

Deposited research article

Negative selection pressure against premature protein truncation is reduced by both alternative splicing and diploidy

Yi Xing and Christopher Lee

Address: Molecular Biology Institute, Center for Genomics and Proteomics, Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095-1570, USA.

Correspondence: Christopher Lee. E-mail: leec@mbi.ucla.edu

Posted: 29 April 2004

Received: 27 April 2004

Genome Biology 2004, 5:P12

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/6/P12>

This is the first version of this article to be made available publicly.

© 2004 BioMed Central Ltd

comment

reviews

reports

deposited research

refereed research

interactions

information



deposited research

AS A SERVICE TO THE RESEARCH COMMUNITY, GENOME **BIOLOGY** PROVIDES A 'PREPRINT' DEPOSITORY TO WHICH ANY ORIGINAL RESEARCH CAN BE SUBMITTED AND WHICH ALL INDIVIDUALS CAN ACCESS FREE OF CHARGE. ANY ARTICLE CAN BE SUBMITTED BY AUTHORS, WHO HAVE SOLE RESPONSIBILITY FOR THE ARTICLE'S CONTENT. THE ONLY SCREENING IS TO ENSURE RELEVANCE OF THE PREPRINT TO GENOME **BIOLOGY**'S SCOPE AND TO AVOID ABUSIVE, LIBELLOUS OR INDECENT ARTICLES. ARTICLES IN THIS SECTION OF THE JOURNAL HAVE **NOT** BEEN PEER-REVIEWED. EACH PREPRINT HAS A PERMANENT URL, BY WHICH IT CAN BE CITED. RESEARCH SUBMITTED TO THE PREPRINT DEPOSITORY MAY BE SIMULTANEOUSLY OR SUBSEQUENTLY SUBMITTED TO GENOME **BIOLOGY** OR ANY OTHER PUBLICATION FOR PEER REVIEW; THE ONLY REQUIREMENT IS AN EXPLICIT CITATION OF, AND LINK TO, THE PREPRINT IN ANY VERSION OF THE ARTICLE THAT IS EVENTUALLY PUBLISHED. IF POSSIBLE, GENOME **BIOLOGY** WILL PROVIDE A RECIPROCAL LINK FROM THE PREPRINT TO THE PUBLISHED ARTICLE.



Negative Selection Pressure Against Premature Protein Truncation Is Reduced by both Alternative Splicing and Diploidy

Yi Xing, Christopher Lee
Molecular Biology Institute
Center for Genomics and Proteomics
Dept. of Chemistry & Biochemistry
University of California, Los Angeles
Los Angeles, CA 90095-1570

EMAIL: leec@mbi.ucla.edu
TEL: 310-825-7374
FAX: 310-267-0248

Draft 5
April 1, 2004

Abstract

The importance of alternative splicing in many genomes has raised interesting questions about its role in evolution. We previously reported that alternative splicing appears to be associated with an increased rate of recent exon creation / loss events, and proposed that it can reduce negative selection pressure on sequence regions that are alternatively spliced. In this paper we test this idea directly, using the occurrence of premature protein truncation events as a metric of selection pressure. We have analyzed 13,384 full-length transcript isoforms from human and 2,227 isoforms from mouse to identify sequences containing premature termination codons (PTC) that are likely targets of mRNA nonsense mediated decay. We found that alternatively spliced isoforms indeed had a much higher frequency of PTCs (11.1%) compared with the major transcript form of each gene (3.7%). However, this effect was strongly influenced by the chromosomal location of the gene: on the X chromosome, which is generally expressed as a single copy, the overall PTC rate was much lower (3.5%, vs. 8.9% on diploid autosomes), and the effect of alternative splicing was enhanced, causing a four-fold reduction in negative selection against PTC. Thus, diploidy and alternative splicing each increased tolerance for PTC by about three-fold, as approximately additive effects. These data may suggest genomic evidence that nonsense mediated decay has itself reduced negative selection pressure during evolution, via rapid degradation of aberrant transcripts that might yield dominant negative phenotypes.

