Gene 553 (2014) 126–129

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Compositional features are potentially involved in the regulation of gene expression of tumor suppressor genes in human tissues

Mohammadreza Hajjari [a,*], Atefeh Khoshnevisan [a], Mehrdad Behmanesh [b]

[a] Department of Genetics, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[b] Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form 2 October 2014
Accepted 6 October 2014
Available online 7 October 2014

Keywords:
Tumor suppressor gene
Codon usage
Expression level
In silico analysis

ABSTRACT

Different mechanisms regulate the expression level of tissue specific genes in human. Here we report some compositional features such as codon usage bias, amino acid usage bias, codon frequency, and base composition which may be potentially related to mRNA amount of tissue specific tumor suppressor genes. Our findings support the possibility that structural elements in gene and protein may play an important role in the regulation of tumor suppressor genes, development, and tumorigenesis. The data presented here can open broad vistas in the understanding and treatment of a variety of human malignancies.

1. Introduction

Much progress has been made toward the identification of molecular pathogenesis of cancers. Several studies have been done to understand cancer related genes including tumor suppressor genes. This class of genes includes many genes which are involved in hereditary and non-hereditary cancers (Haber and Harlow, 1997). Tumor suppressor genes govern a wide range of normal cellular activities such as cell cycle checkpoint and DNA repair which might be deregulated in cancer cells.

Inactivation of these genes in the germ line may lead one to be predisposed to cancer. Tumor suppressor genes are also inactivated by somatic mutations arising during tumor development (Haber and Harlow, 1997; Hajjari et al., 2014). Understanding a detailed characterization of tumor suppressor genes and their cellular function and regulation has the potential to open broad vistas in the treatment of a variety of human malignancies.

A diverse array of mechanisms regulates gene expression level in human. Transcriptional control has been the primary focus of gene regulatory research. Also, deciphering the relationship between transcriptional, posttranscriptional, translational, and transport processes provides new insights into gene regulation. Some important features, which are challenging subjects in human gene regulation, are codon usage, amino acid usage, and base composition of the human genome.

In recent years, numerous studies have attempted to understand the relationship between gene expression levels and compositional features such as biased codon and amino acid usage (Goetz and Fuglsang, 2005; Ingvarsson, 2007; Misawa and Kikuno, 2011; Prabha et al., 2012; Sharp et al., 1986), whole genome regulatory networks (Gao et al., 2004), base composition (Arhondakis et al., 2004), and intron length (Castillo-Davis et al., 2002). Besides many studies, the correlation between these characteristics and expression level of human genes has been controversial. In order to get more precise conclusions, some studies have focused on tissue specific genes (Plotkin et al., 2004; Sémon et al., 2006). These studies are interesting because they can better elucidate the processes involved in differentiation. It has been reported that in most cases the tissue-specific codon usage has been selectively preserved throughout the evolution of human and mice from their common ancestor, yet the biological mechanism and impact of this phenomenon certainly require further study.

Regarding the importance of tumor suppressor genes in the development and progression of cancer, we prompted to understand the potential features which may impact the regulation of these genes in human. In this study, we have correlated the compositional features and expression level of tumor suppressor genes in human tissues. Elucidating any correlation between these factors can lead to a better biological understanding of tumor suppressor genes. Also, the results may help to find some clues about the regulation of these genes in cancer initiation and progression.

Abbreviations: TSG, tumor suppressor genes; CAI, codon adaption index; CDS, coding domain sequence.

* Corresponding author.
E-mail addresses: Mohamad.hajari@gmail.com, M-hajari@scu.ac.ir (M. Hajjari).

http://dx.doi.org/10.1016/j.gene.2014.10.011
0378-1119/

2. Materials and methods

2.1. Sequence retrieval, alignments, and expression data acquisition

In total, 69 tumor suppressor genes (the list are in Supplementary File 1) which code proteins with different molecular functions and localizations in the cell, were extracted along with their normalized mRNA expression directly from SOURCE, 2009 (Stanford Microarray) database (source.stanford.edu). This database links to microarray experiments that included the queried gene. Their CDS and protein sequences were obtained from NCBI database. Because of minimizing the statistical errors, multiple alignments were performed for obtained sequences by ClustalW program (http://www.ebi.ac.uk). Twenty five tissues were selected and in each tissue the average expression level, the number of expressed genes, and the highest expressed gene were recognized.

2.2. Sequence compositional features

Calculating the base content (totally and in each position for all codons) in percent, GC content (totally and GC1, 2, 3) in percent, number of codons and their frequencies and synonymous codon usage features – the percentage of each synonymous codon in each codon family that codes for the same amino acid – was done with FREQSQ program (http://www.bioinfo.hku.hk). Also, for estimating the codon usage bias for each gene, codon adaptation index (CAI) was assessed for each gene (http://genomes.urv.es/CAIcal/).

