



Expression level of human tumor suppressor tissue specific genes correlates with corresponding coding sequence features

Hajgari Mohammadreza*/**, khoshnevisan Atefeh and Saffar Behnaz

Department of genetics, science faculty of shahrekord university, Iran
mohamad.hajari@gmail.com

Abstract:

In respect to importance of gene expression mechanism and its regulation, We have analyzed the correlation between gene expression level of tumor suppressor tissue specific genes with their compositional features including CDS length, codon content (in number and percent), synonymous codon usage (percentage of each synonymous codon in each codon family that codes for the same amino acid), base content (totally and position ally). This may help us to understand the effect of compositional features and feature selection on regulation of gene expression. In our analysis we used NCBI database and Source database (microarray database) and multiple bioinformatics and statistical software. 71 tumor suppressor genes and 21 tissues were selected. Finally we reached the significant features with P_value below than 0.01. Only kidney didn't have any significant features. Some correlations were positive and some were negative. Interestingly most of them were common among tissues. These results may support the opinion that in these tissues there is a common regulation mechanism for the expression of tumor suppressor genes. Some of these significant features have significance to understanding translational selection and other has significance to understanding various effects on mRNA and protein stability. By such analysis, we may explore some aspect of correlation between genomics, transcriptomics and proteomics.

Introduction

It is accepted that regulation of gene information conversion to protein can be occurred in sequential stages including mRNA decay, mRNA translation rate and efficiency. With bioinformatics analysis we can now understand many combinatorial factors that control these phenomena. In recent years numerous studies have been tried to understand the relationship between gene expression levels and compositional features such as biased codon usage and amino acid usage [1-12], whole genome regulatory networks [13-14], base composition [15-18], intron length [19] and gene distribution [20]. In spite of many studies, the correlation between expression level of human and mammalian has been controversial.

Different reports have indicated that there is some evidence of selection pressure on codon usage bias in mammalian genes [21-23], also the relation of intron length and gene expression has been revealed [24], also some studies concluded that highly expressed genes code for small proteins, have little intronic content, high codon and amino acid bias [25].

Recently, in order to get more precise conclusion, some studies have focused on tissue specific genes. Due to these studies Alexander E. Vinogradov found a weak positive correlation between gene GC content and expression level [27]. later plotkin et al reported a significant difference between genes that are selectively expressed in several human tissues. [28]. Interestingly in a study , Se'mon and et al. concluded that in spite of significant difference in synonymous codon usage between tissue-specific genes in different tissue, this effect is linked to regional variations of GC-content, and not to a selective pressure to optimize translation, so it is not a good predictor [29] in a recent study Yan Zhou and et al analyzed the relationship between Human oncogene tissue-specific expression level and sequence compositional features, It was interesting because they analyzed a specific subset of genes, that are very important in cell cycle and tumor formation. They reached some significant correlation which implies the selection pressure to some extent [30]. In this study we have studied on relation between tumor suppressor genes expression level and some of their compositional features.

2) Materials and Methods

Totally 71 tumor suppressor genes and their normalized expression data with different molecular function and biological process, involved in various cellular components extracted directly from SOURCE database [31]. Their CDS were obtained from NCBI database. Because of minimizing the statistical errors we performed multiple alignment for obtained sequences with clustalW [32]. alignment score for each of the two sequences was between 1 and 27 percent. Calculation of bases content (totally and positionally in each



codon) in percent, GC content (totally and GC1, 2,3) in percent, codons frequencies in number and percent and synonymous codon usage features (percentage of each synonymous codon in each codon family that codes for the same amino acid) were done with FREXSQ. 21 tissues were selected and in each tissue average expression, number of expressed genes, and the highest expressed gene was recognized. for each of them the correlation between gene expression level of tumor suppressor genes and CDS compositional features was analyzed by MINITAB program, P values below than 0.01 were considered significant.

3) Results and discussion:

Among the most significant characteristics (given in table) in our analysis there were not any significant correlation with any CDS length, number of codon content. The most significant features in various tissues are codon content in percent and synonymous codon usage. Also in stomach tissue there is a positive correlation with G2. It is important to not that most of the significant features are common between tissues. This may support the opinion that common mechanism may be responsible for the gene expression regulation in different tissues.

The positive correlation between codon content and gene expression in tissues may imply the correlation between mRNA stability dependence on codon content and regulation of expression. Also, the codon content may affect on the amino acid content and this may affect on the protein stability. The correlation between synonymous codon usage and gene expression level can be attributed to tRNA frequency in specific tissues.

