

# Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages

Yedael Y. Waldman<sup>1</sup>, Tamir Tuller<sup>1,2,3</sup>, Tomer Shlomi<sup>4</sup>, Roded Sharan<sup>1</sup> and Eytan Ruppin<sup>1,3,\*</sup>

<sup>1</sup>Blavatnik School of Computer Science, <sup>2</sup>Department of Molecular Microbiology and Biotechnology and <sup>3</sup>School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel, <sup>4</sup>Computer Science Department, Technion – Israel Institute of Technology, Haifa 32000, Israel

Received October 21, 2009; Revised January 1, 2010; Accepted January 5, 2010

## ABSTRACT

Various studies in unicellular and multicellular organisms have shown that codon bias plays a significant role in translation efficiency (TE) by co-adaptation to the tRNA pool. Yet, in humans and other mammals the role of codon bias is still an open question, with contradictory results from different studies. Here we address this question, performing a large-scale *tissue-specific* analysis of TE in humans, using the tRNA Adaptation Index (tAI) as a direct measure for TE. We find tAI to significantly correlate with expression levels both in tissue-specific and in global expression measures, testifying to the TE of human tissues. Interestingly, we find significantly higher correlations in adult tissues as opposed to fetal tissues, suggesting that the tRNA pool is more adjusted to the adult period. Optimization based analysis suggests that the tRNA pool—codon bias co-adaptation is globally (and not tissue-specific) driven. Additionally, we find that tAI correlates with several measures related to the protein functional importance, including gene essentiality. Using inferred tissue-specific tRNA pools lead to similar results and shows that tissue-specific genes are more adapted to their tRNA pool than other genes and that related sets of functional gene groups are translated efficiently in each tissue. Similar results are obtained for other mammals. Taken together,

these results demonstrate the role of codon bias in TE in humans, and pave the way for future studies of tissue-specific TE in multicellular organisms.

## INTRODUCTION

Codon bias, the unequal usage of different codons, has been observed in many organisms. It is believed that in many unicellular organisms codon bias has evolved to optimize translation efficiency (TE) by favoring codons with higher levels of tRNA (1–4). Similar evidence was found in multicellular organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster* (5,6). However, the evolution and functional significance of codon bias in multicellular organisms such as mammals, and specifically in humans, are still largely unknown (7,8). Small scale analyses have shown that expressing a foreign gene in human cells while fitting its codons to the bias of human can increase its protein levels by two orders of magnitude (9), implying a relation between codon bias and TE. However, large-scale analyses of codon bias in humans show contradictory results: some of the studies relate the bias to TE while others do not find a significant correlation between these two variables (2,10–16) [reviewed in (8)].

When analyzing the role of codon bias in TE in humans, two points should be noted: First, a measure for TE should address the translation process directly. However, most previous studies relied on various measures for codon bias that are not directly related to TE. For example, Wright (17) suggested the measure of effective

\*To whom correspondence should be addressed. Tel: +972 3 640 6528; Fax: +972 3 640 9357; Email: ruppin@post.tau.ac.il  
Correspondence may also be addressed to Tamir Tuller. Tel: +972 3 640 6528; Fax: +972 3 640 9357; Email: tamirtul@post.tau.ac.il

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

number of codons, which quantifies how far the codon usage of a gene departs from equal usage of synonymous codons. Urrutia and Hurst (13,14) used the Maximum Likelihood Codon Bias (MCB), which measures the observed bias from a background model of expected bias. Others have defined measures for the similarity between the codon bias of a gene to the usage found in highly expressed genes which their usage is presumably selected for [e.g. Frequency of Optimal Codons (18); Codon Bias Index (19); Codon Adaptation Index (20)]. This assumption is of course plausible, but yet indirect [e.g. highly expressed gene may also undergo selection for increased GC content (21)]. Hence, when studying TE, a more useful approach would go beyond quantifying the codon bias and also aim to measure the codon-bias—tRNA pool co-adaptation. Indeed, the *tRNA Adaptation Index* (tAI) forms such a measure (2), as it assigns for each codon a score based upon the availability of the corresponding tRNAs. tAI was shown to correspond to *in vivo* measurements of translation rate of specific codons (22), as well as being a better predictor of protein abundance in *Saccharomyces cerevisiae* as compared to other measures (3,23). However, tAI has not been used to study TE in depth in human or other multicellular organisms. Second, in difference from previous studies of TE in multicellular organisms that utilized global measures of expression [e.g. (13–16)], an analysis of TE in this setting should consider the distinct gene expression (GE) pattern of each tissue. However, most previous studies either focused on the codon bias of small sets of tissue-specific genes rather than the overall expression patterns of these tissues (24–26) or took global expression measures such as expression breadth (number of tissues in which the gene is expressed), expression rate (mean expression over tissues) (13,14) or maximum expression level (15,16). While a former study indeed used tissue-specific expression levels in humans (12), it used other measures for codon and amino acid (AA) bias rather than direct measures for TE such as tAI. Here we perform the first large-scale study in human that combines both points: a *tissue-specific* direct analysis of TE in human, using the tAI measure. Our results suggest that codon bias plays an important role in TE in humans, with significant differences between tissues and developmental stages.

## MATERIALS AND METHODS

### tAI calculation

For all organisms, tAI was calculated as described in dos Reis *et al.* (2).

Briefly, this measure is based on the fact that each codon is recognized by some anti-codon(s) and the corresponding tRNA(s). According to Crick's wobble rules (27), more than one anti-codon can recognize the codon, with different efficiency, based on the match between them [see dos Reis *et al.* (2) for all the relations between codons and anti-codons].

Let  $n_i$  be the number of tRNA isoacceptors recognizing codon  $i$ . Let  $tGCN_{ij}$  be the copy number of the  $j$ th tRNA

that recognize the  $i$ th codon, and let  $S_{ij}$  be the selective constraint on the efficiency of the codon–anti-codon coupling. We define the *absolute adaptiveness*  $W_i$  for each codon  $i$  as:

$$W_i = \sum_{j=1}^{n_i} (1 - S_{ij}) tGCN_{ij}$$

From  $W_i$  we obtain  $w_i$ , which is the *relative adaptiveness value of codon  $i$*  by normalizing the  $W_i$ 's values (dividing them by the maximal  $W_i$ ). Supplementary Table S1 shows the  $w$  values for all codons according to the human genomic tRNA pool.

The final tAI of a gene,  $g$ , is the following geometric mean:

$$tAI(g) = \left( \prod_{k=1}^{lg} w_{ikg} \right)^{1/lg}$$

where  $i_{kg}$  is the codon defined by the  $k$ 'th triplet on gene  $g$  and  $lg$  is the length of the gene (excluding stop codons).  $S$  values were taken from dos Reis *et al.* (2).

