Bioengineering and Bioscience 3(3): 29-38, 2015
DOI: 10.13189/bb.2015.030301 http://www.hrpub.org

Cross-species Comparison of CpG Density in the Promoter Regions of Protein Kinase Oncogenes

Gabriel G. Pioth[1], Kong-Nan Zhao[2], Guang B. Liu[1,*]

[1]School of Health and Wellbeing, Faculty of Health, Engineering and Sciences, University of Southern Queensland, Australia
[2]Centre for Kidney Disease Research—Venomics Research, The University of Queensland School of Medicine, Translational Research Institute, Australia

Copyright © 2015 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In this report, we investigated CpG density occurred in promoter regions (PRs) and downstream flanking regions (DFR) of 61 human protein kinase oncogenes (PKOs), together with other three species: chimpanzee, mouse and horse. The quantified numbers of CpGs in the PRs of human PKOs were much higher than those of chimpanzee, mouse and horse, suggesting that the CpG density changes among the four species are associated with species evolution. Human PKOs with relatively high number of CpGs in the PRs showed stronger gene expression than the mouse PKOs in tumour tissues, but not in normal tissues. Furthermore, human PKOs with extremely high density of CpGs in the PRs exhibited much lower expression in tumour tissues than in normal tissues. Our data initially suggest that the occurrence and density of CpGs in the PRs of PKOs play an important role in regulating gene expression associated with the tumorigenesis. Thus, further improvement of our understanding of the density and spatial arrangement of CpGs in the PRs of PKOs and other oncogenes involved in tumorigenesis is very important for providing preventive and therapeutic strategies for human cancer.

Keywords CpG Island , Protein Kinase Oncogene , Oncogenesis , Promoter Region, Transcription Start Sites

1. Introduction

Presence of high CpGs density on the gene promoter region (PR) is believed to be associated with high gene expression [1-3]. This PR with high CpG density stretches from 0.5 to 2 kilo base pairs (kbp) upstream of gene translation start sites (TSS) [4-6]. Within the region when a section longer than 200bp is characterised by presence of 70% G+C dinucleotides, then this section is often referred to as a CpG Island [7]. The CpG Islands (CGIs) are found in two-third of house-keeping genes and a quarter of tissue-specific genes, respectively [8].

What makes the presence of CpG dinucleotides in the CGIs significant is the fact that the cytosines in a CGI are not often methylated. High CpG density make the whole CGI serves as a '*Passover*' sign during the *de novo* methylation of blastocyst genome by DNA cytosine methyltransferases (*Dnmt3a* and *Dnmt3b*) [9]. Thus, any mutation, which adds cytosines into the CGI, is more likely to increase the density of CpGs in the gene PRs. While mismatches of dinucleotides in young cells are repaired quickly, the ageing cells do not or never correct their genomic errors. This is because of that majority of genes involved in the DNA repair and methyltransferases (Dmnt1) are rarely transcribed in senescent cells [10-12]. Consequently these genomic changes are passed down from generation to generation, which contribute to high CpG density in the gene PRs. The high CpG density causes high gene expressions reminiscent of cancer genes. It has been shown that CpG island methylation in the PRs is likely to be an important mechanism in modulating gene expression in several cancers [13, 14]. In addition, an apparent drop of CpGs in the PR causes inactivation of tumour suppressor genes, aberrant cellular pathways such as DNA repair, cell cycle, apoptosis and cell division and imprinted genetic disorder [15-18].

Oncogenesis is related to various processes including the accumulation of deoxyribonucleic acid (DNA) nucleotides changes on the precancerous cell genome. The mechanisms by which the nucleotides (genetic/epigenetics) changes cause cancer have been proposed in proto-oncogenes, tumour suppressor and stability genes. However, the relationship between unusual variation of CpG dinucleotides in the human protein kinase genes PRs remains to be deeply investigated. The aim of this study is to investigate whether occurrence and density of CpG in the PRs of protein kinase oncogene (PKO) contributes to human oncogenesis. The research quantified the number of CpG dinucleotides in the 1K base pairs upstream of TSS and downstream flanking region of 61 PKOs in four mammalian species. The information acquired from this study has a potential for a

better understanding of oncogenesis in human.

2. Material and Method

2.1. Screening the Protein Kinase Oncogenes from the Human Kinome

The protein kinase oncogenes (PKOs) used in this study was screened out from the 518 protein kinase (PK) genes of the human kinome, which was published by Manning and colleagues [19]. Each of the 518 PK genes was then searched for in the Online Mendelian Inheritance in Man database (OMIM) at the National Centre for Biotechnology Information (NCBI) website, (http://www.ncbi.nlm.nih.gov/). The information retrieved from the database revealed full description of the gene, its functions (oncogenic or not) and peer reviewed journal articles which confirmed pro-oncogenic potential of the gene. Based on this information, a decision is made to either include or exclude it from the list of possible PKOs that would be investigated. The same process was repeated for the whole of 518 PK genes. This list of thoroughly screened PKOs from human, acted as a template for downloading the DNA sequences of orthologous PKOs from the mouse, horse and chimpanzee.

2.2. Obtaining DNA Sequences from the Ensembl Database

The DNA sequences of human PKOs were searched from *Ensembl* release 54 (GRCh37, http://www.ensembl.org/Homo_sapiens/Info/Index) where 61 human PKOs were selected for intensive analysis and cross-species comparison.

The retrieved data contained gene markers and the real gene links from different assemblies, in order to obtain the accurate gene sequences. To determine the accurate position of the TSS, each downloaded DNA sequence was carefully compared with the alignment of the same gene from Evidence Viewer (EV) and Sequence Viewer (SV) of NCBI website. The down-stream section of the PR (10–25 bases immediately upstream from the TSS) was aligned back with the corresponding section of the same gene illustrated on the EV and SV. Complete alignment confirmed correctness of the downloaded gene-including DNA sequence.

Each gene-including DNA sequence downloaded includes 5K bases (bs) upstream from the transcription start site (TSS), the full length of the gene (extron+intron) and the 1K bs downstream from the 3' end of the gene (stop codon). If the inter-gene distance (the distance between the TSS of the gene being downloaded and the 3' end of the upstream gene) is less than 5K bs, then the downloaded DNA sequence includes the full sequence within the inter-gene distance, the full length of the gene and 1K bs downstream from the 3' end of the gene. In each gene-including DNA sequence, the sequence upstream of the TSS is termed as a putative PR and the sequence downstream of the 3' end of the gene is the downstream flanking region (DFR [20, 21].

