 10% FBS, 1% P/S, 5 µg/ml insulin, 10 ng/ml EGF and 0.5 µg/ml hydrocortisone. All other cell lines were cultured in RPMI 1640 1× medium (Mediatech, 10-040-CV) supplemented with 10% FBS and 1% P/S.

Total RNA for each cell line was obtained at 80–90% confluency using the mirVanaTM miRNA Isolation Kit (Ambion, AM1560). This procedure isolates RNA species as short as 15 nucleotides and is therefore not biased against tRNA. Total RNA quality was verified on an agarose gel.

Normal breast and breast tumor total RNA

Two samples of human breast total RNA were purchased from Ambion (FirstChoice[®] Human Breast Total RNA, AM6952, lot numbers 0808001 and 0812006). An additional sample of human breast total RNA was purchased from Stratagene (MVP Total RNA Human Breast, 540045, lot number 0380023). Both providers certify small RNA content in total RNA samples and provide relevant donor information (Table S1).

Breast tumor samples were obtained from the Human Tissue Resource Center (HTRC) at the University of Chicago (Table S1). Pathology reports were provided for all samples. All breast tumor samples were obtained frozen and embedded in OCT (optimal cutting temperature compound). Following removal of the surrounding OCT and tissue pulverization, total RNA was isolated using TRIzol reagent (Invitrogen, 15596-018). Total RNA quality was checked on agarose gels.

Chemical ligation substrates and templates

All oligodeoxynucleotides were synthesized on an Expedite Nucleic Acid Synthesis System. 5'-Iodo-oligodeoxyribonucleotides were prepared by reacting the 5'-OH of the oligodeoxyribonucleotide which remains on a solid support with methyltriphenoxyposphonium iodide, followed by standard deprotection. 3'-Phosphorothioate oligodeoxyribonucleotides were prepared using a 3'-phosphate CPG column

(Glen Research). After the first coupling with the phosphoramidite, the phosphorus was sulfurized with Sulfurizing Reagent II (Glen research) instead of oxidized with iodine. Synthesis and deprotection proceeded as with standard DNA synthesis.

30-mer RNA template oligonucleotides were designed to contain minimal secondary structure when the nucleotide at the 15th position is U (18). Four chemically synthesized 30-mer RNAs have the same sequence at all positions except the 16th nucleotide (I, U, G, C). Chemically synthesized 30-mer RNAs were purchased from Dharmacon Research, de-protected, and purified according to manufacturer's protocols. When needed, RNA oligonucleotides were purified on denaturing polyacrylamide gels containing 7 M urea. Yeast tRNA^{Phe} was purchased from Sigma-Aldrich (R4018) and used without further purification.

Chemical ligation reactions

For 30-mer model RNAs, reactions were carried out with 0.2 µM RNA template, 0.7 µM 5'-iodo probe, and 0.5 µM 5'-³²P-labeled, 3'-phosphorothioate probe in 50 mM pH 7.5 Tris-HCl containing 1.5 mM EDTA, 100 mM NaCl, and 1 mM TCEP at 30°C for 1 h. tRNA^{Phe} reactions were carried out with 0.05–0.15 µM RNA template, 0.26 µM 5'-iodo substrate, and 0.065 µM 5'-³²P-labeled, 3'-phosphorothioate probe in 20 mM pH 7.5 Tris-HCl containing 1.5 mM EDTA, 200 mM NaCl, and 1 mM TCEP at 60°C for 24 h. Total human RNA reactions were carried out with 1–6 µg total RNA, 0.26 µM each 5'-iodo substrate, and 0.065 µM each Cy3-labeled, 3'-phosphorothioate probe in 20 µl 20 mM pH 7.5 Tris-HCl containing 1.5 mM EDTA, 200 mM NaCl, and 1 mM TCEP. Reactions were incubated at 60°C for 24 h. Following ligation, all reaction mixtures were treated with 0.1 U/µl RNase H (Epicentre Biotechnologies, R0601K). The reactions were quenched with an equal volume of 8 M urea/100 mM EDTA, boiled for 2 min and rapidly cooled on ice prior to loading on 10–15% denaturing polyacrylamide gels containing 7 M urea. Radiolabeled ligation products were visualized on a FUJIX BAS1000 Phosphorimager and quantified using the accompanying software. Fluorescent-labeled ligation products were visualized on a PharosFX Molecular Imager and quantified using QuantityOne software.

Standard tRNA microarrays

The standard tRNA microarray experiment consists of four steps starting from total RNA: (i) deacylation to remove remaining amino acids attached to the tRNA, (ii) selective Cy3/Cy5 labeling of tRNA, (iii) hybridization on commercially printed arrays and (iv) data analysis.

- (i) *Deacylation*: 0.25 µg/µl total RNA premixed with three tRNA standards (*Escherichia coli* tRNA^{Lys}, *E. coli* tRNA^{Tyr}, and yeast tRNA^{Phe}) at 0.2 µM each was incubated in 100 mM Tris-HCl (pH 9.0) at 37°C for 30 min. The solution was neutralized by the addition of an equal volume of 100 mM Na-acetate/acetic acid (pH 4.8) plus 100 mM NaCl, followed by ethanol precipitation. Deacylated total

RNA was dissolved in water, and its integrity verified by agarose gel electrophoresis.

- (ii) *Cy3/Cy5 labeling*: tRNA in the total RNA mixture was selectively labeled with either Cy3 or Cy5 fluorophore using an enzymatic ligation method previously described. The ligation reaction relies on an 8-base pair RNA:DNA hybrid helix containing a Cy3 or Cy5 fluorophore pre-attached to the loop and an overhang complementary to the universally conserved 3'CCA nucleotides present in all tRNAs. The ligation reaction was carried out overnight (~16 h) at 16°C with 1 U/μl T4 DNA ligase (USB Corporation, 70042X) and 9 μM labeling oligonucleotide.
- (iii) *Hybridization*: hybridization was performed in a GeneMachines Hyb4 at 60°C overnight (16 h) with 1–2.5 μg each of Cy3- or Cy5-labeled total RNA as previously described. Multiple arrays were run using the MCF10A reference sample labeled with either Cy3 or Cy5.
- (iv) *Data analysis*: arrays were scanned using a GenePix 4000b scanner (Axon Instruments). For both Cy3 and Cy5 wavelengths, PMT gain was set at 600 and power at 100%. These settings were chosen to provide optimal signal without saturation. Array images were generated and analyzed using GenePix 6.0 software. GenePix adaptive circle spot segmentation was used for image analysis. To account for differences in labeling and hybridization efficiencies, the following normalization procedure was used: first, the median Cy5/Cy3 ratio at each probe spot was divided by the corresponding MCF10A Cy5/Cy3 reference ratio. Second, the obtained value was divided by the averaged Cy5/Cy3 ratio value for the three tRNA standards added in (i) (*E. coli* tRNA^{Lys}, *E. coli* tRNA^{Tyr} and yeast tRNA^{Phe}).

The standard microarray method described above uses a very standard, two-color microarray method extensively used for mRNA analysis. Compared to mRNA arrays, the main difference is the previously developed method for the labeling of tRNA in total RNA samples. The reproducibility of the tRNA array method and result validation by Northern blots have been extensively described in previously published papers (17,19–22).

All microarray data has been deposited in GEOarchive (accession numbers GSE17945, GPL9143).

Chemical ligation tRNA Microarrays

These microarray experiments consist of four steps starting from total RNA: (i) chemical ligation to detect target tRNA isoacceptors, (ii) purification of chemical ligation products, (iii) hybridization, and (iv) data analysis.

- (i) *Chemical ligation*: 5 μg total RNA was premixed with 4 pmol each of three tRNA standards (*E. coli* tRNA^{Lys}, *E. coli* tRNA^{Tyr}, and yeast tRNA^{Phe}). Reactions were carried out with 5 μg total RNA, 0.26 μM of each 5'-iodo probe, and 0.065 μM of each Cy3- or Cy5-labeled, 3'-phosphorothioate

probe in 50 μl 15 mM pH 7.5 Tris-HCl containing 1 mM EDTA, 200 mM NaCl, and 1 mM TCEP. Reactions were incubated at 60°C for 24 h, and then treated with 0.1 U/μl RNase H. The reactions were quenched with an equal volume of 8 M urea/100 mM EDTA, then boiled for 2 min and rapidly cooled on ice prior to loading on 10% denaturing polyacrylamide gels containing 7 M urea.