Codon usage for human, *Mus musculus* (mouse) and *S. cerevisiae* (yeast) genes was downloaded from the Codon Usage Database (28) (<http://www.kazusa.or.jp/codon/>) on 7.2008. The codon usage for *Pan troglodytes* (chimpanzee), *Rattus norvegicus* (rat) and *Escherichia coli*, was calculated from genes' sequences that were downloaded from BioMart (29). The codon usage for *C. elegans* and *D. melanogaster* (fruit fly) was calculated from genes' sequences downloaded from KEGG database (30).

Human, chimpanzee, mouse, rat, *S. cerevisiae*, *C. elegans* and *E. coli* tRNA copy numbers were downloaded from the Genomic tRNA Database (31) (<http://gtrnadb.ucsc.edu>).

### Tissue-specific tRNA levels

Dittmar *et al.* (26) have characterized the expression levels of tRNA across human tissues, but we were not able to use these results here as they are not absolute tissue-specific levels but only ratios relative to the brain. In addition, such data are currently available on only eight human tissues. Nevertheless, we have conducted several additional extensive analyses using the data of Dittmar *et al.* (26), to further support our choice of tRNA copy number and the results reported in the article.

First, we used the tRNA levels in the brain as our reference point, following Dittmar *et al.* (26), i.e. making their levels equal to the genomic copy number provided by Dittmar *et al.* (26) (in their Supplementary Data) and used the ratios reported in their article to infer from them the levels in other tissues. We used the copy number in Dittmar *et al.* (26) though it is slightly different from the current version in Genomic tRNA Database (31), in order to use their original assignment of each probe to the corresponding tRNA and infer the relevant ratios. Nevertheless, the two sets of copy numbers are highly correlated and as can be seen from Supplementary Tables S2 and S8, lead to very similar results, both in

the correlation magnitude and in the ranking of tissues. Next, we measured the tAI–GE correlation in six of the tissues, for which we had expression data in our dataset. Second, in an alternative route for identifying and using tissue-specific tRNA levels for the TE computation, we have employed an optimization procedure to infer the tissue-specific tRNA levels. Briefly, instead of assuming that the tRNA levels in the brain are determined by the tRNA copy numbers, we allowed their levels to vary and, while maintaining the tRNA expression ratios reported in the work of Dittmar *et al.* as constraints, we maximized the overall correlation between tAI and expression levels across the tissues with both expression data and relative tRNA values.

Formally, let  $T_i$  denote the inferred vector of tRNA levels in tissue  $i$  ( $i = 1$  denotes the brain); let  $tAI(T_i)$  denote the corresponding vector of gene tAI when using  $T_i$  as proxy of tRNA levels; let  $GE(T_i)$  denote GE vector in tissue  $i$ ; Let  $R(\cdot, \cdot)$  denote Spearman correlation.

We solved the following optimization problem:

Optimize  $R([tAI(T_1), tAI(T_2), \dots], [GE(T_1), GE(T_2), \dots])$

such that:

- (i) For every  $i$  each entry in  $T_i$  is larger than 0 if the copy number of the corresponding entry is larger than 0; otherwise the entry remains 0.
- (ii) The ratios between the expression levels of tRNAs in each tissue ( $i = 1, 2, \dots$ ) relative to the brain ( $i = 1$ ) are identical to the results reported in (26).

To solve this problem we employed the optimization approach of Nelder–Mead (32,33). Optimization was done simultaneously on all tissues.

As further tissue-specific analysis, we measured the tissue-specific tAI measure (using optimized tissue-specific tRNA levels, as above) of tissue-specific genes in those tissues and found them to be significantly higher than other genes expressed in these tissues ('Discussion' section). To evaluate the significance of TE in tissue-specific genes, we compared their mean tissue-specific tAI to those of 100 000 random gene sets of the same size (all of them expressed in the specific tissue).

### GE data

All expression data was downloaded from Gene Expression Omnibus (34) (<http://www.ncbi.nlm.nih.gov/geo/>).

Human tissues (including fetal tissues): we used the GE of Su *et al.* (35) (GDS596). As the original data set is redundant (i.e. it includes similar tissues; for example, more than 20 of the tissues are from different parts of the brain) we focused our analysis on 30 (out of 79) non-redundant normal adult tissues (Supplementary Table S2).

Other GE sets: fetal and adult circulating blood reticulocytes (GDS2655), Mouse tissues (GDS592), Mouse fetal and adult liver (GSE13149), Mouse embryonic stem cells (GDS2666), Yeast (GDS772, wild type), Chimpanzee (GSE7540), Rat (GDS589, three strains), *E. coli* (GSE6836), *D. melanogaster* (GSE7763) and *C. elegans* (GSE8004).

We averaged technical repeats and probes of the same gene. Following (36), a gene was defined as expressed in a tissue if its expression level in Affymetrix chip was above 200 standard Affymetrix average-difference units. A gene was defined as tissue specific in a tissue if it is expressed in that tissue and its expression level is at least 2 SD above its mean across all tissues.

### Protein–protein interactions

Protein–protein interaction data were downloaded from (37). After mapping to Entrez IDs, it contains 10 024 proteins and 78 799 interactions. For tissue-specific protein interactions networks, we considered only proteins whose genes are expressed in the tissue.

### Essential genes

A list of essential genes, both for human and mouse, was compiled using the Mouse Genome Informatics (38) (<http://www.informatics.jax.org>). We defined a mouse gene to be essential if its phenotype data contained the term 'MP:0005374' (lethality-prenatal/perinatal). Using this criterion, 1844 genes were defined to be essential. Similarly, we defined a human gene to be essential if its mouse homolog was essential, resulting in 1765 human essential genes. We evaluated the significance of their median tAI by comparing it to those of 10 000 random gene sets of the same size.

### GC content

In order to control for GC content, we have used several measures of GC content:

- (i) GC percentage of (a) the gene (exons only) (b) introns only (c) both introns and exons of the gene (d) 3' UTR and (e) 5' UTR. All these measures were downloaded from BioMart (29).
- (ii) GC percentage of the isochore of the gene [which was shown to be highly correlated with the GC content of the gene (8)]. The genomic location of each gene was downloaded from BioMart (29), and the GC content of isochores along the human genome was taken from Constantini *et al.* (39).
- (iii) GC percentage of the probes used in the microarrays of Su *et al.* (35) (Affymetrix GeneChip Human Genome U133 Array Set HG-U133A). Probes sequences were downloaded from Affymetrix website (table HG-U133A\_2.probe\_tab for human) and we have calculated accordingly the GC content. When more than one probe was mapped to a gene, we have averaged the probes' GC content.

In addition, we made nucleotide-preserving permutation significance testing. We permuted the codons for each gene, while preserving its nucleotide content. We divided the 61 codons (ignoring stop codons) into equivalence classes. Codons A and B were in the same class if and only if they share the same nucleotides (e.g. {CGU, CUG, UCG, GCU, GUC, UGC} is an equivalence class). Permutations were allowed only between members of

the same class. After obtaining random (nucleotide preserving) codon distribution, we computed tAI based on these distributions, and repeated our analyses. This test was repeated a 100 times.

