

SHORT REPORT

Open Access

Relationship between amino acid composition and gene expression in the mouse genome

Kazuharu Misawa^{1,2*}, Reiko F Kikuno³

Abstract

Background: Codon bias is a phenomenon that refers to the differences in the frequencies of synonymous codons among different genes. In many organisms, natural selection is considered to be a cause of codon bias because codon usage in highly expressed genes is biased toward optimal codons. Methods have previously been developed to predict the expression level of genes from their nucleotide sequences, which is based on the observation that synonymous codon usage shows an overall bias toward a few codons called major codons. However, the relationship between codon bias and gene expression level, as proposed by the translation-selection model, is less evident in mammals.

Findings: We investigated the correlations between the expression levels of 1,182 mouse genes and amino acid composition, as well as between gene expression and codon preference. We found that a weak but significant correlation exists between gene expression levels and amino acid composition in mouse. In total, less than 10% of variation of expression levels is explained by amino acid components. We found the effect of codon preference on gene expression was weaker than the effect of amino acid composition, because no significant correlations were observed with respect to codon preference.

Conclusion: These results suggest that it is difficult to predict expression level from amino acid components or from codon bias in mouse.

Background

Codon bias is a phenomenon that refers to the differences in the frequencies of occurrence of synonymous codons among different genes [1]. In the translation-selection model, natural selection is considered to be a cause of codon bias, because codon usage in highly expressed genes is biased toward “optimal” codons, i.e., codons corresponding to more abundant tRNAs in many organisms [2-6]. Methods have previously been developed to predict the expression level of genes from their nucleotide sequences, which is based on the observation that synonymous codon usage shows an overall bias toward a few codons called major codons [2-12].

Previous studies have provided clear evidence that the translation-selection model applies to some prokaryotes, such as *Escherichia coli* [6,13], but not to all bacteria

[4]. Additionally, some evidence exists that suggests this model is also applicable to various eukaryotes, including *Saccharomyces cerevisiae* [14-17], *Caenorhabditis elegans* [18,19], and the fruit fly [16,20], and even to the vertebrate *Xenopus laevis* [21]. However, the relationship between codon bias and expression level as proposed by the translation-selection model is less evident in mammals [22-30]. Urrutia and Hurst [23] found a weak correlation between gene expression levels and codon bias in human, but failed to find a relationship between this correlation and tRNA-gene copy numbers.

Amino acid content is also known to be dependent on gene expression level in some bacteria [31,32], as well as in budding yeast [8]. To determine why the relationship between codon bias and gene expression level, as proposed by the translation-selection model, is less evident in mammals, we investigated the correlations between the expression levels of genes and both the amino acid contents of genes and codon preference, in mouse. Subsequently, we compared the effect of gene expression on codon preference to the effect of gene expression on

* Correspondence: kazumisawa@riken.jp

¹Research Program for Computational Science, Research and Development Group for Next-Generation Integrated Living Matter Simulation, Fusion of Data and Analysis Research and Development Team, RIKEN, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Full list of author information is available at the end of the article

amino acid composition. We used the expression data of mouse genes contained in the InGap database [33].

Materials and methods

Mouse Genes

We obtained cDNA sequences of genes of *Mus musculus* from the ROUGE (<http://www.kazusa.or.jp/rouge/index.html>) database [34]. In total, 449,444 codons from 1,182 genes were used. Mouse expression data were retrieved from the InGap database using cDNA microarray [33].

