AD_____

Award Number: W81XWH-07-1-0595

TITLE: tRNAs as Biomarkers and Regulators for Breast Cancer

PRINCIPAL INVESTIGATOR: Tao Pan, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, IL 60637

REPORT DATE: August 2009

TYPE OF REPORT: Annual Report

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

Form Approved
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)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-08-2009	Annual	July 30, 2008–July 29, 2009

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)

Report writing: Tao Pan, Marsha Rosner.
Performing: M. Pavon-Eternod, S. Gomes, R. Geslain, Q. Dai.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The 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:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	19	USAMRMC

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

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 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 produce targets for a new class of cancer drugs.

BODY (Tables and Figures at the end of the text):

Task 1: Development of a "second-generation" tRNA microarray capable of studying single-nucleotide differences in human tRNA. → **Completed when the last annual report was written in August 2008.**

Task 2: Development of a "third-generation" method capable of studying the expression of >90% of human tRNA species. → **In development.**

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 significantly reduced cost. We are continuing to work on this.

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

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 1). Importantly, genome-wide mRNA expression data are available for these cell lines (1) 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. 1A, 1C). 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 tumorigeniccell 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 (Fig. 1C). 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. 1C, 1D). 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 2). 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 (not shown).

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 3). The 9 breast tumor samples were selected from the three major subtypes of breast cancer: luminals (ER+, HER2-), basals (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. 2A, 2B). 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 upregulation 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 upregulation of global mitochondrial-encoded tRNA levels in breast tumors relative to normal breast.

Consistent with our results on cell lines, tRNA over-expresssion in breast tumor samples is also selective: certain individual tRNAs are more strongly over-expressed than others (Fig. 2C). 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. 2D). 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.**

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 (2-4). 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 (5).

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. 3A). 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 (6). 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 (6). 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 4 and 5). 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. 3B). Though there is limited overlap across cell line-specific gene sets, their codon usage was remarkably similar (average r_s =

0.92 ± 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 welll 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 (not shown). The absence of correlations may be explained by the low mRNA level of these genes at the global scale (Fig. 3A). 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 median for the housekeeping genes. Among human tissues examined, a significant tRNA abundance-codon usage correlation was found only in liver (5). 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. 3C-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. 3C). 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. 3D). 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. 3E).

Task 5: Identify correlations between tRNA and active protein synthesis: → **In development.**

Based on the results from Task 4, we have selected several gene targets for protein expression studies. We are still working on refining the methodology for this study.

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.

REPORTABLE OUTCOMES:

- A revised manuscript is now under review in Nucleic Acids Research. Title: tRNA Over-expression in Breast Cancer and Functional Consequences. Authors: Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha Rich Rosner, Tao Pan.

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.

REFERENCES:

1. 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.

2. 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. *Journal of Molecular Biology*, **151**, 389-409.

3. 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. *Journal of Molecular Biology*, **158.**, 573-597.

4. 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.

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

6. Eisenberg, E. and Levanon, E.Y. (2003) Human housekeeping genes are compact. *Trends Genet*, **19**, 362-365.

Appendices:

Figure legends:

Figure 1 – 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 PharosFX 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) Sames as (B), data are grouped from high to low expression.

Figure 2 – tRNA over-expression in breast cancer *in vivo*. Data is shown for 3 normal breast tissue samples (A-01, A-03, S-23), 4 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 is 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. (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) Sames as (B), data are grouped from high to low expression.

Figure 3 - 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-tumorigenic cells are unlikely to positively correlate to the codon usage of all three groups. (B) Codon usage correlation between cell-line specific genes, cancer-related genes and housekeeping genes. The degree of correlation was assessed using Spearman's rho (r_s). Mean r_s values are plotted for the following correlations: cell line vs. cell line, cell line vs. housekeeping, cancer-related vs.cancer-related, and cancer-related vs. housekeeping. Error bars indicate standard deviation from the mean. (C) Correlation of relative tRNA levels to ratios of codon usage between cancer-related and house-keeping genes. As discussed in the text, a positive correlation is only expected for the codons that are over-represented in the cancer-related genes (x >2). (D) Correlation of relative tRNA levels to ratios of codon usage between the top-third (9 genes) and bottom-third (9 genes) transcribed cell cycle genes. Again, a positive correlation is only expected for the codons that are over-represented

in the top third genes (x >1.2). (E) Correlation of relative arginine tRNA isoacceptor levels to the arginine codon frequency of cancer-related genes.

Table 1. 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 2. 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/No = 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 3. 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. 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).

