

Tissue Specific Expression Levels of Apoptosis Involved Genes Have Correlations with Codon and Amino Acid Usage

Mohammadreza Hajjari^{1§*}, Iman Sadeghi^{2§}, Abbas Salavaty^{1§}, Habib Nasiri³,
Maryam Tahmasebi Birgani^{4**}

¹Department of Genetics, Faculty of Sciences, Shahid Chamran University of Ahvaz, Ahvaz 61357-83151, Iran,

²Department of Molecular Genetics, Faculty of Biosciences, Tarbiat Modares University of Tehran, Tehran 14115116, Iran,

³Department of Medical Genetics, Nika Center of Preventive Medicine and Health Promotion, Tehran 1418944711, Iran,

⁴Department of Medical Genetics, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz 61357-15794, Iran

Different mechanisms, including transcriptional and post transcriptional processes, regulate tissue specific expression of genes. In this study, we report differences in gene/protein compositional features between apoptosis involved genes selectively expressed in human tissues. We found some correlations between codon/amino acid usage and tissue specific expression level of genes. The findings can be significant for understanding the translational selection on these features. The selection may play an important role in the differentiation of human tissues and can be considered for future studies in diagnosis of some diseases such as cancer.

Keywords: apoptosis, codon, gene expression regulation, translation

Introduction

Regulation of gene expression level is an important step in different biological processes [1, 2]. The expression levels are the overall results of the transcriptional and post-transcriptional regulations [1, 3]. Codon usage bias usually refers to different frequencies of synonymous codons attributed to one specific amino acid [4, 5]. Synonymous codons translate with different speeds; an event which indicates that the translational efficiency may be specifically modified by using different codons [6]. Furthermore, in some prokaryotes and eukaryotes, amino acid content is also known to be related with gene expression level [7]. Two major factors including natural selection and mutational bias are reported to explain such codon and amino acid usage bias [8]. However, this relationship, as proposed by the translation selection model, is less evident in mammals.

There is a clear relation between the expression levels of genes and codon composition in most organisms [5]. Nonetheless, due to tissue-specific gene expression, identification of factors affecting gene expression in multicellular organisms is much more complicated [9]. Thus, some studies have focused on the relation between tissue specific gene expression and compositional features of the genes. Sémon *et al.* [10] reported that selective pressure influenced the synonymous codon usage of human tissue-specific genes and modify the expression of special proteins. Also, Zhou *et al.* [9] identified some significant correlations between tissues specific oncogenes and sequence compositional features in human tissues. Then, Hajjari *et al.* [11] found that the structural elements in tumor suppressor genes and proteins can play important roles in their regulation. However, it is still unclear whether translational selection affects other human genes specifically based on their tissue expression patterns [12].

Received November 3, 2016; Revised November 25, 2016; Accepted December 1, 2016

*Corresponding author: Tel: +98-6133330010, Fax: +98-6133338965, E-mail: Mohamad.hajjari@gmail.com, m-hajjari@scu.ac.ir

**Corresponding author: Tel: +98-6133330010, Fax: +98-6133338965, E-mail: tahmasebi-ma@ajums.ac.ir

§Mohammadreza Hajjari, Iman Sadeghi, and Abbas Salavaty contributed equally to this work.

Copyright © 2016 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

Apoptosis, the process of programmed cell death, is considered as a crucial component of different processes such as normal cell cycle, development, and function of the immunity system [13]. Different genes controlling the apoptosis process have been identified. Since the expression levels of genes involved in apoptosis are different in human tissues, we aimed to find if the compositional features of these genes are involved in their regulation. Considering the important role of these genes in cellular and molecular function, the results may provide new clues about their tissue specific regulation. Since the dysregulation of these genes is a common event in some diseases especially cancers, the unraveling different mechanisms of their regulation may also reveal novel strategies for therapeutic intervention.

In this study, we tried to elucidate the potential correlation between some compositional features of genes/proteins structures and their expression level in some human tissues. The results may reveal the support for the relation between codon usage, amino acid usage and tissue specific gene expression.