The sequences of 1K bs at either end of the gene were selected for counting the density of CpGs. The gene sequence key elements such as: (1) 15 bs from the start codon (ATG) and the stop codon; (2) 1K bs from either flanking region of the transcript; (3) gene descriptions; (4) location of the gene either on plus (+1) or minus (-1) strand; (5) the number of exons; (6) location of the gene on the chromosome; (7) the coordinate of each exon and intron, were fed into the Matlab program for analysis. The same process was repeated for the 61 PKOs in the human, mouse, horse and chimpanzee. The analysed results were stored in both Matlab and excel *tab-limited* formats for convenient access.

2.3. Downloading Gene Expression Data

In order to investigate whether the CpG density at the 1k bp upstream of TSS was related to high expression level or not, twenty (20) PKOs with promoter region (PR) CpG densities that range from 1-138 were selected. The expression data of these PK oncogenes were downloaded from the GeneHub-GEPIS website. Gene expression data were retrieved in 34 normal and tumour tissues/organs from human and mouse. The gene expression data of the PKOs from both normal and tumour tissues were derived from the human and mouse (the data from other two species are unavailable currently) expressed sequence tag (EST). These data were used to compare the quantitative relationship between the expression levels of PKOs and the CpG density in their PRs.

2.4. How Gene Expression Data Are Determined From GeneHub-GEPIS Database

The algorithms used to calculate gene expression profile were previously described by Zhang and colleagues [22]. These expression data were integrated into human and mouse genes databases by GeneHub, a component of GeneHub-GEPIS database. The merged databases (GEPIS and GeneHub) incorporated gene specificity and biological characteristics relevant to the expression data which has made these data suitable for this research [22].

3. Results

3.1. CpG Density at the Promoter Region (PR) of Protein Kinase Oncogenes

The study screened out 61 protein kinase oncogenes from the human kinome, which were used as template to download the homologous genomic sequences from chimpanzee, mouse and horse. The DNA sequences of these species were fed into self-written program for analysis. The

CpG island is a DNA sequence equal or longer than the starting window (SW), in which the observed/expected (O/E) value of CpGs should be equal or higher than 0.6 and G%+C%>50 [23]. The method used for CpG island identification was according to a previous report [18, 20] with some modifications. A SW of 200 bs was applied from the 5' end of the DNA sequence, with the G%+C% and O/E value of the CpGs being calculated within the SW while the SW was moving step-by-step towards the TSS. To avoid the mathematical CpG island, 7 CpGs/200 bs was selected as a reasonable cutoff for a CpG island identified in the SW [23]. Two individual CpG islands were connected if they were separated by less than 50 bs. The CpG island information that included the number of CpGs in each CpG island, the density of the CpG island (number of CpG/100 bs), the position of each CpG within the CpG island, the starting and ending coordinates of the CpG island along the gene sequence were collected for statistical analysis using two-way ANOVA.

3.1.1. Human

Figure 1 shows the CpG density and distribution in both PRs and DFRs of 61 human PKOs. On an average, 61 human PKOs had a number of 51.5 CpGs in their PRs and 15.8 in their DFRs, respectively. Forty-six PKOs (75.4%) showed very high CpG density with over 20 CpGs identified in their PRs while fifteen PKOs (24.6%) had a low CpG density with the CpGs less than 20 (Figure 1). Five PKOs: HIPK2, IGF1R, PKCE, CAMK1B and PIM1 showed the highest density of CpGs in their PRs, with a number of 139, 134, 119, 116 and 103 CpGs quantified while they had a number of 40, 13, 20, 18 and 23 CpGs identified in their DFRs. Three PKOs had a very low density of CpGs in their PRs, with MER only containing 1, FRK and TRKA having 10 each while they also had 10, 5 and 8 CpGs quantified in their DFRs (Fig 1).