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 4 – 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 5 – 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

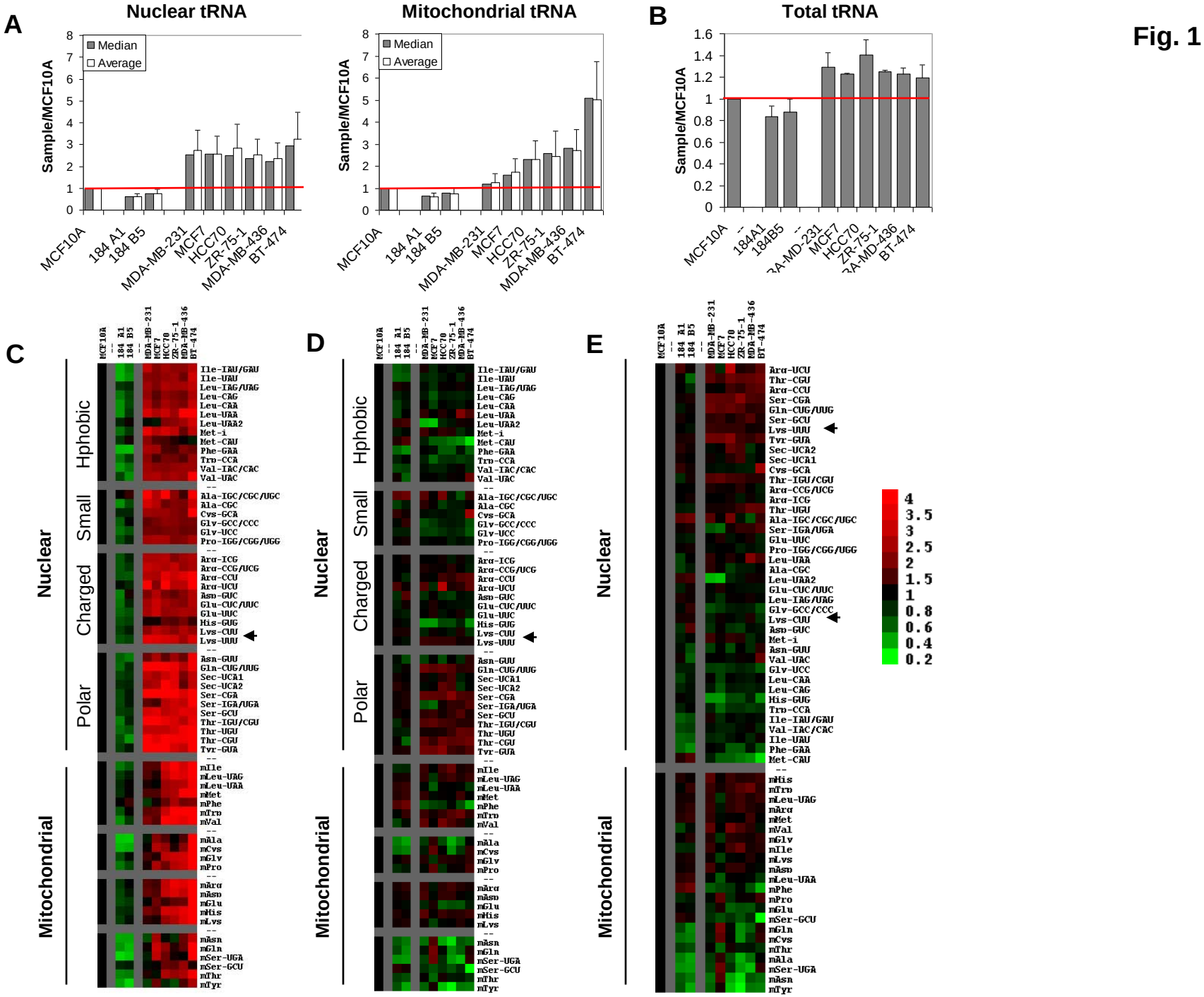

A
Nuclear tRNA
Mitochondrial tRNA
Median
Average
Sample/MCF10A
MCF10A
184 A1
184 B5
MDA-MB-231
MCF7
HCC70
ZR-75-1
MDA-MB-436
BT-474
B
Total tRNA
C
D
E
Nuclear
Mitochondrial
Hphobic
Small
Charged
Polar
Ile-IAU/GAU
Ile-UAU
Leu-IAG/UAG
Leu-CAG
Leu-CAA
Leu-UAA
Leu-UAA2
Met-i
Met-CAU
Phe-GAA
Trp-CCA
Val-IAC/CAC
Val-UAC
Ala-IGC/CGC/UGC
Ala-CGC
Cys-GCA
Gly-GCC/CCC
Gly-UCC
Pro-IGG/CGG/UGG
Arg-ICG
Arg-CCG/UCG
Arg-CCU
Arg-UCU
Asn-GUC
Glu-CUC/UUC
Glu-UUC
His-GUG
Lys-CUU
Lys-UUU
Asn-GUU
Gln-CUG/UUG
Sec-UCA1
Sec-UCA2
Ser-CGA
Ser-IGA/UGA
Ser-GCU
Thr-IGU/CGU
Thr-UGU
Thr-CGU
Tyr-GUA
mIle
mLeu-UAG
mLeu-UAA
mMet
mPhe
mTrp
mVal
mAla
mCys
mGly
mPro
mArg
mAsn
mGlu
mHis
mLys
mAsn
mGln
mSer-UGA
mSer-GCU
mThr
mTyr
Arg-UCU
Thr-CGU
Arg-CCU
Ser-CGA
Gln-CUG/UUG
Ser-GCU
Lys-UUU
Tyr-GUA
Sec-UCA2
Sec-UCA1
Cys-GCA
Thr-IGU/CGU
Arg-CCG/UCG
Arg-ICG
Thr-UGU
Ala-IGC/CGC/UGC
Ser-IGA/UGA
Glu-UUC
Pro-IGG/CGG/UGG
Leu-UAA
Ala-CGC
Leu-UAA2
Glu-CUC/UUC
Leu-IAG/UAG
Gly-GCC/CCC
Lys-CUU
Asn-GUC
Met-i
Asn-GUU
Val-UAC
Gly-UCC
Leu-CAA
Leu-CAG
His-GUG
Trp-CCA
Ile-IAU/GAU
Val-IAC/CAC
Ile-UAU
Phe-GAA
Met-CAU
mHis
mTrp
mLeu-UAG
mArg
mMet
mVal
mGly
mIle
mLys
mAsn
mLeu-UAA
mPhe
mPro
mGlu
mSer-GCU
mGln
mCys
mThr
mAla
mSer-UGA
mAsn
mTyr
4
3.5
3
2.5
2
1.5
1
0.8
0.6
0.4
0.2