Introduction

Recently, it was proposed that alternative splicing may play a special role in evolution, by reducing negative selection pressure against large-scale mutations such as exon creation and loss¹⁻³. Whereas most aspects of orthologous gene structures in human, mouse and rat are highly similar, alternatively spliced exons frequently are not conserved between these genomes^{3,4}. Such divergent features indicate an exon creation or loss event subsequent to the separation of these genomes during evolution. The fact that such exon creation / loss events are observed at a much higher rate for alternatively spliced exons than constitutive exons suggests that alternative splicing increases tolerance for such large-scale alterations of gene structure. Intuitively, this makes sense. Insertion or

deletion of an exon is likely to disrupt a protein's reading frame, structure or function. However, if a new exon is added as an alternatively spliced exon that is included in only 10% of transcripts of the gene, 90% of transcripts will still encode the original product, greatly reducing negative selection pressure against the new form.

To assess the validity of this hypothesis, it would be useful to find measures of negative selection pressure that are applicable to large-scale mutations such as exon creation. This requires somewhat different metrics of selection pressure than are typically applied to small-scale mutations (e.g. K_a/K_s for single nucleotide polymorphisms⁵). One simple metric that can be applied to large-scale mutations is the frequency of occurrence of truncated protein reading frames that are likely targets of nonsense-mediated decay (NMD)⁶. If a transcript's open reading frame (ORF) has a premature termination codon (PTC; a STOP codon more than 50 nt upstream from its last exon-exon junction), it is likely to be degraded by NMD⁷. Thus, NMD-candidate transcripts are likely to have reduced function relative to the wildtype transcript, and their rate of occurrence (as a fraction of observed transcript forms) gives a simple measure of this negative selection pressure.

A second major question is the validity of the above hypothesis for diploid genes that are present in two copies in each cell (one from each of the two copies of its chromosome, in a diploid genome). Even if a new mutation completely eliminated production of the original transcript form from its gene, in a heterozygote the wildtype copy of the gene would ensure that the original transcript form would still be produced at 50% of its original level (instead of 0%, as would be the case if there was only one copy of the gene). This might alleviate most of the negative effects of the mutation, so that alternative splicing of the mutation would not produce much additional relief of its negative selection pressure. This question was not addressed in our original model³, but is critical for evaluating this model.

Fortunately, this question can be tested by measuring negative selection pressure against PTC-containing isoforms on diploid chromosomes (e.g. autosomes) vs. on haploid

chromosomes (e.g. sex chromosomes). The human X chromosome is haploid in males, and due to X inactivation, its gene expression is typically limited to a single chromosome in female cells as well⁸. In this paper we analyze the frequency of PTCs in the canonical splice form for each gene (which we will refer to as the “major” splice form) and in alternative splice forms for these genes (which we will refer to as “minor” splice forms), on both autosomal and X chromosomes in human and mouse.

Methods

We detected alternative splice forms for human and mouse by mapping mRNA and EST sequences onto genomic sequences as previously described⁹, using the following data: (1) January 2002 download of UniGene EST data¹⁰ (2) January 2002 human and mouse genome sequence downloaded from NCBI. A database of alternatively spliced transcript isoforms from human and mouse was constructed from mRNA-EST-genomic multiple sequence alignments using our isoform generation algorithm described before¹¹. Major isoforms were characterized as isoforms with maximum number of EST evidences for any given gene. Premature transcripts that are likely targets of Nonsense Mediated Decay were identified by checking for a STOP codon located over 50bp upstream of the last exon-exon junction sites.

Results

In our set of 13,384 transcript isoforms for 4,422 human genes, we found that 3.7% of major transcript isoforms (165 out of 4,422) had premature termination codons (PTCs), similar to the percentage reported for human mRNAs^{11,12}. However, 11.1% of the alternative-splicing (minor) isoforms had PTCs, a nearly 3-fold increase. We observed the same pattern in the mouse genome (3.6% for major isoforms vs. 9.3% for minor isoforms, see Table 1). These data indicate that alternative splicing is indeed associated with a substantial reduction in selection pressure against PTCs.

Analyzing the human data by chromosome, we found a large decrease in the frequency of PTCs on the X chromosome (3.5%) compared with autosomal chromosomes(8.9%) (see Table1B). X had the lowest PTC rate of all 23 chromosomes (Fig. 1), and its difference vs. autosomal chromosomes was statistically significant ($P < 0.000006$). Moreover, we

found the same result on the mouse X chromosome (2.9%) vs. mouse autosomal chromosomes (7.2%). Thus diploidy also is associated with a significant reduction in selection pressure against PTCs, compared with haploid chromosomes (X). This pattern held true even when we limited our analysis to major splice forms: in human, 1.2% had PTCs on the X chromosome, vs. 3.8% on autosomal chromosomes.