Methods

Sequence retrieval, alignments, and expression data acquisition

The list of apoptosis involved genes coding proteins with different molecular functions and localizations in the cell were drawn from the Deathbase (<http://deathbase.org/>) (Supplementary Table 1). In order to minimize the statistical errors, multiple alignments were performed for the sequences by ClustalW program (<http://www.ebi.ac.uk>), and finally, 65 genes were selected for our study. Normalized expressions of genes were directly drawn from SOURCE, 2014 (Stanford Microarray) database (Source.stanford.edu) (Supplementary Table 1). The normalized gene expression in this database presents the relative expression level of a gene (defined as a UniGene Cluster) in different tissues and is “normalized” for the number of clones from each tissue that are included in UniGene (<http://source-search.princeton.edu/help/SOURCE/normalization.html>). This database links to the microarray experiments that included the queried genes. Thirty four different tissues were selected and in each tissue, the average expression level, the number of expressed genes, and the highest expressed gene were recognized.

Sequence compositional features

The codon usage table of *Homo sapiens* was obtained (<http://www.kazusa.or.jp/codon/>), and used to calculate Codon Adaptation Index (CAI) (<http://genomes.urv.es/CAIcal/>) for each coding domain sequence (CDS). The tRNA Adaptation Index (tAI) was also measured for each CDS

using Visual Gene Developer 1.4 Build 750. The frequencies per thousand of all of 61 codons were calculated for all 65 CDSs by Countcodon program (<http://www.kazusa.or.jp/codon/countcodon.html>). Also, The Relative Synonymous Codon Usage (RSCU) of all amino acids was measured (http://www.bioinformatics.org/sms2/codon_usage) for all desired proteins.

Amino acid sequence characteristics

The protein sequences of all 65 genes were obtained from NCBI Resource Portal. Then, the percentage of each amino acid was obtained for all of the proteins using Protein Information Resource (<http://pir.georgetown.edu>).

Statistical analyses

For each tissue, the correlation between gene expression level of desired genes and compositional features of CDS/protein was analyzed by Graphpad. p-values less than 0.01 were considered as significant. However, in some tissues, because of lacking the p-value less than 0.01, we considered the characteristics with p-values less than 0.05 as the most significant features. Finally, in order to determine truly significant features, false discovery rate (FDR) was analyzed through the FDR online calculator (<http://www.sdmproject.com/utilities/?show=FDR>) which its method coincides with the R code of the version proposed by Benjamini and Hochberg.

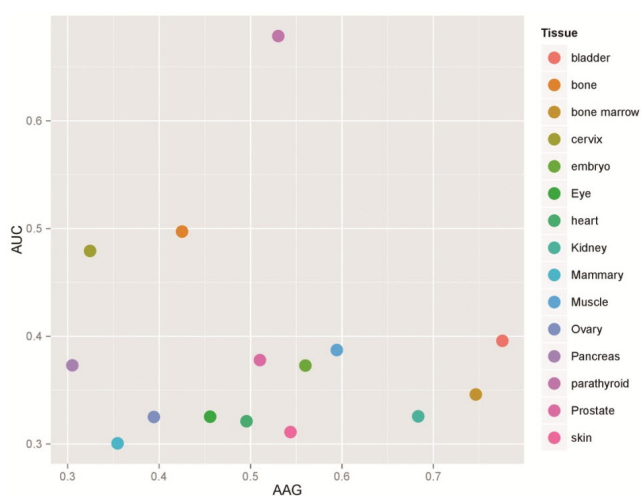


Fig. 1. The tissues in which both AAG and AUC have significant correlations with the gene expression level. The bar plot is drawn by the R statistical software. $p < 0.05$ is considered statistically significant.

Results

Correlation between gene compositional features and expression level

The correlation analyses showed that the gene compositional features are associated with the expression levels of desired genes. The results indicated that the expression levels have significant correlation with the frequencies of some codons such as AAG (in 20 tissues), AUC (in 17 tissues), and GAC (in 12 tissues) (Supplementary Table 2). Among these codons, AAG and AUC showed the most significant correlation coefficients. In order to identify the tissues in which both AAG and AUC have significant correlations with the gene expression level, a bar plot was drawn by R software for these codons (Fig. 1). In Supplementary Table 2, we can see that the most frequent and attributed tRNAs are attributed to the AAG-Lys, AUC-Ile and GAC-Asn codons.

