

Research Article

A Comparative Analysis of Synonymous Codon Usage Bias Pattern in Human Albumin Superfamily

Hoda Mirsafian,¹ Adiratna Mat Ripen,² Aarti Singh,¹ Phaik Hwan Teo,¹ Amir Feisal Merican,^{1,3} and Saharuddin Bin Mohamad^{1,3}

¹ *Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

² *Allergy and Immunology Research Centre, Institute for Medical Research, Jalan Pahang, Kuala Lumpur, Malaysia*

³ *Crystal, Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

Correspondence should be addressed to Saharuddin Bin Mohamad; saharuddin@um.edu.my

Received 25 November 2013; Accepted 11 January 2014; Published 20 February 2014

Academic Editors: Y. Lai, S. Ma, and Z. Su

Copyright © 2014 Hoda Mirsafian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Synonymous codon usage bias is an inevitable phenomenon in organismic taxa across the three domains of life. Though the frequency of codon usage is not equal across species and within genome in the same species, the phenomenon is non random and is tissue-specific. Several factors such as GC content, nucleotide distribution, protein hydrophathy, protein secondary structure, and translational selection are reported to contribute to codon usage preference. The synonymous codon usage patterns can be helpful in revealing the expression pattern of genes as well as the evolutionary relationship between the sequences. In this study, synonymous codon usage bias patterns were determined for the evolutionarily close proteins of albumin superfamily, namely, albumin, α -fetoprotein, afamin, and vitamin D-binding protein. Our study demonstrated that the genes of the four albumin superfamily members have low GC content and high values of effective number of codons (ENC) suggesting high expressivity of these genes and less bias in codon usage preferences. This study also provided evidence that the albumin superfamily members are not subjected to mutational selection pressure.

1. Introduction

Amino acids, the monomeric unit of proteins, are encoded by triplet of nucleotides called codons. Most of the amino acids have alternative codons which are known as synonymous codons. The frequencies with which these synonymous codons are used are unequal [1], some codons being used preferentially than others. Furthermore, Plotkin et al. [2] reported that codon usage is tissue-specific. The phenomenon of codon usage bias, which can be interpreted as an outcome of either mutational bias or translational selection, is an essential feature of most genomes across all the three domains of life [3]. The patterns of codon usage within the mammalian genomes are markedly different from other taxa. In mammals, the codon usage bias is found to be influenced by the variation in isochores (GC content) or variation in tRNA pool of the cell [4, 5]. The differences in codon usage or the variation in tRNA abundance can elicit varied responses to

the environmental changes, in terms of regulation of translation mechanism and cell phenotype [6]. Urrutia and Hurst [7] reported that, in humans, the codon usage bias is positively related to gene expression but is inversely related to the rate of synonymous substitution. Several factors contribute to synonymous codon usage bias such as gene expression level, protein hydrophathy, protein secondary structure, and translational selection [8–11]. Information on the synonymous codon usage pattern can provide significant insights pertaining to the prediction, classification, and molecular evolution of genes and design of highly expressed genes and cloning vectors [12]. It may be useful in better understanding of host-pathogen interactions as information on synonymous codon usages can reveal about the host-pathogen coevolution and adaptation of pathogens to specific hosts [13].

The evolutionarily close proteins of albumin superfamily are comprised of albumin (ALB), α -fetoprotein (AFP), vitamin D-binding protein (VDBP), and afamin (AFM). In

human, the genes encoding these proteins are mapped to chromosome 4. These proteins are synthesized primarily and predominantly in liver but the expression pattern varies temporally. One common functional property amongst all the members of albumin superfamily is their tendency to serve as transporters to various cellular components, metabolites, and so forth. ALB, an abundant serum protein of MW of ~66 KDa, binds and transports a variety of ligands such as steroids, fatty acids, bilirubin, lysolecithin, prostaglandins, thyroid hormones, and drugs. In addition to this, ALB is known to be involved in various cellular functions including oxygen-free radicals scavenging, anticoagulation, and maintenance of physiological pH and oncotic pressure of the plasma [14]. AFP (MW ~67 KDa), a serum glycoprotein which is expressed at high levels by fetal liver and visceral yolk sac [15, 16], is critical for the female fertility rather than embryonic development [17]. VDBP or Gc globulin (MW ~58 KDa) is synthesized by various tissues, namely, liver, kidneys, gonads, and fat, and also by neutrophils [18]. Apart from binding and transporting vitamin D sterols, VDBP's physiological functions include scavenging of G-actin [19], macrophage activation [20], and enhancement of chemotactic activity of C5a and C5a des-Arg molecules [21, 22]. AFM or α -albumin (MW ~87 KDa) is synthesized by liver and brain capillary endothelial cells. It mediates the transport of α -tocopherol across the blood-brain barrier [23].