Fig. 1

Fig. 2

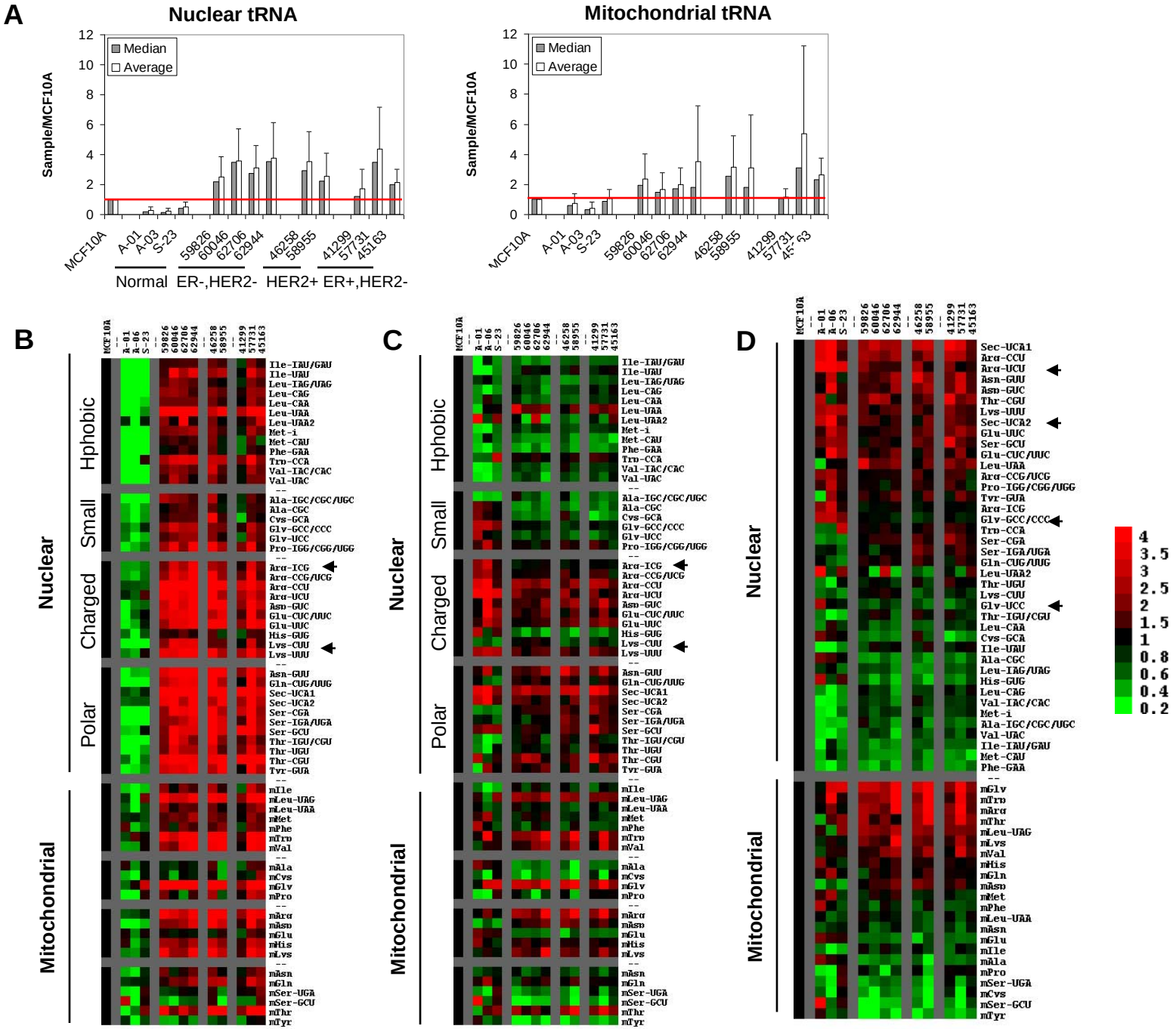
