International Journal of Computational Biology
ISSN: 2229-6700 & E-ISSN:2229-6719, Volume 5, Issue 1, 2014, pp.-061-067.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000220

NEUTRAL MUTATION AND SELECTIVE CONSTRAINT PLAY MAJOR ROLE IN DICTATING THE CODON BIAS IN BREAST CANCER RISK GENES

MAZUMDER T.H. AND CHAKRABORTY S.*

Department of Biotechnology, Assam University, Silchar- 788 011, Assam, India.
*Corresponding Author: Email- supriyoch_2008@rediffmail.com

Received: September 24, 2014; Accepted: October 30, 2014

Abstract- Breast cancer is ranked the first cancer in women worldwide and the second leading cause of death after cervical cancer particularly in developing countries. Specific genes are associated with breast cancer risk in human genome. Understanding the codon usage bias (CUB) with compositional dynamics of coding sequence is of great importance in gaining clues to predict the level of gene expression and genome characterization. In this study, we have analyzed the complete nucleotide coding sequences of fifteen breast cancer risk genes with the help of various genetic indices. Our analysis revealed that both neutral mutation and selective constraint play major roles in the codon usage patterns of breast cancer risk genes. Our results further show that gene expression level is linked with alterations in the nucleotide skewness. Breast cancer gene products showed the dominance of three amino acids namely serine, leucine and glutamate in their composition but the least usage of two amino acids namely tryptophan and methionine. Moreover, the level of breast cancer gene expression measured by RCBS revealed a significant negative correlation with highly used amino acids. In addition, highly used amino acids had negative impact on the expression of breast cancer risk genes.

Keywords- Codon usage, Synonymous codon, Breast cancer, Gene expression, Mutation pressure, Amino acid usage, Nucleotide skewness, Relative codon usage bias

Citation: Mazumder T.H. and Chakraborty S. (2014) Neutral Mutation and Selective Constraint Play Major Role in Dictating the Codon Bias in Breast Cancer Risk Genes. International Journal of Computational Biology, ISSN: 2229-6700 & E-ISSN:2229-6719, Volume 5, Issue 1, pp.-061-067.

Introduction

Codon usage bias (unequal usage of synonymous or alternative codons for encoding the same amino acid in a protein) is a unique property of genome and shows species specific deviation [1-4]. Synonymous codons are often used in unequal frequencies between genomes, genes from the same genome and within a single gene [3-5]. Previous studies have revealed that nucleotide compositional constraints under natural selection or mutation pressure are considered as a major factor in the continuation of codon usage variation among different organisms [6-9]. Breast cancer is the topmost prevalent cancer in women worldwide and the second leading cause of death after cervical cancer particularly in the developing countries [10].

In this study, an attempt has been made to analyze the codon usage bias across the nucleotide coding sequences of genes associated with breast cancer risk in human and to explore any possible significance of translational selection on the sequence features of these genes. Following this work, we employed several genetic indices namely, the relative codon usage bias (*RCBS*), relative synonymous codon usage (*RSCU*), skewness and compositional dynamics for the background nucleotide constraints. The major objective of this study is to unravel the codon usage patterns and to estimate the strength of selection on the expression of genes associated with breast cancer risk.

Materials and Methods

Sequence Data

Genes associated with breast cancer were retrieved from the web site Breast cancer- Genetics (http://ghr.nlm.nih.gov/condition/breast-cancer). We have considered only the complete nucleotide coding sequences (cds) having perfect initiator codon, terminator codon, devoid of any unknown bases (N) and are perfect multiple of three bases. The cds for each of the concerned gene satisfying the aforementioned criteria were retrieved from National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov). Finally, we selected fifteen genes associated with breast cancer: *BRCA1*, *BRCA2*, *CDH1*, *STK11*, *TP53*, *AR*, *ATM*, *BARD1*, *BRIP1*, *CHEK2*, *ERBB2*, *NBN*, *PALB2*, *RAD50*, and *RAD51* for CUB analysis [Table-1].

Prediction of Base Composition Bias

We have calculated the occurrence of overall frequency of the nucleotide (G+C) % at first (GC_1%), second (GC_2%) and third (GC_3%) position of synonymous codon to quantify the extent of base com-

position bias. In addition, we have analyzed the skewness values for the AT, GC, purine, pyrimidine, keto and amino content for each cds to estimate the base composition bias particularly in relation to transcription processes.

Table 1- Complete nucleotide coding sequences (cds) with accession number and length (bp) of 15 human breast cancer risk genes used in this study

Sl. No.	Genes	Accession no.	Length (bp)
1	*BRCA1*	U14680.1	5592
2	*BRCA2*	U43746.1	10257
3	*CDH1*	GU371438.1	2649
4	*CHEK2*	CU012979.1	1761
5	*PALB2*	BC044254.2	3561
6	*STK11*	BC019334.1	1302
7	*TP53*	U94788.1	1182
8	*ATM*	U33841.1	9171
9	*ERBB2*	AY208911.1	3768
10	*AR*	M20132.1	2760
11	*BARD1*	U76638.1	2334
12	*BRIP1*	BC101472.1	3750
13	*NBN*	BC146797.1	2265
14	*RAD50*	U63139.1	3939
15	*RAD51*	D14134.1	1020

Computation of Gene Expression

Gene expression was estimated through *RCBS* which can be defined as the overall score of a gene indicating the influence of relative codon bias (RCB) of each codon in a gene [11]. The RCBS value of each gene was calculated as follows:

$$w_c^{RCB} = \frac{O_C - E[O_C]}{E[O_C]}$$

where, *Oc* is the observed number counts of codon *c* of the query sequence and $E[O_C]$ is the expected number of codon occurrences given the nucleotide distribution at three codon positions (b1b2b3) [11].

$$RCB = \exp\left(\frac{1}{O_{tot}} \sum_{c \in C} \log w_c^{RCB}\right) - 1$$

Where, O_{tot} is the total number of codons [12].