The members of albumin superfamily have been found to act as markers in various disease states in humans. AFP in maternal serum is an indicative of Down's syndrome and neural tube defects in the fetus [24, 25]. AFP levels are elevated in patients with high risk for hepatocellular carcinoma. In some patients, an increase in AFP levels manifests liver metastasis with gastric cancer and the condition is termed as α -fetoprotein producing gastric cancer (AFPGC) [26, 27]. VDBP may serve as a biomarker for vascular injury as predicted by proteomic identification [28]. AFM may act as a potential adjunct marker to cancer antigen 125 (CA125) for the diagnosis of ovarian cancer [29]. A vast array of research has been done on the members of albumin superfamily; however, so far, studies related to the usage of synonymous codon and the factors influencing the codon usage in this gene family have not been done. In this study, we applied bioinformatics approaches to elucidate the pattern of synonymous codon usage bias and its consequences on the expression level of genes in the albumin superfamily.

2. Materials and Methods

2.1. Sequences. The mRNA reference sequences of human serum albumin (ALB), afamin (AFM), α -fetoprotein (AFP), and vitamin D-binding protein (VDBP) in FASTA format were retrieved from GenBank of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/genbank/>). Open Reading Frame (ORF) of the mRNA sequences of human albumin superfamily was obtained by using ExPASy Translate tool (<http://web.expasy.org/translate/>).

2.2. Hydrophobicity Analysis. Grand average of hydrophobicity score (Gravy score) was calculated to quantify the general average hydrophobicity for the translated gene product found in albumin superfamily. It was calculated as the arithmetic mean of the sum of the hydrophobic indices of each amino acid as shown in

$$\frac{1}{N} \sum_{i=1}^N K_i, \quad (1)$$

where N corresponds to the number of amino acids, while K_i represents hydrophobic index of amino acid. The Gravy score of a protein can be either negative or positive depending on the frequency of amino acids with distinct properties. Negative Gravy score implies that the protein is hydrophilic and is soluble in water. In contrast, protein with positive Gravy is considered as hydrophobic and is water soluble [30].

2.3. Codon Usage Analysis. The nucleotide distribution for albumin superfamily was analyzed using ExPASy ProtParam tool (<http://web.expasy.org/protparam/>). The quantities of individual nucleotide (A, T, G, and C) were determined and used to sum up the AT and GC content for each protein in the albumin superfamily.

2.4. Rare Codon (RC) Analysis. Rare codon (RC) is considered as low-usage codon in the genome such as synonymous codon or stop codon [31]. The RC analysis was performed using the GenScript web server (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis/) to examine the number of highest-usage and lowest-usage codons in the human albumin superfamily.

2.5. Indices of Codon Usage Deviation. Indices of codon usage deviation were calculated using CodonW (J Peden, version 1.4.2 <http://codonw.sourceforge.net/>) [32] to measure deviation between the observed codon usage and expected codon usage. Based on that, two internal measures were applied including identification of GC variation and third nucleotide preference in codon [33, 34]. These were obtained by calculating the number of GC nucleotides and number of G or C nucleotides at the third position of synonymous codon (GC_3), except the start and termination codons. In addition, the expected effective number of codons (ENC) for each albumin superfamily protein was calculated. ENC is the measure of codon usage affected only by the GC_3 as a consequence of mutation pressure and genetic drift. The ENC was calculated according to [35]

$$ENC = 2 + s + \frac{29}{s^2 + (1 - s)^2}, \quad (2)$$

where s corresponds to the GC_3 value ranging from 0 to 100%.