Amino Acid Contents and Codon Preference

We calculated the proportion of the amino acid contents of all genes. In order to examine the translation-selection model, we classified amino acids into two classes, i.e., C- and T-adapted, on the basis of tRNA-gene copy numbers in the mouse genome, since tRNA-gene copy numbers can be considered as a rough estimate of tRNA abundance [26]. If natural selection is a cause of codon bias, codon usage in C-adapted amino acids of highly expressed genes will be biased toward C-ending codons and vice versa. First, we defined C-ending and T-ending codons; for instance, AGC is a C-ending codon and AGT is a T-ending codon. However, both encode the Ser residue. In the mouse genome, when the number of tRNAs complementary to C-ending codons for an amino acid is larger than the number of tRNAs that are complementary to T-ending codons, the amino acid is defined as a C-adapted amino acid. If the opposite is true, the amino acid is instead classified as a T-adapted amino acid. Furthermore, an amino acid is classified as T-adapted when the number of tRNAs that are complementary to C-ending codons is the same as the number of tRNAs that are complementary to T-ending codons. We obtained the number of tRNAs in the mouse genome from the GtRNAdb database [35]. Ser, Leu, Pro, Arg, Ile, Thr, Val, and Ala are T-adapted amino acids, whereas Phe, Tyr, Cys, His, Asn, Ser, Asp, and Gly are C-adapted amino acids. Of note, Ser is encoded by TCT, TCC, TCA, TCG, AGC, and AGT. The number of tRNAs that are complementary to TCT is larger than the number that are complementary to TCC, whereas the number of tRNAs that are complementary to AGT is smaller than the number that are complementary to AGC. We considered the two types of codons that specifically encoded Ser. We compared the expression levels of genes to the nucleotide composition at the 3rd position of the codons. We conducted this comparison for all amino acids, including the for T-adapted, and C-adapted amino acids

Because of CpG hypermutability, the mutation rates of codons are affected by the 3' adjacent codon [36,37]. Thus, the frequency of codon occurrence is dependent

on the adjacent amino acid [36,38,39]. We analyzed the effect of adjacent nucleotides on amino acid composition. Specifically, we calculated the correlation between the proportion of the first and third nucleotides of the 3' adjacent codon in genes and the expression levels of those genes.

Analysis

The Pearson product-moment correlation coefficients were calculated using *R* software [40]. Because the probability density functions of the amino acid contents, codon preference, and expression levels are not known, we used a Kendall test, which is a nonparametric correlation test. Some of highly expressed genes might have specific sequences and functions. Thus, we eliminated the outliers from the data; we defined outliers as both the 5% of genes with the highest expression levels and the 5% of genes with the lowest expression levels. We also conducted multiple-regression analysis.

Results

Correlation between amino acid contents and gene expression level

Figure 1 shows a scatter plot of amino acid contents and gene expression levels from the analysis of the mouse genome. Genes were sorted into bins of 50 genes by their expression level when this scatter plot was prepared. Subsequently, the 50 genes were concatenated as a single large gene, and the amino acid contents of the proteins were calculated. Each point on the plot represents a bin. For correlation analyses, we did not use these bins, but instead used each gene as a single entity. Figure 1 shows that the bin with highest expression level was exceptional. We also examined the sequence lengths, GC contents, and gene functions of highly expressed genes by using the ROUGE database [34], but we did not find any specific features.

Table 1 shows the correlation between amino acid composition and the gene expression level. The italicized numbers are the values that were calculated after eliminating the outliers. After the outliers were eliminated, the correlation test was performed. The contents of both Cys and His showed significant negative correlations with the expression level, whereas the content of Ile showed a significant positive correlation with the expression level. Multiple regression analysis showed that the multiple R^2 is 0.0797 and the adjusted R^2 is 0.06465 when all of amino acid components were used as predictors. After the outliers were eliminated, the multiple R^2 is 0.07193 and the adjusted R^2 is 0.0550 when all of amino acid components were used as predictors. These values of R^2 indicate that more than 90% of variation of expression levels cannot be explained by amino acid components.