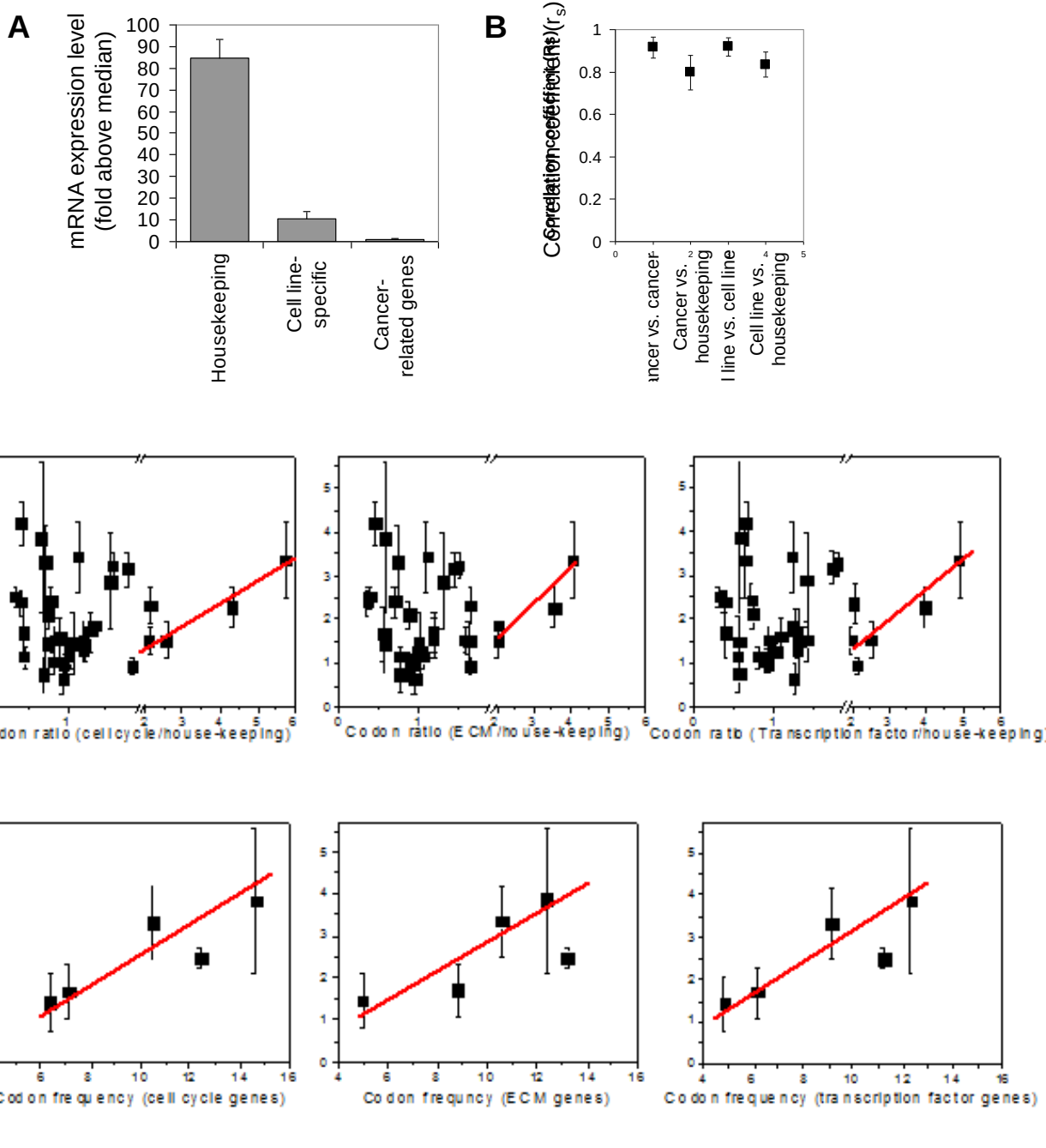