- (ii) *Purification*: chemical ligation products were purified on 10% denaturing polyacrylamide gels containing 7 M urea. Ligation products were visualized on a PharoFX Molecular Imager, and the corresponding band cut out for purification.
- (iii) *Hybridization*.
- (iv) *Data analysis*: these are identical to those for the standard arrays described above.

Cell proliferation

Proliferation rates were measured over four days using the Promega CellTiter Blue metabolic assay. Cells were plated at low (1500 cells/well) and high (5000 cells/well) cell density in 100 μl medium in 96 well-plates. A total of 20 μl CellTiter Blue reagent was added to each well at the indicated timepoints. Fluorescence (560ex/590em) was measured after 2 h incubation at 37°C. Fluorescence is proportional to the number of cells present and is normalized to fluorescence at day 0 to account for differences in metabolic efficiency across cell lines. Doubling times were obtained from the equation $N/N_0 = e^{(kt)}$. All measurements were performed in triplicate.

RESULTS

Over-expression of tRNAs in breast cancer cell lines

To our knowledge, no systematic genome-wide studies of tRNA expression have been carried out in cancer. To explore the potential of tRNAs to define breast cancer signatures, we used tRNA microarrays to generate comparative tRNA profiles for three non-cancer-derived breast epithelial cell lines (MCF10A, 184 A1, 184 B5) and six breast cancer cell lines (BT-474, HCC70, MCF7, MDA-MB-231, MDA-MB-436, ZR-75-1). The breast cancer cell lines cover a range of physiological and molecular properties (Table S2). Importantly, genome-wide mRNA expression data are available for these cell lines (23) to allow tRNA expression and codon usage association analysis.

We used a previously described array method in the studies of tissue-specific expression of human tRNAs (17). Arrays used for this study contain 40 probes for human nuclear-encoded tRNAs and 22 probes for human mitochondrial-encoded tRNAs; each probe is repeated 21 times. The array also includes 42 probes for bacterial tRNAs and 275 probes complementary to short regions in yeast and human rRNA as hybridization and specificity controls. After isolation of total RNA from cell culture, tRNAs in the sample are labeled by selective ligation to a fluorophore-containing oligonucleotide.

The labeled samples are hybridized directly onto the array (Figure 1A). The MCF10A (a non-cancer-derived breast epithelial cell line) sample was included as a reference in all array hybridizations to correct for variations in fluorescence labeling and array manufacturing. Since tRNA constitutes up to 30% of the total RNA, the method requires no amplification and only 1–2 µg of total RNA per array. The specificity of the array is illustrated by the fluorescence signals derived from human tRNA probes compared to the non-human tRNA probes (Figure 1B and C).

Significant differences are observed in the expression levels of tRNA among non-cancer-derived and cancer-derived breast cell lines (Figure 2A and C). The global level of nuclear- and mitochondrial-encoded tRNAs can be approximated separately by the median and mean sample-to-MCF10A ratio. For nuclear-encoded tRNAs, this ratio is 0.7–0.8 for the other non-cancer-derived cell lines and 2–3 for the cancer-derived cell lines. These differences are even more pronounced for mitochondrial-encoded tRNA levels: the sample-to-MCF10A ratio for non-cancer-derived cell lines is ~0.75, but as much as 5 for the cancer-derived cell lines.

The over-expression of tRNA in cancer-derived relative to non-cancer-derived cell lines is also selective: certain individual tRNAs are more strongly over-expressed than others (Figure 2C). Variations in the relative expression of nuclear-encoded tRNAs carrying certain amino acid types are readily observed across cancer-derived cell lines. For example, nuclear-encoded tRNAs carrying polar amino acids (e.g. Ser, Thr, and Tyr) are up-regulated 3- to 4-fold in breast cancer cell lines relative to MCF10A, while nuclear-encoded tRNAs carrying small amino acids (Ala, Cys, Gly, Pro) are up-regulated only 1.5- to 2-fold. These differences can be observed more clearly when the tRNA expression is normalized to the median value for either nuclear- or mitochondrial-encoded tRNAs within each cell line (Figure 2C and D). Differences in the relative expression of tRNA isoacceptors also become more apparent after normalization. For example, tRNA^{Lys(UUU)} is expressed strongly above the median, while tRNA^{Lys(CUU)} is expressed below the median. All breast cancer cell lines exhibit a remarkable similarity in relative tRNA expression patterns: tRNA^{Arg(UCU)}, tRNA^{Arg(CCU)}, tRNA^{Thr(CGU)}, tRNA^{Ser(CGA)}, and tRNA^{Tyr(GTA)} are among the most over-expressed tRNAs, while tRNA^{His(GTG)}, tRNA^{Phe(GAA)}, and tRNA^{Met(CAT)} are among the least over-expressed tRNAs. Selective up-regulation of tRNA levels is also observed for mitochondrial-encoded tRNAs: certain tRNAs are expressed up to 2-fold above median, and others are expressed 4-fold below median. While nuclear-encoded tRNA expression patterns are remarkably similar across cell lines, mitochondrial-encoded tRNA expression patterns exhibit greater variations.

tRNA over-expression determined by the microarray method does not necessarily mean that the amount of total tRNA in the cell is increased by the same extent. The median increase for each tRNA species is weighted equally in the microarray measurements regardless of their

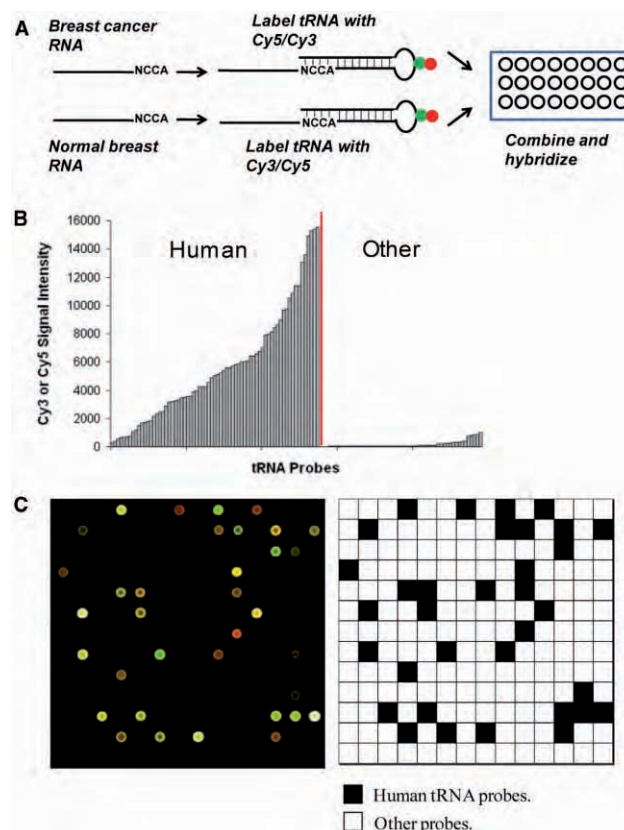


Figure 1. tRNA overexpression in breast cancer cells. (A) Fluorescence labeling scheme of tRNA in total RNA samples. (B) Fluorescence intensity of a human total RNA sample hybridized to the array. “Human” indicates signals for the designated human tRNA probes, “other” signals for bacterial tRNA probes (no signal expected). (C) One of the 48 blocks from a standard tRNA microarray hybridized with MCF10A (Cy5) and MCF7 (Cy3). The schematic indicates the position of human tRNA probes (black) and other probes (white, include ribosomal RNA and bacterial tRNA probes).

absolute abundance, whereas abundant tRNAs are over-represented in the total tRNA. Quantitative comparison of total RNA samples by gel electrophoresis shows that the cancer-derived cells still have globally elevated tRNA levels (1.2- to 1.4-fold, Figure 2B), although this increase in total tRNA is smaller compared to the microarray measurements (2- to 3-fold over the median increase). This result highlights the over-expression of low abundant tRNAs in breast cancer cells. Because each tRNA species functions separately in decoding, changes in low-abundant tRNA levels as measured by microarrays should have greater functional significance.