### Replication time

In order to control for replication time bias, which was shown to correlate with GC content (40), we have used replication time data of the human genome, with a resolution of 1 MB (downloaded from <http://www.sanger.ac.uk/PostGenomics/replication-timing/>).

### Recombination rate

Recombination rate was shown to be correlated with various sequence features, including GC content (41). Recombination rate data were taken from Jensen-Seaman *et al.* (41).

### AA bias

We defined two measures of AA bias for control:

- (i) For each gene, its AA bias was defined as the entropy of the frequencies of the AAs along its corresponding protein [an accepted measure for non-uniformity; e.g. (42)]. Let  $p_i = (n_i / \sum_i n_i)$  denote the fraction of AA  $n_i$  ( $1 \leq i \leq 20$ ) in a protein. The AA bias for the protein is:

$$-\sum_i p_i \log(p_i)$$

- (ii) We made an additional control for AA bias using a new tAI measure, tAI<sub>aa</sub>. The tAI<sub>aa</sub> of a gene was defined as its tAI normalized by the maximal possible tAI value for a sequence with the same AAs. Briefly, as in (2), each codon is given a weight  $w$ , according to the genomic copy numbers of its associated tRNAs. For each AA we define an optimal codon, being the one with the highest  $w$ . Let tAI( $g$ ) be the tAI of gene  $g$ , and let tAI<sub>max</sub>( $g$ ) be the tAI for the same sequence of AAs, when using optimal codons only. We define tAI<sub>aa</sub> to be tAI( $g$ )/tAI<sub>max</sub>( $g$ ). Following this definition, tAI<sub>aa</sub> controls for AA bias, as it considers the actual AA sequence as well.

### Evolutionary rate

Evolutionary rate (ER) is defined as the ratio between synonymous and non-synonymous mutations ( $dN/dS$ ) of a gene. We used two measures of ER: (i) between human and mouse and (ii) along the human lineage. ER values between human and mouse were downloaded from BioMart (29). ER values along the human lineage were computed as describe in (43).

### tRNA-codon bias co-adaptation significance test

In order to check the significance of the tRNA-codon bias co-adaptation found in human tissues we made two

randomization tests, one for the codon bias and the other for the tRNA pool:

- (i) In the codon bias test, we made permutations between different codons of the same AA. Each permutation was done for each gene separately.
- (ii) In the tRNA pool test, we made permutations between the copy numbers of different tRNAs of the same codon.

Thus, both tests preserve the original protein sequence and the total copy number of tRNAs for each of the AAs.

Repeating each test 100 times, we measured the tAI-GE correlation for both tests, and obtained an empirical  $P$ -value for the current tRNA codon bias co-adaptation (being the fraction of random sets above the correlation observed in the original data). tAI-GE correlation was measured for each tissue separately and for all the tissues (by concatenating all GE data into one vector) as well.

### tRNA pool optimality

The weaker tAI-GE correlation in humans (e.g. as compared to yeast) may be the result of multi-cellular constraints—i.e. the need to determine the composition of a common global pool (the tRNA copy number) for the different expression levels observed in different tissues. Examining this hypothesis further, we performed a heuristic hill climbing search (simulated annealing) from numerous random starting states, to find a distribution of tRNA copy numbers that yields both a minimal and maximal tAI-expression correlations (separately, of course) given a single tissue (in our case, the heart, the tissue with the highest correlation). We then repeated this computation as we incrementally added additional tissues to the set whose correlations need to be maximized/minimized. We added tissues in the following order: heart, PB-BDCA4+ dendritic cells, skeletal muscle, bronchial epithelial cells, whole brain and whole blood. When considering more than one tissue, the total correlation was computed for the concatenation of the GE of all the tissues being optimized.

In each step of the optimization the overall sum of all tRNA copy numbers present in the genome was preserved (additionally preserving the copy numbers that are equal to zero)—i.e. the underlying assumption is that there is a limited given amount of tRNA genes in the genome, but evolution is free to choose the ‘best’ combination of specific tRNA copies given this allocation.

### Functional enrichment

Functional enrichment was done using the DAVID bioinformatics Functional Annotation tool (44). For each analysis we ranked the genes according to their tAI and took the top genes (taking the top 5, 10 and 20% genes for genome-wide, tissue-specific and housekeeping analyses, respectively. Differences are due to different sample sizes to allow comparable sets of several hundred genes for each analysis). In analyzing differential TE (i.e. genes that are efficient in a specific tissue as compared to their efficiency in other tissues) we measured

for each tissue its tAI (and normalized its with the mean tAI in the tissue). Next, for each gene in each tissue we calculated the ratio between its tAI in that tissue and the mean tAI in other tissue, and took the top 5% genes among all genes. We report here only results that are significant after false discovery rate ('Benjamini' in DAVID results). The results (Supplementary Table S7) are grouped according to the clusters generated by DAVID. In cases where the same gene ontology (GO) term appears in more than one cluster, we present it only once.

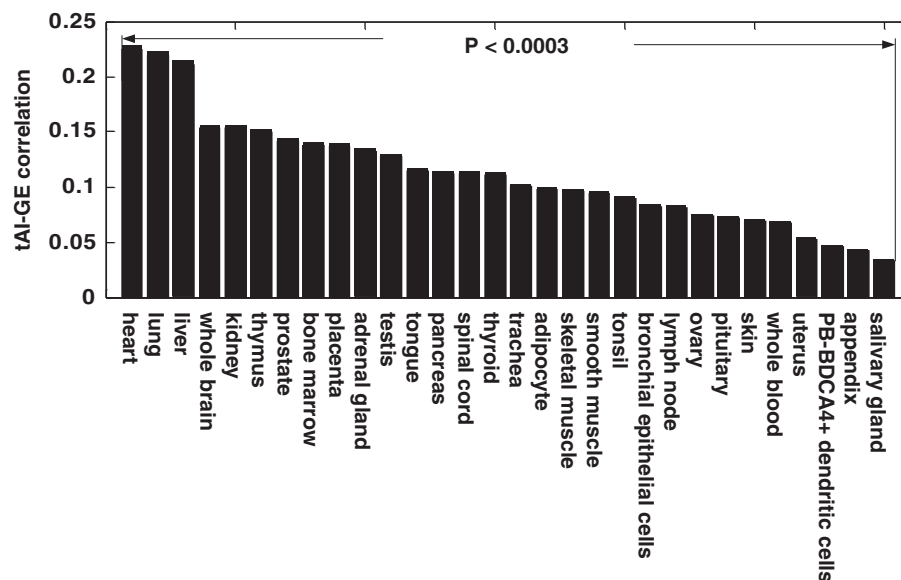
### Analysis

Analysis was done using Matlab<sup>®</sup> software (MathWorks Inc.). All correlations are the non-parametric Spearman correlation (and partial correlation).