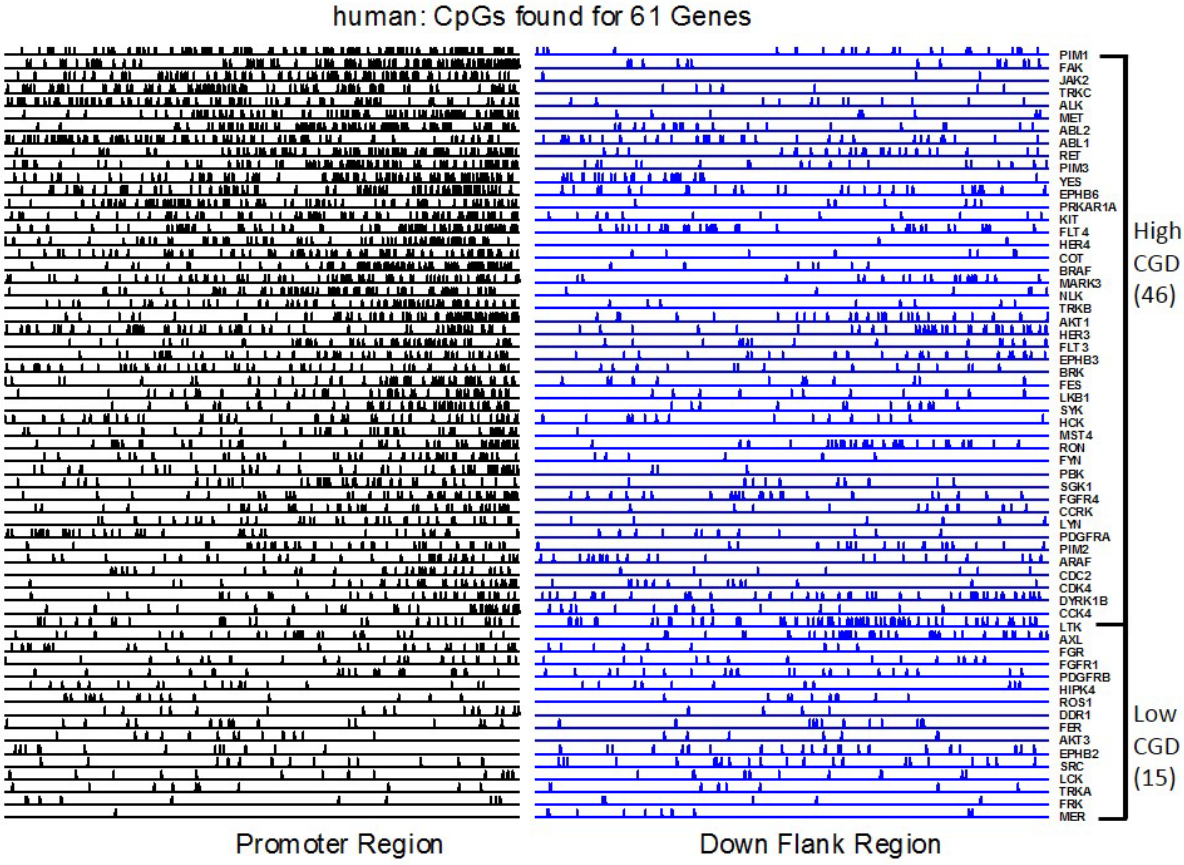

Figure 1. Human PK oncogene CpG density at the promoter regions (PRs) and downstream flanking regions (DFRs). The CpG density clusters (see closed brace on the right hand side) are labelled as high and low CpG Density (CGD). Low CGD means less than 20 dinucleotides on 1kbp upstream of TSS; High CGD is greater than 20 dinucleotides on 1kbp upstream of TSS. Each of the dark and blue vertical spikes represents a pair of CpG dinucleotides and horizontal line represents the length of 1kb of the PR and DFR respectively. PK genes are plotted in a descending order of the CpG density in their PRs. The number in bracket next to the low or high CGD represents the number of genes in that cluster. The result shows 75% of the human PKOs have high CGD and 25% have low CGD in their PRs respectively.

3.1.2. Chimpanzee

Figure 2 shows the CpG density and distribution in both PRs and DFRs of 61 chimpanzee PKOs. The average number of CpGs in the PRs and DFRs of the 61 chimpanzee PKOs were 44.7 and 14.6, respectively. Among the 61 PKOs, 42 (68.9%) had a high CpG density while 31.1% had a low CpG density in their PRs. Only two Chimpanzee PKOs showed to have more than 100 CpGs in their PRs, with SGK1 having 121 and BRK 104 (Fig 2). Both genes had a low density of CpGs in their DFRs too, with 8 CpGs detected (Fig 2). Furthermore, seven PKOs had a very low density of CpGs with less than 10 CpGs identified in their PRs, which were LCK (0), FER (5), FAK (5), EPHB6 (8), DYRK1B (9), FRK(9) and FGR (9) (Fig 2). Two genes EPHB6 (28) and DYRK1B (44) had a high CpG density in their DFRs while the other five PKOs had low CpG density ranging from 4-13 in their DFRs (LCK, 4; FER, 13; FAK, 7; FRK, 3 and FGR, 10) (Fig 2).

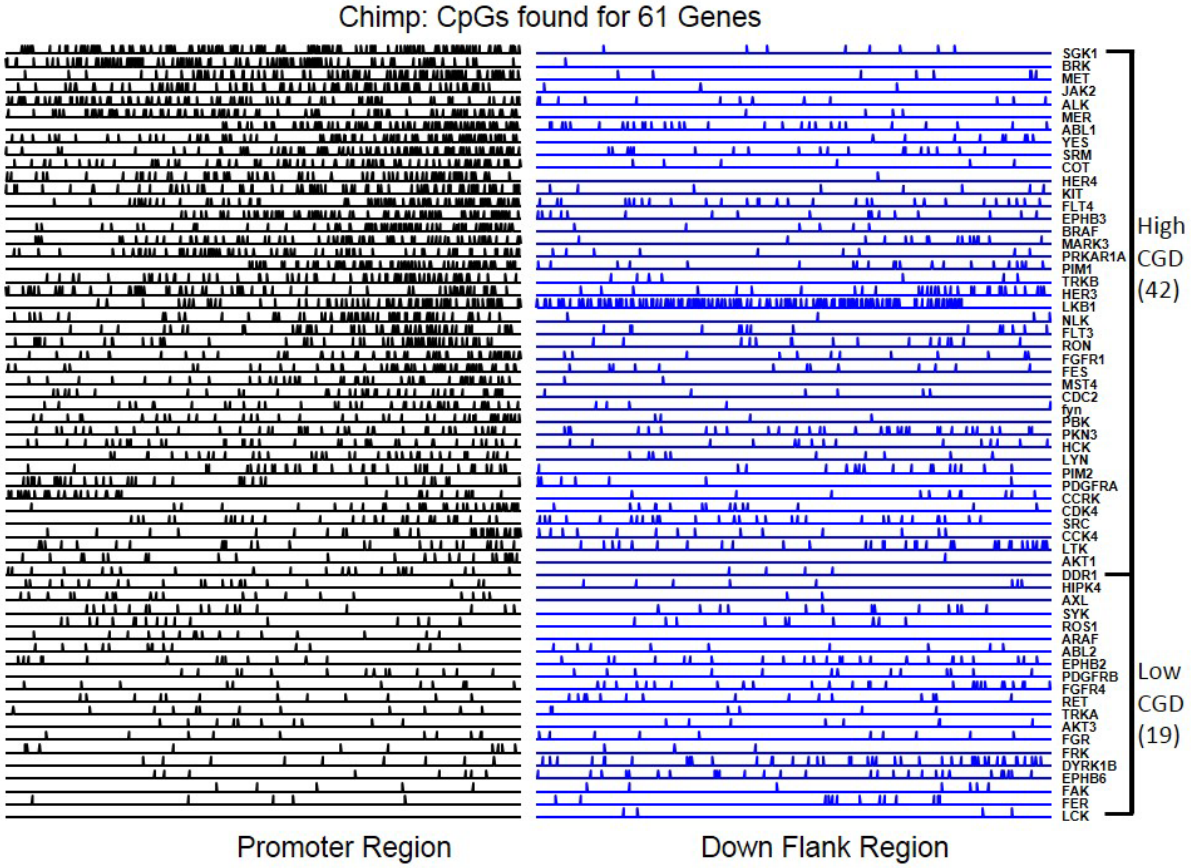

Figure 2. Chimpanzee PK oncogene CpG density at the promoter and downstream flanking regions. In their PRs 69% of the chimpanzee PKOs have high CGD, and 31% have low CGD respectively.

3.1.3. Mouse

Figure 3 illustrates the CpG density and distribution in both PRs and DFRs of 61 mouse PKOs. Results showed that the 61 mouse PK genes had an average of 38.6 CpGs in their PRs and 13.0 in their DFRs. 65% mouse PKOs had high CpG density and 35% had low density in their PRs respectively (Fig 3). Furthermore, three mouse PKOs exhibited to have over 100 CpGs in their PRs, with HIPK2 (137), IGF1R (133) and PKCE (116). These genes had 14, 21 and18 CpGs identified in their DFRs (Fig 3). Six mouse PKOs had less than 10 CpGs occurred in their PRs, which were ROS1 (5), AKT3 (5), FRK (7), FER (7), LTK (7) and MET (8). Three genes also had less than 10 CpGs occurred in their DFRs with AKT3 (5), FRK (3), and MET (7) while the other three genes: ROS1 (13), FER (15) and LTK (24) had more than 10 CpGs in their DFRs.