We also determined that tRNA over-expression in breast cancer cell lines does not simply reflect an increase in proliferation rate (Table S3). Under the culture conditions used in this study, doubling times range from 21 to 47 h for the non-cancer-derived cell lines, versus 15–35 h for the cancer-derived cell lines. Thus, the proliferation of all three non-cancer-derived cell lines with low tRNA content is comparable to that of breast cancer cell lines with much higher tRNA content. Doubling times do not correlate with either nuclear or mitochondrial global tRNA levels (not shown).

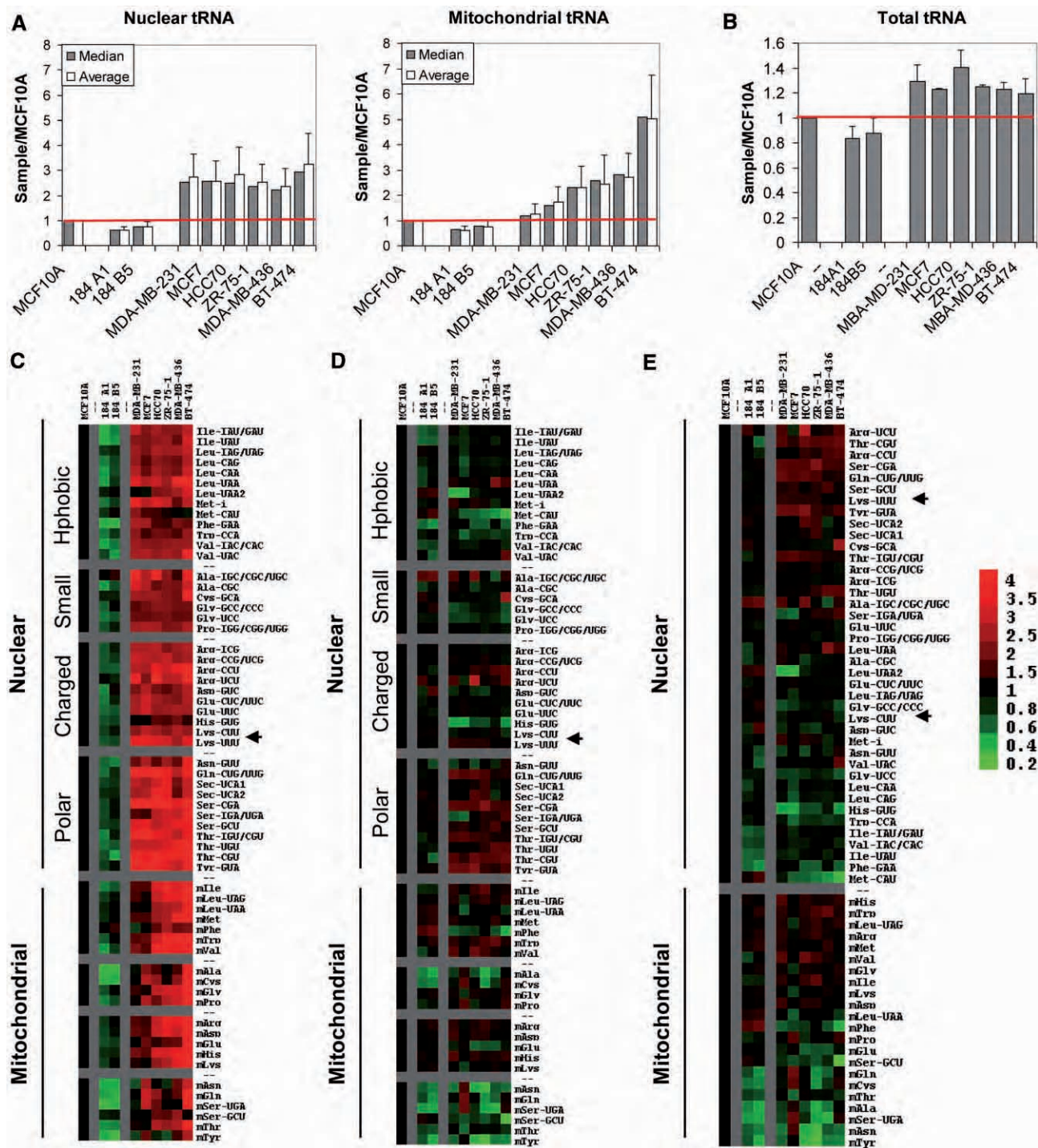


Figure 2. Relative abundance of nuclear and mitochondrial encoded tRNAs in breast cancer cells. Data is shown for three breast epithelial cell lines (MCF10A, 184 A1, 184 B5) and six breast cancer cell lines (MDA-MB-231, MCF7, HCC70, ZR-75-1, MDA-MB-436, BT-474), all relative to MCF10A. (A) Mean and median values of the nuclear (left) and mitochondrial (right) encoded tRNAs. (B) Total tRNA quantified by agarose gel electrophoresis for all samples. All RNAs are detected by ethidium bromide staining and quantified using a PharoFX Molecular Imager. Fraction of total tRNA was measured relative to the non-tRNA bands in the same lane, and then normalized to that of MCF10A. (C) Expression of nuclear and mitochondrial encoded tRNAs shown as TreeView image. All values are relative to MCF10A. Green indicates a decreased level of expression; red indicates an increased level of expression relative to MCF10A. Data are grouped according to amino acid type. (D) Expression of nuclear and mitochondrial encoded tRNAs normalized to median shown as TreeView image. All values are relative to MCF10A and normalized to the median value for each sample. Green indicates a decreased level of expression; red indicates an increased level of expression relative to median. Data are grouped according to corresponding amino acid type. (E) Same as (D), data are grouped from high to low expression.

tRNA over-expression in patient-derived breast tumor samples

In all breast cancer cell lines examined, we observe global tRNA over-expression and differential tRNA isoacceptor expression. To generalize our results in cell lines to breast tumor tissue, we measured tRNA expression patterns in nine patient-derived breast tumor samples and three normal breast tissue samples (Table S3). The nine breast tumor samples were selected from the three major subtypes of breast cancer: luminal (ER+, HER2-), basal (ER-, HER2-), and HER2+. For consistency of data analysis, all samples were run using MCF10A as a reference in array experiments.

As in the breast cancer cell lines, we observe significant differences in global tRNA expression levels among breast tumor and normal breast samples (Figure 3A and B). For nuclear-encoded tRNAs, the mean sample-to-MCF10A ratio is 0.2 to 0.5 for normal breast tissue samples, versus 2 to 4 for breast tumor samples. This result indicates that global nuclear-encoded tRNA levels are increased up to 20-fold in breast tumors relative to normal breast. For mitochondrial-encoded tRNAs, the mean sample-to-MCF10A ratio is 0.4 to 1 for normal breast tissue samples, versus 1.2 to 5 for breast tumor samples. This result indicates that global mitochondrial-encoded tRNA levels are increased up to 13-fold in breast tumors relative to normal breast.

Consistent with our results on cell lines, tRNA over-expression in breast tumor samples is also selective: certain individual tRNAs are more strongly over-expressed than others (Figure 3C). Among the top 10 over-expressed, nuclear-encoded tRNAs, six tRNA species overlap between cell lines and tumor tissues: tRNA^{Arg(CCU)}, tRNA^{Thr(CGU)}, tRNA^{Leu(UAA)}, tRNA^{Tyr(GUA)}, tRNA^{Ser(GCU)} and tRNA^{Arg(UCU)}. We readily observe variations in the relative expression of both nuclear- and mitochondrial-encoded tRNAs. These variations correlate with both the cognate amino acid type and tRNA isoacceptor identity. For example, tRNAs carrying polar (Ser, Thr, Tyr) and charged (Arg, Glu, Lys) amino acids are upregulated 2- to 8-fold, while tRNAs carrying small amino acids (Ala, Cys, Gly, Pro) are upregulated only 1.5- to 2-fold. We also observe variations in the relative expression of tRNA isoacceptors. For example, tRNA^{Arg(CCU)} and tRNA^{Lys(UUU)} are over-expressed to a greater extent than tRNA^{Arg(ICG)} and tRNA^{Lys(CUU)}, respectively. These variations become more clear when tRNA expression levels are normalized to the median value within each sample (Figure 3D). Interestingly, the tRNA expression patterns in the three major sub-types of breast cancer are very similar, suggesting that tRNA over-expression is a general consequence for all breast tumors.