A
Nuclear tRNA
Mitochondrial tRNA
Median
Average
Sample/MCF10A
MCF10A
A-01
A-03
S-23
59826
60046
62706
62944
46258
58955
41299
57731
45163
Normal
ER-,HER2-
HER2+
ER+,HER2-
B
C
D
Nuclear
Mitochondrial
Hphobic
Small
Charged
Polar
Ile-IAU/GAU
Ile-UAU
Leu-IAG/UAG
Leu-CAG
Leu-CAA
Leu-UAA
Leu-UAA2
Met-i
Met-CAU
Phe-GAA
Trp-CCA
Val-IAC/CAC
Val-UAC
Ala-IGC/CGC/UGC
Ala-CGC
Cys-GCA
Gly-GCC/CCC
Gly-UCC
Pro-IGG/CGG/UGG
Arg-ICG
Arg-CCG/UCG
Arg-CCU
Arg-UCU
Asn-GUC
Glu-CUC/UUC
Glu-UUC
His-GUG
Lys-CUU
Lys-UUU
Asn-GUU
Gln-CUG/UUG
Sec-UCA1
Sec-UCA2
Ser-CGA
Ser-IGA/UGA
Ser-GCU
Thr-IGU/CGU
Thr-UGU
Thr-CGU
Tyr-GUA
mIle
mLeu-UAG
mLeu-UAA
mMet
mPhe
mTrp
mVal
mAla
mCys
mGly
mPro
mArg
mAsp
mGlu
mHis
mLys
mAsn
mGln
mSer-UGA
mSer-GCU
mThr
mTyr
Sec-UCA1
Arg-CCU
Arg-UCU
Asn-GUU
Asp-GUC
Thr-CGU
Lys-UUU
Sec-UCA2
Glu-UUC
Ser-GCU
Glu-CUC/UUC
Leu-UAA
Arg-CCG/UCG
Pro-IGG/CGG/UGG
Tyr-GUA
Arg-ICG
Gly-GCC/CCC
Trp-CCA
Ser-CGA
Ser-IGA/UGA
Gln-CUG/UUG
Leu-UAA2
Thr-UGU
Lys-CUU
Gly-UCC
Thr-IGU/CGU
Leu-CAA
Cys-GCA
Ile-UAU
Ala-CGC
Leu-IAG/UAG
His-GUG
Leu-CAG
Val-IAC/CAC
Met-i
Ala-IGC/CGC/UGC
Val-UAC
Ile-IAU/GAU
Met-CAU
Phe-GAA
mGly
mTrp
mArg
mThr
mLeu-UAG
mLys
mVal
mHis
mGln
mAsp
mMet
mPhe
mLeu-UAA
mAsn
mGlu
mIle
mAla
mPro
mSer-UGA
mCys
mSer-GCU
mTyr
4
3.5
3
2.5
2
1.5
1
0.8
0.6
0.4
0.2

Fig. 3


A
mRNA expression level (fold above median)
Housekeeping
Cell line-specific
Cancer-related genes
B
Correlation coefficients (r_s)
Cancer vs. housekeeping
Cell line vs. housekeeping
C
tRNA isoacceptor ratio Cancer/MCF10A
Codon ratio (cell cycle/house-keeping)
Codon ratio (ECM/house-keeping)
Codon ratio (Transcription factor/house-keeping)
D
tRNA isoacceptor ratio Cancer/MCF10A
Codon ratio (top third/bottom third)
E
Arg-tRNA isoacceptor ratio Cancer/MCF10A
Codon frequency (cell cycle genes)
Codon frequency (ECM genes)
Codon frequency (transcription factor genes)