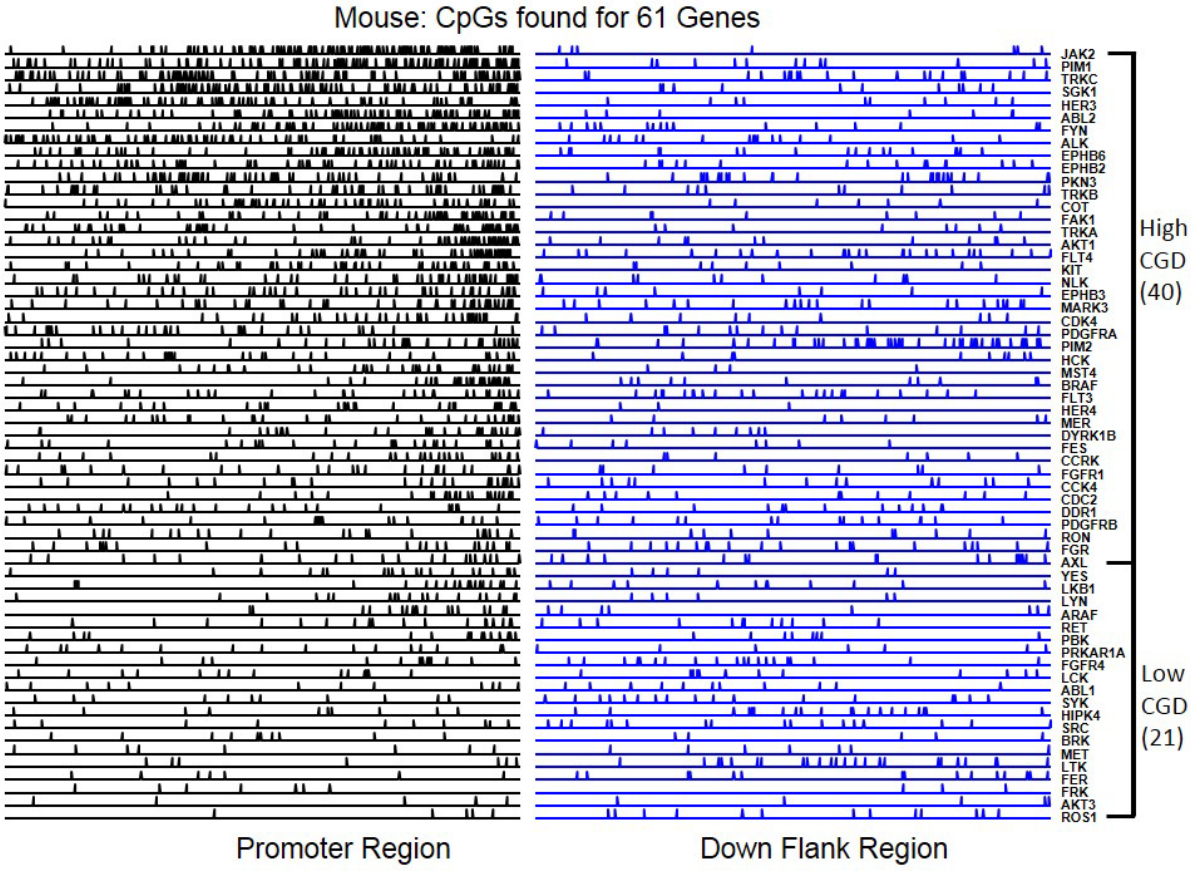

Figure 3. Mouse PK oncogenes CpG density at the promoter and downstream flanking regions. In their PRs 66% of the mouse PKOs have high CGD and 34% have low CGD respectively.

3.1.4. Horse

Figure 4 illustrates the CpG density and distribution in both PRs and DFRs of 61 horse PKOs. The 61 horse PKOs had an average density of 31.9 CpGs in their PRs and 17.9 in their DFRs, with 32 genes (52%) having high CpG density and 29 (48%) having low density in their PRs. However, none of the 61 horse PKOs had over 100 CpGs occurred in the PR. In contrast, six genes had only 10 CpGs detected in their PRs and seven had less than 10 CpGs, with CDC2 (6), HER4 (6), AKT3 (6), PBK (8), ABL2 (9), FLT4 (9) and COT(9). The CpG density in the DFRs of these genes in their PRs was alos low, but varied among them.

The results clearly suggests that human PKOs have the highest density of CpGs occurred in their PRs, then the chimpanzee and mouse, and the horse the lowest. In addition, the CpG density in PRs is not associated with that in DFRs.