2.3. Amino acid sequence characteristics

For each protein, "Protéines: courbe de titrage (ABIM)" program (http://sites.univ-provence.fr) was used in calculating the amino acid composition in number and in percent, aliphatic amino acids percentage, theoretical Pi, and molecular mass.

2.4. Statistical analyses

For each tissue, the correlation between gene expression level of tumor suppressor genes and compositional features of CDSs and proteins was analyzed with MINITAB and GraphPad. *P*-Values below 0.01 were considered significant. However, in some tissues, because of the lack of *p*-value less than 0.01, we considered the characteristics with *p*-values below 0.05 as the most significant features. Finally, to pinpoint which features are truly significant, Bonferroni correction was done to decrease the statistical errors. Bonferroni correction was taken into account for each parameter separately by *SISA* program (http://www.quantitativeskills.com/sisa/calculations/bonhlp.htm) with the α-value = 0.05 in the whole correlation analyses.

3. Results

3.1. Correlation between gene expression and compositional features

Besides analyzing the correlation between tumor suppressor gene expression level and their compositional features in 25 tissues, 6475 correlations were obtained, in which 117 are with *p*-value < 0.01 (Supplementary Table 1). The most significant features that have the lowest *p*-value for each tissue are listed in Table 1 and presented in Fig. 1. It should be pointed out that the adopted threshold for a significant level ($p < 0.01$) is not extremely low. Therefore, approximately 64 (1% of the total number 6475) features may be significant by chance that probably lead to some artifacts of the present results. Nevertheless, the majority of these significant features are informative.

After the Bonferroni correction, 22 significant features were achieved (Supplementary Table 2). It should be noticed that the most significant features in each tissue listed in Table 1 also appear in supplementary Table 2 if that they have passed the Bonferroni correction.

Table 1
Most significant features of tumor suppressor genes which have relation with their expression levels in human tissues.

Tissue	Number of genes	Feature	*p*-Value	Correlation coefficient
Bladder	40	ATG[a]	<0.001	0.407
Kidney	58	AAC[a]	0.019	0.278
Embryonic tissue	56	AGG[a]	<0.001	0.520
Liver	52	AAG[a]	0.002	0.357
skin	58	ATC[a]	<0.001	0.477
Uterus	61	ATC[a]	0.001	0.302
Mammary-gland	55	AAG[a]	0.001	0.381
Testis	64	TTG[c] (L)	0.001	0.383
Prostate	60	R[b]	0.002	0.363
Ovary	54	T[b]	0.001	−0.376
Nerve-tissue	32	CGG[a]	<0.001	0.466
Salivary-glands	28	CTA[a]	0.001	−0.371
Stomach	51	R[b]	<0.001	0.456
Lung	66	CGT[a]	0.003	0.352
Brain	67	GCT[a]	0.003	0.351
Bone	53	ATC[a]	<0.001	0.411
Pancreas	62	ATC[a]	0.001	0.379
Muscle	59	TGT[a]	0.003	0.345
Heart	54	ATC[a]	0.001	0.379
Placenta	58	T[c]	0.023	−0.269
Spleen	37	GCT[a]	0.007	0.317
Thyroid	18	CGT[a]	0.001	0.401
Lymph node	48	AGG[a]	<0.001	0.427
Eye	64	R[b]	<0.001	0.435
Bone marrow	35	R[b]	<0.001	0.404

CAI: codon adaptation index.
[a] Codon frequency.
[b] Amino acid frequency.
[c] Codon Usage feature.

3.2. Common features in different tissues

Among the characteristics in our analyses (given in Supplementary Table 1), the significant features in various tissues include codon content and amino acid content in percent (65 and 25 features respectively), Pi (9 features), and synonymous codon usage (16 features). It is important to note that most of the significant features are common between different tissues. This result supports the conclusion that the common mechanisms may be responsible for regulating gene expression in different tissues.

The results show that the significant codon contents include some codons in which the number of codons coding arginine has more proportions to other amino acids (26/65) (Supplementary Table 1). On the other hand, the amino acid contents, which correlate significantly with the expression level of tumor suppressor genes, are arginine (in 13 tissues), tryptophan (in one tissue), threonine (in five tissues), asparagine (in three tissues), isoleucine (in one tissue), methionine (in one tissue), and lysine (in one tissue).

The results indicate that the only significant feature for nucleotide compositional features is G2 (frequency of guanine in the second base of codons) in the stomach. Furthermore, among synonymous codon usage features, which correlate with the expression level of tumor suppressor genes, three codons are common between some tissues (AGC, CTA and CTC).

To find the level of codon bias, codon adaptation index (CAI) for each gene was measured. We found some correlations between CAI and expression level of tumor suppressor genes in some tissues (*p*-value less than 0.05). These correlations were in skin, ovary, nerve tissues, salivary glands, lung, and pancreas.

3.3. Bonferroni correction

Since multiple correlations were done in the present study, the Bonferroni correction was done to decrease the statistical errors (Supplementary Table 2). The most common features in these results are codon content for AUC and amino acid content for arginine.