Relative Synonymous Codon Usage (*RSCU*) Analysis

RSCU is defined as the observed frequency of a codon divided by the expected frequency if all codons are used equally for any particular amino acid [13]. An RSCU value of codons for each of the selected cds was calculated as follows:

$$RSCU = \frac{g_{ij}}{\sum_j^{ni} g_{ij}} n_i$$

Where, g_{ij} is the observed number of the *i*th codon for the *j*th amino acid which has n_i kinds of synonymous codons [14].

Software Used

A critical drawback of current bioinformatics tools is the lack of single software that can reproduce all the results at a time. To resolve this problem, a PERL program was developed by SC (corresponding author) to analyze all the genetic parameters simultaneously.

Correlations

All the correlation analyses were based on Pearson correlation coefficient. All statistical analysis was carried out by using SPSS software.

Results

Codon usage Patterns in Breast Cancer Genes

The heat maps of correlation coefficient values between codon usage and GC bias [Fig-1a] were performed in order to investigate the relationship between the codon usage variation and GC constraints among the selected cds sequences of breast cancer genes. In our analysis, nearly 17 codons with G/C -ending base showed positive correlation with GC bias but strong negative correlation for nearly all A/T –ending codons with GC bias. However, 12 codons of which 6 are G/C –ending (AAC, AAG, ACC, AGC, AGG, TTC, TTG, TCC, TCG, CAG, GAG, and GTC) showed negative correlation with GC bias. Two G-ending codons, AGG for arginine and TTG for leucine amino acid displayed a strong negative correlation ($p<0.05$) between usage and GC bias [Fig-1b] and [Fig-1c]. These codons exhibit non-linear upward usage profiles due to the effects of GC bias.

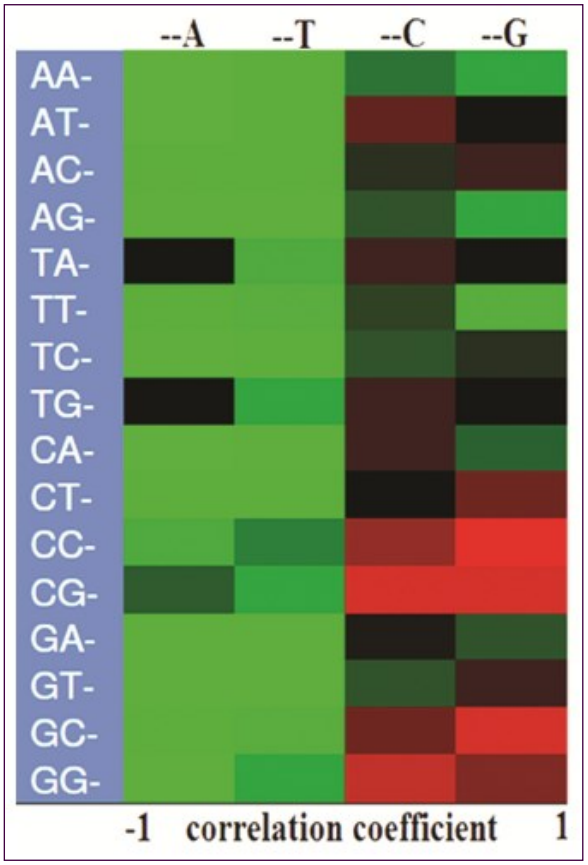

Fig. 1a- Relationship between codon usage patterns and GC_{3s} in human breast cancer genes. Heat maps of correlation coefficient values *vs* GC_3. The color coding and intensity towards red represents the degree of positive correlation, green as negative correlation. The black fields are stop codons (TAA, TAG, TGA) and non-degenerate codons (ATG, TGG) together with two codons viz. CTC for leucine and GAC for asparagine representing very weak positive correlation. AAC, AAG, ACC, AGC, AGG, TTC, TTG, TCC, TCG, CAG, GAG, GTC are the G/C ending codons showing the negative correlation with GC3 values.