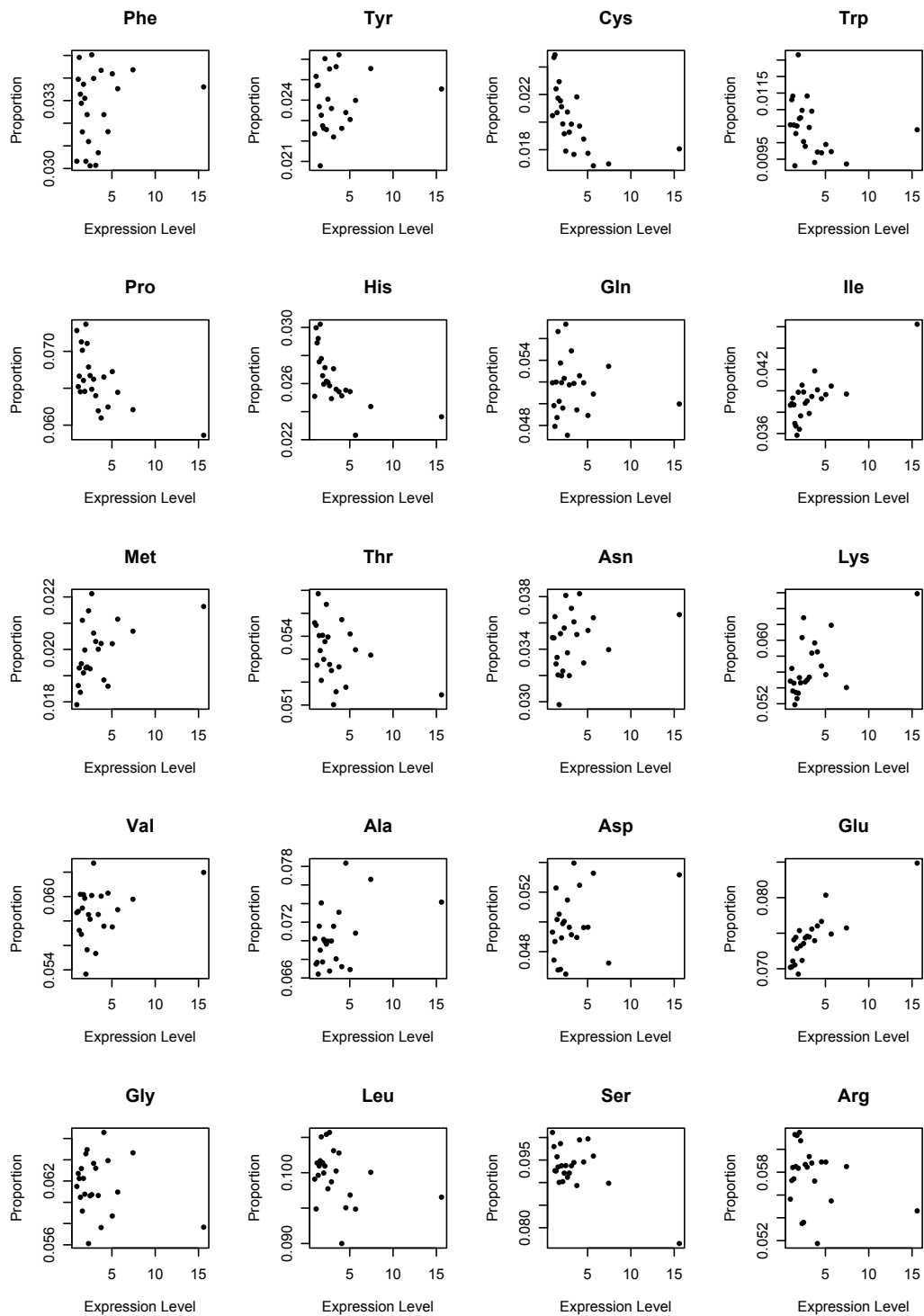


Figure 1 Correlation between amino acid composition and gene expression level. For preparation of this plot, we sorted the genes by their expression levels. The genes were sorted into bins of 50. Subsequently, the 50 genes were concatenated as a single large gene for analysis. Thereafter, the amino acid contents of the proteins were calculated. Each point on the plot represents a bin.

Table 1 Correlation between amino acid abundance and gene expression level

Amino Acid	Overall		After Elimination of Outliers	
	Correlation coefficient	P-value	Correlation coefficient	P-value
Phe	0.091	0.002 *	0.067	0.030
Leu	-0.060	0.040	-0.047	0.124
Ser	-0.123	0.000 **	-0.055	0.071
Tyr	0.046	0.114	0.059	0.055
Cys	-0.090	0.002 *	-0.118	0.000 **
Trp	-0.052	0.074	-0.031	0.319
Pro	-0.080	0.006 *	-0.062	0.045
His	-0.125	0.000 **	-0.169	0.000 **
Gln	-0.053	0.067	-0.010	0.757
Arg	-0.061	0.036	-0.033	0.281
Ile	0.135	0.000 **	0.084	0.006 *
Met	0.060	0.040	0.054	0.081
Thr	-0.040	0.174	-0.045	0.141
Asn	0.065	0.025	0.030	0.328
Lys	0.091	0.002 *	0.031	0.308
Val	0.091	0.002 *	0.069	0.024
Ala	0.037	0.199	0.072	0.018
Asp	0.101	0.000 **	0.054	0.079
Glu	0.084	0.004 *	0.066	0.032
Gly	0.004	0.884	0.018	0.567

* Significant at 1% level.

