



Structural Peculiarities of Human Genes Which Expression Increases in Response to Stress

Riabenko E.A.^{1*}, Tonevitsky E.A.², Tonevitsky A.G.², Grigoriev A.I.³

¹ Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia

² Department of Sport Science, Moscow State University, Moscow, Russia

³ Faculty of Fundamental Medicine, Moscow State University, Moscow, Russia

*** Corresponding author:**

Evgeniy Riabenko

Department of Mathematical Methods of Forecast,

Faculty of Computational Mathematics and Cybernetics,

Moscow State University,

Leninskie Gory, 2nd Educational Building, 119991, Moscow, Russia

E-mail: riabenko.e@gmail.com

Received: 3 December 2010; / Revised: 18 January 2011; / Accepted: 16 February 2011

Abstract

We studied structural peculiarities of 123 human genes which were previously identified as genes of early response to exercise stress. Compared to protein-coding genes in general, their structure revealed some peculiarities: they are shorter in both coding and non-coding regions and contain fewer introns. This result is in good agreement with the similar results reported for stress-responding genes in other species (yeasts, thale cress, and mice).

Keywords: RNA Splicing, Introns, Gene Expression Regulation.

1. Introduction

In higher eukaryotes coding regions of genes alternate with non-coding ones, which are not used for protein translation, and therefore most of the pre-mRNAs have to undergo the modification process called splicing. This is a necessary stage in the gene expression process, during which intronic (non-coding) sequences are removed and exonic (coding) are ligated. The total length of intronic sequences in one human gene can be more than tens of thousands of nucleotides.

The transcription process requires significant amount of both time and energy: it takes approximately 50 milliseconds [1] and at least two adenine triphosphate (ATP) molecules [2] to transcribe a nucleotide. That's why it is coherent to expect the presence of some selection pressure for shortening or against elongation of genes (of course, this pressure could be compensated by the opposite tendency related to other functional reasons). Besides, the energy value required for the subsequent cutting out of the untranslatable nucleotides does not depend on the length of the introns but only on their amount [3]. Therefore,

one could additionally assume that selection pressure is acting for intron loss or against intron gain. In [4] it is shown that this assumption seems reasonable and intron loss is a common tool of adaptation at short evolutionary distances.

It is already known that for some groups of genes selection pressure leads to significant deviations from the average length and intron number. The relative compactness of highly expressed human genes was demonstrated in a number of works [3, 6-9]. Such genes revealed to be shorter and to have fewer introns than the average. The same was shown for antisense genes, which are expected to regulate their sense partners [8], and housekeeping genes, which expression level is continuously maintained on the almost constant level [9]. Recently genes which expression is increased in response to physical stress were described [10, 11]. One could expect that they will show similar regularities in total length and intron number, as their transcription takes place under harsh conditions. We verified the hypothesis that so-called genes of early response are more compact and found that they are shorter in both coding and non-coding regions and have lower number of introns.

2. Methods and Materials

Our research was based on the results of the microarray analysis obtained in [10, 11]. In that work 11 well-trained skiers performed ramp-test for about 15 minutes, after which changes in gene expression in blood mononuclear cells were analyzed with Affymetrix GeneChip 1.0 ST Arrays. The resulting list of genes recognized as differentially expressed included genes with fold change at least 1.3 and FDR-adjusted ($FDR \leq 0.05$) p-value no more than 0.05. As a result, 123 genes with increased level of expression were found.

To identify peculiar structural properties of the 123 genes with increased expression level, we compared distributions of their number of introns, total intron length and total exon length with the norm. A collection of 21634 human protein-

coding genes was chosen as a reference set representing this norm (all protein-coding genes with Ensembl IDs excluding the TTN gene (ENSG00000155657), which is too complex in its intron-exon structure). Information about the structure of genes was obtained from Ensembl database (Homo Sapiens GRCh37) with biomaRt R package. The intron-exon structure of each gene was determined based on its canonical transcript.

Distributions of structural attributes in the considered groups of genes were compared by Mann-Whitney statistical test. It was also used for location shifts estimation and their corresponding 90% confidence intervals construction.

For a more correct estimation of selection towards shorter genes with lower number of introns all protein-coding genes longer than some threshold could be excluded from the reference set. Since, according to the properties of the FDR-controlling procedure used in determination of the up-regulated genes, that list of genes could contain up to 5% of false positives, a 0.95-quantile of considered sample's genes length was used as that threshold. The same statistical technique was applied then to compare distributions of attributes among up-regulated genes with the truncated set of all genes.

3. Results

Results are demonstrated in Table 1. The analysis revealed systematical deviations in the structure of genes with increased expression level. It was shown with a high significance level that they have lower number of introns (location shift 2 units); moreover, their total length is shorter (location shift 5467 base pairs). Total length reduction is significant in both introns and exons, while less significant in 5' and 3' UTR since they are generally shorter and probably could not reduce noticeably. These data support our hypothesis that genes of early response ought to be more compact in comparison to other protein coding genes.