We conclude that elevated tRNA levels are a hallmark of breast cancer. Our results suggest that tRNAs can be used as biomarkers for breast cancer, although it is best suited to determine cancer versus normal breast cells. Furthermore, tRNA expression measured by microarrays more accurately reflects the “functional abundance” of individual tRNA species.

Detecting single-nucleotide differences between tRNA isoacceptors

tRNA over-expression across breast cancer cells suggests that elevated tRNA levels may be of functional significance. For instance, differential expression of tRNA isoacceptors may enhance translation of specific cancer-related genes via their codon usages. In order to perform tRNA expression and codon usage association analysis, more comprehensive profiling of all tRNA isoacceptors is desired. Because the microarray method used above works by hybridization alone, it can only distinguish between tRNAs that differ by at least eight nucleotides (17). However, several human tRNA isoacceptors differ by less than eight nucleotides (Table S4). For example, the proline isoacceptors tRNA^{Pro(IGG)}, tRNA^{Pro(CGG)} and tRNA^{Pro(UGG)} differ by only one nucleotide; they all hybridize to the same probe on the array above.

We therefore modified a previously described non-enzymatic, nucleic acid-templated, phosphorothioate-iodide autoligation chemistry (24,25) to detect single-base differences between tRNA isoacceptors. This method involves the reaction of a phosphorothioate at the 3'-end of one DNA oligo with an iodide leaving group on the 5'-end of the other oligo (Figure 4A).

In order to determine whether the above set-up allows single base discrimination, we performed chemical ligation using a series of model 30-mer RNAs that differ by a single nucleotide at the 16th position (U₁₅I₁₆, U₁₅C₁₆, U₁₅U₁₆, U₁₅G₁₆) (Figure 4B). While the 5'-iodo substrate was always fully complementary to the RNA template, the base at the 3'-end of the 3'-phosphorothioate substrate was varied (X = A, G, C. X = T was not examined because no tRNA isoacceptor contains A34). For example, when ligations are carried out in defined mixtures of U₁₅I₁₆ and U₁₅C₁₆ templates, the amount of product obtained with the X = C 3'-phosphorothioate substrate increases linearly as the amount of U₁₅I₁₆ template increases. Conversely, the amount of product obtained with the X = G 3'-phosphorothioate substrate decreases linearly as the amount of U₁₅C₁₆ template decreases (Figure 4C). Similar results are obtained with other template-substrate combinations.

In order to confirm that the chemical ligation method works with tRNA, we first applied this method to purified yeast tRNA^{Phe} which has G34 (Figure 4D). The ligation product obtained with mismatched 3'-phosphorothioate substrates (X = A, G) is less than 10% the ligation product obtained with the matched 3'-phosphorothioate substrate (X = C). This confirms that phosphorothioate-iodide autoligation chemistry successfully discriminates in favor of complementary nucleotides at the ligation junction on full-length tRNA templates. Finally, we examined whether human tRNAs could be detected and quantified in a total RNA sample using this autoligation chemistry (Figure 4E). The amount of ligation product obtained with 5'-iodo and 3'-phosphorothioate substrates complementary to tRNA^{Pro(UGG)}, tRNA^{Pro(IGG)}, and tRNA^{Arg(UCG)} increases linearly with the amount of total human RNA, and therefore reflects the amount of tRNA present in the reaction.

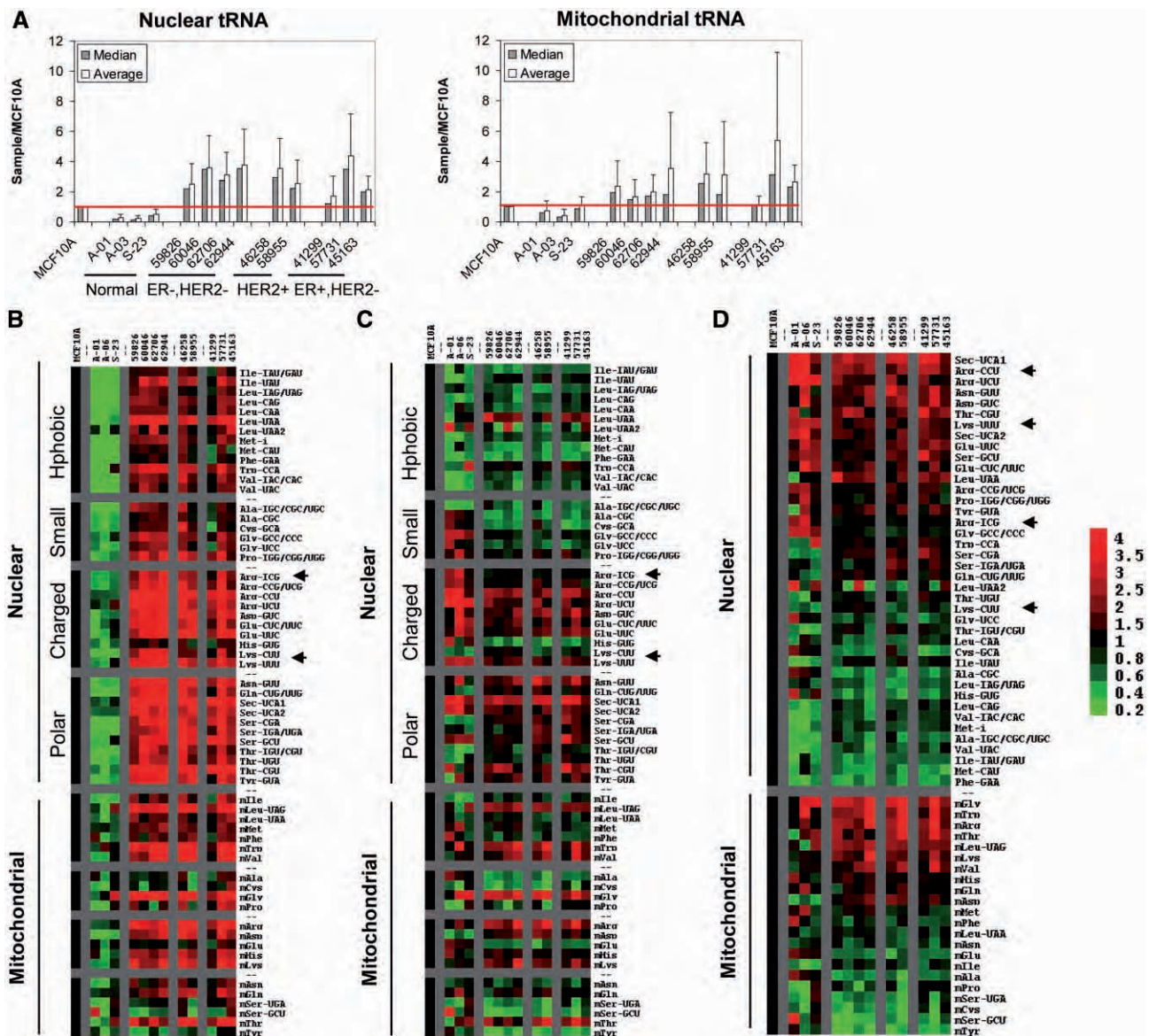


Figure 3. tRNA over-expression in breast cancer *in vivo*. Data is shown for three normal breast tissue samples (A-01, A-03, S-23), four ER[−]/HER2[−] tumor samples (59826, 60046, 62706, 62944), 2 ER[−]/HER2⁺ tumor samples (46258, 58955), and three ER⁺/HER2[−] tumor samples (41299, 57731, 45163). All data are relative to MCF10A. (A) Mean and median values of the nuclear (left) and mitochondrial (right) encoded tRNAs. (B) Expression of nuclear and mitochondrial encoded tRNAs shown as TreeView image. All values are relative to MCF10A. Green indicates a decreased level of expression; red indicates an increased level of expression relative to MCF10A. Data are grouped according to amino acid type. (C) Expression of nuclear and mitochondrial encoded tRNAs normalized to median shown as TreeView image. All values are relative to MCF10A and normalized to the median value for each sample. Green indicates a decreased level of expression; red indicates an increased level of expression relative to median. Data are grouped according to corresponding amino acid type. (D) Same as (C), data are grouped from high to low expression.