Furthermore, our data showed that the expression levels of apoptosis genes have significant correlations with the relative synonymous codon usage features such as CCC (Prolin) and TCC (Serine) (Supplementary Table 3). Since the aforementioned codons have the most significant correlation coefficients, a bar plot was drawn by R software for these features in order to identify the tissues in which both codons have significant correlation coefficients with gene expression (Fig. 2).

To find the level of codon bias, CAI for each gene was measured. We found some correlations between these parameters and the expression level of desired genes in 8

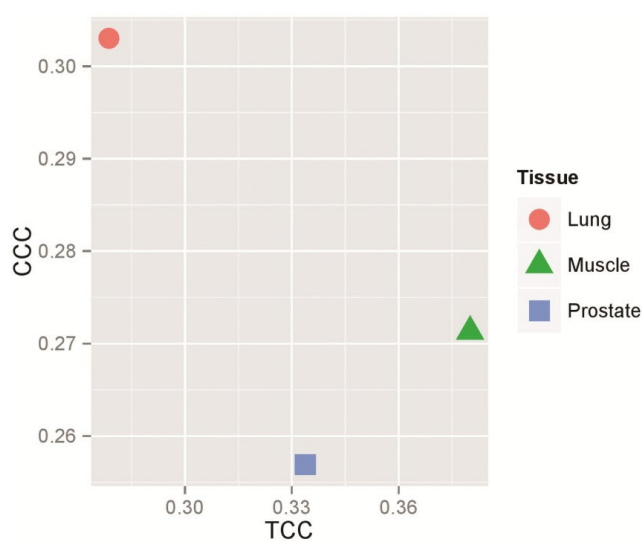


Fig. 2. The tissues in which the codons CCC (proline) and TCC (serine) have the most significant correlation coefficients with gene expression levels. The bar plot is drawn by R statistical software. $p < 0.05$ is considered statistically significant.

tissues ($p < 0.05$). Furthermore, we calculated the tAI, which is a measure of the tRNA usage by coding sequences. Significant correlations between tAI and the expression level of genes were reported in 20 tissues (Supplementary Table 4). Also, the results indicated that nucleotide compositional features (including CG percent in codon bases) have significant correlations with gene expression levels in some tissues (Supplementary Table 4).

Correlation between protein compositional features and the expression levels

Our data indicated that different amino acids including Arg, Asp, Lys, Glu, Gln, Ser, and Ile have significant correlations with the expression level of genes in 15, 14, 12, 11, 11, 9, and 8 tissues respectively (Supplementary Table 5). Among these notable amino acids, Asp and Lys have the most significant positive correlation coefficients. So, a bar plot was drawn by R software for these amino acids in order to identify the tissues in which both Asp and Lys have significant correlation coefficients with gene expression level (Fig. 3).

Most significant and truly significant features in different tissues

In order to find out which feature has the most significant correlation with the gene expression level in different tissues, the features with the least p-value were recognized (Table 1, Fig. 4). Also, since multiple correlations were done in the current study, the FDR analysis was done to decrease