Table 1. Comparison of structural features between all protein-coding genes and genes up-regulated in an early response to exercise. In the “Mean values” column: first row – mean \pm standard error of the mean, second row – median. In “Location shift” column: first row – point estimation of location shift, second – 90% confidence interval for it (Mann-Whitney test).

	Mean values of characteristics		P-value	Location shift
	Up-regulated genes (n=121)	Protein-coding genes (n=21634)		
Number of introns, units	5.7 \pm 0.5 4	8.8 \pm 0.1 6	6.58 $\times 10^{-4}$	2 (1;2)
Total gene length, bp	35356 \pm 4891 10699	59326 \pm 820 21635	4.10 $\times 10^{-4}$	5467 (2620; 8846)
Total exon length, bp	1715 \pm 119 1472	2332 \pm 14 1734	4.38 $\times 10^{-4}$	363 (191; 540)
Total intron length, bp	25368 \pm 3979 6740	48296 \pm 727 14935	5.54 $\times 10^{-5}$	4389 (2175; 7143)
3' UTR length, bp	903 \pm 91 513	1183 \pm 10 697	1.31 $\times 10^{-2}$	124 (41; 221)
5' UTR length, bp	149 \pm 12 132	192 \pm 2 130	1.23 $\times 10^{-1}$	15 (-1; 31)

As seen in Figure 1, modes of distributions of all features in the two groups of genes are quite close, which means that most common values of intron number, gene length, total intron and exon length and 3' and 5' UTR length of up-regulated genes are almost the same as of all protein-coding genes. Detected significant differences between distributions (Table 1) are explained by the rare occurrence of long and intron rich genes between up-regulated genes — in other words, genes with long coding and non-coding sequences and genes with high number of introns are rarer among them.

Comparison of up-regulated genes with truncated reference with Mann-Whitney test showed that their number of introns is still significantly lower (location shift 1 unit, 90% confidence interval for location shift – (0.8; 2) units, p-value=0.01048).

The obtained results are in a good accordance with the results presented in [3], where structural features of genes expressed in stress conditions in yeasts (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*), thale cress (*Arabidopsis thaliana*) and mice (*Mus musculus*) were examined. The most significant of all the identified effects in the article was that genes with rapidly changing expression levels in response to stress contain significantly lower intron densities across all the examined organisms. Thus, one can conclude that this effect might have strong impact on survival of species during evolution.

Acknowledgement

The authors are supported by the Russian Ministry of Science Grant No. 14.740.11.0117.