We then applied this chemical ligation method to simultaneously detect and compare selected tRNA isoacceptors in total RNA samples on a microarray platform (Figure 5A). 5'-Iodo and 3'-phosphorothioate oligonucleotide substrates were designed against all tRNA isoacceptors in pairs differing by less than eight nucleotides (Table S4). For each pair, two 3'-phosphorothioate substrates were synthesized with the first favoring one and the second favoring the other isoacceptor. The ligation junction was placed between U33 of the tRNA and the first position of the anticodon.

Our chemical ligation microarray includes 10 sense tRNA probes complementary to expected ligation products, 6 sense tRNA probes not complementary to expected ligation products, and 79 antisense tRNA probes complementary to human tRNAs; each probe is repeated eight times. We established the specificity of the human tRNA array by examining the cross-hybridization of chemical ligation products to complementary and control probes present on the same array (Figure 5B).

Chemical ligation results also show overall elevated tRNA levels in breast cancer cell lines relative to breast

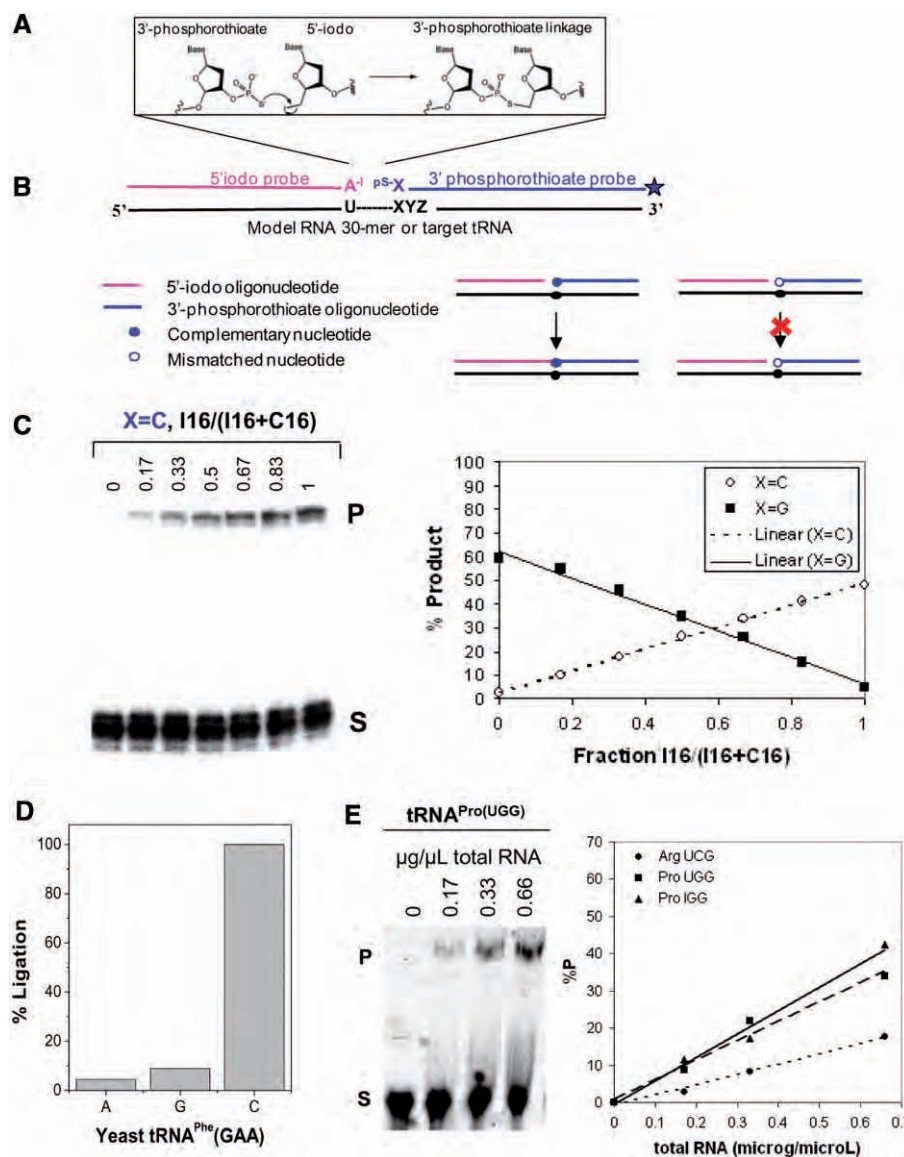


Figure 4. The chemical ligation strategy for detection of single base differences between tRNA isoacceptors. (A) Oligonucleotide reactants complementary to target sequence and mechanism of chemical auto-ligation. For tRNA templates, the ligation junction is between a conserved U and the first position of the anticodon. (B) Overview of the strategy. Chemical ligation proceeds efficiently when the nucleotide at the ligation junction is complementary to the template, but poorly when the nucleotide is mismatched. (C) Ligation yield is proportional to the amount of RNA template. Left: ligation reactions using a defined mixture of X = I and C 30-mer RNA templates, and X = C 3'-phosphorothioate substrate. Right: the percent product from gel analysis is plotted against the fraction of X = I 30-mer RNA template. (D) Relative ligation efficiency with matched and mismatched 3'-phosphorothioate substrates on yeast tRNA^{Phe}(GAA) template. (E) Percent ligation product as a function of total RNA. Left: ligation reaction using varying amounts of human total RNA and substrates for tRNA^{Pro}(UGG). Right: the percent product from gel analysis is plotted versus the amount of human total RNA.

epithelial cell lines (Figure 5C and D). Again, the up-regulation of tRNA levels in cancer-derived relative to non-cancer-derived cells is selective. For instance, tRNA^{Val}(CAC) is decreased 2- to 3-fold in the breast cancer cell lines relative to MCF10A while tRNA^{Val}(IAC) remains unchanged. Similar results are observed for tRNA^{Ser}(UGA)/tRNA^{Ser}(IGA) and tRNA^{Ala}(CGC)/tRNA^{Ala}(UGC).

In conclusion, we successfully developed a microarray method that has a single nucleotide resolution for tRNA isoacceptors. Its application provides comprehensive

profiles of all tRNA isoacceptors for breast cancer cell lines.

Relative tRNA abundance and codon usage associations

The differential expression of tRNA isoacceptors can be used to regulate translational efficiency via the codon usage of specific genes. In prokaryotes and fungi, differences in the abundance of tRNA isoacceptors are correlated with codon preferences of genes encoding highly expressed proteins and impact the synthesis of