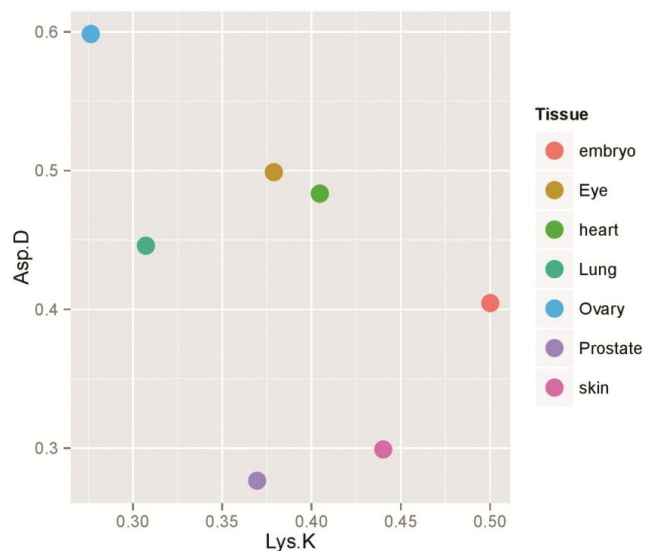


Fig. 3. The tissues in which both Asp and Lys have the most significant positive correlation coefficients with gene expression levels. The bar plot is drawn by R statistical software. $p < 0.05$ is considered statistically significant.

Table 1. Features with the most significant correlation with gene expression levels in different tissues

Tissue	Feature	p-value	Correlation coefficient
Bladder	AAG ^a	2.99E-07	0.7755243
Blood	CTA ^a	3.35E-05	0.528577
Bone marrow	AAG ^a	2.63E-07	0.7464197
Bone	Asp ^b	6.56E-05	0.5536896
Brain	AAG ^a	0.00262809	0.3815733
Connective	Asp ^b	0.00989447	0.3546238
Embryo	AAG ^a	4.26E-05	0.5600275
Esophagus	GAC ^a	0.01866616	0.4415246
Eye	Asp ^b	0.00030731	0.4988698
Heart	AAG ^a	0.00062401	0.4956311
Intestine	Asp ^b	0.0030745	0.3853588
Kidney	AAG ^a	4.67E-09	0.6834499
Liver	CGC ^c	0.00145176	0.4303177
Lung	Asp ^b	0.0004521	0.4458651
Lymph node	GUC ^a	0.01219837	-0.3452
Lymph	ATC ^a	0.00045155	0.5760362
Mammary	Asp ^b	0.00021225	0.5101331
Muscle	AAG ^a	2.70E-06	0.5943742
Nerve	UGC ^a	0.00860171	0.5238741
Ovary	Asp ^b	7.04E-06	0.5984444
Pancreas	Asp ^b	9.47E-07	0.6105691
Parathyroid	AUC ^a	0.0005163	0.6786512
Placenta	CCC ^c	0.02260426	0.3070281
Prostate	AAG ^a	5.88E-05	0.5102705
Skin	AAG ^a	1.78E-05	0.5437638
Spleen	TGC ^a	0.00221411	0.4699773
Stomach	Glu ^b	0.01671794	0.363057
Testis	ATT ^a	0.00091835	0.4382404
Thymus	CCC ^a	0.05080796	0.2835639
Thyroid	Gly ^b	0.00057388	-0.8010838
Uterus	Asp ^b	0.00820454	0.3468847
Vascular	GCC ^c	0.00576327	0.4051663

FDR, false discovery rate.

^aCodon frequency; ^bAmino acid frequency; ^cCodon usage feature.

the statistical errors (Table 2). Through this analysis, we calculated the corrected p-values in order to find truly significant feature. By FDR analysis, we found some truly significant features such as AAG (coding lysine) and GAC (coding aspartate) codons (drawn from codon frequencies), lysine and aspartate (drawn from amino acid frequencies), and CTC^L (from synonymous codon usage features) which are common among some tissues (Table 2).

Discussion

The relation between the expression level and the gene/protein compositional features has been reported in different organisms [5]. However, little is known about the factors affecting the tissue specific gene expression level in

multicellular organisms [9]. In the last years, the effects of codon bias and other compositional features on the tissue-specific translational control of genes have been reported [1, 11]. In the current study, in order to elucidate this level of regulation on the expression of apoptosis involved genes, we investigated the relation between the gene expression and structural features. The results can help us understand the molecular mechanism and regulation of this type of genes. Our data suggest that synonymous codon usage in human genes may not be the result of isochore organization of the genome or neutral evolutionary processes as well. We found some common features correlated with gene expression in different tissues, a finding that may indicate common mechanisms in gene regulation. The current results suggest several hypotheses about the mechanisms of gene regulation and tissue differentiation in human.