## RESULTS

### Tissue-specific translation is efficient

We calculated tAI values for 12271 human genes for which expression data across tissues was available (35), and detected significant correlations between tAI and GE levels in all 30 human adult tissues examined. This testifies to the efficient translation in those tissues, where highly expressed genes have higher TE. The correlations varied among the tissues, with the heart, lung and liver showing the highest correlation scores ( $R > 0.21$ ) and the appendix and salivary gland showing the lowest scores ( $R < 0.05$ , Figure 1 and Supplementary Table S2). These results, as well as others reported below, remain significant after controlling for numerous potentially confounding features reported earlier in the literature (GC content, AA bias, protein length, replication time, recombination rate and more) (Note 1 in Supplementary Data and Supplementary Tables S2 and S3).

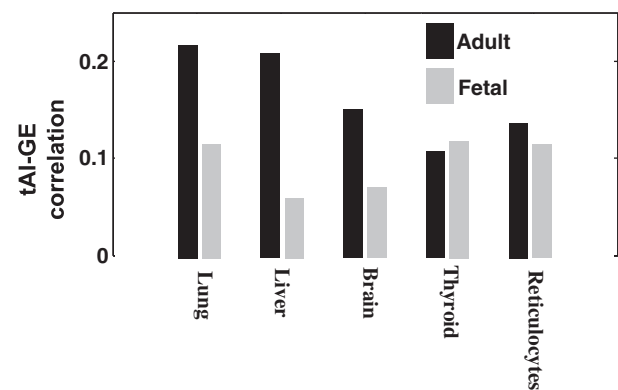


**Figure 1.** Tissue-specific TE in humans. tAI-GE correlation values in 30 normal adult tissues. See Supplementary Table S1 for additional information.

Analyzing tissue-specific TE in a developmental perspective, we studied whether there is a difference in TE between fetal and adult tissues. Comparison between adult and fetal circulating blood reticulocytes showed that adult tissues exhibit a significantly higher correlation ( $P = 2.62 \times 10^{-5}$ , Wilcoxon test). Additional comparison of four tissues further supported this hypothesis (where three of them exhibit 2–3-fold increase in correlation, Figure 2), but we were unable to test the significance of these results due to the small number of samples. Taken together, these results suggest that the current tRNA genomic copy number is more adapted to the adult period rather than the developmental/fetal one.

### The co-adaptation of codon bias and tRNA pool is globally driven

The human tAI-expression correlations shown above are markedly lower than the levels that have been typically obtained in unicellular organisms [in *E. coli* these correlation values varies between 0.30 and 0.44 across 121



**Figure 2.** tAI-GE correlation in adult versus fetal tissues.

different conditions (Supplementary Table S4) and in *S. cerevisiae* being  $R = 0.596$ ,  $P = 4.54 \times 10^{-264}$  (Supplementary Figure S1)]. This may represent a lower selection pressure on multi-tissue organisms, perhaps due to smaller effective population size (45). Yet, another possibility for explaining this gap is that the need to cater for more tissues in multicellular organisms makes the task of optimizing tAI-expression correlation more challenging. Additionally, it is interesting to observe that expression patterns across human tissues have a higher variation than that observed across growth environments in unicellular organisms (Note 2 in Supplementary Data). In light of these comparatively lower human TE levels, we turned to further study the co-adaptation of the tRNA pool-codon bias in both global and in tissue-specific terms ('Materials and Methods' section). First, we randomly permuted the current tRNA pool, without changing the translation code (by permuting the tRNAs of the same AA), finding it to be significantly adapted to the given codon usage human genes over all tissues (empirical  $P = 0.01$ ). Next, we randomly permuted each gene's codon usage, without changing its protein sequence, again finding the current codon bias more adapted to the given tRNA pool over all tissues (empirical  $P < 0.01$ ). However, when we measured these correlations in each tissue separately we found a small number of tissues that were not significantly more correlated than the random sets, for both kinds of tests (Supplementary Table S5). This can be a result for a global, rather than tissue specific, co-adaptation; as the genomic tRNA pool is shared by all tissues, the tRNA pool-codon bias co-adaptation is globally driven, resulting in rather low correlations in specific tissues (in lieu of putative 'compromises' that evolution has to make due to the multiple constraints). Examining this hypothesis further, optimization based analysis shows that, as expected, the tAI-expression correlation decreases as we add more tissues in a simulated scenario ('Materials and Methods' section). Furthermore, this simulation demonstrates that the current correlation observed in humans is closer to being maximal than to being minimal (Supplementary Figure S2).

### Functionally important genes are translated more efficiently

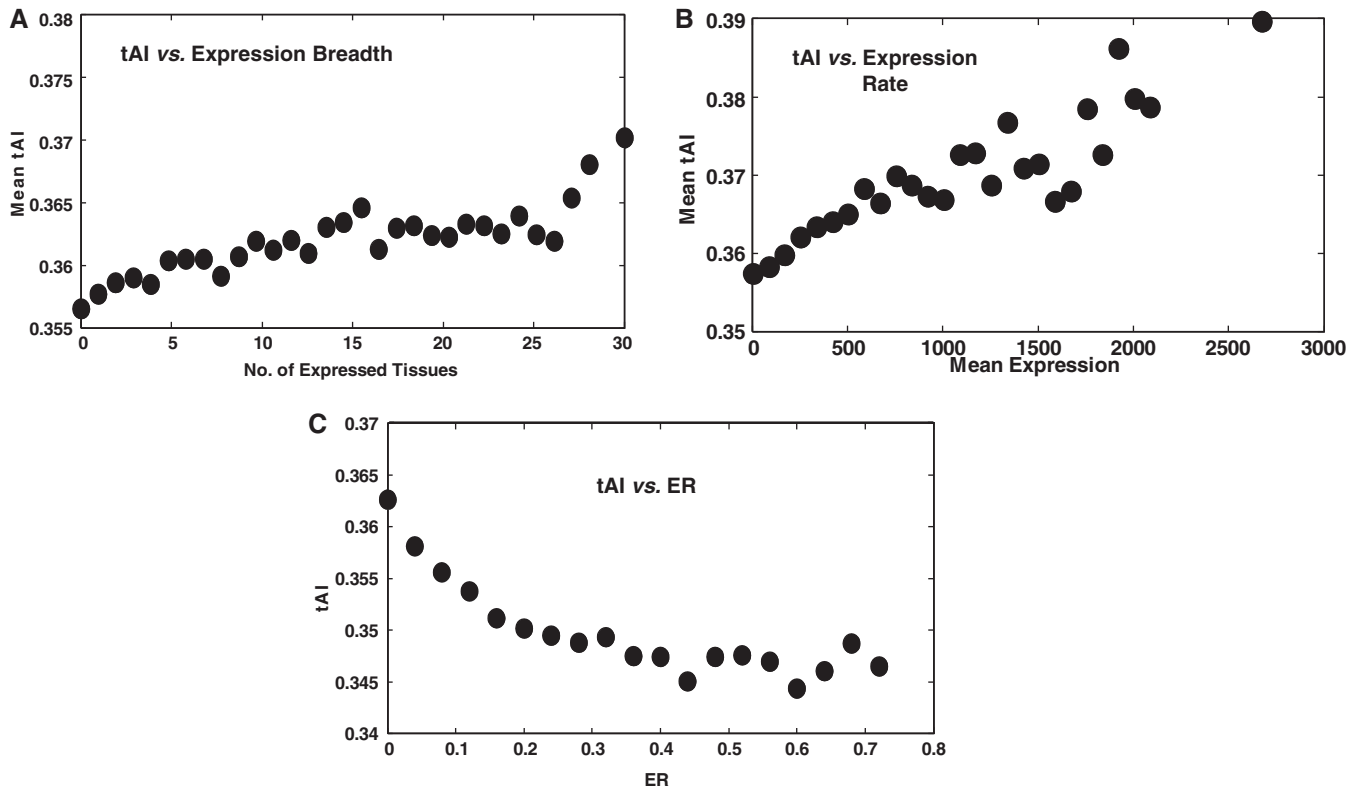
Turning to study global features of human TE, we find that shorter genes are translated more efficiently: there is a significant negative correlation between tAI and protein length ( $R = -0.12$ ,  $P = 1.05 \times 10^{-40}$ ). This result, which is in concordance with previous studies showing (for other organisms) a correlation between codon bias and protein length (46–48), may reflect that both mechanisms for TE (protein length and codon bias) work in concert. In addition, we find tAI to significantly correlate with both expression rate and breadth [ $R = 0.182$ ,  $P = 2.34 \times 10^{-92}$  and  $R = 0.144$ ,  $P = 7.14 \times 10^{-58}$ , respectively (Figure 3A and B)]. Moreover, we find that the tAI reflects previously reported trends more robustly than the MCB measure (13). Whereas the correlation between MCB and expression breadth is not significant after controlling for protein length, the tAI remains significantly correlated with both