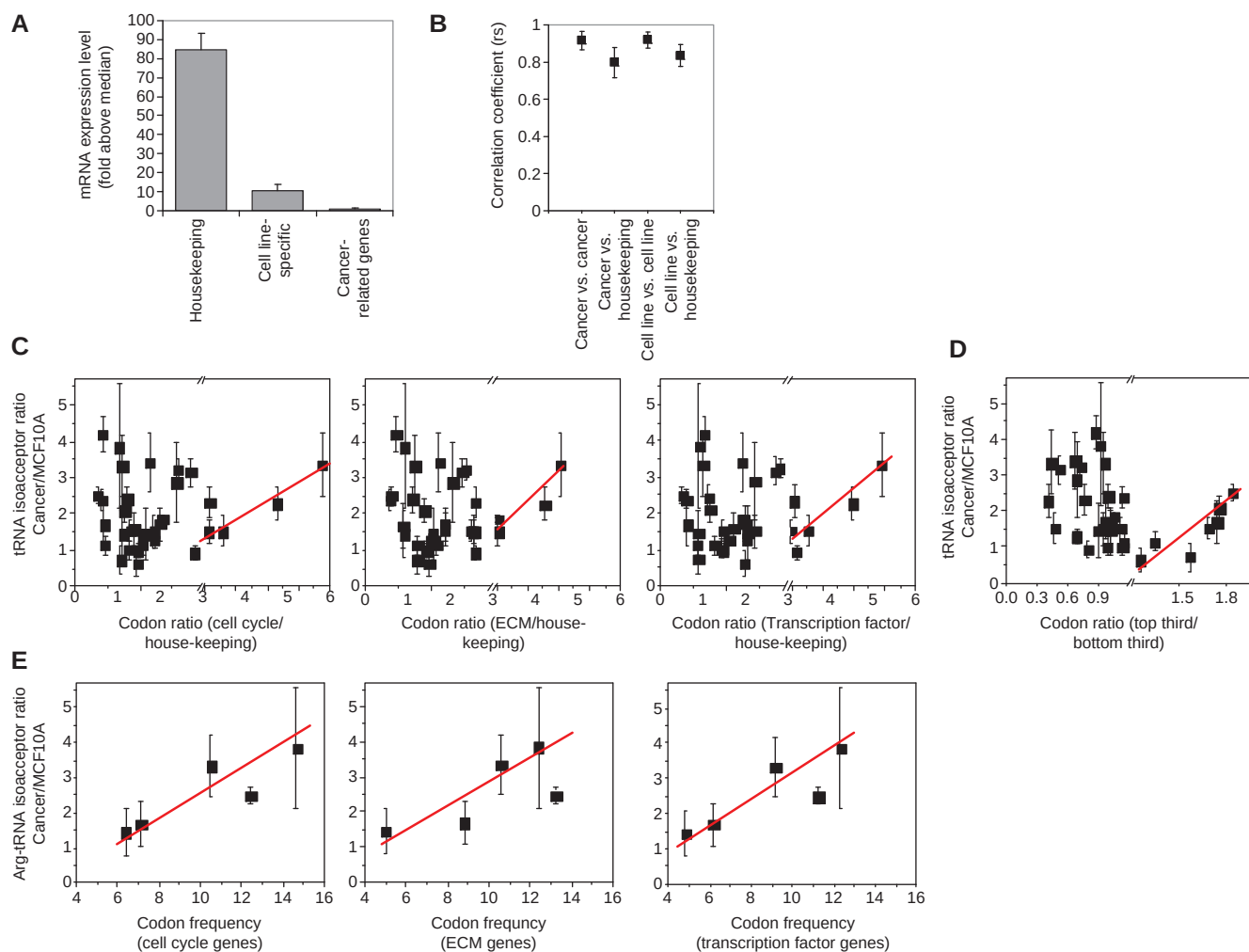


Figure 6. Analysis of codon usage versus tRNA over-expression. **(A)** Three gene groups are relevant in this analysis (mRNA expression level is derived from signals on Affymetrix mRNA arrays). tRNA expression or over-expression when comparing cancer and non-cancer cells are unlikely to positively correlate to the codon usage of all three groups. **(B)** Codon usage comparison between cell-line specific genes, cancer-related genes and house-keeping genes. The degree of association was assessed using Spearman's rho (r_s). Mean r_s values are plotted for the following pairs: cell line versus cell line, cell line versus house-keeping, cancer-related versus cancer-related, and cancer-related versus house-keeping. Error bars indicate standard deviation from the mean. **(C)** Association of relative tRNA levels to ratios of codon usage between cancer-related and house-keeping genes. As discussed in the text, a positive association is only expected for the codons that are over-represented in the cancer-related genes ($x > 2$). **(D)** Association of relative tRNA levels to ratios of codon usage between the top-third (nine genes) and bottom-third (nine genes) transcribed cell cycle genes. Again, a positive association is only expected for the codons that are over-represented in the top third genes ($x > 1.2$). **(E)** Association of relative arginine tRNA isoacceptor levels to the arginine codon frequency of cancer-related genes.

protein kinases, and transcription factors/regulators (Table S6). To identify house-keeping genes, we selected the 30 most highly expressed house-keeping genes as defined in a previous report (30). This list includes ribosomal proteins, actin, ubiquitin, and others (Table S7). In all cases, gene sequences were compiled and analyzed for codon content at <http://www.bioinformatics.org/sms2/>.

For each gene set, we compiled gene sequences and analyzed them for codon usage (expressed as number per one thousand codons, www.bioinformatics.org/sms2) (Tables S8 and S9). Because certain tRNAs read more than one codon, we converted the obtained codon usage into tRNA-based codon usage. For example,

tRNA^{Arg(ICG)} reads both CGU and CGC. Its tRNA-based codon usage is therefore equal to the sum of the CGU and CGC codon usages. For simplicity, we refer to tRNA-based codon usage as codon usage throughout our analysis.

We first compared the codon usages of each gene set in the three gene groups (cell line-specific, cancer-related, and house-keeping) against each other (Figure 6B). Though there is limited overlap across cell line-specific gene sets, their codon usage was remarkably similar (average $r_s = 0.92 \pm 0.04$). The codon usage of cell-line specific gene sets also correlates with the codon usage of the house-keeping genes (average $r_s = 0.80 \pm 0.08$), but to a significantly lesser extent ($P < 0.01$). Similarly, the codon

usages of cancer-related gene sets correlate remarkably well with each other (average $r_s = 0.92 \pm 0.05$) but to a lesser extent with the house-keeping genes (average $r_s = 0.83 \pm 0.06$, $P < 0.01$). These results suggest different gene groups have significantly different codon usages.

Is the codon usage of cell-line specific genes related to the over-expression patterns of nuclear-encoded tRNAs in breast cancer cells? We plotted relative isoacceptor levels (derived from comparative tRNA measurements) versus the codon frequency of cell-line specific genes (derived from comparative mRNA analysis) to determine whether the changes in tRNA levels favor the codon usage of these genes. No obvious associations were observed (not shown). The absence of associations may be explained by the low mRNA level of these genes at the global scale (Figure 6A). The cell line-specific genes identified for this study are expressed only 7- to 15-fold above the median expression level of all genes, compared to ~80-fold above median for the house-keeping genes. Among human tissues examined, a significant tRNA abundance-codon usage association was found only in liver (17). The mRNA levels of liver-specific genes approach those of house-keeping genes (200-fold above median), sufficiently high for tRNAs to adjust to their respective codon usages. Because the codon usage of line-specific genes is different from that of house-keeping genes, adjusting tRNA levels to favor expression of cell line-specific genes would be unfavorable for translation of house-keeping genes.

The codon usage of cancer-related genes, however, appears to have a positive association with the relative tRNA over-expression in breast cancer cell lines (Figure 6C–E). Since the tRNA over-expression pattern is similar for the nuclear-encoded tRNAs across all lines examined, we used the average tRNA over-expression for all six cancer lines for this analysis. Bearing in mind that tRNA over-expression cannot favor all codons of cancer-related genes because such tRNA adjustment would diminish the translational efficiency of house-keeping genes, we reasoned that a positive association should reveal itself only for codons that are strongly over-represented in cancer-related relative to house-keeping genes. A positive association between tRNA over-expression and codon usage is indeed observed for codons over-represented by 2-fold or more for the cell cycle, extracellular matrix, and transcription factor groups (Figure 6C). Among the group of 28 cell cycle genes ranked by their average mRNA expression levels in all lines, a similar relationship can be seen when comparing the codons that are over-represented in the nine genes with the highest mRNA expression versus the nine genes with the lowest mRNA expression (Figure 6D). Finally, tRNA^{Arg} isoacceptors seem to be particularly tuned to increase the translational efficiency of Arg-codons of the cell cycle, extracellular matrix and transcription factor genes (Figure 6E). Consistent with this observation, two of the five tRNA^{Arg} isoacceptors are among the highest over-expressed tRNAs in cancer cells (Figure 2E).