There are some reports indicating that codon usage in mammals has notable effects on translation rate [14], especially during cell differentiation [15]. The systematic tissue-specific codon usage may indicate that human tissues can differ in the relative tRNA amount [16]. This may influence on the expression of the desired proteins. The significant correlation between gene expression level and some codon frequencies especially AAG codon (truly significant in nine tissues) may indicate the importance of the abundance of attributed tRNAs pairing with this codon. The correlation between gene expression and frequency of AAG codon was interestingly approved in a previous study by Hajjari *et al.* [11] on Tumor suppressor genes.

Our data can support the Plotkin *et al.*'s study on synonymous codon usage features [3]. Our data on the significant correlations between tAI and expression levels of genes may support this hypothesis. The relation between synonymous codon usage bias and gene expression levels might be the result of the weak selection on synonymous sites which generate translationally optimal codons [17, 18]. Among our data, the correlation between the synonymous codon usage of leucin and gene expression level is of noted. In a similar study on tumor suppressor genes, this type of correlation was also observed [11]. So, these data altogether may indicate the importance of synonymous codon usage of Leucine in the structure of genes. Altogether, some correlations between relative synonymous codon usage, codon frequency, tAI, CAI, and gene expression show that there may be a translational selection on apoptosis involved genes. Since some of these features are common between different tissues, we assume that common mechanisms may be involved in some specific tissues in this level of regulation.

