UCLA
UCLA Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/1kg7s6b9

Journal

Trends in Genetics, 20(10)

Author

Lee, Christopher

Publication Date

2004

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

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 7
June 15, 2004
Published as TIG, **20**, 472-5 (2004)
http://dx.doi.org/10.1016/j.tig.2004.07.009

Abstract
The importance of alternative splicing in many genomes has raised interesting questions about its role in evolution. We 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 have a much higher frequency of PTCs (11.1%) compared with the major transcript form of each gene (3.7%). 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. Thus, diploidy and alternative splicing each increased tolerance for PTC by about three-fold, as approximately additive effects. These data may suggest 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 loss [1-3]. Whereas 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 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 small fractions of the total transcripts (e.g.10%), the majority 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. Ka/Ks for single nucleotide polymorphisms[5]). 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; defined in connection with NMD as a STOP codon more than 50 nt upstream from its last exon-exon junction[7]), it is likely to be degraded by NMD[8]. Many alternatively spliced isoforms are potential candidates of NMD based on this "50 nucleotide rule" [9,10]. 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 model [3], but is critical for evaluating our 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 well [11]. 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 described [12], using the following data: (1) January 2002 download of UniGene EST data [13] (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 before [10]. Major isoforms were characterized as isoforms with the largest number of ESTs for a 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 site.

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 [9,10]. 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 compared(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. Y chromosome has many fewer genes and is not included in this comparison. Still, we obtained similar results there as none of the seven transcript isoforms we constructed from Y-chromosome genes have PTC.

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 genomes [3], 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) [14]. 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). The percentage of transcripts with PTCs is probably an underestimate since the transcript isoforms we analyzed are those which have escaped NMD, and we had stringent criteria during the isoform generation process to filter out potential EST artifacts (e.g. retention of an entire intron) [10]. In this analysis, we have generally assumed that the occurrence of PTC (and thus the likelihood of NMD) can cause a negative impact on phenotype, namely, failure to produce a functional protein product. There is good evidence for this, but it has not yet been broadly demonstrated. Brenner and colleagues have also advanced the interesting hypothesis that NMD may constitute a functional form of regulation: that is, using the coupling of alternative splicing and NMD (instead of transcriptional control) to regulate the amount of protein product[9].

These data may reflect genomic evidence for an important role for nonsense-mediated decay during mammalian evolution[15,16]. 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.

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

1 Boue, S., Letunic, I. and Bork, P. (2003) *Bioessays* Alternative splicing and evolution.25, 1031-4

2 Kan, Z., States, D. and Gish, W. (2002) *Genome Res* Selecting for functional alternative splices in ESTs.12, 1837-45

3 Modrek, B. and Lee, C.J. (2003) *Nat Genet* Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss.34, 177-80

4 Nurtdinov, R.N., Artamonova, II, Mironov, A.A. and Gelfand, M.S. (2003) *Hum Mol Genet* Low conservation of alternative splicing patterns in the human and mouse genomes.12, 1313-20

5 Li, W.H., Wu, C.I. and Luo, C.C. (1985) *Mol Biol Evol* A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes.2, 150-74

6 Maquat, L.E. (2004) *Nat Rev Mol Cell Biol* Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics.5, 89-99

7 Nagy, E. and Maquat, L.E. (1998) *Trends Biochem Sci* A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance.23, 198-9

8 Maquat, L.E. (2002) *Curr Biol* Nonsense-mediated mRNA decay.12, R196-7

9 Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) *Proc Natl Acad Sci U S A* Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans.100, 189-92

10 Xing, Y., Resch, A. and Lee, C. (2004) *Genome Res* The Multiassembly Problem: Reconstructing Multiple Transcript Isoforms From EST Fragment Mixtures.14, 426-41

11 Heard, E., Clerc, P. and Avner, P. (1997) *Annu Rev Genet* X-chromosome inactivation in mammals.31, 571-610

12 Modrek, B., Resch, A., Grasso, C. and Lee, C. (2001) *Nucleic Acids Res* Genome-wide detection of alternative splicing in expressed sequences of human genes.29, 2850-9