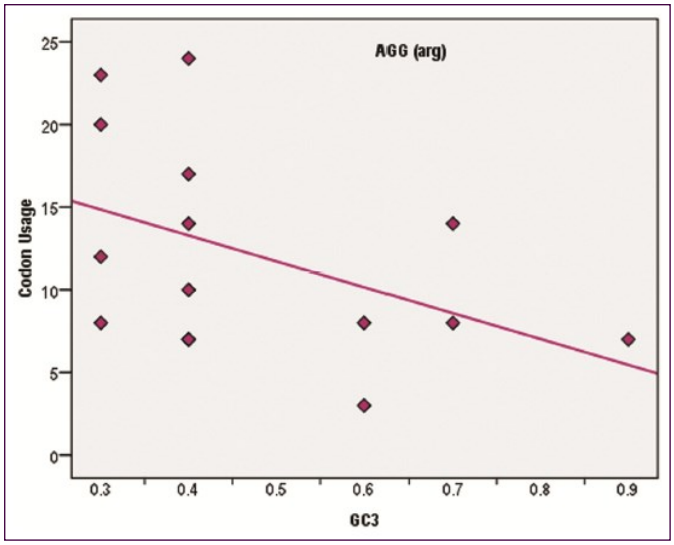

Fig. 1b- Scatter plots of codon usage frequency for the arginine codon AGG *vs* GC_{3s}

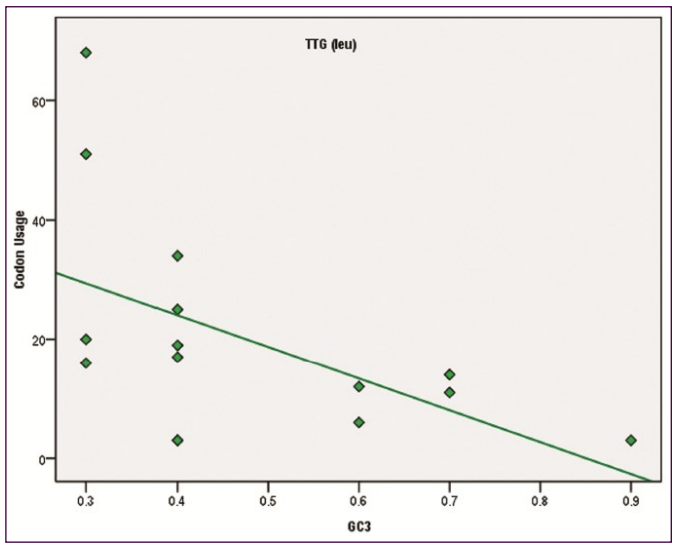

Fig. 1c- Scatter plots of codon usage frequency for the leucine codon TTG *vs* GC_{3s}

GC_{3s} Dictate the Extent of Base Composition Bias in Breast Cancer Genes

Zhou, et al [15], suggested that GC contents at three codon positions (GC_{3s}) are good indicators of the degree of base composition bias. Overall percentage of GC contents and GC_{3s} across the genes associated with breast cancer was 46.7±8.98 and 46.6±17.74 respectively [Table-2]. The GC contents at different codon positions varied among the breast cancer genes. We compared the values of GC contents at first (GC_1) and second (GC_2) codon positions with that of synonymous third codon positions (GC_3), and observed significant positive correlation (Pearson r=0.780, p<0.01). In addition, a neutrality plot was drawn to investigate the relationship between codon bias and mutation pressure [Fig-2] [16].

Gene Expression Level Affects Synonymous Codon usage Bias in Breast Cancer

The expression level of gene associated with breast cancer risk was assessed through RCBS values [11,12], which ranged from 0.013 to 0.144 with a mean value of 0.0528 and a standard deviation of 0.0429. The frequency of G+C contents over all the codons in the cds sequences (GC) and the third variable codon positions GC_3 revealed a strong positive correlation with *RCBS* measure [Fig-3], which suggested that mutational bias might be involved in genes transcribed at different levels [17].

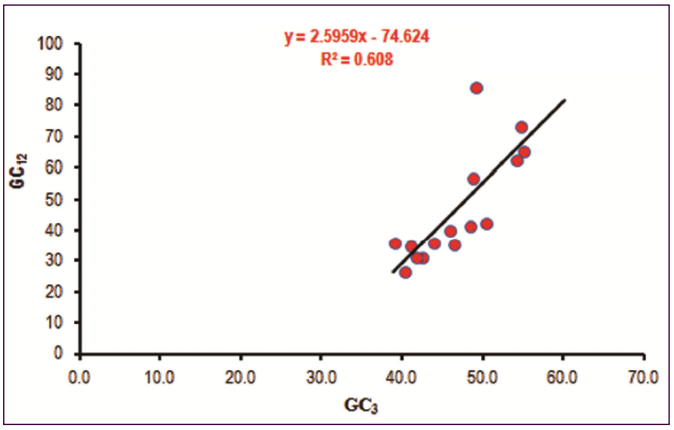

Fig. 2- Neutrality plot of GC_{12} *vs* GC_{3s}. GC_{12}: average GC contents at first and second codon positions.