Some previous studies declare that the amino acid

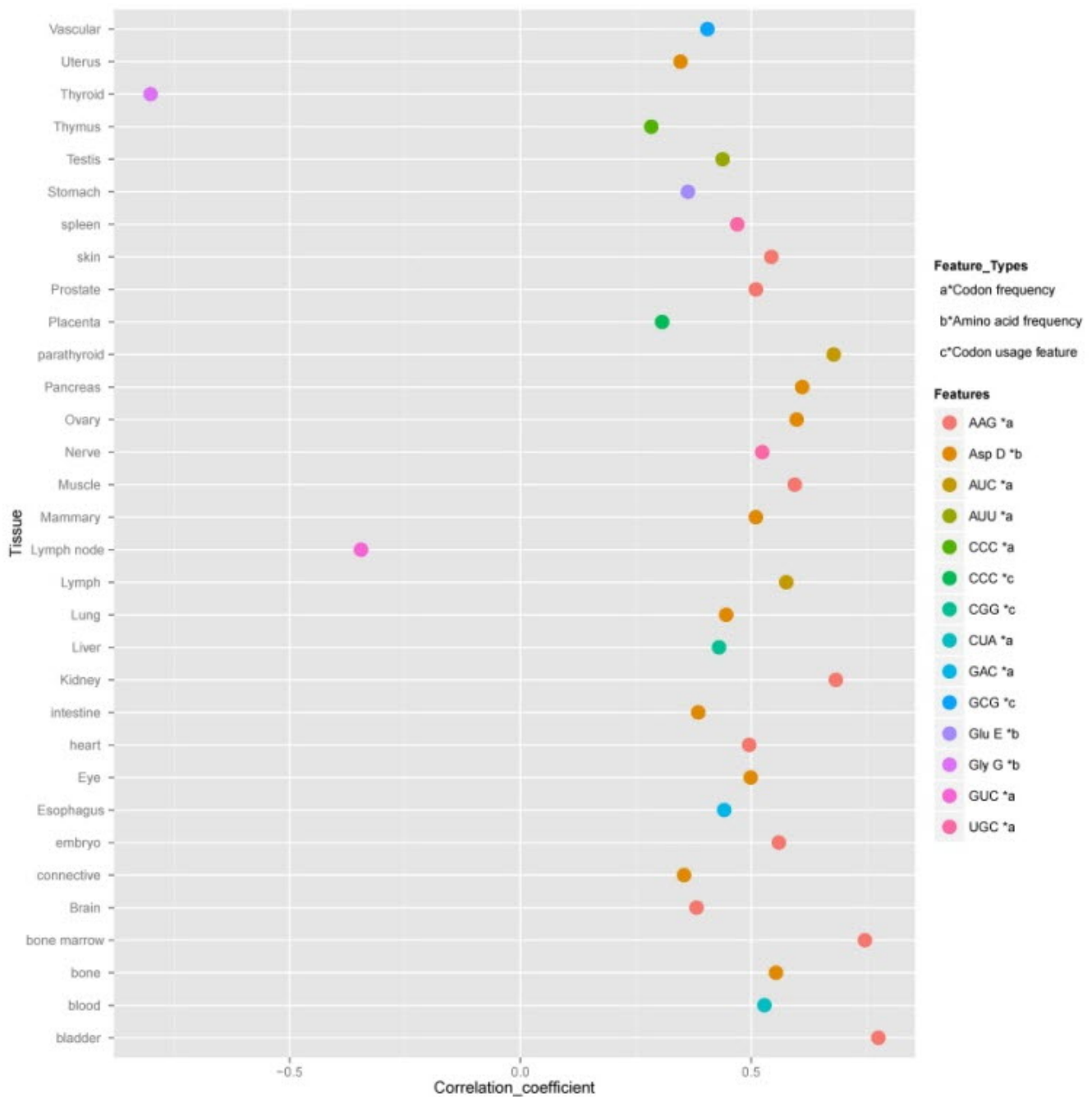


Fig. 4. The features which have the most significant correlations with gene expression levels in different tissues. The scatter plot is drawn by the R statistical software.

composition of proteins varies with increasing levels of gene expression in different tissues [19-21]. This result might suggest translational selection at the amino acid level. Based on this, the amino acids that are used more frequently in highly expressed genes may correspond to the most abundant tRNAs [19, 22]. Furthermore, some studies also demonstrate that amino acid composition is affected by selection to decrease the metabolic cost of protein

production [23]. In this study, we showed tissue specific correlations between amino acid usage and expression levels of apoptosis genes. Altogether, it may be inferred that the content of amino acids in tissues are very important for the regulation of the expression level of apoptosis involved genes. According to our results, some amino acids have tissue specific correlations with the expression of apoptosis genes. Our results showed that different amino acids

Table 2. The FDR corrected features which have the correlation with genes expression levels

Tissue	Feature	Corrected p-value
Bladder	AAG ^a	4.34E-05
	Gly ^b	0.009978
	Lys ^b	0.002843
	CTC (L) ^c	0.008334
Blood	CTA ^a	0.004387
	AGA ^a	0.004387
	CTA(L) ^c	0.011475
Bone marrow	AAG ^a	3.81E-05
	Lys ^b	0.002462
	CTC(L) ^c	0.000565
Bone	Asp ^b	0.009519
	AUC	0.02669
	GAC	0.02669
Embryo	AAG ^a	0.00617
	Lys ^b	0.02493
Eye	Asp ^b	0.04456
Kidney	AAG ^a	4.67E-09
	Lys ^b	1.40563E-05
Lung	Asp ^b	0.033489
	AAG ^a	0.033489
	GAC ^a	0.033489
Mammary	Asp ^b	0.030776
Muscle	AAG ^a	0.000391
	Gly ^b	0.011784
	CTC(L) ^c	0.001518
Ovary	Asp ^b	0.00102
	GAC ^a	0.009138
Pancreas	Asp ^b	0.000137
	GAC ^a	0.002486
Prostate	AAG ^a	0.008527
Skin	AAG ^a	0.002585

^aCodon frequency; ^bAmino acid frequency; ^cCodon usage feature.

including Arg, Asp, Lys, Glu, Gln, Ser, and Ile have significant correlations with the expression level of genes in 15, 14, 12, 11, 11, 9, and 8 tissues respectively (Supplementary Table 5). If we consider all of the codons attributed to one specific amino acid in one group for correlation analysis, we also get the same correlation coefficient and p-value as the amino acid features. So, all of the codons of arginine including CGT, CGA, CGC, CGG, AGA, and AGG (as one group) has the significant correlation with gene expression level. Some previous studies have demonstrated the role of these amino acids in apoptosis progression [24, 25]. So, it seems that the concentration of these amino acids in specific cell types can play an important role in cell survival and apoptosis.