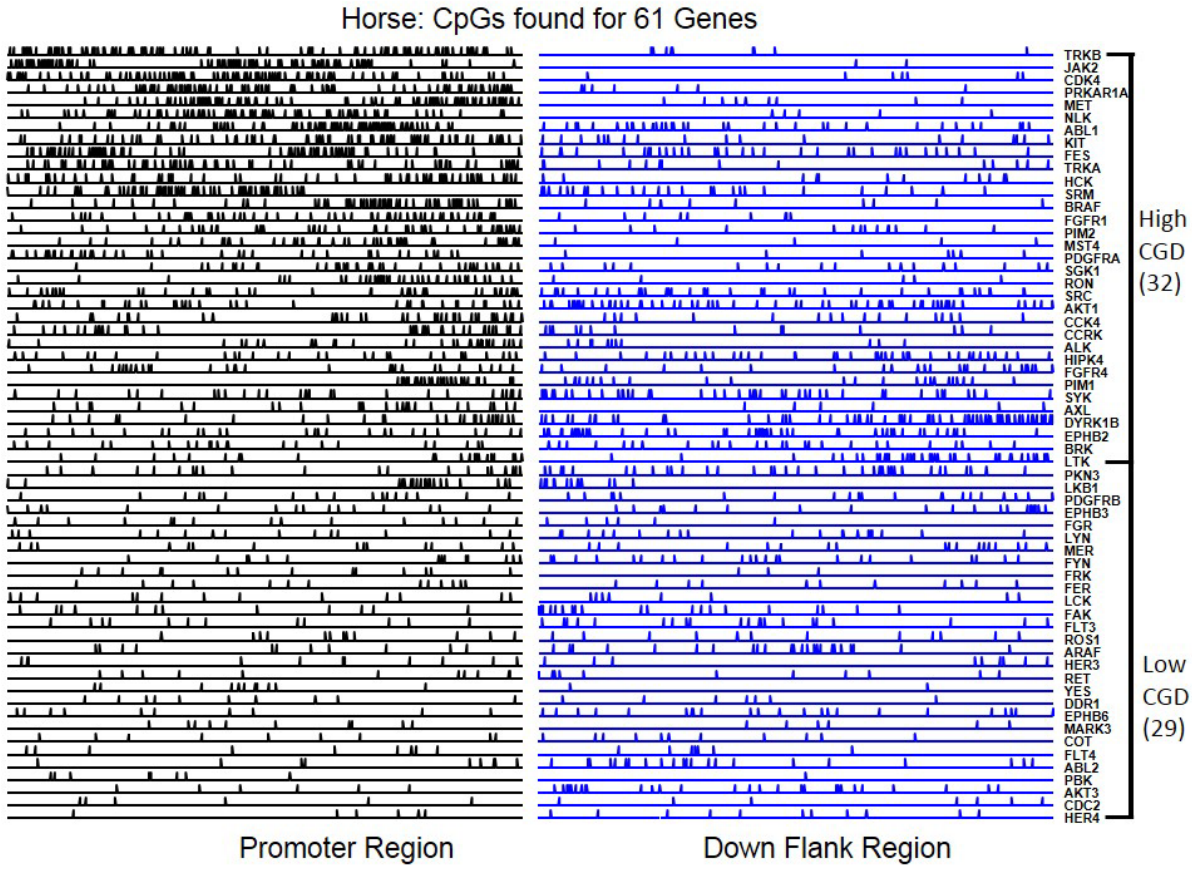

Figure 4. Horse PK oncogene CpG density at the promoter region and downstream flanking regions. In their PRs 52% of the horse PKOs have high CGD, and 48% have low CGD respectively.

3.2. Comparison of the CpG Distribution in Selected PKOs

We compared the CpG distribution in the PR and DFR of selected PKOs of four species in order to estimate their evolutionary clue [19, 24]. Five PKOs were selected where substantial cross-species changes seemed to have taken place in their PRs and DFRs. For instance, the human AKT1 oncogene showed a higher number of CpGs in the PR than in any of the other three mammalian species (Fig 5, top group). Similarly, the CpG density in the PRs of FAK1 and HIPK2 has remained high in human and mouse but become low in chimpanzee and horse (Fig 5, 2nd and 3rd groups from top). However, The TRKA oncogenes appeared to have lost majority of CpGs in the PRs of human and mouse but remained high in the chimpanzee and horse (Fig 5, 2nd group from bottom). The LKB1 showed an apparent difference to the previous observations. The Chimpanzee LKB1 has a similarly high density of CpGs in both PR and DFR, making an apparent different pattern from those observed in human, mouse and horse PKOs which are comparatively low in both the PRs and DFRs (Fig 5, bottom group)

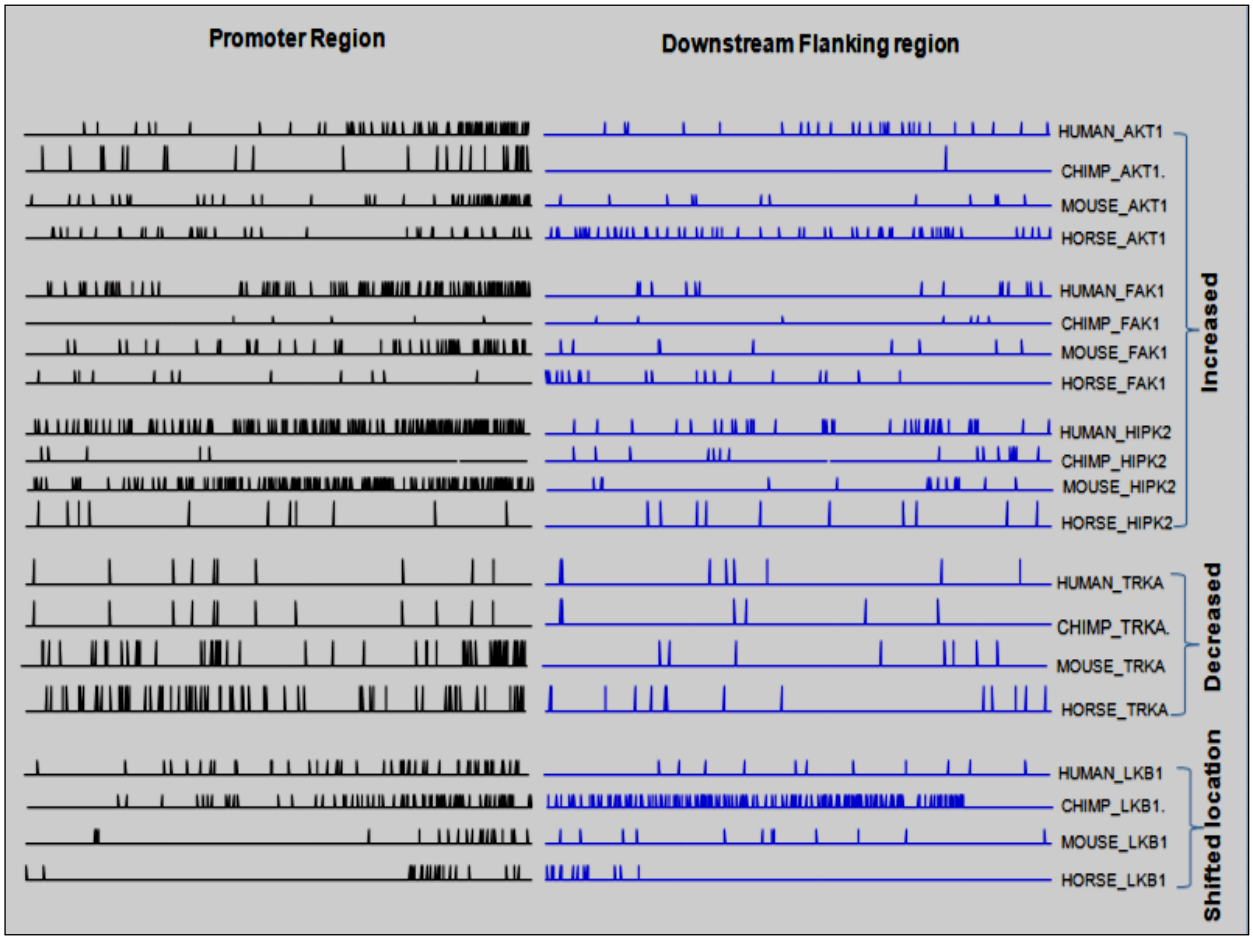

Figure 5. Comparison of CpG density in the PR and DFR of 5 selected PKOs among four species. The labels on right hand side shows clusters of PKOs that have decreased / increased or shifted CpG density at the PRs. The human AKT1 shows a higher number of CpGs in the PR than in any of the other three mammalian species (top group). The CpG density in the PRs of FAK1 and HIPK2 has remained high in human and mouse but become low in chimpanzee and horse (2nd and 3rd groups from top). The TRKA appears to have lost majority of CpGs in the PRs of human and mouse but remained high in the chimpanzee and horse (2nd group from bottom). The LKB1 showed an apparent difference to the previous observations. The Chimpanzee LKB1 has a similarly high density of CpGs in both PR and DFR, making an apparent different pattern from those observed in human, mouse and horse PKOs which are comparatively low in both the PRs and DFRs (bottom group).