Table 2- Nucleotide guanine-cytosine contents at different codon positions and RCBS values of 15 breast cancer risk genes

Sl. No.	GC%	GC1%	GC2%	GC3%	RCBS
1	30.6	48.4	39.3	35.9	0.0211
2	42.5	44	36.6	26.7	0.0134
3	29	56.4	41	57	0.0587
4	14.1	44.5	52.1	41.6	0.0985
5	20.8	50.8	41.8	35.3	0.0504
6	18.7	58.5	39.4	85.9	0.1295
7	13.1	59.1	49	62.4	0.1441
8	48.1	47.5	34.2	34.8	0.0214
9	49.1	63	46.1	73.2	0.0551
10	31.7	61.6	48.6	65.4	0.0187
11	14.2	49	42.8	40	0.0227
12	18.2	46.6	38.4	31.2	0.0175
13	1	46.1	37.2	31.4	0.0335
14	17.3	50.8	27.2	35.9	0.0204
15	0.7	57.9	42.6	42.6	0.0872
Mean	46.67	52.28	41.09	46.6	0.0211
SD	8.984	6.481	6.325	17.744	0.0134

SD: Standard deviation, RCBS: Relative codon usage bias

Relative Synonymous Codon usage across the Breast Cancer Genes

In order to investigate the non random usage of synonymous codons in breast cancer risk genes, relative synonymous codon usage values were analyzed. The RSCU value equal to unity represents that the codons are used equally or randomly for the corresponding amino acid and no bias for that amino acid. The RSCU value greater than one represents positive codon usage bias with greater usage of the most abundant codons whereas the RSCU value less than one represents a negative codon usage bias using the least abundant codons. The overall relative synonymous codon usage values revealed that most predominantly used codons were T and A/C ending, of which 12 codons were T –ending and 7 were A or C -ending codons, respectively [Table-3].

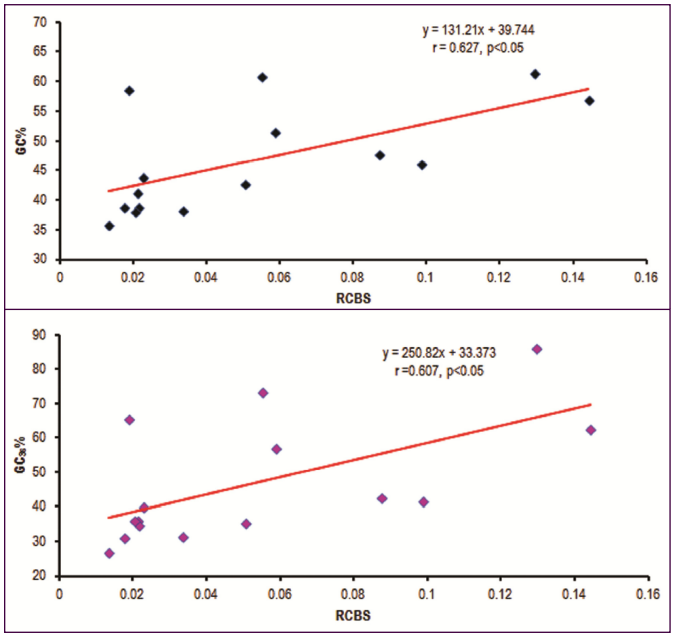

Fig. 3- Correlation between percentage of GC and GC_{3s} with RCBS values. RCBS and GC% (Pearson r=0.627, p<0.05); RCBS and GC_{3s}% (Pearson r=0.607, p<0.05).

Table 3- Overall relative synonymous codon usage pattern (RSCU) for 15 breast cancer risk genes

Amino Acid	Codon	Total No.	RSCU[a]	Amino Acid	Codon	Total No.	RSCU[a]
Ala	GCA*	319	1.14	Leu	TTA	325	0.81
	GCC*	286	1.26		TTG	302	0.95
	GCG	56	0.28		CTA	191	0.59
	GCT*	345	1.32		CTC	209	0.82
Arg	CGT	84	0.52		CTG*	427	1.76
	CGC	79	0.65		CTT*	342	1.08
	CGA	87	0.61	Lys	AAA*	836	1.09
	CGG	92	0.78		AAG	527	0.91
	AGA*	338	1.97	Phe	TTT*	404	1.11
	AGG*	182	1.38		TTC	214	0.85
Asn	AAC	316	0.88	Pro	CCA*	319	1.35
	AAT*	591	1.12		CCC*	226	1.02
Asp	GAT*	576	1.1		CCG	79	0.41
	GAC	381	0.91		CCT*	318	1.23
Cys	TGC	176	0.82	Ser	TCA*	351	1.14
	TGT*	307	1.18		TCC*	220	1.02
Gln	CAA	380	0.73		TCG	38	0.22
	CAG*	557	1.27		TCT*	425	1.36
Glu	GAA*	973	1.17		AGC*	259	1.1
	GAG	517	0.83		AGT*	379	1.16
Gly	GGA*	340	1.43	Thr	ACA*	384	1.31
	GGC	234	0.99		ACC*	233	1.09
	GGG	170	0.76		ACG	68	0.31
	GGT	200	0.82		ACT*	376	1.29
His	CAT*	265	1.07	Tyr	TAC*	211	1.01
	CAC	174	0.93		TAT	247	0.99
Ile	ATA	257	0.66	Val	GTA	207	0.71
	ATC*	226	1.11		GTC	193	0.85
	ATT*	430	1.23		GTG*	331	1.5
					GTT	283	0.95

[a] mean values of RSCU based on the synonymous codon usage frequencies of 15 breast cancer genes. *RSCU>1

Furthermore, the clustering analysis of RSCU values of each codon among breast cancer risk genes [Fig-4] showed that, the number of frequently and the rarely used codons differed among the genes. The codon AGA was displayed as the over represented codon (RSCU>1.6) in nine breast cancer genes out of fifteen selected in this study. Our results suggested that, compositional constraint was one of the factors in shaping the codon usage variation among these genes.