13 Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) *Nat Genet* dbEST--database for "expressed sequence tags".4, 332-3

14 Sorek, R., Ast, G. and Graur, D. (2002) *Genome Res* Alu-containing exons are alternatively spliced.12, 1060-7

15 Lynch, M. (2002) *Proc Natl Acad Sci U S A* Intron evolution as a population-genetic process.99, 6118-23

16 Lynch, M. and Kewalramani, A. (2003) *Mol Biol Evol* Messenger RNA surveillance and the evolutionary proliferation of introns.20, 563-71

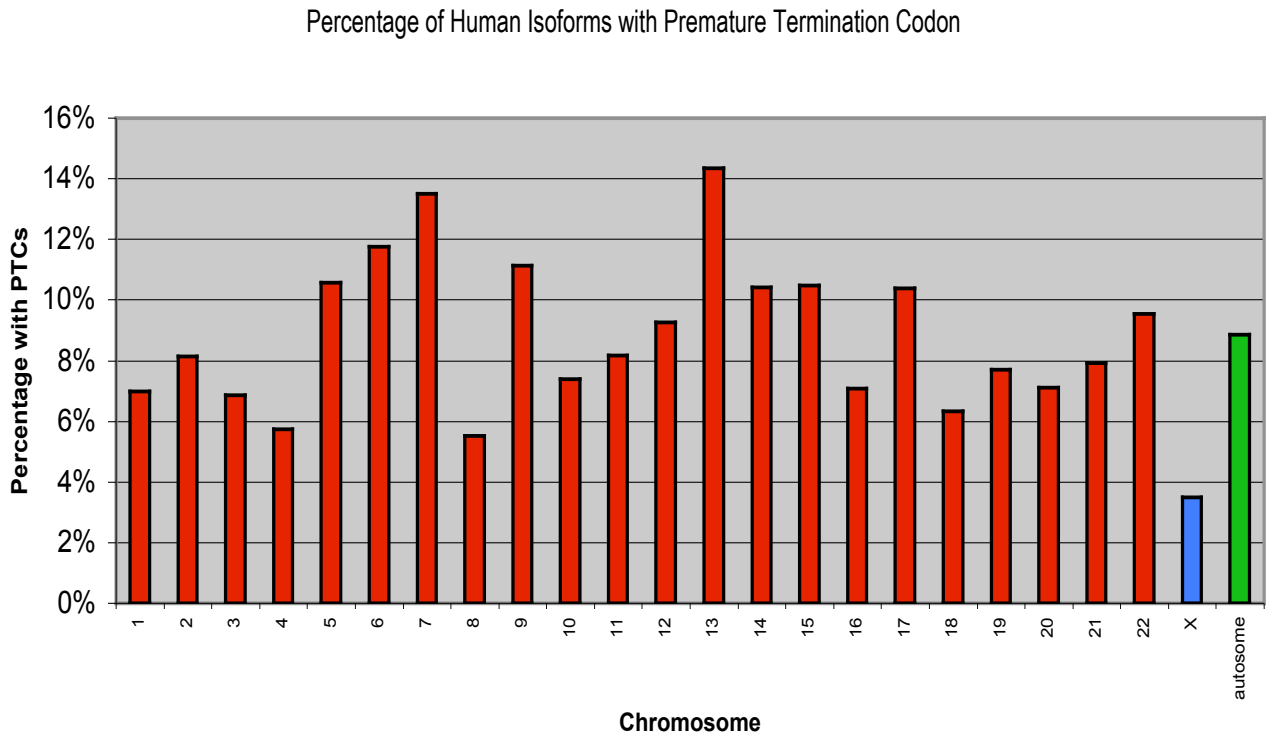
Percentage of Human Isoforms with Premature Termination Codon
Percentage with PTCs
16%
14%
12%
10%
8%
6%
4%
2%
0%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
autosome
Chromosome

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