3.3. The CpG-rich PKO Promoters and Their Expression Level in Tumour and Normal Tissues

We next determined whether and how the CpG density in the PRs of the PKOs is associated with gene expression in tumour and normal tissues in both mouse and human because of no data available for Chimpanzee and horse. In human tumour tissues, the PKOs having more than 57 CpGs in their PRs did not show high levels of gene expression, suggesting that very high density of CpGs in the PR does not favour the PKO expression in tumour tissues (Fig 6a). Four peaks of the gene expression of the PKOs having a number of 12, 25, 36 and 46 CpGs were observed when the CpG density ranged from 10 to 46, (Fig 6a). In human normal tissues, peaked gene expression was observed for the PKOs which have a number of 10, 15, 25, 35, 46, 57, 68 and 94 CpGs in their PRs (Fig 6b). Three PKOs containing higher numbers of 57, 68 and 94 CpGs had much higher gene expression, especially the 57 CpGs-containing PKO showed extremely high gene expression (Fig 6b), in contrast, a very low expression of these PKOs was in tumour tissues (Fig 6a). Generally, the PKOs that are highly expressed in tumour tissues tend to have lower expression in normal tissues, associating with the CpG density in their PRs (Fig 6a, b).

PKOs with the CpG density range of 20-76 in their PRs have much higher expression level in human than in mouse in tumour tissues (Fig 6a). But the PKOs of the two species have similar expression pattern and level in normal tissues, excepted for two genes (Fig 6b). The 57 CpGs-containing human PKO showed extremely high expression in normal tissues (Fig 6b), which was at least five times higher than the 57 CpGs-containing mouse PKO. In contrast, a 25 CpGs-containing human PKO had only one of third expression level, compared with the 25 CpGs-containing mouse PKO (Fig 6b). On an average, the PKO with high CpG density in the PRs have high expression in human than in mouse in both tumour and normal tissues (see table 1), indicating that CpG density in the PRs associated with the PKO gene expression is specie-related.

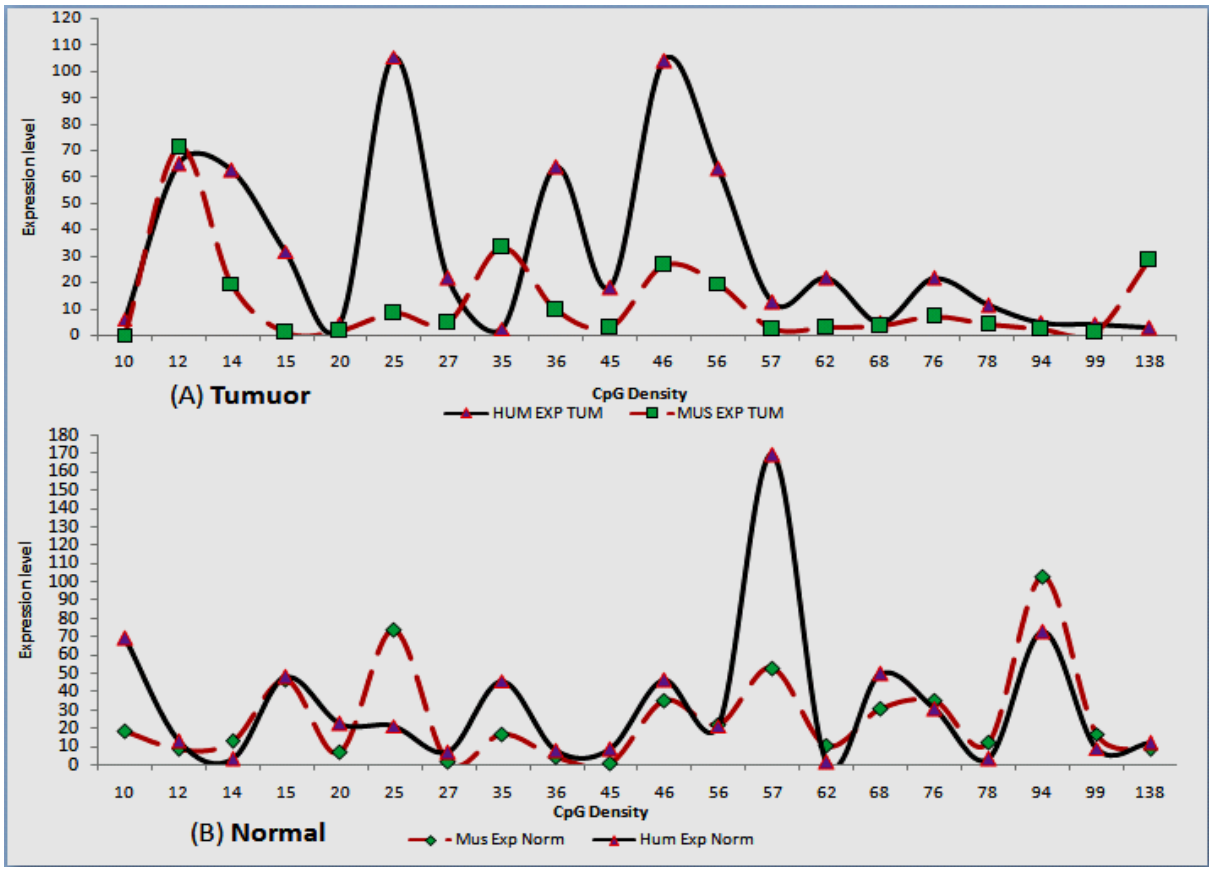

Figure 6. Comparison of the relationship between gene expression level and CpG density in protein kinase oncogene (PKO) in human and mouse. The graph (A) compares the expression level of PKO in tumour tissues whereas graph (B) contrast them in normal tissues. The vertical coordinates are expression levels and the horizontal coordinates are the densities of CpG dinucleotides in the promoter region.