Relationship between Different Nucleotide Skewness and Codon Bias

The frequency of nucleotide usage varies across the genes and significantly different in exons and introns of human genes [18]. In this study, the variation in base composition within each cds was calculated from their differences in usage for the GC, AT, keto, amino, purine, and pyrimidine bases of the fifteen genes associated with breast cancer risk. Skewness has revealed that the base composition bias is related to transcription processes [19-21]. Positive GC skew represents richness of G over C and the negative GC skew represents richness of C over G [22]. Negative GC skew was found in case of *CDH1*, *PALB2*, *TP53*, *ERBB2* and *AR* genes indicating abundance of C over G, where as all other genes had positive GC skew. GC_3 skew was calculated [23] and the value ranged from 0.155 to -0.198, which suggested that GC composition at the third position of codon played an important role in the codon usage bias. Negative values of keto skew (k_{skew}) was observed as reported earlier by Zhang & Gerstein in human genome [24]. The k_{skew} values were negative for the cds sequences of *STK11*, *TP53*, *ERBB2* and *AR*. Wide variation was observed in purine, pyrimidine and amino skew [Fig-5], which might affect the gene expression patterns. Furthermore, we performed correlation analysis between gene expression level measured by *RCBS* with keto skew, purine skew and amino skew to elucidate the relationship on gene expression for each selected cds associated with breast cancer risk. Significant negative correlation was observed between keto contents (Pearson, r= -0.614, p<0.05), purine contents (Pearson, r= -0.605, p<0.05) and amino contents (Pearson, r= -0.596, p<0.05) with the expression level. These results suggested that, gene expression was linked with the substitution of A to C/G and T to G in the coding sequences due to observed keto, purine & amino skew respectively.

Amino Acid usage Contributes to Level of Gene Expression in Breast Cancer

Singer & Hickey [25], reported that background nucleotide GC contents affects the amino acid composition of proteins. The overall amino acid frequency of the selected genes associated with breast cancer risk [Fig-6] depicted that three amino acids namely serine, leucine and glutamate were mostly used. Conversely the least usage of two amino acids namely tryptophan and methionine was recorded. We compared these values with level of gene expression measured by RCBS and observed significant negative correlation between highly used amino acid and RCBS whereas non significant correlation between the least used amino acids and RCBS.

Discussion

The advent of whole genome sequencing of many organisms from virus to multicellular eukaryotes and the easily accessible DNA sequences from NCBI database have invited much attention of researchers to study the codon usage bias among different organ-

isms. Synonymous codons are not used with equal frequencies during translation in most sequenced genomes and show species specific deviation [1,9]. Several studies were carried out by different workers on synonymous codon usage bias in a wide variety of organisms including prokaryotes and eukaryotes [4,26-30], and till date in many organisms the codon usage patterns have been interpreted for diverse reasons.