2.6. Relative Synonymous Codon Usage (RSCU). Relative synonymous codon usage (RSCU) was calculated in order to examine the frequency of each synonymous codon that encoded the same amino acid without confounding effect

TABLE 1: Genomic information of the reference sequences, grand average hydrophobicity score, ENC_s, GC content, and GC₃ of human albumin superfamily members.

	Human albumin superfamily			
	Albumin (ALB)	Afamin (AFM)	Alpha-fetoprotein (AFP)	Vitamin D-binding protein (VDBP)
GenBank accession number	NM_000477.5	NM_001133.2	NM_001134.1	NM_000583.3
Gene length (bp)	2264	1997	2032	2024
Grand average of hydrophobicity score (Gravy score)	-0.354	-0.248	-0.388	-0.336
GC content	42.95	42.02	39.28	44.63
Effective number of codons (ENC)	53.91	51.65	54.78	56.62
GC ₃	38.00	37.10	37.30	42.80

on the composition of amino acid. The index was calculated as follows [36]:

$$RSCU_{ij} = \frac{X_{ij}}{(1/n_i) \sum_{j=1}^{n_i} X_{ij}}, \quad (3)$$

where X_{ij} is the amount of j th codon to represent the i th amino acid that can be encoded by n_i synonymous codons.

3. Results and Discussion

Genomic information of mRNA sequences of the four members of human albumin superfamily is shown in Table 1. The mRNA sequences of albumin superfamily were translated into protein sequences using the ExpASY Translate Tool. Only the ORF with no intermediate stop codon was selected for codon usage analysis. The similarity of nucleotide and amino acid sequences of the albumin superfamily members is summarized in Figure 1. The results showed that ALB and AFP are more closely related compared to AFM and VDBP. AFP and VDBP have almost similar gene length of 2032 bp and 2024 bp, respectively. ALB possesses the longest (2264 bp), while AFM has the shortest gene length (1997 bp). Moreover, human ALB and AFP possessed exactly the same length of ORF (1830 bp), while AFM (1800 bp) has similar length of the ORF compared to that of ALB and AFP. VDBP (1425 bp) has the shortest length of ORF within the albumin superfamily. The similarity pattern of ORF among ALB, AFM, and AFP indicated that they may carry out similar biological functions, especially AFM, since its function is not well-known.

The solubility of protein for the members of the albumin superfamily was assessed through Gravy score (Table 1). All the family members are found to have negative Gravy score, suggesting that these proteins are water soluble. This is in accordance with the biological role of these proteins as serum transporters.

The nucleotide distribution of albumin superfamily is shown in Table 2. The members of this superfamily exhibit low GC content (<44.63%). ALB and AFP shows similar nucleotide distribution pattern implying that they share similarity in their structures and biological functions. There is a close relationship between the nucleotide composition

TABLE 2: Nucleotide distribution of human albumin superfamily members.

	ALB (%)	AFP (%)	AFM (%)	VDBP (%)
A	30.4 (556)	32.6 (596)	32.8 (591)	29.9 (426)
T	26.7 (488)	25.4 (465)	27.9 (502)	25.5 (363)
G	23.0 (421)	21.7 (397)	20.1 (361)	21.4 (305)
C	19.9 (365)	20.3 (372)	19.2 (346)	23.2 (331)
AT	57.049	57.978	60.722	55.368
GC	42.951	42.022	39.278	44.632

The values in parenthesis represent the number of individual nucleotides in the genes of human albumin superfamily members.

and gene function [37]. AFM has the highest AT content, whereas VDBP has the lowest AT content. Although AFM and VDBP are grouped in the same superfamily, they show differential nucleotide composition suggesting variation in their biological functions compared to the other members of albumin superfamily.