Table 1. Comparison of expression levels in high and low CpG density at promoter region of PKO in tumour and normal tissues

Oncogene	Expression of PKO with high CpG density at the promoter region in tumour tissues		Expression of PKO with high CpG density at the promoter region in normal tissues		Expression of PKO with low CpG density at the promoter regions in tumour tissues		
	Human	Mouse	Human	Mouse	Oncogene	Human	Mouse
ABL2	14.03	0	7.05	1.32	ABL1	61.91	99.92
AKT1	227.72	12.67	60.84	31.6	AKT3	0.91	0.8
FAK1	60.63	18.98	67.59	18.1	DDR1	103.97	25.73
FYN	21.25	4.28	44.79	34.3	FER	4.25	1.92
PIM1	36.22	0.8	21.8	6.64	FRK	2.8	27.7
SGK1	63.32	69.15	187.02	213.85	LCK	12.43	14.32
SRM	44.67	7.86	44.67	17.95	MER	10.85	3.98
VEGFR1	4.45	3.21	20.9	71.59	SRC	20.83	6.87
Average	59.04	14.62	56.83	49.42	Average	27.24	22.66

Highlighted numbers (red) are expression levels with statistical significance (P<0.02).

4. Discussion

Previous studies have reported that unmethylated CpGs in the PR served as a genomic footprint of replication origin [25, 26]. In this report, we used human 61 PKOs as templates to compare the number of CpGs in the PR of the same genes in chimpanzee, horse and mouse genomes. As a result, the human PKOs had significantly higher density of CpGs occurred in their PRs, compared with the chimpanzee, mouse and horse PKOs. For instance, human focal adhesive kinase (*FAK2*) has more than 50 CpGs in its PR while the horse *FAK2* only has 11 (see appendix 1). However, the human genome did not show that all the examined PKOs consistently have higher density of CpGs in their PRs than the other three species. A typical sample is that human *TRKA* has substantially decreased the CpGs density in the PR with

only 11 CpGs detected, compared with the horse *TRKA* gene that has 41 CpGs in its PR. The significant difference of the CpG density in the PKO PRs among the four compared species is probably due to their evolutionary impacts and changes. It is well documented that evolutionary tree of human genome is distinctly from that of the other mammalian species [27]. It has been reported that the evolutionary impacts and changes often target gene promoter region leading to a diverse arrangement of the nucleotides in this region [5, 28]. Molecularly, genetic transposition and recombination during the gene replication may play an important role in generating the high density of CpGs in both PRs and DFRs of human PKOs [29, 30]. The genetic transposition also contributed to the chimpanzee *LKB1* gene to have very high density of CpGs in its DFR [31, 32]. Thus, it is likely that genetic transposition plays a key role in determining that chimpanzee *LKB1* gene has 63 CpGs in PR and 133 CpGs in DFR while human *LKB1* gene only 47 and 11 in these two regions in the present study.

In the present study, we observed first that human PKOs with a relatively high CpG density in the PRs are highly expressed in tumour tissues, consistent with previous studies that high CpGs density in the gene PR is associated with high gene expression [1-3]. This is probably due to the accumulations of methylated cytosines in the CpG-rich promoter regions that increase potentially the turnover rates of proto-oncogenes [33]. However, we observed further that human PKOs with a very high CpG density in the PRs exhibit very high expression in normal tissues. Thus, it appears that CpG occurrence and density in the PKO PRs may play a differential role in regulating their gene expression in human tumour and normal tissues. It has been reported that the occurrence and density of CpGs in the PRs of estrogen receptor genes and insulin-like growth factor 2 gene are age-dependently varied in human tumours [34-37]. The variability of CpG density in the PR contributed in part to the high cancer risk associated with age over 70 in human to [38]. Thus, the occurrence and density of CpGs in the PR of PKOs and other oncogenes may be potentially served as an indicator of oncogene expression associated with the development of sporadic cancers in human.

5. Conclusions

The occurrence and density of CpGs in the PRs of the 61 PKOs differ from one gene to another within individual species. On an average, human PKOs have much higher density of CpGs in the PRs than the other three species (chimpanzee, horse and mouse). The human PKO with relatively high density of CpGs (between 12-46) in the PRs are highly expressed in tumour tissues than in normal tissues. Differently, the human PKOs with extremely high density of CpGs (>57) in the PRs are highly expressed in normal tissues than in tumour tissues. The mechanism that CpGs density is linked to gene expression of the PKOs in human remains to be investigated.

Authors' Contributions

Gabriel Pioth: data collection, analysis and first draft production.

Kong-Nan Zhao: manuscript improvement.

Guang B Liu: Project designation, software design and development, refinement of manuscript.

Conflict of Interest

There is no conflict of interest in publishing this manuscript.

REFERENCES

[1] Ponger, L., L. Duret, and D. Mouchiroud, (2001). *Determinants of CpG islands: expression in early embryo and isochore structure.* Genome Res, **11**(11): 1854-60.

[2] Kundu, T.K. and M.R. Rao, (1999). *CpG islands in chromatin organization and gene expression.* J Biochem (Tokyo), **125**(2): 217-22.

[3] Yu, G., X. Zhang, H. Wang, D. Rui, et al., (2008). *CpG island methylation status in the EpCAM promoter region and gene expression.* Oncol Rep, **20**(5): 1061-7.

[4] Einstein, D.B., B. Wessels, B. Bangert, J. Lewin, et al., (2009). *Phase II Trial Combining Conventional Radiotherapy with Stereotactic Radiosurgery to High-risk Tumor Regions as Determined by MR Spectroscopy for Patients with Glioblastoma Multiforme.* International Journal of Radiation Oncology Biology Physics, **75**(3): S124-S124.

[5] Yan, Q., R. Masson, Y. Ren, B. Rosati, et al., (2012). *Evolution of CpG island promoter function underlies changes in KChIP2 potassium channel subunit gene expression in mammalian heart.* Proc Natl Acad Sci U S A, **109**(5): 1601-6.

[6] Zhao, X.M., J.J. Ren, W.H. Du, H.S. Hao, et al., (2013). *Effect of vitrification on promoter CpG island methylation patterns and expression levels of DNA methyltransferase 1o, histone acetyltransferase 1, and deacetylase 1 in metaphase II mouse oocytes.* Fertil Steril, **100**(1): 256-61.

[7] Draht, M.X., R.R. Riedl, H. Niessen, B. Carvalho, et al., (2012). *Promoter CpG island methylation markers in colorectal cancer: the road ahead.* Epigenomics, **4**(2): 179-94.