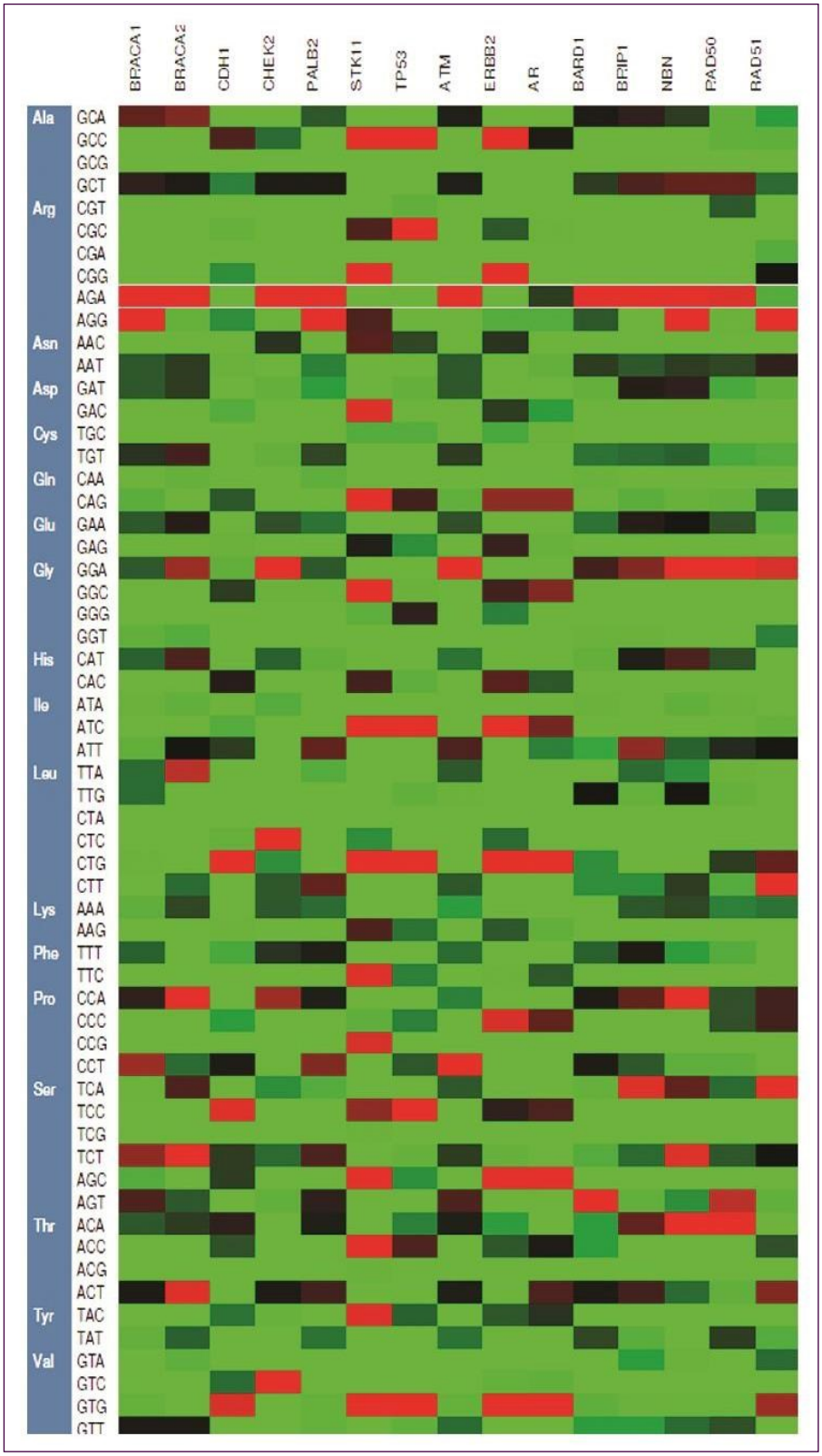

Fig. 4- Clustering of RSCU values of each codon among breast cancer genes. Each rectangular box on the map represents the RSCU value of a codon (shown in rows) corresponding to the breast cancer genes (shown in columns). The color coding indicates different RSCU values: green indicates RSCU<1, dark green & black RSCU>1 and red RSCU>1.6.