Rare codon analysis was carried out using the GenScript web server as described in Materials and Methods. A graph of codon frequency distribution was plotted to identify the quantities of rare codons present in each albumin superfamily protein (Figure 2). Frequency of codon usage with a value of 100 indicates that the codons are highly used for a given amino acid. Conversely, the frequency of codon usage with a value of less than 30 is determined as low-frequency codon, which is likely to affect the expression efficiency. Percentages of low-frequency codon present in protein ALB, AFM, AFP, and VDBP are 4%, 3%, 4%, and 4%, respectively. This result suggested that members of the albumin superfamily contain a significantly small number of rare codons that may reduce translational efficiency of the genes.

Indices of codon usage deviation are used to determine the differences between the observed and expected codon usage. The results for the effective number of codon (ENC), GC content, and G or C nucleotides at the third position of synonymous codon are summarized in Table 1. The effective number of codons (ENC) for each member of human albumin superfamily was calculated in order to examine the pattern of synonymous codon usage independent of the gene length. The ENC value ranges from 20 to 61, in which value

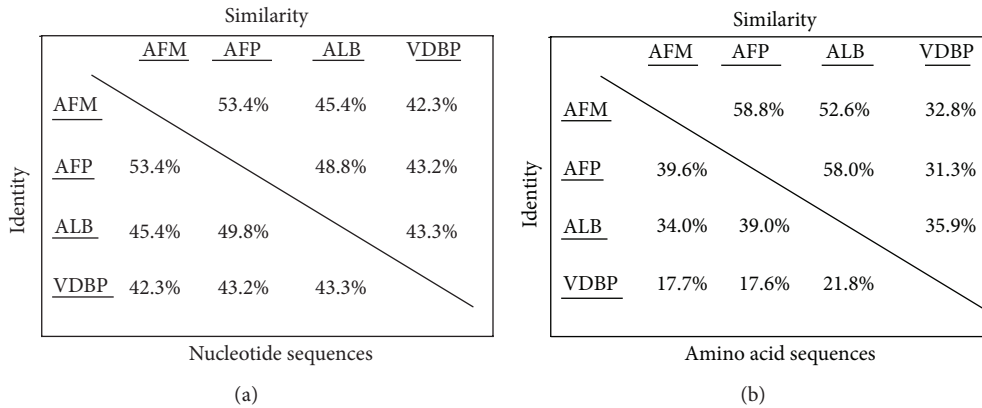


FIGURE 1: Comparison of percent similarity and identity of nucleotide sequences and amino acid sequences of human albumin superfamily members.

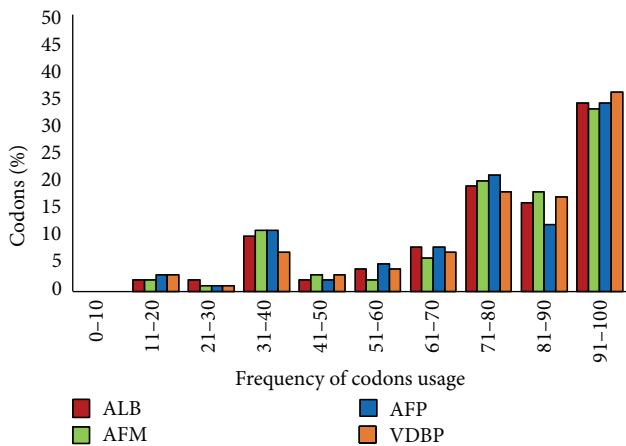


FIGURE 2: Codon frequency distribution of human albumin superfamily members.

of 20 indicates extreme bias toward the usage of one codon, while value of 61 represents equal usage of the synonymous codons [35, 38]. Result from this analysis revealed that the ENC value of albumin superfamily varies from 51.65 to 56.62. The overall ENC value of albumin superfamily is greater than 50. The high ENC value suggested that the synonymous codons of albumin superfamily were equally used and hence displayed less biased synonymous codon usage.

The GC content of albumin superfamily is given in Table 1. GC content may affect the thermostability, bendability, and the ability of DNA helix transition from B to Z form. GC content can be related to the ability of coding region to be in an open chromatin state, leading to active transcription [39]. It is evident that all the members of albumin superfamily genes have low GC content, indicating that these family members are highly expressed. Furthermore, it has been reported that highly transcribed genes may have low mutation rates because they are subjected to DNA repair [40]. However, within the albumin superfamily, VDBP contains the highest GC content indicating that it has the lowest expressivity level.