expression rate and breadth even after controlling to gene length, as well as many other features [ $R = 0.139$ ,  $P = 8.58 \times 10^{-33}$  and  $R = 0.10$ ,  $P = 1.27 \times 10^{-17}$ , for expression rate and breadth, respectively (Note 1 in Supplementary Notes; Supplementary Table S6)]. The correlation between tAI and expression breadth may indicate that housekeeping genes are translated more efficiently. To examine this hypothesis further we measured the correlation between tAI and a number of measures that are conventionally considered to be related to the functional importance of genes. We find that essential human genes have higher values of tAI as compared to those of all genes (empirical  $P < 10^{-5}$ ); conserved genes show higher tAI values manifesting a significant negative correlation between tAI and ER ( $R = -0.263$ ,  $P = 5.81 \times 10^{-231}$ , Figure 3C); and finally a gene's tAI significantly correlates with its connectivity in a human protein-protein interaction network ( $R = 0.055$ ,  $P = 4.73 \times 10^{-8}$ ). Taken together, these results indicate that functionally important genes are translated more efficiently in humans.

In addition, to identify whether the most efficient genes (in tAI terms) are enriched in some functional classes, we ranked all the genes according to their tAI values and looked whether the top genes were enriched in specific GO (49) categories ('Materials and Methods' section). We find that the most translationally efficient genes are enriched in functional classes such as ribosomal activity, metal (copper, cadmium) and calcium ion binding, GTPase activity, development and more (Supplementary Table S7). When focusing on housekeeping genes (i.e. those expressed in all tissues), we find enrichment in genes related to ribosomal activity (Supplementary Table S7).

### TE analysis based on inferred tissue-specific tRNA levels

All results above are based on global copy numbers of tRNAs. Yet, when analyzing tissue-specific aspects of TE, a more adequate measure would be the actual tissue-specific tRNA levels, which varies between tissues (26). However, as such data are currently unavailable on a broad scope, we used tRNA genomic copy number as approximation for their actual levels. Nevertheless, as additional analysis of tissue-specific TE, we used published data of *relative* tRNA levels in various human tissues (26). Being relative measurements we could not use the data directly, but, using an optimization procedure we inferred tissue-specific tRNA levels from the data ('Materials and Methods' section). Interestingly, the results obtained using tRNA genomic copy numbers remain fairly similar when using inferred tissue-specific tRNA levels instead of the tRNA copy numbers, for a set of six tissues for which both relative tRNA levels (26) and expression data was available (35) (Figure 5 and Supplementary Table S8), suggesting that tRNA genomic copy numbers can be used for first approximation. In addition, for all six tissues we find significant correlation between tAI and expression for the genes expressed in these tissues. Moreover, focusing on tissue-specific genes, we find that in three of the six tissues



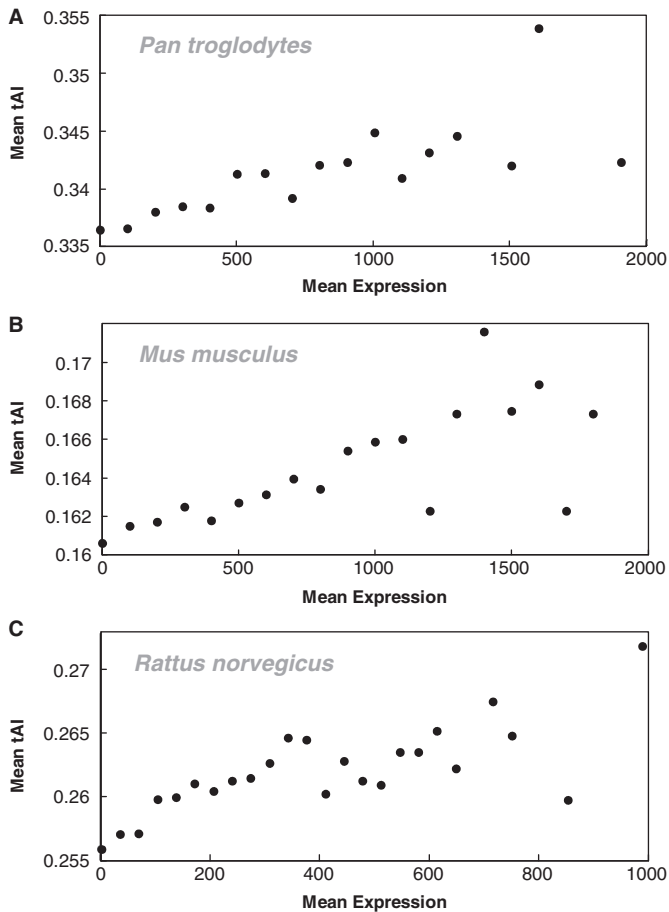
**Figure 3.** Correlations between TE and expression patterns and between TE and ER. (A) tAI versus expression breadth ( $R = 0.869$ ,  $P = 3.48 \times 10^{-7}$ ). Genes are binned according to the number of tissues in which they are expressed. For each bin we show its mean tAI. (B) tAI versus expression rate ( $R = 0.872$ ,  $P = 1.41 \times 10^{-6}$ ). We divided the genes into bins according to their mean expression over 30 tissues. Each bin is 200 standard Affymetrix average-difference units wide, using a minimal bin size of at least 10 genes. For each bin we show its mean tAI. (C) tAI versus ER ( $R = -0.887$ ,  $P = 3.97 \times 10^{-7}$ ) across human genes. ER values are the  $dN/dS$  ratio between the human and mouse. We divided the ER values in the range [0,1] to 25 intervals (each 0.04 wide), taking the mean tAI for each interval. We considered only intervals with at least 10 genes.

examined, their inferred tissue-specific tAI measure is significantly higher than that of other genes expressed in the tissue (empirical  $P < 0.002$ , Supplementary Table S9).