[8] Antequera, F. and A. Bird, (1993). *Number of CpG islands and genes in human and mouse.* Proc Natl Acad Sci U S A, **90**(24): 11995-9.

[9] Dodge, J.E., B.H. Ramsahoye, Z.G. Wo, M. Okano, et al., (2002). *De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation.* Gene, **289**(1-2): 41-8.

[10] Vertino, P.M., R.W. Yen, J. Gao, and S.B. Baylin, (1996). *De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase.* Mol Cell Biol, **16**(8):

4555-65.

[11] Wada, Y., H. Ohya, Y. Yamaguchi, N. Koizumi, et al., (2003). *Preferential de novo methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants.* J Biol Chem, **278**(43): 42386-93.

[12] Grandjean, V., R. Yaman, F. Cuzin, and M. Rassoulzadegan, (2007). *Inheritance of an epigenetic mark: the CpG DNA methyltransferase 1 is required for de novo establishment of a complex pattern of non-CpG methylation.* PLoS One, **2**(11): e1136.

[13] Tan, L.W., T. Bianco, and A. Dobrovic, (2002). *Variable promoter region CpG island methylation of the putative tumor suppressor gene Connexin 26 in breast cancer.* Carcinogenesis, **23**(2): 231-6.

[14] Nishimura, M., T. Saito, H. Yamasaki, and R. Kudo, (2003). *Suppression of gap junctional intercellular communication via 5′ CpG island methylation in promoter region of E-cadherin gene in endometrial cancer cells.* Carcinogenesis, **24**(10): 1615-23.

[15] Hanahan, D. and R.A. Weinberg, (2000). *The hallmarks of cancer.* Cell, **100**(1): 57-70.

[16] Baylin, S.B. and P.A. Jones, (2011). *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, **11**(10): 726-34.

[17] Jones, P.A. and S.B. Baylin, (2007). *The epigenomics of cancer.* Cell, **128**(4): 683-92.

[18] Takai, D. and P.A. Jones, (2002). *Comprehensive analysis of CpG islands in human chromosomes 21 and 22.* Proc Natl Acad Sci U S A, **99**(6): 3740-3745.

[19] Manning, G., D.B. Whyte, R. Martinez, T. Hunter, et al., (2002). *The protein kinase complement of the human genome.* Science, **298**(5600): 1912-34.

[20] Liu, G.B., H. Yan, Y.F. Jiang, R. Chen, et al., (2005). *The properties of CpG islands in the putative promoter regions of human immunoglobulin (Ig) genes.* Gene, **358**(1): 127-138.

[21] Doyle, M.C. and I.S. Han, (2001). *The roles of two TATA boxes and 3'-flanking region of soybean beta-tubulin gene (tubB1) in light-sensitive expression.* Mol Cells, **12**(2): 197-203.

[22] Zhang, Y., D.A. Eberhard, G.D. Frantz, P. Dowd, et al., (2004). *GEPIS--quantitative gene expression profiling in normal and cancer tissues.* Bioinformatics, **20**(15): 2390-8.

[23] Gardiner-Garden, M. and M. Frommer, (1987). *CpG islands in vertebrate genomes.* J Mol Biol, **196**(2): 261-282.

[24] Manning, G., G.D. Plowman, T. Hunter, and S. Sudarsanam, (2002). *Evolution of protein kinase signaling from yeast to man.* Trends Biochem Sci, **27**(10): 514-20.

[25] Antequera, F., (2003). *Structure, function and evolution of CpG island promoters.* Cellular and Molecular Life Sciences (CMLS), **60**(8): 1647-1658.

[26] Antequera, F. and A. Bird, (1999). *CpG islands as genomic footprints of promoters that are associated with replication origins.* Current Biology, **9**(17): R661-R667.

[27] Oppenheimer, S., (2012). *Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map.* Philos Trans R Soc Lond B Biol Sci, **367**(1590): 770-84.

[28] Antequera, F., (2003). *Structure, function and evolution of CpG island promoters.* Cell Mol Life Sci, **60**(8): 1647-58.

[29] Stone, J.R. and G.A. Wray, (2001). *Rapid Evolution of cis-Regulatory Sequences via Local Point Mutations.* Mol Biol Evol, **18**(9): 1764-1770.

[30] Dunnick, W.A., J. Shi, K.A. Graves, and J.T. Collins, (2004). *Germline transcription and switch recombination of a transgene containing the entire H chain constant region locus: effect of a mutation in a STAT6 binding site in the gamma 1 promoter.* J Immunol, **173**(9): 5531-9.

[31] Dupuy, F., T. Griss, J. Blagih, G. Bridon, et al., (2013). *LKB1 is a central regulator of tumor initiation and pro-growth metabolism in ErbB2-mediated breast cancer.* Cancer Metab, **1**(1): 18.

[32] Lutzner, N., J. De-Castro Arce, and F. Rosl, (2012). *Gene expression of the tumour suppressor LKB1 is mediated by Sp1, NF-Y and FOXO transcription factors.* PLoS One, **7**(3): e32590.

[33] Xiang, C., H. Gao, L. Meng, Z. Qin, et al., (2012). *Functional variable number of tandem repeats variation in the promoter of proto-oncogene PTTG1IP is associated with risk of estrogen receptor-positive breast cancer.* Cancer Sci, **103**(6): 1121-8.

[34] Issa, J.P., Y.L. Ottaviano, P. Celano, S.R. Hamilton, et al., (1994). *Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon.* Nat Genet, **7**(4): 536-40.

[35] Issa, J.P., B.A. Zehnbauer, C.I. Civin, M.I. Collector, et al., (1996). *The estrogen receptor CpG island is methylated in most hematopoietic neoplasms.* Cancer Res, **56**(5): 973-77.

[36] Ramezani, F., S. Salami, M.D. Omrani, and D. Maleki, (2012). *CpG island methylation profile of estrogen receptor alpha in Iranian females with triple negative or non-triple negative breast cancer: new marker of poor prognosis.* Asian Pac J Cancer Prev, **13**(2): 451-7.

[37] Itoh, M., S. Ide, S. Takashima, S. Kudo, et al., (2007). *Methyl CpG-binding protein 2 (a mutation of which causes Rett syndrome) directly regulates insulin-like growth factor binding protein 3 in mouse and human brains.* J Neuropathol Exp Neurol, **66**(2): 117-23.

[38] Lensch, M.W., (2009). *Cellular reprogramming and pluripotency induction.* Br Med Bull, **90**: 19-35.