GC content at the third position of codons (GC_3) is a putative indicator of the extent of base composition bias. Table 1 revealed that the albumin superfamily has low GC_3 values ranging from 37.1% to 42.8%. The albumin superfamily has low GC_3 value because the majority of genes in this superfamily are located in AT-rich region. Genes in AT-rich regions within the genome would prefer to use A or T ending codon. The low usage of codons ending with G or C signifies less GC codon usage bias in albumin superfamily. In other words, it proved the homogeneity of synonymous codon usage pattern in albumin superfamily.

The synonymous codon bias usage of each albumin superfamily protein was computed and tabulated in Table 3. The most preferentially used codon for a given amino acid is highlighted in red. Asn of AFP and His, Cys, and Arg of VDBP have equal usage of the synonymous codons. The variation of relative synonymous codon usage (RSCU) values not only indicated the different frequency of occurrence of each codon for a given amino acid in different albumin superfamily protein but also revealed the preference of either A + U or G + C codon usage as listed in Table 3. The results of RSCU analysis (Table 3) are summarized in Table 4. Preferential codon usage in albumin superfamily indicates that the codons with A or U at the third position are more preferred compared to G or C ending codons. Table 4 also shows that the total score of A + U and G + C codon usage in the proteins of albumin superfamily is not equal to 20. It is because some amino acid residues are encoded in equal frequencies by both A or U and G or C ending codons and hence are excluded from the analysis. The tendency of albumin superfamily to use high A + U and low G + C indicated that the mutational bias does not play a significant role in synonymous codon usage.

4. Conclusions

The members of albumin superfamily, namely, ALB, AFP, AFM, and VDBP, exhibit sequence and structural similarities. The proteins possess three homologous folding domains as a result of conserved pattern of cysteine residues in

TABLE 3: Relative synonymous codon usage in human albumin superfamily members. The value in bold indicates the codons used with high frequency.

Amino acid	Codons	RSCU ¹	Number	RSCU ²	Number	RSCU ³	Number	RSCU ⁴	Number
Phe	UUU	1.43	25	1.06	17	1.30	28	1.16	11
	UUC	0.57	10	0.94	15	0.70	15	0.84	8
Leu	UUA	0.94	10	1.00	10	1.20	11	0.53	5
	UUG	1.13	12	1.10	11	0.87	8	0.63	6
	CUU	1.78	19	0.90	9	1.20	11	1.16	11
	CUC	0.66	7	0.40	4	0.98	9	0.95	9
	CUA	0.38	4	0.90	9	0.65	6	1.05	10
	CUG	1.13	12	1.70	17	1.09	10	1.68	16
Ile	AUU	1.33	4	1.32	15	1.07	10	1.13	3
	AUC	1.33	4	0.71	8	0.75	7	1.13	3
	AUA	0.33	1	0.97	11	1.18	11	0.75	2
Val	GUU	1.12	12	1.47	11	1.44	13	0.89	6
	GUC	0.65	7	0.80	6	0.67	6	1.19	8
	GUA	0.74	8	0.80	6	0.67	6	1.04	7
	GUG	1.49	16	0.93	7	1.22	11	0.89	6
Ser	UCU	0.64	3	1.26	8	2.06	12	1.43	10
	UCC	1.50	7	0.47	3	0.86	5	1.29	9
	UCA	1.29	6	1.58	10	1.03	6	1.71	12
	UCG	0.64	3	0.47	3	0.00	0	0.00	0
Pro	CCU	1.67	10	1.71	9	1.71	12	1.38	9
	CCC	1.00	6	0.76	4	0.71	5	0.92	6
	CCA	1.17	7	1.52	8	1.57	11	1.54	10
	CCG	0.17	1	0.00	0	0.00	0	0.15	1
Thr	ACU	0.97	7	1.78	16	1.18	10	1.25	10
	ACC	1.24	9	0.67	6	0.82	7	1.38	11
	ACA	1.52	11	1.33	12	1.53	13	1.13	9
	ACG	0.28	2	0.22	2	0.47	4	0.25	2
Ala	GCU	1.90	30	1.20	15	1.57	11	1.82	15
	GCC	0.89	14	0.88	11	0.71	5	1.09	9
	GCA	1.08	17	1.68	21	1.29	9	0.85	7
	GCG	0.13	2	0.24	3	0.43	3	0.24	2
Tyr	UAU	1.37	13	1.06	9	1.06	9	1.13	9
	UAC	0.63	6	0.94	8	0.94	8	0.88	7
His	CAU	1.38	11	1.63	13	1.23	8	1.00	4
	CAC	0.63	5	0.38	3	0.77	5	1.00	4
Gln	CAA	1.10	11	1.15	23	1.26	17	1.33	8
	CAG	0.90	9	0.85	17	0.74	10	0.67	4
Asn	AAU	1.29	11	1.00	10	1.03	17	1.33	12
	AAC	0.71	6	1.00	10	0.97	16	0.67	6
Lys	AAA	1.33	40	1.29	33	1.33	28	0.93	20
	AAG	0.67	20	0.71	18	0.67	14	1.07	23
Asp	GAU	1.39	25	1.27	21	1.30	15	1.23	16
	GAC	0.61	11	0.73	12	0.70	8	0.77	10
Glu	GAA	1.23	38	1.24	34	1.36	40	1.26	27
	GAG	0.77	24	0.76	21	0.64	19	0.74	16
Cys	UGU	0.86	15	1.06	18	0.81	13	1.00	14
	UGC	1.14	20	0.94	16	1.19	19	1.00	14