** Significant at 0.1% level.

Codon preference and gene expression level

Table 2 shows the correlation between gene expression levels and codon preference among mouse genes. This data revealed no significant correlations between codon preference and gene expression level. Thus, the effect of

Table 2 Correlation between the nucleotide composition at the 3rd position of codons and gene expression level

3' adjacent nucleotide	Overall	After elimination of outliers
All amino acids		
T	-0.063	-0.064
C	-0.061	-0.054
A	-0.068	-0.048
G	0.020	0.019
T-adapted amino acids		
T	-0.069	-0.068
C	-0.051	-0.033
A	-0.065	-0.050
G	0.012	0.000
C-adapted amino acids		
T	-0.049	-0.061
C	-0.061	-0.061
A	-0.054	-0.032
G	0.023	0.034

codon preference on gene expression was weaker than the effect of amino acid composition.

Table 3 shows the observed number of combinations of nucleotides at the third position and their 3' adjacent nucleotides. From this table, we can see the number of Cs at the third position of codons is significantly smaller than that of Ts when the 3' adjacent nucleotide is G ($p < 0.1\%$, chi-square test). This finding indicates that codon preference in the mouse genome is affected by CpG hypermutability.

Discussion

Variation of gene expression level

There is a large variation among gene expression level [33]. More than 90% of variation of expression levels cannot be explained by amino acid components. Gene expression levels are known to be affected by many factors, such as 3'UTR lengths [41]. Further study must be necessary.

Amino acid contents and gene expression level

To our knowledge, this is the first study that showed amino acid composition depends on the gene expression level in mouse. Previous study has shown that, in the case of budding yeast, some residues showed a positive correlation, and most of these residues were small [8]. Furthermore, Akashi and Gojobori [31] showed an increase in the abundance of less energetically costly amino acids in highly expressed proteins. This study also suggested that natural selection for energetic efficiency appears to constrain the primary structures of the proteins of *Bacillus subtilis* and *E. coli* [31]. Amino acid mutations that do not cause changes in protein functions may result in subtle, but evolutionarily important, fitness consequences through their effects on translation and metabolism.

We compared the estimates of the cost of amino acid synthesis from the above mentioned study [31] to the

Table 3 Observed number of combinations of nucleotide at the third position and their 3' adjacent nucleotide

Codon type	Third nucleotide	The 3' adjacent nucleotide				
		T	C	A	G	H
All	T	45922	71142	48193	135728	165257
	C	88111	113015	141307	45671	** 342433
T-adapted	T	24963	42070	22892	64855	89925
	C	44589	54213	72472	21423	** 171274
C-adapted	T	20959	29072	25301	70873	75332
	C	43522	58802	68835	24248	** 171159

H stands for A, C, or T.

** Significantly lower than the number of Cs when the 3' adjacent nucleotide is H ($p < 0.1\%$) after Bonferroni correction when a chi-square test was used.

correlation coefficients presented in Table 1 (data not shown); however, we determined that the correlation was insignificant. It may be difficult to estimate the accurate metabolic cost of each amino acid, because the mouse obtains amino acids from food. Furthermore, the cost may depend on the environment. In the case of mouse, 10 amino acids, namely, Arg, His, Ile, Leu, Lys, Met, Phe, Thr, Try, and Val, are essential for natural growth [42]. Thus, sparing the incorporation of His and Ile in highly expressed proteins may be advantageous to the mouse. However, Cys is not an essential amino acid, and is negatively correlated with gene expression. Of note, both amino acid composition and gene expression level may be influenced by protein functions. Furthermore, adaptive changes in protein sequences may overcome the increases in the metabolic cost, and the amino acid sequences may not be optimized for metabolic cost. Further study is necessary to elucidate these issues. Our results show that the coefficient of determination is very small so that it would be hard to predict expression level from amino acid contents in mammals.

Codon preference and gene expression level

We determined that the effect of codon preference on gene expression was weaker than the effect of amino acid composition, because no significant correlations were observed with respect to codon preference. This result is consistent with the relationship between codon bias and expression level, as proposed by the translation-selection model, is less evident in mammals [22-30]. In mammals, it would also be hard to predict expression level from codon bias.