Does alternative splicing still provide relief from negative selection even on diploid chromosomes? For human autosomes, the frequency of PTCs was 3.8% for major splice forms, and 11.3% for minor splice forms (a three-fold increase). For mouse autosomal chromosomes, 3.7% of major forms vs. 9.7% of minor forms had PTCs. However, the strength of this effect appears to be strongest on the X chromosome. On the human X chromosome, the frequency of PTCs was 1.2% for major splice forms, vs. 4.6% for minor alternative splice forms, a nearly four-fold increase.

Discussion

These data provide independent evidence for the hypothesis that alternative splicing can relieve negative selection pressure. Whereas the original evidence for this hypothesis was based on comparative genomics analysis of exon creation / loss in mammalian genomes³, here we have focused on a different phenomenon—the incidence of premature termination codons that are likely to be targets for nonsense-mediated decay. Sorek et al. have reported that Alu sequences sometimes occur in exons, but were always associated with alternative splicing (i.e. Alu were only found in alternatively-spliced exons)¹³. This also indicates reduced negative selection pressure in alternatively spliced exons.

Alternative splicing and diploidy appear to give about the same magnitude reduction (three-fold) in negative selection pressure against PTCs during recent mammalian evolution. In both cases, the effect of “having another functional copy” appears to increase tolerance for PTC isoforms by about three-fold. The effects of alternative splicing and diploidy appear to be independent and additive. Alternative splicing still relieves negative selection pressure even on diploid chromosomes, but this effect appears to be stronger on haploid chromosomes (e.g. nearly four-fold on the human X

chromosome). The combined effect of alternative splicing and diploidy yields a more than nine-fold increase in tolerance for PTCs (from 1.2% in major splice forms on the X chromosome, to 11.3% in minor splice forms on autosomes).

These data may reflect genomic evidence for an important role for nonsense-mediated decay during mammalian evolution^{14,15}. The increased incidence of NMD-candidate forms observed on diploid chromosomes is consistent with NMD's function of degrading aberrant transcript forms that might produce dominant negative phenotypes. Since this function is useful for genes on diploid chromosomes (where the second copy of the chromosome can supply a working copy of the gene), but not on haploid chromosomes, this would be expected to give rise to more NMD-candidate forms on diploid chromosomes.

Acknowledgements

We wish to thank Drs. D. Black, C. Grasso, Y. Marahrens, B. Modrek, A. Resch, Q. Xu for their discussions and comments on this work. C.J.L. was supported by NIMH/NINDS Grant MH65166, and DOE grant DEFG0387ER60615.

- 1 Boue, S., Letunic, I. and Bork, P. (2003) *Bioessays* 25, 1031-4
- 2 Kan, Z., States, D. and Gish, W. (2002) *Genome Res* 12, 1837-45
- 3 Modrek, B. and Lee, C.J. (2003) *Nat Genet* 34, 177-80
- 4 Nurtdinov, R.N., Artamonova, II, Mironov, A.A. and Gelfand, M.S. (2003) *Hum Mol Genet* 12, 1313-20
- 5 Li, W.H., Wu, C.I. and Luo, C.C. (1985) *Mol Biol Evol* 2, 150-74
- 6 Maquat, L.E. (2004) *Nat Rev Mol Cell Biol* 5, 89-99
- 7 Maquat, L.E. (2002) *Curr Biol* 12, R196-7
- 8 Heard, E., Clerc, P. and Avner, P. (1997) *Annu Rev Genet* 31, 571-610
- 9 Modrek, B., Resch, A., Grasso, C. and Lee, C. (2001) *Nucleic Acids Res* 29, 2850-9
- 10 Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) *Nat Genet* 4, 332-3
- 11 Xing, Y., Resch, A. and Lee, C. (2004) *Genome Res* 14, 426-41
- 12 Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) *Proc Natl Acad Sci U S A* 100, 189-92
- 13 Sorek, R., Ast, G. and Graur, D. (2002) *Genome Res* 12, 1060-7
- 14 Lynch, M. (2002) *Proc Natl Acad Sci U S A* 99, 6118-23
- 15 Lynch, M. and Kewalramani, A. (2003) *Mol Biol Evol* 20, 563-71

Figure Legends

Figure 1: *Percentage of isoforms with PTCs on individual human chromosomes*

A bar chart for the percentage of isoforms with PTCs on individual human chromosomes. Red: autosomes ; Blue: X chromosome ; Green: average percentage for all human autosomes.