TABLE 3: Continued.

Amino acid	Codons	RSCU ¹	Number	RSCU ²	Number	RSCU ³	Number	RSCU ⁴	Number
Arg	CGU	0.67	3	0.50	2	0.55	2	0.00	0
	CGC	0.22	1	0.25	1	0.27	1	0.00	0
	CGA	0.67	3	0.50	2	0.55	2	0.92	2
	CGG	0.44	2	0.25	1	0.00	0	0.46	1
Ser	AGU	1.29	6	1.54	9	0.79	5	0.86	6
	AGC	0.64	3	0.51	3	1.42	9	0.71	5
Arg	AGA	2.89	13	2.75	11	3.27	12	2.31	5
	AGG	1.11	5	1.75	7	1.36	5	2.31	5
Gly	GGU	0.92	3	0.75	3	0.62	4	0.29	1
	GGC	0.92	3	0.75	3	0.77	5	1.43	5
	GGA	1.85	6	1.50	6	2.00	13	1.43	5
	GGG	0.31	1	1.50	4	0.62	4	0.86	3

RSCU¹: RSCU values for ALB; RSCU²: RSCU values for AFP; RSCU³: RSCU values for AFM; RSCU⁴: RSCU values for DBP.

TABLE 4: A + U and G + C preferential codon usage of human albumin superfamily members.

	A + U	G + C
ALB	17	3
AFP	17	1
AFM	18	2
VDBP	11	4

the members of albumin superfamily [41, 42]. Our study on codon usage bias in the members of the albumin gene family revealed that they are also similar in terms of their low GC content, low GC₃, and high ENC values. In addition, they are not having a bias in the usage of synonymous codons and are highly expressible genes. Furthermore, low GC and GC₃ values revealed that mutational bias and translational selection do not play a significant role in shaping the codon usage pattern in the albumin superfamily.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank Director General of Ministry of Health, Malaysia, for granting permission to publish this paper. This research is cosupported by the High Impact Research Grant UM-MOHE UM.C/625/1/HIR/MOHE/30 from the Ministry of Higher Education, Malaysia, and University of Malaya Research Grant (UMRG), Grant no. RP004C-BAFR.