Next, we obtained *tissue-specific* protein-protein interaction networks, based upon the genes expressed in each tissue ('Materials and Methods' section). We find that in five of the six tissues, hubs (proteins with connectivity above the 90th percentile) are translated significantly more efficiently compared to other genes expressed in the tissue. Similarly, we find that tAI significantly correlates with network connectivity in four of the tissue-specific networks (Supplementary Table S9).

Similar to our analysis of global tAI measures, we used the tissue-specific tAI measure to look for functional enrichment in translationally-efficient genes. We find ribosomal proteins show significant tissue-specific TE in all tissues. Interestingly, we also find that genes translated efficiently in the brain are enriched in calcium ion-binding genes. Calcium is known to have a major essential role in various neuronal activities such as neurotransmitter release and dendritic propagation (50). In addition, we find genes translated efficiently in the thymus to be enriched for genes related to chemotaxis and locomotion, which play a key role in T cells maturation in the thymus (51) (Table 1, Supplementary Table S7). As each tissue has its own tRNA pool, we identified the genes that are

significantly efficient in a given tissue according to its own tRNA pool, compared to their TE (tAI values) in other tissues. Interestingly, we find that the set of genes in the gonads (testis and ovary) is overlapping considerably compared to other tissues, as well as the set of genes in the immune system tissues of the thymus and lymph nodes (Jaccard index being 0.253 and 0.156 for the gonads and immune system, respectively, compared to mean Jaccard index 0.055 across all tissues. See also Supplementary Table S7 for enriched GO terms in the gene set of each tissue). Notably, this distinction cannot be inferred from GE alone (e.g. expression in the testis is more correlated with the brain and thymus than the ovary). Additionally, we find that genes related to calcium ion binding are translated more efficiently in the brain and that hormonal genes [including growth hormone 2 (Entrez ID: 4262), insulin (Entrez ID: 3630) and more (GO:0005179)] are translated more efficiently in the liver, as compared to other tissues (Table 1, Supplementary Table S7). Similar tissue-specific analysis of the ovary shows enrichment for extracellular matrix related genes (GO:0031012), specifically for collagen genes (GO:0005581), both are known to have an important role in normal ovary function (52) as well as in ovarian cancer progression, outcome and treatment (53–55) (Table 1, Supplementary Table S7). In addition



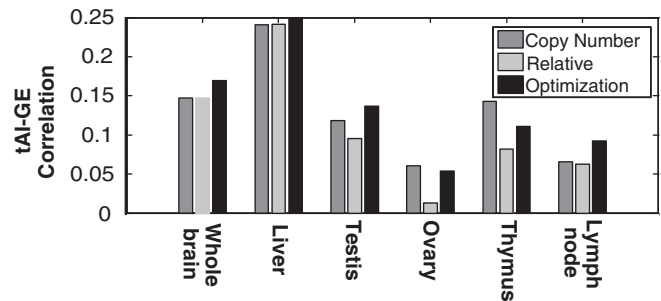
**Figure 4.** TE analysis in other mammals. tAI versus expression rate in other mammals: (A) in *Pan troglodytes* (chimpanzee)  $R = 0.841$ ,  $P = 5.49 \times 10^{-6}$ , (B) in *Mus musculus* (mouse)  $R = 0.758$ ,  $P = 2.59 \times 10^{-4}$  and (C) in *Rattus norvegicus* (rat, Spargue strain)  $R = 0.692$ ,  $P = 2.60 \times 10^{-4}$ . We divided the genes into bins according to their mean expression over 30 tissues. Each bin is 150 standard Affymetrix average-difference units wide, using a minimal bin size of at least 10 genes. For each bin we show its mean tAI. In addition, we performed further analyses, such as tissue-specific analysis in these mammals with several controls and more, obtaining significant results. See Notes 3–4 (Supplementary Data) and Supplementary Tables S10–S12 for detailed results.

we find that cadmium ion binding is enriched in the testis; this may be related to the sensitivity of this tissue to elevated levels of cadmium [e.g. (56)].

Taken together, these results suggest that inferred tissue-specific tAI values reveal additional tissue-specific features of TE.

## DISCUSSION

Using tAI as a measure for TE and large-scale tissue-specific expression data, we analyzed the role of codon bias in TE in humans, finding it to be important. As opposed to most previous studies in multicellular organisms, our tissue-specific approach enabled us to find significant differences across tissues and developmental stages. Computation based analysis implies that the codon bias–genomic tRNA pool co-adaptation is



**Figure 5.** Comparison between genomic tAI values and inferred tissue-specific tAI values. tAI-GE correlation values in six human tissues, based on tAI-GE correlation for (i) copy number—genomic copy numbers (shared by all tissues), (ii) Relative tRNA levels—obtained by fixing the brain tRNA levels as a base point equal to the genomic copy numbers and using the ratios reported in (26) to infer the levels in the five other tissues, and (iii) Optimized tRNA levels, where we allowed variance in the brain tRNA levels (while maintaining the tRNA expression ratios in other tissues) when optimizing the overall tAI-GE in tissues (\*Materials and Methods\* section). As expected from such an optimization procedure, the resulting tAI-GE correlations were higher than those previously obtained using the gene copy number, but remarkably, in average less than 5% higher (Supplementary Table S8).

globally driven, thus explaining why the tAI-GE correlations observed in human, although significant, are lower than those obtained in unicellular organisms. In addition, using various measures, we find that functionally important genes are translated more efficiently, and identify functional groups that are translated more efficiently. Extending our analyses to mouse, chimpanzee and rat, we obtain qualitatively similar results (Figures 4; Notes 3–4 in Supplementary Data; Supplementary Tables S10–S12). These findings suggest that codon bias plays an important functional role in TE in humans as well as in other mammals.

Recently, employing some of methods described here, we have shown that codon bias plays a significant role in TE not only in normal tissues but also in cancerous tissues. Analyzing the distribution of cancerous mutations in the tumor suppressor gene TP53, we found them to be under selection for their TE effect (22). This result generalizes the results reported here for normal tissues.