DISCUSSION

tRNA over-expression in breast cancer cells

Consistent with previous reports on individual tRNAs, our genome-wide results show that elevated tRNA levels are characteristic of all breast cancer cells analyzed. To our knowledge, however, our study is the first to analyze expression levels for all tRNAs in cancerous versus non-cancerous cells. Our results reveal an unexpected selectivity in the over-expression of individual tRNA species, based on cognate amino acid properties and isoacceptor identities. Each breast cancer cell line generates tRNA profiles that are markedly different from that of non-cancer-derived breast epithelial cell lines. These observations hold true for breast tumor samples compared to normal breast tissue. However, we find no significant differences in tRNA expression patterns across the three major subtypes of breast cancer we examined. The small sample size used in this study is likely insufficient to reveal subtle but significant differences between breast cancer subtypes. Also, tRNA expression at different tumor stages remains to be examined. Overall, our results highlight the potential of using both nuclear- and mitochondrial-encoded tRNAs as biomarkers for malignancy, tumor type, or tumor progression.

Remarkably, tRNAs carrying the amino acids serine, threonine, and tyrosine are among the most over-expressed across all breast cancer cell lines and breast tumors analyzed. Since these amino acid residues are targets for protein kinases and phosphatases, this result suggests these tRNAs might be part of a potential mechanism for potentiating post-translational regulation of proteins involved in signal transduction. How such a mechanism may work remains obscure at this time.

Furthermore, significant differences are observed in the relative expression levels of tRNA isoacceptors. Differential expression of tRNA isoacceptors may provide an additional level of translational regulation for key genes involved in tumorigenesis. Tumors display both quantitative and qualitative changes in protein expression, with preferential translation of growth factors, cell-cycle promoters, and proto-oncogenes, such as VEGF, cyclin D1, and c-Myc (2). It is possible that selective over-expression of tRNA isoacceptors enhances translational efficiency of some of these genes.

Another mechanism linking translational control by tRNA and oncogenic transformation has recently been reported (9). In that work, the translation initiator tRNA_i^{Met} was shown to have oncogenic capacity. A modest over-expression of initiator tRNA_i^{Met} (<2-fold) results in malignant transformation of immortalized mouse fibroblasts, whereas over-expression of the elongator tRNA_e^{Met} to a similar level does not. Consistent with these observations, our results indicate that tRNA_i^{Met} is also over-expressed by 2- to 3-fold in all cancer-derived versus non-cancer-derived breast cell lines. However, over-expression of tRNA_i^{Met} is below that of other tRNAs such as the Ser/Thr/Tyr-tRNAs in the breast cancer cells. Over-expression of these tRNAs may simply be the consequence of over-expression of tRNA_i^{Met}, although a more exciting possibility is that

some of these tRNAs also have oncogenic potential and may be regulated independently. Further studies are needed to elucidate the regulatory relationships between tRNA expression and cancer.

Variations of tRNA-codon usage associations

A main regulatory function of tRNA is to tune the translation efficiency of individual mRNAs via their codon usages (10). A common hypothesis is that during active cell growth, the amount of tRNA isoacceptors correlates with the codon usage of genes translated at high levels. This hypothesis has been validated for bacteria and yeast (26–28). In mammalian cells, this hypothesis also holds for isolated instances where tRNA amounts have been measured in specialized cells, e.g. red blood cells (31).

The common analysis of tRNA-codon usage correlation/association relies on knowing the amount of tRNA isoacceptors in each sample. Information on absolute tRNA abundance levels, however, is difficult to obtain in microarray measurements due to variations in labeling and hybridization efficiencies. Therefore, a direct assessment of tRNA-codon usage association in breast cancer cells cannot be carried out in this study.

Instead, the array results compare the relative abundance of each isoacceptor between two samples. Among the three groups of genes relevant for codon usage-tRNA over-expression association analysis, positive associations are found only for certain codons among cancer-related genes. Our results suggest that during active cell growth (as examined in this work), a complete adjustment of tRNA abundance to fit the codon usage of regulatory genes would be very difficult. tRNAs in the same cell must balance promoting translation of regulatory genes versus demoting translation of house-keeping genes with different codon usage. From the codon usage perspective, tRNA over-expression can only benefit the translation of a subset of regulatory genes or a subset of codons in the regulatory genes. Thus, over-expression of tRNA in breast cancer cells likely affects other regulatory processes that are yet to be explored or understood.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr. Lucy Godley for help with the procurement of tumor samples.

FUNDING

Congressionally Directed Medical Research Programs grant [W81XWH-07-1-0595] to M.R. and T.P.; National Institutes of Health training grant [T32GM007183-33] to M.P.-E. Funding for open access charge: W81XWH-07-1-0595.

Conflict of interest statement. None declared.

REFERENCES

1. Bjornsti, M.A. and Houghton, P.J. (2004) Lost in translation: dysregulation of cap-dependent translation and cancer. *Cancer Cell*, **5**, 519–523.
2. Pandolfi, P.P. (2004) Aberrant mRNA translation in cancer pathogenesis: an old concept revisited comes finally of age. *Oncogene*, **23**, 3134–3137.
3. Ruggero, D. and Pandolfi, P.P. (2003) Does the ribosome translate cancer? *Nat Rev Cancer*, **3**, 179–192.
4. White, R.J. (2004) RNA polymerase III transcription and cancer. *Oncogene*, **23**, 3208–3216.
5. Winter, A.G., Sourvinos, G., Allison, S.J., Tosh, K., Scott, P.H., Spandidos, D.A. and White, R.J. (2000) RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. *Proc. Natl Acad. Sci. USA*, **97**, 12619–12624.
6. Marshall, L. and White, R.J. (2008) Non-coding RNA production by RNA polymerase III is implicated in cancer. *Nat. Rev. Cancer*, **8**, 911–914.
7. Chen, W., Bocker, W., Brosius, J. and Tiedge, H. (1997) Expression of neural BC200 RNA in human tumours. *J. Pathol.*, **183**, 345–351.
8. Chen, W., Heierhorst, J., Brosius, J. and Tiedge, H. (1997) Expression of neural BC1 RNA: induction in murine tumours. *Eur. J. Cancer*, **33**, 288–292.
9. Marshall, L., Kenneth, N.S. and White, R.J. (2008) Elevated tRNA(iMet) synthesis can drive cell proliferation and oncogenic transformation. *Cell*, **133**, 78–89.
10. Zouridis, H. and Hatzimanikatis, V. (2008) Effects of codon distributions and tRNA competition on protein translation. *Biophys. J.*, **95**, 1018–1033.
11. Bertucci, F. and Goncalves, A. (2008) Clinical proteomics and breast cancer: strategies for diagnostic and therapeutic biomarker discovery. *Future Oncol.*, **4**, 271–287.
12. Payne, S.J., Bowen, R.L., Jones, J.L. and Wells, C.A. (2008) Predictive markers in breast cancer – the present. *Histopathology*, **52**, 82–90.
13. Rennstam, K. and Hedenfalk, I. (2006) High-throughput genomic technology in research and clinical management of breast cancer. Molecular signatures of progression from benign epithelium to metastatic breast cancer. *Breast Cancer Res.*, **8**, 213.
14. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
15. Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
16. Soll, D. (1993) Transfer RNA: an RNA for all seasons. *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 157–184.
17. Dittmar, K.A., Goodenbour, J.M. and Pan, T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, e221.
18. Saikia, M., Dai, Q., Decatur, W.A., Fournier, M.J., Piccirilli, J.A. and Pan, T.A. (2006) Systematic, ligation-based approach to study RNA modifications. *RNA*, **12**, 2025–2033.
19. Dittmar, K.A., Sorensen, M.J., Elf, J., Ehrenberg, M. and Pan, T. (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.*, **6**, 151–157.
20. Dittmar, K.A., Mobley, E.M., Radek, A.J. and Pan, T. (2004) Exploring the regulation of tRNA distribution on the genomic scale. *J. Mol. Biol.*, **337**, 31–47.
21. Zaborske, J.M., Narasimhan, J., Jiang, L., Wek, S.A., Dittmar, K.A., Freimoser, F., Pan, T. and Wek, R.C. (2009) Genome-wide analysis of tRNA charging and activation of the eIF2 kinase Gcn2p. *J. Biol. Chem.*, **284**, 25254–25267.
22. Zhou, Y., Goodenbour, J.M., Godley, L.A., Wickrema, A. and Pan, T. (2009) High levels of tRNA abundance and alteration of tRNA charging by bortezomib in multiple myeloma. *Biochem. Biophys. Res. Commun.*, **385**, 160–164.
23. Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F. et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, **10**, 515–527.