References

- [1] J. L. Bennetzen and B. D. Hall, "Codon selection in yeast," *Journal of Biological Chemistry*, vol. 257, no. 6, pp. 3026–3031, 1982.
- [2] J. B. Plotkin, H. Robins, and A. J. Levine, "Tissue-specific codon usage and the expression of human genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 34, pp. 12588–12591, 2004.
- [3] L. Duret, "Evolution of synonymous codon usage in metazoans," *Current Opinion in Genetics and Development*, vol. 12, no. 6, pp. 640–649, 2002.
- [4] G. Bernardi, B. Olofsson, and J. Filipiński, "The mosaic genome of warm-blooded vertebrates," *Science*, vol. 228, no. 4702, pp. 953–958, 1985.
- [5] K. A. Dittmar, J. M. Goodenbour, and T. Pan, "Tissue-specific differences in human transfer RNA expression," *PLoS Genetics*, vol. 2, no. 12, pp. 2107–2115, 2006.
- [6] H. S. Najafabadi, H. Goodarzi, and R. Salavati, "Universal function-specificity of codon usage," *Nucleic Acids Research*, vol. 37, no. 21, Article ID gkp792, pp. 7014–7023, 2009.
- [7] A. O. Urrutia and L. D. Hurst, "Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection," *Genetics*, vol. 159, no. 3, pp. 1191–1199, 2001.
- [8] S. K. Gupta and T. C. Ghosh, "Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*," *Gene*, vol. 273, no. 1, pp. 63–70, 2001.
- [9] Q. Liu, "Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*," *BioSystems*, vol. 85, no. 2, pp. 99–106, 2006.
- [10] G. D'Onofrio, T. C. Ghosh, and G. Bernardi, "The base composition of the human genes is correlated with the secondary structures of the encoded proteins," *Gene*, vol. 300, no. 1-2, pp. 179–187, 2002.
- [11] H. Naya, H. Romero, N. Carels, A. Zavala, and H. Musto, "Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*," *FEBS Letters*, vol. 501, no. 1-3, pp. 127–130, 2001.
- [12] X.-S. Liu, Y.-G. Zhang, Y.-Z. Fang, and Y.-L. Wang, "Patterns and influencing factor of synonymous codon usage in porcine circovirus," *Virology Journal*, vol. 9, article 68, pp. 1–9, 2012.
- [13] A. Pandit and S. Sinha, "Differential trends in the codon usage patterns in HIV-1 genes," *PLoS ONE*, vol. 6, no. 12, Article ID e28889, 2011.