Our results indicate that there is a significant difference between the TEs at distinct developmental stages and that the tRNA pool is more adjusted to the expression regime of adult versus fetal period. In order to see whether we observe this difference in other mammals as well, we compared adult and fetal liver tissues in mouse (57). Again, we find the adult tissue showing higher correlation values ( $P = 0.0158$ , Wilcoxon test; Supplementary Table S13). As there are GE measurements in various developmental stages of mouse embryonic stem cells (58), we examined how this correlation changes during development. Interestingly, we find a significant correlation between development and tAI-GE correlation—as the stem cells develop, the correlation increases ( $R = 0.864$ ,  $P = 0.0013$ , over 11 developmental time points. See Supplementary Table S14). Notably, we also find that as the fetal liver develops, its GE regime becomes more

Table 1. List of functional groups showing enrichment for tissue-specific TE

	Brain	Liver	Testis	Ovary	Thymus	Lymph node
<b>Tissue-specific TE</b>	<b>ribosome, calcium ion binding, GTP binding</b>	<b>ribosome, structural molecule activity, cytoplasm</b>	<b>ribosome, copper ion binding, actin binding</b>	<b>ribosome, copper ion binding, extracellular region</b>	<b>ribosome, chemotaxis, locomotory behavior</b>	<b>ribosome, transporter activity, cation transport</b>
<b>Differential tissue-specific TE</b>	<b>calcium ion binding, protein folding, actin binding</b>	immune system response, extracellular region, <b>hormonal activity</b>	DNA binding, Zinc ion binding, <b>cadmium ion binding</b>	DNA binding, <b>collagen, extracellular matrix</b>	transcription factor activity, signal transducer activity, amine transport	transporter activity, amine transport, peptide binding

For each tissue we show the top three significant functional groups (based upon GO classification) exhibiting significant TE (as compared to other genes in the tissue [tissue-specific TE] and as compared to their TE in other tissues [differential tissue-specific TE]). The bold groups are those discussed in the text. For the full list of functional groups exhibiting significant tissue-specific TE see Supplementary Table S7.

similar to the adult liver ( $R > 0.97$ ,  $P = 3.96 \times 10^{-4}$  over eight prenatal developmental stages), which may partially explain the increase in co-adaptation to the genomic tRNA pool during developmental stages.

Nevertheless, the lower correlation observed in fetal tissues does not necessarily indicate that their translation is less efficient. It may be a result of post-transcriptional regulation such as RNA-binding proteins (59,60) or ncRNAs (61,62). Indeed, various mechanisms of such regulation are known to be abundant during development (63,64).

As mammals are viviparous organisms, some proteins are translated in maternal tissues and transported to the fetus. Therefore, the expression in fetal tissues may not necessarily represent the actual demands and may account for the different correlation values. To analyze this hypothesis, we examined whether this phenomenon is also observed in oviparous organisms which lay eggs and where the embryo translates all its proteins by itself. For that purpose, we have examined expression data from the central nervous system of the oviparous nematode *C. elegans*. We found that also in this organism, embryonic tissue exhibited significant lower correlation than the corresponding larval tissue ( $P = 0.0159$ , Wilcoxon's test; Supplementary Table S15). To assess our assumption that in our context larval tissues resemble adult tissues (and differ from fetal tissues), we compared between larval and adult tissues in the oviparous fruit fly *D. melanogaster*, not finding significant differences in tAI-GE correlations (Supplementary Table S16). These results suggest that also in some of the oviparous the tRNA pool is more adjusted to the expression regime at the adulthood. An interesting future work would be to further analyze the differences in TE between various developmental stages in different organisms.

Nevertheless, some limitations of the current analysis should be pointed out.

First, the tAI measure encapsulates within its definition the wobble rules (27). However, there are some post-transcriptional modifications which alter the specific binding of some tRNAs (65,66). When more data on these modifications will be available it will obviously be interesting to repeat the analysis shown here.

In addition, following previous studies in humans and other multicellular organisms (6,11,15,16), we used the genomic copy numbers of tRNA genes as a first approximation of their levels; indeed, tRNA copy numbers were previously shown to highly correlate with their expression levels in several unicellular organisms (18,67–69). This global measure, shared by all tissues, is adequate when analyzing global features of TE (such as the fetal/adult tissue comparison, multi-tissue optimization, the functional importance of genes and more). Yet, when analyzing tissue-specific aspects of TE, a more adequate measure would be the actual tissue-specific tRNA levels, which varies between tissues (26). However, such data are currently unavailable on a broad scope, therefore we used tRNA genomic copy number as approximation for their actual levels. While tRNA copy numbers are obviously identical across tissues, their usage leads to distinct tissue-specific TE estimations since each tissue

has a different GE pattern. Nevertheless, analysis based on inferred on tissue-specific tRNA levels for six tissues leads to results similar to those reported using the common pool, suggesting that tRNA genomic copy numbers can be used for first approximation. As tissue-specific tRNA levels data accumulates, it will be highly interesting to repeat our analysis: the tissue-specific TE scores presented here should possibly be viewed as lower bounds on the final TE values in humans, as the actual tissue-specific tRNA levels are probably further transcriptionally regulated to optimize TE.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank the anonymous referees for their valuable comments and suggestions.

## FUNDING

Edmond J. Safra Bioinformatics program at Tel-Aviv University (to Y.Y.W.); Koshland Scholar award from the Weizmann Institute of Science (to T.T.); Israeli Science Fund (ISF), the Ministry of Science and Technology (MOST) and the Tauber Fund (to E.R.); The Israel Science Foundation (grant number 385/06 to R.S.). Funding for open access charge: Tauber Fund.

*Conflict of interest statement.* None declared.

## REFERENCES

- Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
- dos Reis, M., Savva, R. and Wernisch, L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.
- Man, O. and Pilpel, Y. (2007) Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat. Genet.*, **3**, 415–421.
- Lithwick, G. and Margalit, H. (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res.*, **13**, 2665–2673.
- Duret, L. and Mouchiroud, D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **96**, 4482–4487.
- Duret, L. (2000) tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.*, **16**, 287–289.
- Duret, L. (2002) Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.*, **12**, 640–649.
- Chamary, J.V., Parmley, J.L. and Hurst, L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.*, **7**, 98–108.
- Gustafsson, C., Govindarajan, S. and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.*, **22**, 346–353.
- Kanaya, S., Yamada, Y., Kinochi, M., Kudo, Y. and Ikemura, T. (2001) Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J. Mol. Evol.*, **53**, 290–298.
- 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.
- Comeron, J.M. (2004) Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics*, **167**, 1293–1304.
- Urrutia, A.O. and Hurst, L.D. (2001) Codon usage covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics*, **159**, 1191–1199.
- Urrutia, A.O. and Hurst, L.D. (2003) The signature of selection mediated by expression on human genes. *Genome Res.*, **13**, 2260–2264.
- Lavner, Y. and Kotlar, D. (2005) Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene*, **345**, 127–138.
- Kotlar, D. and Lavner, Y. (2006) The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genomics*, **7**, 67.
- Wright, F. (1990) The effective number of codons used in a gene. *Gene*, **87**, 23–29.
- 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* system. *J. Mol. Biol.*, **151**, 389–409.
- Bennetzen, J.L. and Hall, B.D. (1982) Codon selection in yeast. *J. Mol. Chem.*, **257**, 3026–3031.
- Sharp, P.M. and Li, W.H. (1987) The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.
- Lercher, M.J., Urrutia, A.O., Pavlicek, A. and Hurst, L.D. (2003) A unification of mosaic structures in the human genome. *Hum. Mol. Genet.*, **12**, 2411–2415.
- Waldman, Y.Y., Tuller, T., Sharan, R. and Rupp, E. (2009) TP53 cancerous mutations exhibit selection for translation efficiency. *Cancer Res.*, **69**, 8807–8813.
- Tuller, T., Kupiec, M. and Rupp, E. (2007) Determinants of protein abundance and translation efficiency in *S. cerevisiae*. *PLoS Comput. Biol.*, **3**, e248.
- Plotkin, J.B., Robins, H. and Levine, A.J. (2004) Tissue-specific codon usage and the expression of human genes. *Proc. Natl Acad. Sci. USA*, **101**, 12588–12591.
- Sémon, M., Lobry, J.R. and Duret, L. (2006) No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol. Biol. Evol.*, **23**, 523–529.
- Dittmar, K.A., Goodenbour, J.M. and Pan, T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, e221.
- Crick, F.H. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
- Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.*, **28**, 292.
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. and Huber, W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
- Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.*, **37**, D93–D97.
- Nelder, J.A. and Mead, R. (1965) A simplex method for function minimization. *Computer J.*, **7**, 308–313.
- Lagarias, J.C., Reeds, J.A., Wright, M.H. and Wright, P.E. (1998) Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, **9**, 112–147.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M. and Edgar, R. (2006) NCBI GEO: mining tens of millions of