Table:

Tissue	Gene number	Average expression	Highest expressed tumor suppressor gene	P-Value	r _s -Value	feature
Bladder	41	1.871	RPL10	0.005 0.008	0.333 0.311	AUG AAG
Kidney	60	13.648	CTNNA1	0.029 No significant feature	0.260	AAG
Embryonic tissue	58	6.577	GLTSCR2	0.001 0 0	0.390 0.441 0.520	GCU CGG AGG
Liver	54	10.239	RPL10	0.002 0.004 0.005	0.355 0.334 0.331	AAG UGU CGU
Skin	59	19.63	CAV1	0 0.008 0.001	0.485 0.313 0.371	AUC AAG CTC(L)
Stomach	53	7.324	RPL10	0.006 0.002 0 0.002 0.008	0.326 0.360 0.436 0.355 0.312	AAG UGU CGG AGG G2



Spleen	38	3.958	RPL10	0.007	0.317	GCU
Mammary-gland	37	6.533	RPL10	0.007 0.003 0.008	0.308 0.348 0.313	AUG AAG® UGU
Testis Lymph node	66 48	20.577 8.84	CTNNA1 GLTSCR2	0.002 0.008 0.007 0.001 0.004 0.006 0.001	0.354 0.312 0.319 0.382 0.382 0.307 0.307 0.383	UUG CGU GUU AGU GCU UCA UCA TTC(L)
Prostate	62	9.8632	RPL10	0.009 0.002 0.004 0.005	0.368 0.438 0.349 0.327	AUC AAG AGU CGU
Bonemarrow	36	4.088	RPL10	0.009 0.005 0.003 0.004 0.004 0.002	0.308 0.408 0.358 0.358 0.348 0.386	UGG AGU CGC GAA(Q) GAG(Q) AGC(S)
Salivary-glands	29	1.239	RPL10	0.001 0.008 0.005	-0.371 -0.313 -0.330	CUA ACU CTA(L)
Lung	68	29.394	RPL10	0.003 0.006	0.345 0.321	CGU CGG
Brain	69	69.056	CTNNA1	0.002	0.368	GCU
Bone	55	6.324	RPL10	0 0.009 0.007	0.411 0.307 0.315	AUC AAG CGG
Muscle	62	7.859	RPL10	0.006 0.004 0.004	0.326 0.338 -0.339	AAG UGU AGC(S)
Heart	56	5.803	RPL10	0.003 0.001 0.008 0.002 0.007 0.004 0.006 0.009 0.007	-0.350 0.391 0.315 0.367 0.315 0.337 -0.320 0.307 -0.316	CUA AUC UAA CGU CGG CTC(L) CTA(L) ACC(T) AGA®



References:

- 1) P.M. Sharp, T.M. Tuohy and K.R. Mosurski, Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes, *Nucleic Acids Res.* 14 (1986), pp. 5125–5143
- 2) M. Jia and Y. Li, The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*, *FEBS Letter.* 24(2005), 5333-5337
- 3) P.K. Ingvarsson, Gene Expression and Protein Length Influence Codon Usage and Rates of Sequence Evolution in *Populus tremula*, *Mol. Biol. Evol.* 24(3) (2007) ,836–844
- 4) Powell JR, Moriyama EN, Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci USA.* 94 (1997), 7784–7790
- 5) Peixoto, L., Zavala, A., Romero, H., Musto, H., The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium melioli*. *Gene.* 320 (2003), 109–116
- 6) Romero, H., Zavala, A., Musto, H., Bernardi, G., The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene* 317 (2003), 141–147
- 7) Sharp, P.M., Li, W.H., An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24 (1986), 28–38
- 8) R. M. Goetz and A. Fuglsang, Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli* , [Biochemical and Biophysical Research Communications](#). 327 (2005), 4-7
- 9) A. Coghlan and K.H. Wolfe, Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*, *Yeast* 16 (2000), 1131–1145
- 10) R. Jansen, H.J. Bussemaker and M. Gerstein, Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models, *Nucleic Acids Res.* 31 (2003), 2242–2251
- 11) H. Akashi, Translational selection and yeast proteome evolution, *Genetics* 164 (2003), 1291–1303.
- 12) H. Akashi, Translational selection and yeast proteome evolution, *Genetics* 164 (2003), 1291–1303.
- 13) M. Caselle, F. Di Cunto and P. Provero, Correlating overrepresented upstream motifs to gene expression: a computational approach to regulatory element discovery in eukaryotes, *BMC Bioinformatics* 3 (2002), 7.
- 14) F. Gao, B.C. Foat and H.J. Bussemaker, Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data, *BMC Bioinformatics* 5 (2004), 31.
- 15) D'Onofrio, 2002. G. D'Onofrio, Expression patterns and gene distribution in the human genome. *Gene* 300 (2002), 155–160
- 16) Lercher et al., 2002. M.J. Lercher, A.O. Urrutia and L.D. Hurst, Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* 31 (2002), 1801–1831.
- 17) Konu and Li, 2002. O. Konu and M.D. Li, Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. *J. Mol. Evol.* 54 (2002), 35–41
- 18) S. Arhondakis, F. Auletta, G. Torelli and G. D'Onofrio, Base composition and expression level of human genes, *Gene* 325 (2004), pp. 165-169
- 19) Castillo-Davis et al., 2002. C.I. Castillo-Davis, S.L. Mekhedov, D.L. Hartl, E.V. Koonin and F.A. Kondrashov, Selection for short introns in highly expressed genes. *Nat. Genet.* 31 (2002), 415–418.
- 20) Ponger et al., 2001. L. Ponger, L. Duret and D. Mouchiroud , Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Res.* 11 (2001), 1854–1860
- 21) Eyre-Walker, A. An analysis of codon usage in mammals: Selection or mutation bias? *J. Mol. Evol.* 33(1991), 442–449.
- 22) Debry, R.W. and Marzluff, W.F., Selection on silent sites in the rodent H3 histone gene family. *Genetics* 138(1994) 191–202.