- [14] J. P. Nicholson, M. R. Wolmarans, and G. R. Park, "The role of albumin in critical illness," *British Journal of Anaesthesia*, vol. 85, no. 4, pp. 599–610, 2000.
- [15] R. R. Meehan, D. P. Barlow, R. E. Hill, B. L. Hogan, and N. D. Hastie, "Pattern of serum protein gene expression in mouse visceral yolk sac and foetal liver," *The EMBO Journal*, vol. 3, no. 8, pp. 1881–1885, 1984.
- [16] M. A. Dziadek and G. K. Andrews, "Tissue specificity of alpha-fetoprotein messenger RNA expression during mouse embryogenesis," *The EMBO Journal*, vol. 2, no. 4, pp. 549–554, 1983.
- [17] P. Gabant, L. Forrester, J. Nichols et al., "Alpha-fetoprotein, the major fetal serum protein, is not essential for embryonic development but is required for female fertility," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12865–12870, 2002.
- [18] L. Chishimba, D. R. Thickett, R. A. Stockley, and A. M. Wood, "The vitamin D axis in the lung: a key role for vitamin D-binding protein," *Thorax*, vol. 65, no. 5, pp. 456–462, 2010.
- [19] L. R. Otterbein, C. Cosio, P. Graceffa, and R. Dominguez, "Crystal structures of the vitamin D-binding protein and its complex with actin: structural basis of the actin-scavenger system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 8003–8008, 2002.
- [20] S. B. Mohamad, H. Nagasawa, Y. Uto, and H. Hori, "Preparation of Gc protein-derived macrophage activating factor (GcMAF) and its structural characterization and biological activities," *Anticancer Research*, vol. 22, no. 6, pp. 4297–4300, 2002.
- [21] R. R. Kew and R. O. Webster, "Gc-globulin (vitamin D-binding protein) enhances the neutrophil chemotactic activity of C5a and C5a des Arg," *Journal of Clinical Investigation*, vol. 82, no. 1, pp. 364–369, 1988.
- [22] S. J. DiMartino, A. B. Shah, G. Trujillo, and R. R. Kew, "Elastase controls the binding of the vitamin D-binding protein (Gc-Globulin) to neutrophils: a potential role in the regulation of C5a co-chemotactic activity," *Journal of Immunology*, vol. 166, no. 4, pp. 2688–2694, 2001.
- [23] I. Kratzer, E. Bernhart, A. Wintersperger et al., "Afinin is synthesized by cerebrovascular endothelial cells and mediates α -tocopherol transport across an in vitro model of the blood-brain barrier," *Journal of Neurochemistry*, vol. 108, no. 3, pp. 707–718, 2009.
- [24] M. J. Seller, "Alphafetoprotein in midtrimester Down's syndrome fetal serum," *Journal of Medical Genetics*, vol. 27, no. 4, pp. 240–243, 1990.
- [25] J. H. Brock, "Alphafetoprotein and neural tube defects," *Journal of Clinical Pathology*, vol. 29, no. 10, pp. 157–164, 1976.
- [26] S. Hirajima, S. Komatsu, D. Ichikawa et al., "Liver metastasis is the only independent prognostic factor in AFP-producing gastric cancer," *World Journal of Gastroenterology*, vol. 19, no. 36, pp. 6055–6061, 2013.
- [27] X.-D. Li, C.-P. Wu, M. Ji et al., "characteristic analysis of a-fetoprotein-producing gastric carcinoma in China," *World Journal of Surgical Oncology*, vol. 11, no. 246, 2013.
- [28] N. F. Huang, K. Kurpinski, Q. Fang, R. J. Lee, and S. Li, "Proteomic identification of biomarkers of vascular injury," *The American Journal of Translational Research*, vol. 3, no. 2, pp. 139–148, 2011.
- [29] H. Dieplinger, D. P. Ankerst, A. Burges et al., "Afinin and apolipoprotein A-IV: novel protein markers for ovarian cancer," *Cancer Epidemiology Biomarkers and Prevention*, vol. 18, no. 4, pp. 1127–1133, 2009.
- [30] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [31] T. Ikemura, "Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system," *Journal of Molecular Biology*, vol. 151, no. 3, pp. 389–409, 1981.
- [32] J. F. Peden, *Analysis of Codon Usage*, University of Nottingham, 2000.
- [33] J. A. Novembre, "Accounting for background nucleotide composition when measuring codon usage bias," *Molecular Biology and Evolution*, vol. 19, no. 8, pp. 1390–1394, 2002.
- [34] J. M. Comeron and M. Aguadé, "An evaluation of measures of synonymous codon usage bias," *Journal of Molecular Evolution*, vol. 47, no. 3, pp. 268–274, 1998.
- [35] F. Wright, "The 'effective number of codons' used in a gene," *Gene*, vol. 87, no. 1, pp. 23–29, 1990.
- [36] M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*, Oxford University Press, New York, NY, USA, 2000.
- [37] J. A. L. Garcia, A. Fernández-Guerra, and E. O. Casamayor, "A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes," *Molecular Phylogenetics and Evolution*, vol. 61, no. 3, pp. 650–658, 2011.
- [38] A. Fuglsang, "The effective number of codons," *Biochemical and Biophysical Research Communications*, vol. 317, no. 3, pp. 957–964, 2004.
- [39] A. E. Vinogradov, "DNA helix: the importance of being GC-rich," *Nucleic Acids Research*, vol. 31, no. 7, pp. 1838–1844, 2003.
- [40] O. G. Berg and M. Martelius, "Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure," *Journal of Molecular Evolution*, vol. 41, no. 4, pp. 449–456, 1995.
- [41] H. S. Lichenstein, D. E. Lyons, M. M. Wurfel et al., "Afinin is a new member of the albumin, α -fetoprotein, and vitamin D-binding protein gene family," *Journal of Biological Chemistry*, vol. 269, no. 27, pp. 18149–18154, 1994.
- [42] H. Nishio and A. Dugaiczyk, "Complete structure of the human α -albumin gene, a new member of the serum albumin multi-gene family," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 15, pp. 7557–7561, 1996.