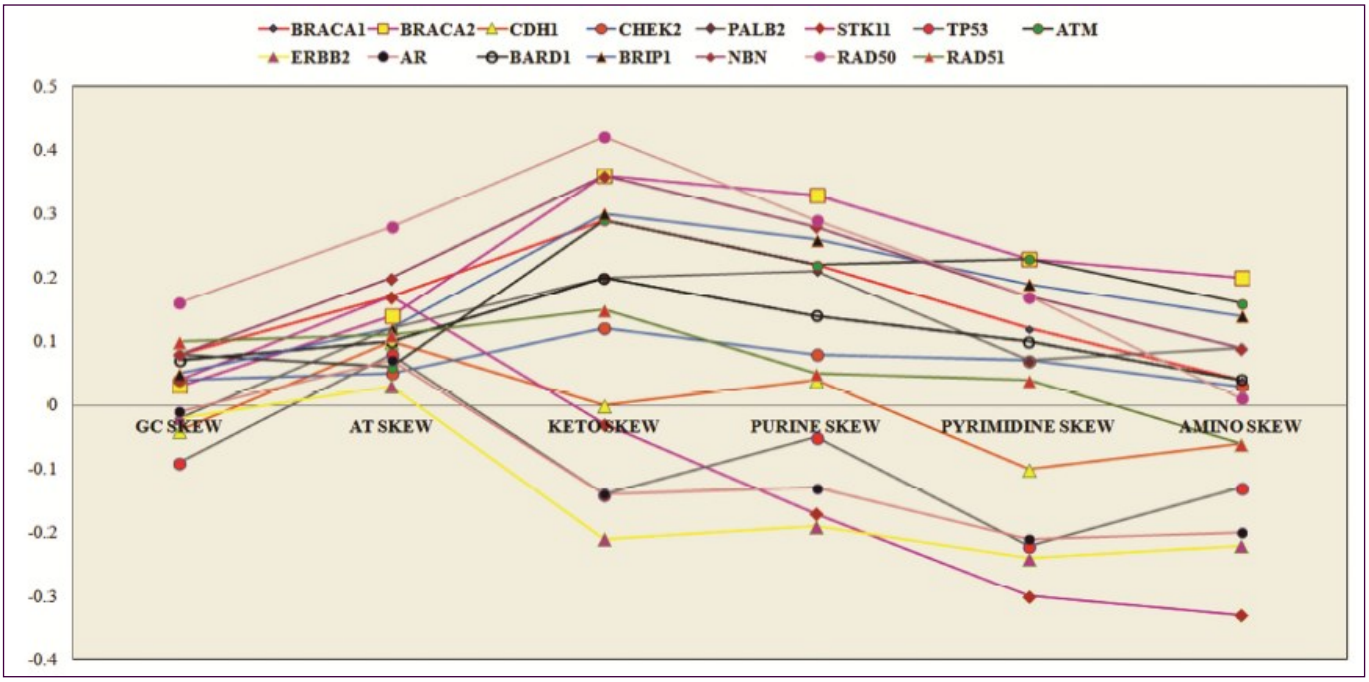

Fig. 5- Alterations in the nucleotide skewness among fifteen breast cancer genes

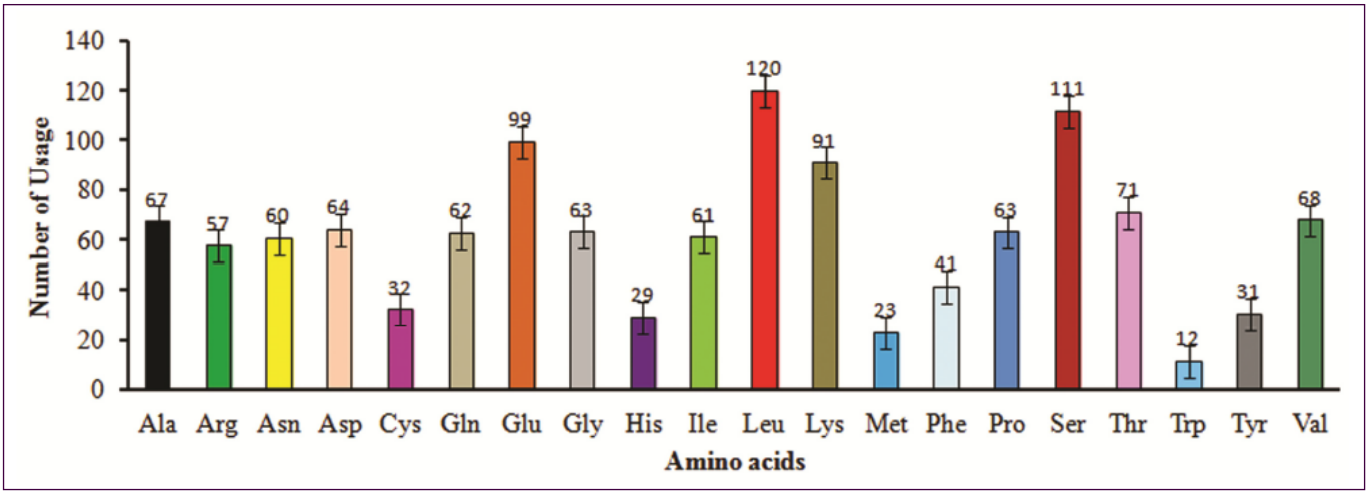

Fig. 6- Overall amino acids usage across fifteen genes associated with breast cancer. Number of usage of each amino acid is represented by particular color coding

Previously, it was reported that mutation pressure influences the codon usage pattern and codon bias is correlated to GC content of genes in human [31]. In this study to investigate the relationship between mutation pressure and codon bias, a neutrality plot of GC_{12} *vs* GC_{3s} [Fig-2] was carried out [16] and showed that all the fifteen selected breast cancer genes had narrow distribution of GC_{3s}. Further a significant positive correlation ($p<0.01$) between GC_{3s} and GC_{12} was also observed and the regression coefficient of GC_{3s} on GC_{12} was 2.595 [Fig-2], representing that both neutral mutation and selective constraint play major roles in the codon usage patterns of these genes [9,30].

Correlation between codon usage frequency and GC_{3s} showed that 12 codons with G/C ending base had negative correlation with GC_{3s} but not statistically significant ($p>0.05$) except for AGG and TGG. The codons, AGG for arginine and TTG for leucine amino acid had significant negative correlation ($p<0.05$) with respect to GC_{3s} [Fig-1b], [Fig-1c] indicated codon usage decreases with increase in GC bias due to GC_{3s} which was similar to the work done by Palidwor, et al [32] in prokaryotes, plants and humans.

In addition we observed wide variation of frequently and rarely used codons from the relative synonymous codon usage values of each codon among breast cancer risk genes [Fig-4] which suggested that compositional constraint was one of the factors in shaping the codon usage variation among these genes [14,33]. It should also be noted that AGA codon was over represented (RSCU>1.6) in nearly all breast cancer risk genes and codons with T -ending base was mostly favored in the coding sequences of these genes.

Previous studies reported that codon usage is correlated with gene expression level in a variety of organisms [6,34-38]. In general, codon bias and gene expression may be associated with mutational bias within genes transcribed at different levels [17]. Our results, showed that GC_{3s} had significant linear correlation with gene ex-

pression level (r = 0.607, p <0.05) as measured by *RCBS* [Fig-3], which was similar to the findings reported by Ma, et al [39] in human housekeeping genes, though the gene expression level was measured by *CDC* (Codon Deviation Index). These results suggested that, in addition to translational selection other factors like transcriptional selection, mRNA stability or biased gene conversion might be involved in CUB and gene expression. Moreover, the correlation of gene expression with keto, purine and amino contents was significantly negative which suggested that gene expression was linked with the substitution of A to C/G and T to G in the coding sequences due to observed keto, purine and amino skew respectively.

Our results also revealed that amino acids usage contribute to gene expression in breast cancer. A strong negative correlation was observed between highly used amino acids (serine, leucine, and glutamate) and gene expression measured by *RCBS* which indicated that increased usage of these amino acids leads to a decrease in breast cancer risk gene expression.

Conclusion

Our analysis revealed that both neutral mutation and selective constraint play major roles in the codon usage patterns of breast cancer risk genes. Our results further show that gene expression level is linked with alterations in the nucleotide skewness. Moreover, highly used amino acids had negative impact on the expression of gene as measured by *RCBS*. Compositional analysis of breast cancer gene products revealed that three amino acids namely serine, leucine and glutamate are mostly preferred whereas two amino acids namely tryptophan and methionine are least preferred. Furthermore, the level of breast cancer gene expression (measured by RCBS) revealed a significant negative correlation with the highly used amino acids but a non-significant correlation with the least used amino acids.