- 23) Iida, K. and Akashi, H., A test of translational selection at 'silent' sites in the human genome: Base composition comparisons in alternatively spliced genes. *Gene* 261(2000), 93-105.
- 24) Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., and Kondrashov, F.A. Selection for short introns in highly expressed genes. *Nat. Genet.* 31(2002), 415-418.
- 25) A. O. Urrutia and Laurence D. Hurst, The Signature of Selection Mediated by Expression on Human Genes, *Genome Res.* 13 (2003), 2260-2264
- 26) G. Kudla, L. Lipinski, F. Caffin, A. Helwak, and M. Zylicz, High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells, *PLoS biology* 4: e180
- 27) A. E. Vinogradov, Isochores and tissue-specificity, *Nucleic Acids Research.* 31 (2003), 5212-5220
- 28) Plotkin, J.B., Robins, H. and Levine, A.J. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. USA* 101(2004), 12588-12591
- 29) Se' mon, M., Lobry, J.R. and Duret, L. No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol. Biol. Evol.* 23(2006), 523-529.
- 30) Y. Zhou, B. Ma, H. Zhang, Human oncogene tissue-specific expression level significantly correlates with sequence compositional features, *FEBS Letters* 581(2007), 4361-4365
- 31) Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J.C., Hernandez-Boussard, T., Rees, C.A., Cherry, J.M., Botstein, D., Brown, P.O. and Alizadeh, A.A. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res.* 31(2003) 219-223
- 32) Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice *Nucleic Acids Res.* Nov 11;22(22) (1994), 4673-80

چکیده

باتوجه به اهمیت مکانیسمهای تنظیمی بیان ژن، در این مطالعه بر آن شدید تا به بررسی ارتباط بین سطح بیان ژن های مهار کننده تومور دارای بیان اختصاصی در بافت ها و ویژگی های ساختمانی در این ژنها از جمله طول توالی کدگذار، محتوی کدون (به تعداد و درصد نسبت به کل) استفاده از کدون های هم معنی (درصد هر کدون به کل خانواده کدونی کدگذار یک اسیدامینه)، محتوی بازی (در کل و در هر جایگاه) بپردازیم. این امر می تواند به درک هر چه بهتر اثرات خصوصیات ساختاری و اثر انتخاب ترجمه ای بر تنظیم بیان ژن کمک نماید. در این آنالیز از پایگاه اطلاعاتی NCBI برای گرفتن توالی های ژنی و از پایگاه SOURCE برای گرفتن اطلاعات بیانی بهره بردیم. همچنین از برنامه های بیوانفورماتیکی و آماری متعددی نیز استفاده نمودیم. در این مطالعه ۷۱ ژن مهارگر تومور و ۲۱ بافت انسانی انتخاب گردید. در نهایت به ارتباطاتی معنادار با ارزش P کمتر از ۰,۰۱ دست یافتیم. تنها در بافت کلیه هیچ ارتباط معناداری حاصل نشد. برخی از این ارتباطات مثبت و برخی منفی بودند. جالب آنکه اکثر آنها بین بافت ها مشترک بودند. این نتیجه می تواند از این فرضیه حمایت کند که برای این بافت ها یک نوع مکانیسم تنظیمی مشترک برای بیان ژنهای مهارگر تومور می تواند وجود داشته باشد. برخی از این ارتباطات معنی دار از آن حیث با اهمیتند که می توان با استفاده از آنها اثر انتخاب ترجمه ای را نمایان ساخت و بقیه نیز از آن حیث که می توانند پایداری رونوشت و نسخه پروتئینی را آشکار سازند. با چنین آنالیز هایی قادر خواهیم بود برخی از جنبه های ارتباط بین ژنومیکس، ترانسکریپتومیکس و پروتئومیکس را بهتر نمایان سازیم.

واژه های کلیدی: تنظیم بیان ژن، ژن مهارگر تومور، بیوانفورماتیک، ویژگی ساختمانی، ارتباط معنی دار