Altogether, our results demonstrate the potential translational selection on sequence features of human apoptosis genes. Our data support the previous studies performed by Hajjari *et al.* [11] and Zhou *et al.* [9] who showed that the

expression of cancer involved genes are correlated with compositional features in different human tissues. The current findings have implications for the optimal design of gene therapies targeted in specific tissues and will promote a better biological understanding of translational selection of tissue specific gene expression in human.

Supplementary materials

Supplementary data including five tables can be found with this article online at <http://www.genominfo.org/src/sm/gni-14-234-s001.pdf>.

Acknowledgments

We acknowledge Shahid Chamran University of Ahvaz for supporting this study.

References

- Angov E. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* 2011;6:650-659.
- Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, *et al.* Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One* 2009;4:e7002.
- Plotkin JB, Robins H, Levine AJ. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci U S A* 2004;101:12588-12591.
- Mazumder TH, Chakraborty S, Paul P. A cross talk between codon usage bias in human oncogenes. *Bioinformatics* 2014;10:256-262.
- Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet* 2012;28:574-581.
- Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* 2012;8:e1002603.
- Dimitrieva S, Anisimova M. Unraveling patterns of site-to-site synonymous rates variation and associated gene properties of protein domains and families. *PLoS One* 2014;9:e95034.
- Lavner Y, Kotlar D. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2005;345:127-138.
- Zhou Y, Ma BG, Zhang HY. Human oncogene tissue-specific expression level significantly correlates with sequence compositional features. *FEBS Lett* 2007;581:4361-4365.
- Sémon M, Lobry JR, Duret L. No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol Biol Evol* 2006;23:523-529.
- Hajjari M, Khoshnevisan A, Behmanesh M. Compositional features are potentially involved in the regulation of gene expression of tumor suppressor genes in human tissues. *Gene* 2014;553:126-129.
- Ma L, Cui P, Zhu J, Zhang Z, Zhang Z. Translational selection in human: more pronounced in housekeeping genes. *Biol Direct* 2014;9:17.

13. Portt L, Norman G, Clapp C, Greenwood M, Greenwood MT. Anti-apoptosis and cell survival: a review. *Biochim Biophys Acta* 2011;1813:238-259.
14. Levy JP, Muldoon RR, Zolotukhin S, Link CJ Jr. Retroviral transfer and expression of a humanized, red-shifted green fluorescent protein gene into human tumor cells. *Nat Biotechnol* 1996;14:610-614.
15. Calkhoven CF, Müller C, Leutz A. Translational control of gene expression and disease. *Trends Mol Med* 2002;8:577-583.
16. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* 2006;2:e221.
17. Comeron JM. Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics* 2004;167:1293-1304.
18. Williford A, Demuth JP. Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, *Tribolium castaneum*. *Mol Biol Evol* 2012;29:3755-3766.
19. 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;16:287-289.
20. Heizer EM Jr, Raiford DW, Raymer ML, Doom TE, Miller RV, Krane DE. Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. *Mol Biol Evol* 2006;23:1670-1680.
21. Raiford DW, Heizer EM Jr, Miller RV, Akashi H, Raymer ML, Krane DE. Do amino acid biosynthetic costs constrain protein evolution in *Saccharomyces cerevisiae*? *J Mol Evol* 2008;67:621-630.
22. Akashi H. Translational selection and yeast proteome evolution. *Genetics* 2003;164:1291-1303.
23. Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 2002;99:3695-3700.
24. Zhang FX, Rubin R, Rooney TA. N-Methyl-D-aspartate inhibits apoptosis through activation of phosphatidylinositol 3-kinase in cerebellar granule neurons: a role for insulin receptor substrate-1 in the neurotrophic action of n-methyl-D-aspartate and its inhibition by ethanol. *J Biol Chem* 1998;273:26596-26602.
25. Verzola D, Fama A, Villaggio B, Di Rocco M, Simonato A, D'Amato E, *et al.* Lysine triggers apoptosis through a NADPH oxidase-dependent mechanism in human renal tubular cells. *J Inherit Metab Dis* 2012;35:1011-1019.