CpG hypermutability

Hypermutability of CpG dinucleotides [43] is one of major causes of codon substitution in mammalian genes [44-48]. CpG dinucleotides are often methylated at sites of cytosine (C); subsequently, the methylated C spontaneously deaminates to thymine (T) with a higher frequency than that of other types of point mutations [49]. It has previously been estimated that approximately 14% of codon substitutions are caused by hypermutations at CpG sites [36]. Furthermore, CpG hypermutation has been shown to affect the rate of amino acid substitution [39].

Table 2 shows that gene expression levels do not significantly affect codon preference in mouse. Furthermore, Table 3 indicates that the effect of codon preference is weaker than that of CpG hypermutability. Thus, the relationship between codon bias and gene expression level can be explained on the basis of the translation-selection model [2]. This model proposes that codon usage in highly expressed genes is biased toward "optimal" codons, i.e., codons corresponding to more abundant tRNAs. This bias has been demonstrated

to affect both elongation rate and accuracy [50,51]. As shown in Table 2, the calculated negative correlation indicates that the codons used in this study are not optimal. In human and mouse genomes, the most frequently used codons [52] are not those with the most abundant tRNAs [35].

Recent studies [36,39] have shown that CpG mutation rates in the non-coding regions of the human genome negatively correlate with the local GC content [53-56]. Isochores of the human genome [57] appear to be an influential factor that affects codon composition [53-56], and several studies have shown that this factor is related to gene expression levels [58,59]. However, additional studies are necessary to confirm the relationship between codon bias and the positional effect of genes.

Plotkin *et al.* [24] showed that codon usage for tissue-specific genes varies among the tissues in which such genes are expressed, thereby suggesting that this variation may be affected by differential tRNA-gene copy numbers in different tissues. However, this variability in codon usage among tissues is still under debate [22,60,61]. Nevertheless, it is noteworthy that codon substitutions are affected by adjacent codons [36,39], and are therefore indirectly affected by adjacent amino acids [38]. Amino acid frequencies may also be tissue specific, although additional studies are necessary to investigate the effect of CpG hypermutability on tissue-specific codon usage. Furthermore, codon bias in mammalian genomes should also be investigated with regard to the presence of CpG nucleotides [27,28,30].

Conclusion

In mouse, the effect of gene expression level on codon bias is weaker than both the effect of gene expression level on amino acid composition and the effect of CpG hypermutability on codon bias. However, to detect the effect of gene expression level on codon bias in mouse, a study of more genes is necessary.

Acknowledgements

We thank Dr. Osamu Ohara and all members of the Human Genetics Laboratory of Kazusa DNA research institutes for their encouragement and useful comments on the manuscript. The present study was supported by the National Project on "Next-generation Integrated Living Matter Simulation" of the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Author details

¹Research Program for Computational Science, Research and Development Group for Next-Generation Integrated Living Matter Simulation, Fusion of Data and Analysis Research and Development Team, RIKEN, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan. ²Chiba Industry Advancement Center, 2-6 Nakase, Mihama-ku, Chiba 261-7126, Japan. ³Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan.

Authors' contributions

KM wrote the software and the manuscript. RFK supervised the project. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 25 June 2010 Accepted: 27 January 2011