Conflicts of Interest: None declared.

References

[1] Grantham R., Gautier C., Gouy M., Jacobzone M. & Mercier R. (1981) *Nucleic Acids Res.,* 9(1), 43-74.

[2] Marin A., Bertranpetit J., Oliver J.L. & Medina J.R. (1989) *Nucleic Acids Res.,* 17(15), 6181-6189.

[3] Hooper S.D. & Berg O.G. (2000) *Nucleic Acids Res*., 28(18), 3517-3523.

[4] Lavner Y. & Kotlar D. (2005) *Gene*, 345(1), 127-138.

[5] Supek F. & Vlahovicek K. (2005) *BMC Bioinformatics,* 6, 182.

[6] Duret L. & Mouchiroud D. (1999) *Proc. Natl. Acad. Sci. USA.,* 96(8), 4482-4487.

[7] Li W.H. (1987) *J. Mol. Evol.*, 24(4), 337-345.

[8] Xu C., Cai X., Chen Q., Zhou H., Cai Y. & Ben A. (2011) *Evol. Bioinform. Online*, 7, 271-278.

[9] Nair R.R., Nandhini M.B., Sethuraman T. & Doss G. (2013) *Springerplus*, 2, 492.

[10] Febriansah R., Putri D.D., Sarmoko, Nurulita N.A., Meiyanto E., & Nugroho A.E. (2014) *Asian Pac. J. Trop. Biomed.*, 4(13), 228-233.

[11] Roymondal U., Das S. & Sahoo S. (2009) *DNA Res*., 16(1), 13-30.

[12] Fox J.M. & Erill I. (2010) *DNA Res*., 17(3), 185-196.

[13] Sharp P.M. & Li W.H. (1986) *Nucleic Acids Res*., 14(19), 7737-7749.

[14] Butt A.M., Nasrullah I. & Tong Y. (2014) *PLoS One*, 9(3), e90905.

[15] Zhou T., Gu W., Ma J., Sun X. & Lu Z. (2005) *Biosystems*, 81 (1), 77-86.

[16] Sueoka N. (1988) *Proc. Natl. Acad. Sci. USA.*, 85(8), 2653-2657.

[17] Francino M.P. & Ochman H. (2001) *Mol. Biol. Evol*., 18(6), 1147 -1150.

[18] Louie E., Ott J. & Majewski J. (2003) *Genome Res.*, 13(12), 2594-2601.

[19] Beletskii A. & Bhagwat A.S. (2001) *J. Bacteriol*., 183(21), 6491-6493.

[20] Fujimori S., Washio T. & Tomita M. (2005) *BMC Genomics,* 6, 26.

[21] Touchon M. & Rocha E.P. (2008) *Biochimie*., 90(4), 648-659.

[22] Tillier E.R. & Collins R.A. (2000) *J. Mol. Evol*., 50(3), 249-257.

[23] Tatarinova T.V., Alexandrov N.N., Bouck J.B. & Feldmann K.A. (2010) *BMC Genomics*, 11, 308.

[24] Zhang Z. & Gerstein M. (2003) *Nucleic Acids. Res*., 31(18), 5338-5348.

[25] Singer G.A. & Hickey D.A. (2000) *Mol. Biol. Evol*., 17(11), 1581-1588.

[26] McInerney J.O. (1998) *Proc. Natl. Acad. Sci. USA.,* 95(18), 10698-10703.

[27] Ermolaeva M.D. (2001) *Curr. Issues Mol. Biol*., 3(4), 91-97.

[28] Jenkins G.M. & Holmes E.C. (2003) *Virus Res*., 92(1), 1-7.

[29] Gu W., Zhou T., Ma J., Sun X. & Lu Z. (2004) *Virus Res*., 101 (2), 155-161.

[30] Liu H., Huang Y., Du X., Chen Z., Zeng X., Chen Y. & Zhang H. (2012) *Genet. Mol. Res*., 11(4), 4695-4706.

[31] Eyre-Walker A.C. (1991) *J. Mol. Evol*., 33(5), 442-449.

[32] Palidwor G.A., Perkins T.J. & Xia X. (2010) *PLoS One,* 5(10), e13431.

[33] Zhang Z., Dai W. & Dai D. (2013) *PLoS One*, 8(11), e81469.

[34] Duret L. (2002) *Curr. Opin. Genet. Dev.*, 12(6), 640-649.

[35] Gouy M. & Gautier C. (1982) *Nucleic Acids. Res.*, 10(22), 7055-7074.

[36] Ikemura T. (1985) *Mol. Biol. Evol.*, 2(1), 13-34.

[37] Castillo-Davis C.I. & Hartl D.L. (2002) *Mol. Biol. Evol.*, 19(5), 728-735.

[38] Goetz R.M. & Fuglsang A. (2005) *Biochem. Biophys. Res. Commun*., 327(1), 4-7.

[39] Ma L., Cui P., Zhu J. & Zhang Z. (2014) *Biol. Direct.*, 9(1), 17.