Table 1: *Percentage of human and mouse isoforms with PTCs*

- A) Percentage of PTCs for major and minor isoforms of human or mouse
- B) Percentage of PTCs for human isoforms from autosomes or X chromosome
- C) Percentage of PTCs for mouse isoforms from autosomes or X chromosome

Percentage of Human Isoforms with Premature Termination Codon

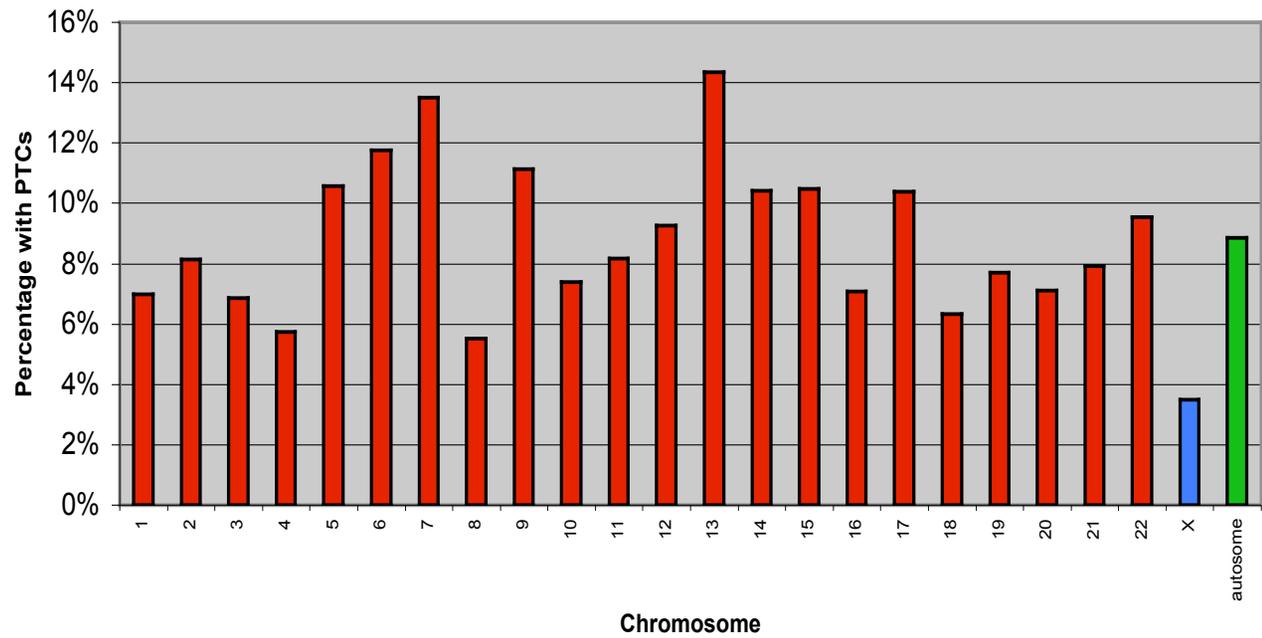


FIGURE 1

A. Human and mouse isoforms with PTCs

% with PTCs	Major	Minor
Human	3.7%	11.1%
Mouse	3.6%	9.3%

B. Human isoforms with PTCs on autosomes and X chromosome

% with PTCs	Major	Minor	All isoforms
autosomes	3.8% (160/4221)	11.3%(975/8602)	8.9%(1135/12823)
X chromosome	1.2% (2/161)	4.6%(14/299)	3.5%(16/460)

C. Mouse isoforms with PTCs on autosomes and X chromosome

% of PTCs	Major	Minor	All isoforms
autosomes	3.7% (32/871)	9.7% (121/1253)	7.2%(153/2124)
X chromosome	2.6%(1/39)	3.1%(2/64)	2.9%(3/103)

TABLE 1