- expression profiles—database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.
35. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
  36. Su,A.I., Cooke,M.P., Ching,K.A., Hakak,Y., Walker,J.R., Wiltshire,T., Orth,A.P., Vega,R.G., Sapinosa,L.M., Moqrich,A. *et al.* (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA*, **99**, 4465–4470.
  37. Bossi,A. and Lehner,B. (2009) Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.*, **5**, 260.
  38. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Blake,J.A. (2008) Mouse Genome Database Group: The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
  39. Costantini,M., Clay,O., Auletta,F. and Bernardi,G. (2006) An isochore map of human chromosomes. *Genome Res.*, **16**, 5536–5541.
  40. Woodfine,K., Fiegler,H., Beare,D.M., Collins,J.E., McCann,O.T., Young,B.D., Debernardi,S., Mott,R., Dunham,I. and Carter,N.P. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, **13**, 191–202.
  41. Jensen-Seaman,M.I., Furey,T.S., Payseur,B.A., Lu,Y., Roskin,K.M., Chen,C.F., Thomas,M.A., Haussler,D. and Jacob,H.J. (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.*, **14**, 528–538.
  42. Cover,T.M. and Thomas,J.A. (1991) *Elements of Information Theory*. J. Wiley and sons, NY.
  43. Tuller,T., Kupiec,M. and Ruppin,E. (2008) Evolutionary rate and gene expression across different brain regions. *Genome Biol.*, **9**, R142.
  44. Dennis,G. Jr, Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.*, **4**, P3.
  45. Sharp,P.M., Averof,M., Lloyd,A.T., Matassi,G. and Peden,J.F. (1995) DNA sequence evolution: the sounds of silence. *Philos. Trans. R. Soc. Lond. B.*, **349**, 241–247.
  46. Eyre-Walker,A. (1996) Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? *Mol. Biol. Evol.*, **13**, 864–872.
  47. Duret,L. and Mouchiroud,D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **96**, 4482–4487.
  48. Coghlan,A. and Wolfe,K.H. (2000) Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast*, **16**, 1131–1145.
  49. The Gene Ontology Consortium. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
  50. Kandel,E.R., Schwartz,J.H. and Jessel,T.M. (2000) Principles of neural science, 4th edn. McGraw-Hill companies, New York, NY.
  51. Takahama,Y. (2006) Journey through the thymus: stromal guides for T-cell development and selection. *Nat. Rev. Immunol.*, **6**, 127–135.
  52. Rodgers,R.J., Irving-Rodgers,H.F. and Russell,D.L. (2003) Extracellular matrix of the developing ovarian follicle. *Reproduction*, **126**, 415–424.
  53. Santala,M., Simojoki,M., Risteli,J., Risteli,L. and Kauppila,A. (1999) Type I and III collagen metabolites as predictors of clinical outcome in epithelial ovarian cancer. *Clin. Cancer Res.*, **5**, 4091–4096.
  54. Sherman-Baust,C.A., Weeraratna,A.T., Rangel,L.B., Pizer,E.S., Cho,K.R., Schwartz,D.R., Shock,T. and Morin,P.J. (2003) Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. *Cancer Cell*, **3**, 377–386.
  55. De Cecco,L., Marchionni,L., Gariboldi,M., Reid,J.F., Lagonigro,M.S., Caramuta,S., Ferrario,C., Bussani,E., Mezzanzanica,D., Turatti,F. *et al.* (2004) Gene expression profiling of advanced ovarian cancer: characterization of a molecular signature involving fibroblast growth factor 2. *Oncogene*, **23**, 8171–8183.
  56. Haouem,S., Najjar,M.F., El Hani,A. and Sakly,R. (2008) Accumulation of cadmium and its effects on testis function in rats given diet containing cadmium-polluted radish bulb. *Exp. Toxicol. Pathol.*, **59**, 307–311.
  57. Li,T., Huang,J., Jiang,Y., Zeng,Y., He,F., Zhang,M.Q., Han,Z. and Zhang,X. (2009) Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. *Genomics*, **93**, 235–242.
  58. Hailesellasse Sene,K., Porter,C.J., Palidwor,G., Perez-Iratxeta,C., Muro,E.M., Campbell,P.A., Rudnicki,M.A. and Andrade-Navarro,M.A. (2007) Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics*, **8**, 85.
  59. Glisovic,T., Bachorik,J.L., Yong,J. and Dreyfuss,G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.
  60. Mittal,N., Roy,N., Babu,M.M. and Janga,S.C. (2009) Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proc. Natl Acad. Sci. USA*, **106**, 20300–20305.
  61. Mercer,T.R., Dinger,M.E. and Mattick,J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
  62. Mattick,J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.
  63. Amaral,P.P. and Mattick,J.S. (2008) Noncoding RNA in development. *Mamm. Genome*, **19**, 454–492.
  64. Dinger,M.E., Amaral,P.P., Mercer,T.R., Pang,K.C., Bruce,S.J., Gardiner,B.B., Askarian-Amiri,M.E., Ru,K., Soldà,G., Simons,C. *et al.* (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.*, **18**, 1433–1445.
  65. Agris,P.F. (2004) Decoding the genome: a modified view. *Nucleic Acids Res.*, **32**, 223–238.
  66. Agris,P.F., Vendeix,F.A. and Graham,W.D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.*, **366**, 1–13.
  67. Dong,H., Nilsson,L. and Kurland,C.G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.*, **260**, 649–663.
  68. Percudani,R., Pavesi,A. and Ottonello,S. (1997) Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J. Mol. Biol.*, **268**, 322–330.
  69. 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.