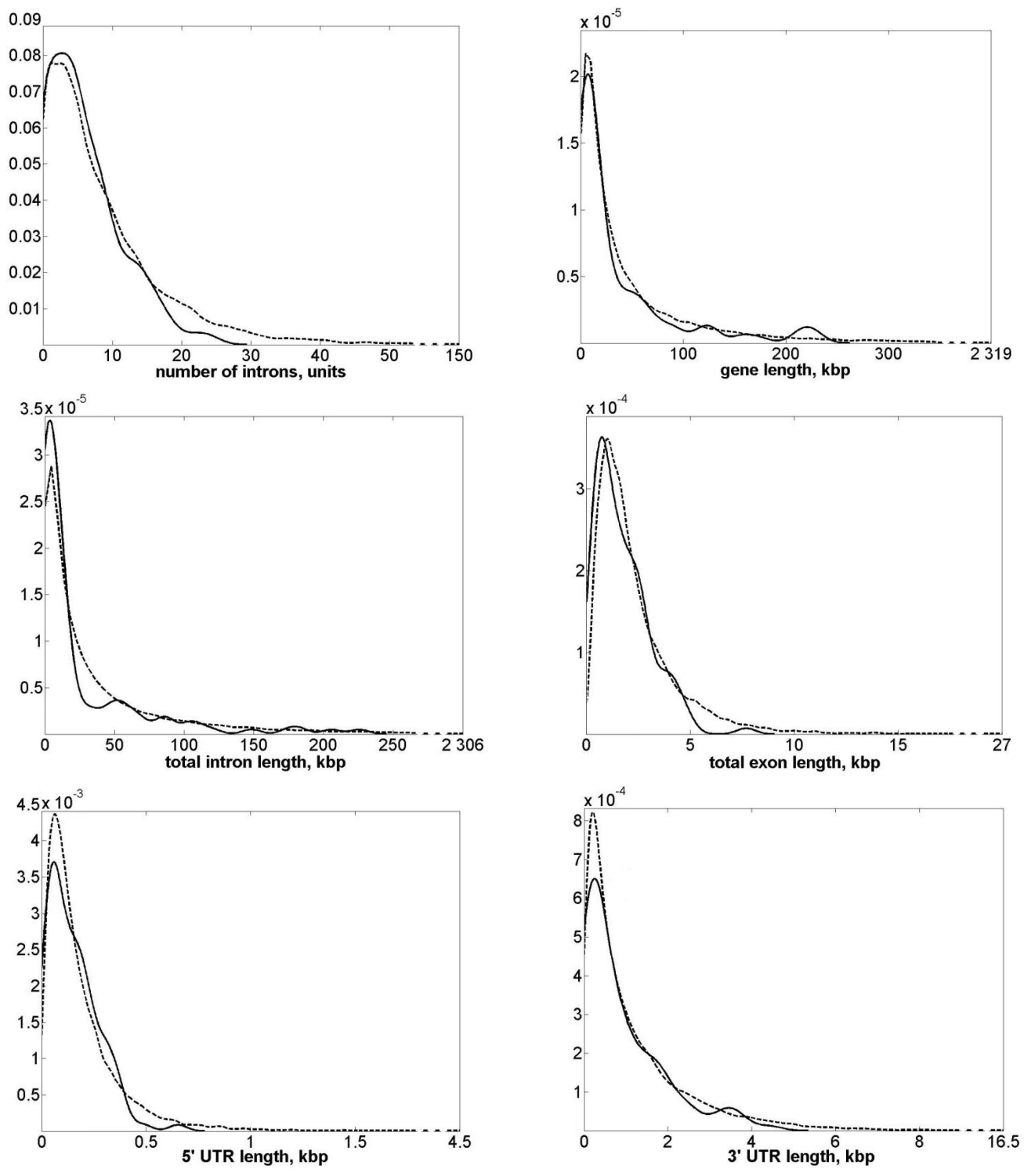


Figure 1. Probability density functions of structural features of genes in two groups: solid line – genes with increased expression level, dotted line – all protein-coding genes. Dashed parts of horizontal axes are omitted; densities of features of protein-coding genes there are monotonic and close to zero.

References

- [1] Izban, M.G.; Luse, D.S. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates, *The Journal of Biological Chemistry*, 1992, 267, 13647–13655.
- [2] Lehninger, A.L.; Nelson, D.L.; Cox, M.M. *Principles of biochemistry, third edition*; Worth Publishers: New York, 2000; pp. 615–644.
- [3] Jeffares, D.C.; Penkett, C.J.; Bähler, J. Rapidly regulated genes are intron poor, *Trends in genetics*, 2008, 24(8), 375–378. DOI: [10.1016/j.tig.2008.05.006](https://doi.org/10.1016/j.tig.2008.05.006)
- [4] Rogozin, I.B.; Sverdlov, A.V.; Babenko, V.N.; Koonin, E.V. Analysis of evolution of exon-intron structure of eukaryotic genes, *Briefings in bioinformatics*, 2005, 6(2), 118–134. DOI: [10.1093/bib/6.2.118](https://doi.org/10.1093/bib/6.2.118)
- [5] Carmel, L.; Koonin, E.V. A universal nonmonotonic relationship between gene compactness and expression levels in multicellular eukaryotes, *Genome biology and evolution*, 2009, 1, 382–390. DOI: [10.1093/gbe/evp038](https://doi.org/10.1093/gbe/evp038)
- [6] Li, S.W., Feng, L., Niu, D.K. Selection for the miniaturization of highly expressed genes, *Biochemical and Biophysical Research Communications*, 2007, 360, 586–592. DOI: [10.1016/j.bbrc.2007.06.085](https://doi.org/10.1016/j.bbrc.2007.06.085)
- [7] Urrutia, A.O., Hurst, L.D. The signature of selection mediated by expression on human genes, *Genome Research*, 2003, 13, 2260–2264. DOI: [10.1101/gr.641103](https://doi.org/10.1101/gr.641103)
- [8] Chen, J., Sun, M., Hurst, L.D., Carmichael, G.G., Rowley, J.D. Human antisense genes have unusually short introns: evidence for selection for rapid transcription, *Trends in genetics*, 2005, 21, 203–207. DOI: [10.1016/j.tig.2005.02.003](https://doi.org/10.1016/j.tig.2005.02.003)
- [9] Eisenberg, E., Levanon, E.Y. Human housekeeping genes are compact, *Trends in Genetic*, 2003, 19(7), 362–365. DOI: [10.1016/S0168-9525\(03\)00137-9](https://doi.org/10.1016/S0168-9525(03)00137-9)
- [10] Tonevitsky, E.A.; Trushkin, E.V.; Shkurnikov, M.U.; Akimov, E.B.; Sakharov, D.A. Changed Profile of Expression of Splicing Regulator Genes in Response to Exercise, *Bulletin of Experimental Biology and Medicine*, 2009, 147(6), 733–736. DOI: [10.1007/s10517-009-0593-0](https://doi.org/10.1007/s10517-009-0593-0)
- [11] Shleptsova, V.A.; Trushkin, E.V.; Bystryh, O.A.; Davydov, J.I.; Obrazcova, N.P.; Grebenuk, E.S.; Tonevitsky, A.G. Expression of Early Immune Response Genes during Physical Exercise, *Bulletin of Experimental Biology and Medicine*, 2010, 149(1), 89–91. DOI: [10.1007/s10517-010-0883-6](https://doi.org/10.1007/s10517-010-0883-6)