Published: 27 January 2011

References

- Miyata T, Hayashida H, Yasunaga T, Hasegawa M: **The preferential codon usages in variable and constant regions of immunoglobulin genes are quite distinct from each other.** *Nucleic Acids Res* 1979, **7**(8):2431-2438.
- Akashi H, Eyre-Walker A: **Translational selection and molecular evolution.** *Curr Opin Genet Dev* 1998, **8**(6):688-693.
- Willie E, Majewski J: **Evidence for codon bias selection at the pre-mRNA level in eukaryotes.** *Trends Genet* 2004, **20**(11):534-538.
- Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE: **Variation in the strength of selected codon usage bias among bacteria.** *Nucleic Acids Res* 2005, **33**(4):1141-1153.
- Karlin S, Barnett MJ, Campbell AM, Fisher RF, Mrazek J: **Predicting gene expression levels from codon biases in alpha-proteobacterial genomes.** *Proc Natl Acad Sci USA* 2003, **100**(12):7313-7318.
- Roymondal U, Das S, Sahoo S: **Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome.** *DNA Res* 2009, **16**(1):13-30.
- Henry I, Sharp PM: **Predicting gene expression level from codon usage bias.** *Mol Biol Evol* 2007, **24**(1):10-12.
- Raghava GP, Han JH: **Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein.** *BMC Bioinformatics* 2005, **6**:59.
- Raghava GP, Han JH, Hwang DJ: **ECGpred: Correlation and prediction of gene expression from nucleotide sequence.** *The Open Bioinformatics Journal* 2008, **2**:64-71.
- Jansen R, Bussemaker HJ, Gerstein M: **Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models.** *Nucleic Acids Res* 2003, **31**(8):2242-2251.
- Beer MA, Tavazoie S: **Predicting gene expression from sequence.** *Cell* 2004, **117**(2):185-198.
- Coghlan A, Wolfe KH: **Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae.** *Yeast* 2000, **16**(12):1131-1145.
- Ikemura T: **Codon usage and tRNA content in unicellular and multicellular organisms.** *Mol Biol Evol* 1985, **2**(1):13-34.
- Bennetzen JL, Hall BD: **Codon selection in yeast.** *J Biol Chem* 1982, **257**(6):3026-3031.
- Ikemura T: **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* 1982, **158**(4):573-597.
- Akashi H: **Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA.** *Genetics* 1995, **139**(2):1067-1076.
- Akashi H: **Translational selection and yeast proteome evolution.** *Genetics* 2003, **164**(4):1291-1303.
- Duret L: **tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes.** *Trends Genet* 2000, **16**(7):287-289.
- Marais G, Duret L: **Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans.** *J Mol Evol* 2001, **52**(3):275-280.
- Moriyama EN, Powell JR: **Codon usage bias and tRNA abundance in Drosophila.** *J Mol Evol* 1997, **45**(5):514-523.
- Musto H, Cruveiller S, D'Onofrio G, Romero H, Bernardi G: **Translational selection on codon usage in Xenopus laevis.** *Mol Biol Evol* 2001, **18**(9):1703-1707.
- Urrutia AO, Hurst LD: **Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection.** *Genetics* 2001, **159**(3):1191-1199.
- Urrutia AO, Hurst LD: **The signature of selection mediated by expression on human genes.** *Genome Res* 2003, **13**(10):2260-2264.
- Plotkin JB, Robins H, Levine AJ: **Tissue-specific codon usage and the expression of human genes.** *Proc Natl Acad Sci USA* 2004, **101**(34):12588-12591.
- Lavner Y, Kotlar D: **Codon bias as a factor in regulating expression via translation rate in the human genome.** *Gene* 2005, **345**(1):127-138.
- Kotlar D, Lavner Y: **The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids.** *BMC Genomics* 2006, **7**:67.
- Hellmann I, Zollner S, Enard W, Ebersberger I, Nickel B, Paabo S: **Selection on human genes as revealed by comparisons to chimpanzee cDNA.** *Genome Res* 2003, **13**(5):831-837.
- Kondrashov FA, Ogurtsov AY, Kondrashov AS: **Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites.** *J Theor Biol* 2006, **240**(4):616-626.
- Subramanian S: **Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes.** *Genetics* 2008, **178**(4):2429-2432.
- dos Reis M, Wernisch L: **Estimating translational selection in eukaryotic genomes.** *Mol Biol Evol* 2009, **26**(2):451-461.
- Akashi H, Gojobori T: **Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis.** *Proc Natl Acad Sci USA* 2002, **99**(6):3695-3700.
- Herbeck JT, Wall DP, Wernegreen JJ: **Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont Wigglesworthia.** *Microbiology* 2003, **149**(Pt 9):2585-2596.
- Koga H, Yuasa S, Nagase T, Shimada K, Nagano M, Imai K, Ohara R, Nakajima D, Murakami M, Kawai M, et al: **A comprehensive approach for establishment of the platform to analyze functions of KIAA proteins II: public release of inaugural version of InGaP database containing gene/protein expression profiles for 127 mouse KIAA genes/proteins.** *DNA Res* 2004, **11**(4):293-304.
- Kikuno R, Nagase T, Nakayama M, Koga H, Okazaki N, Nakajima D, Ohara O: **HUGE: a database for human KIAA proteins, a 2004 update integrating HUGEpPi and ROUGE.** *Nucleic Acids Res* 2004, **32** Database: D502-504.
- Chan PP, Lowe TM: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Res* 2009, **37** Database: D93-97.
- Misawa K, Kikuno RF: **Evaluation of the effect of CpG hypermutability on human codon substitution.** *Gene* 2009, **431**(1-2):18-22.
- Eyre-Walker AC: **An analysis of codon usage in mammals: selection or mutation bias?** *J Mol Evol* 1991, **33**(5):442-449.
- Wang GZ, Chen LL, Zhang HY: **Neighboring-site effects of amino acid mutation.** *Biochem Biophys Res Commun* 2007, **353**(3):531-534.
- Misawa K, Kamatani N, Kikuno RF: **The universal trend of amino acid gain-loss is caused by CpG hypermutability.** *J Mol Evol* 2008, **67**(4):334-342.
- R Development Core Team: **R: A language and environment for statistical computing.** Vienna, Austria; 2008.
- Okazaki N, Imai K, Kikuno RF, Misawa K, Kawai M, Inamoto S, Ohara R, Nagase T, Ohara O, Koga H: **Influence of the 3'-UTR-length of mKIAA cDNAs and their sequence features to the mRNA expression level in the brain.** *DNA Res* 2005, **12**(3):181-189.
- John AM, Bell JM: **Amino acid requirements of the growing mouse.** *J Nutr* 1976, **106**(9):1361-1367.
- Bird AP: **DNA methylation and the frequency of CpG in animal DNA.** *Nucleic Acids Res* 1980, **8**(7):1499-1504.
- Jukes TH: **Codons and nearest-neighbor nucleotide pairs in mammalian messenger RNA.** *J Mol Evol* 1978, **11**(2):121-127.
- Karlin S, Mrazek J: **What drives codon choices in human genes?** *J Mol Biol* 1996, **262**(4):459-472.
- Krajewski C, Blacket M, Buckley L, Westerman M: **A multigene assessment of phylogenetic relationships within the dasyurid marsupial subfamily Sminthopsinae.** *Mol Phylogenet Evol* 1997, **8**(2):236-248.
- Huttley GA: **Modeling the impact of DNA methylation on the evolution of BRCA1 in mammals.** *Mol Biol Evol* 2004, **21**(9):1760-1768.
- Lunter G: **Probabilistic whole-genome alignments reveal high indel rates in the human and mouse genomes.** *Bioinformatics* 2007, **23**(13):i289-296.
- Scarano E, Iaccarino M, Grippo P, Parisi E: **The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos.** *Proc Natl Acad Sci USA* 1967, **57**(5):1394-1400.
- Bulmer M: **The selection-mutation-drift theory of synonymous codon usage.** *Genetics* 1991, **129**(3):897-907.
- Akashi H: **Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy.** *Genetics* 1994, **136**(3):927-935.
- Nakamura Y, Gojobori T, Ikemura T: **Codon usage tabulated from international DNA sequence databases: status for the year 2000.** *Nucleic Acids Res* 2000, **28**(1):292.