24. Xu, Y. and Kool, E. T. (1999) High sequence fidelity in a non-enzymatic DNA autoligation reaction. *Nucleic Acids Res.*, **27**, 875–881.
25. Xu, Y., Karalkar, N. B. and Kool, E. T. (2001) Nonenzymatic autoligation in direct three-color detection of RNA and DNA point mutations. *Nat. Biotechnol.*, **19**, 148–152.
26. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.*, **151**, 389–409.
27. Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *J. Mol. Biol.*, **158**, 573–597.
28. Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, **238**, 143–155.
29. Hu, Y. *et al.* (2003) Analysis of genomic and proteomic data using advanced literature mining. *J. Proteome Res.*, **2**, 405–412.
30. Eisenberg, E. and Levanon, E. Y. (2003) Human housekeeping genes are compact. *Trends Genet.*, **19**, 362–365.
31. Hatfield, D. and Rice, M. (1986) Aminoacyl-tRNA(anticodon): codon adaptation in human and rabbit reticulocytes. *Biochem. Int.*, **13**, 835–842.

Table S1. Characteristics of breast tumor and normal breast tissue samples.

Relevant characteristics of breast tumor and normal breast tissue samples analyzed in this study are summarized here. All samples were obtained from female patients. Race, age, gross description, diagnosis, estrogen receptor (ER), progesterone receptor (PR), and HER2 status are listed here. For the breast tumor samples, this table is based on pathology reports provided by the University of Chicago Human Tissue Resource Center. For the normal breast tissue samples, this table is based on data sheets provided by the vendor (see Materials and Methods). In addition to the data listed below, the vendor specified cause of death for A-01 (intracranial hemorrhage) and A-06 (lung cancer, metastasized to bones).

Sample ID	Race	Age	Gross Description	Diagnosis	ER	PR	HER2
59826	Unknown or not reported		Unknown	Invasive Ductal Carcinoma, Not Otherwise Specified	-	-	-
60046	Unknown or not reported		Tumor	Ductal Breast Carcinoma in Situ	-	-	-
62706	Unknown or not reported	48	Tumor	Invasive Ductal Carcinoma, Not Otherwise Specified	-	-	-
62944	African American or Black	51	Tumor	Invasive Ductal Carcinoma, Not Otherwise Specified	-	-	-
46258	Unknown or not reported	34	Tumor	Ductal Breast Carcinoma	-	-	+
58955	Indian	75	Tumor	Ductal Breast Carcinoma	-	-	+
41299	White or Caucasian	55	Unknown	/	+	+	-
57731	White or Caucasian	59	Tumor	Ductal Breast Carcinoma in Situ	+	+	-
45163	Unknown or not reported	62	Tumor	Ductal Breast Carcinoma in Situ	+	-	+
A-01	White or Caucasian	27	Normal	/			
A-06	White or Caucasian	78	Normal	/			
S-23	White or Caucasian	41	Normal	Fibrocystic change			

Table S2. Breast Cell Line Characteristics.

Relevant characteristics of tumorigenic and non-tumorigenic cell lines used in this study are summarized here. Estrogen receptor (ER), progesterone receptor (PgR), primary tumor type, tissue source, tumorigenicity, and tumor classification are indicated. This table is based on previously published data (Lacroix and Leclercq, 2004; Neve et al., 2006).

Cell Line	ER	PgR	HER2	Tumor Type	Tissue Source	Tumorigenic	Tumor Classification
MCF10A	-	-		Fibrocystic Disease	Mammary gland	No	N/A
184 A1	-	-		Normal	Mammary gland	No	N/A
184 B5	-	-		Normal	Mammary gland	No	N/A
BT-474	+	+	+	Invasive ductal carcinoma	Primary	Yes	Luminal
HCC70	-	-		Ductal carcinoma	Primary	Unknown	Basal A
MCF7	+	+		Invasive ductal carcinoma	Metastasis - Pleural effusion	Yes	Luminal
MDA-MB-231	-	-		Adenocarcinoma	Metastasis - Pleural effusion	Yes	Basal B
MDA-MB-436	-	-		Adenocarcinoma	Metastasis - Pleural effusion	Yes	Basal B
ZR-75-1	+	+		Invasive ductal carcinoma	Metastasis - Ascites	Yes	Luminal

Table S3. Breast Cell Line Doubling Times

Cell proliferation was measured over four days using the Promega CellTiter Blue metabolic assay. Cells were plated at low cell density (1,500 cells/well) and high cell density (5,000 cells/well). Doubling times were calculated from the equation: $N/N_0 = e(kt)$. Values are averages of three replicates.

	Doubling time (h)	
	Low cell density	High cell density
MCF10A	ND	27 ± 5
184 A1	31 ± 5	47 ± 6
184 B5	21 ± 5	38 ± 5
BT-474	21 ± 5	35 ± 3
HCC70	27 ± 4	ND
MCF7	28 ± 6	15 ± 1
MDA-MB-231	26 ± 2	25 ± 3
MDA-MB-436	29 ± 3	25 ± 6
ZR-75-1	18 ± 0.3	ND

Table S4 – Target tRNA isoacceptors. Target tRNA isoacceptor pairs (or triplets) differ by less than 8 nucleotides in their body sequence and cannot be distinguished by the standard tRNA microarray method. The first position of the anticodon is highlighted in red.

Amino Acid	Anticodons
Arg	C CG/UCG
Gln	C UG/UUG
Glu	C UC/UUC
Gly	C CC/GCC
Leu	I AG/UAG
Ser	I GA/UGA
Thr	I GU/CGU
Val	I AC/CAC
Ala	I GC/CGC/UGC
Pro	I GG/CGG/UGG

Table S5 – List of cell line-specific genes. Cell line-specific genes were selected based on mRNA expression data (see text). Cell line-specific gene names and sequences are listed in the accompanying excel file.

Table S6 – List of cancer-related genes. Cancer-related genes were selected from a list of potential diagnostic genes known to be highly associated with breast cancer. Cancer-related genes fall into the following functional categories: cell cycle, growth and proliferation, extra-cellular matrix, protein kinases, transcription factors and regulators. Cancer-related gene names and sequences are listed in the accompanying excel file.

Table S7 – List of housekeeping genes. The 30 most highly expressed housekeeping genes were selected for this study, as defined in a previous report. Our set of housekeeping genes includes ribosomal proteins, actin, ubiquitin, and others. Housekeeping gene names and sequences are listed in the accompanying excel file.

Table S8 – Codon usage compilation of breast cancer cell line-specific genes.

a. To define highly-expressed genes in breast cancer cell lines, we set a threshold of 2.6- to 2.8-fold above the median expression value for all transcripts.

b. The top 20 to 30 transcripts based on cell line/MCF10A expression ratio were selected for codon usage-tRNA correlation analysis. The median expression level for these transcripts is indicated here, relative to the median value for all transcripts.

Cell Line-Specific Genes	Threshold^a (fold over median)	Median expression level^b (fold over median)	Number of genes	Number of codons compiled
MDA-MB-231	2.6	7.0	26	36468
MCF7	2.7	8.2	26	26931
HCC70	2.8	14.8	24	47364
ZR-75-1	2.6	11.0	26	28950
MDA-MB-436	2.8	7.5	27	31494
BT-474	2.6	14.6	22	38133

Table S9 – Codon usage compilation for house-keeping and cancer-related genes.

Cancer-related genes	Number of genes	Number of codons compiled
Housekeeping	29	19835
Cell cycle	28	56850
Cell growth and proliferation	35	60831
ECM molecules	21	52353
Protein kinases	17	45882
Transcription factors and regulators	16	40893