53. Fryxell KJ, Moon WJ: CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol* 2005, **22**(3):650-658.
54. Taylor J, Tyekucheva S, Zody M, Chiaromonte F, Makova KD: Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol Biol Evol* 2006, **23**(3):565-573.
55. Tyekucheva S, Makova KD, Karro JE, Hardison RC, Miller W, Chiaromonte F: Human-macaque comparisons illuminate variation in neutral substitution rates. *Genome Biol* 2008, **9**(4):R76.
56. Walser JC, Ponger L, Furano AV: CpG dinucleotides and the mutation rate of non-CpG DNA. *Genome Res* 2008, **18**(9):1403-1414.
57. Bernardi G: The vertebrate genome: isochores and evolution. *Mol Biol Evol* 1993, **10**(1):186-204.
58. Vinogradov AE: Isochores and tissue-specificity. *Nucleic Acids Res* 2003, **31**(17):5212-5220.
59. Vinogradov AE: Noncoding DNA, isochores and gene expression: nucleosome formation potential. *Nucleic Acids Res* 2005, **33**(2):559-563.
60. Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J, Dillon W, Lee KF, Clark KE, Haverly P, et al: A compendium of gene expression in normal human tissues. *Physiol Genomics* 2001, **7**(2):97-104.
61. Semon M, Lobry JR, Duret L: No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol Biol Evol* 2006, **23**(3):523-529.

doi:10.1186/1756-0500-4-20

Cite this article as: Misawa and Kikuno: Relationship between amino acid composition and gene expression in the mouse genome. *BMC Research Notes* 2011 **4**:20.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

