Mutation Research 784-785 (2016) 34–38

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Estimation of the minimum mRNA splicing error rate in vertebrates

A. Skandalis

Brock University, St. Catharines, Ontario, Canada

ARTICLE INFO

Article history:
Received 6 July 2015
Received in revised form 6 January 2016
Accepted 10 January 2016
Available online 13 January 2016

Keywords:
mRNA splicing
Aberrant splice variants
Mutagenesis
Vertebrates

ABSTRACT

The majority of protein coding genes in vertebrates contain several introns that are removed by the mRNA splicing machinery. Errors during splicing can generate aberrant transcripts and degrade the transmission of genetic information thus contributing to genomic instability and disease. However, estimating the error rate of constitutive splicing is complicated by the process of alternative splicing which can generate multiple alternative transcripts per locus and is particularly active in humans. In order to estimate the error frequency of constitutive mRNA splicing and avoid bias by alternative splicing we have characterized the frequency of splice variants at three loci, *HPRT*, *POLB*, and *TRPV1* in multiple tissues of six vertebrate species. Our analysis revealed that the frequency of splice variants varied widely among loci, tissues, and species. However, the lowest observed frequency is quite constant among loci and approximately 0.1% aberrant transcripts per intron. Arguably this reflects the "irreducible" error rate of splicing, which consists primarily of the combination of replication errors by RNA polymerase II in splice consensus sequences and spliceosome errors in correctly pairing exons.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Errors by processes such as DNA replication and repair that can degrade heritable genetic information have been the subject of long and intense study [1,2]. On the other hand, little is known about errors by other molecular processes, such as constitutive mRNA splicing, that can transiently corrupt the flow of genetic information. mRNA splicing is a process that depends on the spliceosome, a complex of at least 145 core factors, to discriminate between introns and exons present in many metazoan protein coding primary transcripts [3]. In the human genome the spliceosome targets for excision over 200,000 introns ranging in size from under 50 nucleotides to over 500,000 nucleotides utilizing ill-conserved, degenerate splice sites to identify the introns [4,5]. Because of these complexities, it is reasonable to assume that splicing is an error-prone process that generates a significant number of aberrant transcripts. This assumption has been confirmed by recent genomic analyses that suggest that over 90% of protein-coding genes produce some non-canonical transcripts [6]. Human genome studies by the GENCODE project identified 140,066 annotated, non-canonical transcripts from 20,687 protein-coding loci. On average, 6.3 splice variant transcripts (SVs) were characterized per locus [7,8]. Many of these SVs contained premature stop codons and appeared unable to code for proteins. Moreover, analysis of more than 650,000 intronic and exonic variants using machine-learning methods revealed widespread patterns of allele specific aberrant splicing linked to disease [9].

Clearly splicing can be a significant source of degradation of genetic information with a large impact [9]. Aberrant transcripts may not be permanent since most transcripts have limited half-lives, but they do have the potential to transiently affect the efficiency of translation, alter gene expression, contribute to disease, and perhaps to temporarily increase the DNA mutation rate [10,11]. Despite some efforts, assessing the constitutive spliceosome error rate has proven challenging because of the many factors and processes contributing to SV frequency [12,13].

An important source of SVs is alternative splicing (AS), the process by which the spliceosome can generate multiple mRNA isoforms from a single pre-mRNA, thus regulating the spatial and temporal expression of a large number of genes [14]. In practice, determining which transcripts are the result of AS and which ones are the result of error has proven difficult because RNA has many functions in addition to protein coding. Consequently, the existence of functional alternative transcripts mis-identified as aberrant transcripts would tend to overestimate the error rate. This is a significant problem since attempts to measure the error rate that depend on data from genome-wide sequencing analyses are more likely to detect major SVs generated by the alternatively spliced genes rather than the potentially much rarer SVs generated by constitutive splicing errors. Further complicating the estimation of splicing error rates is the process of nonsense mediated decay

E-mail address: askandalis@brocku.ca

http://dx.doi.org/10.1016/j.mrfmmm.2016.01.002
0027-5107/© 2016 Elsevier B.V. All rights reserved.

(NMD), a pathway targeting for destruction some, but not all, transcripts that contain a premature stop codon as a result of splicing that altered the reading frame [15]. Aberrant transcript removal by this pathway may lead to an underestimate of the true spliceosome error rate, but the abundance of transcripts with premature stop codons identified by GENCODE make its overall impact unclear.

In order to estimate the spliceosome error rate we took an approach that addresses the challenges outlined above. Our analysis was based on the characterization of SV frequencies at three loci, *HPRT*, *POLB*, and *TRPV1*, using a cloning methodology capable of detecting low SV frequencies generated by constitutive splicing errors. These loci had been previously tested and found to be largely unaffected by NMD degradation [16]. SV frequencies were assessed in multiple tissues and vertebrate species in order to identify tissue or species-specific alternative splicing biases. In this communication, we report that the frequency of SVs varies widely among loci, tissues, and species. However, each locus exhibits a very similar minimal SV frequency ranging from 0.1 to 0.3% which was not affected by NMD. Consequently we propose that the constitutive splicing error rate in vertebrates is no less than 0.1% per exon.

2. Materials and methods

2.1. Cell description, source and growth

Four cell lines were used. SW480 and RKO are human colon-derived cell lines and the kind gift of Dr. D. Sedwick at Case Western Reserve University. MRC5 cells are untransformed, normal fetal lung fibroblasts obtained from Coriell Cell Repositories (repository number AG05965). These cells were grown in Eagle's minimal essential medium with Hank's BSS, 26 mM HEPES, 10% non-activated fetal bovine serum, 2 mM L-glutamine at 37 °C in 5% CO2. The medium was supplemented with 1×10^{-1} mM hypoxanthine, 4×10^{-4} mM aminopterin and 1.6×10^{-2} mM thymidine to ensure that the cells maintained a functional HPRT gene. TK6 lymphoblastoid cells were a gift from Dr. Howard Liber. The cells were grown in RPMI 1640 supplemented with 10% heat-inactivated horse serum at 37 °C in 5% CO2. All reagents were obtained from Sigma Canada.

2.2. Tissues and tissue sources

AS was investigated in six vertabrates: Human (*Homo sapiens*), rat (*Rattus norvegicus*), rabbit (*Oryctolagus cuniculus*), chicken (*Gallus gallus*), gull (*Larus glaucescens*), and frog (*Xenopus laevis*). Except for human samples, tissues were recovered from young animals before the first quarter of their maximal lifespan. All tissue samples were collected following Brock University Animal Care and Use Committee (ACUC) guidelines except tissue samples from the gull that were collected according to University of British Columbia ACUC. Recovered tissues were placed in RNALater (Ambion) and stored at −80 °C. Human liver RNA (pooled, average age 54 years) was purchased from Clonetech (Liver total RNA, 636,531), human brain RNA (23 individuals pooled, average age 68 years) was purchased from Ambion (FirstChoice human brain reference RNA, AM6050), and tumor RNA (infiltrative breast ductal carcinoma isolated from a 63-year-old Caucasian female) was purchased from Clontech (636635).

2.3. RNA isolation from samples

Total RNA was isolated either from 5×10^6 cells or from tissues using the RNeasy kit (Qiagen) according to the vendor's instructions. Cells were lysed directly in the lysis buffer and tissues were initially homogenized either by mortar and pestle or Dounce glass tissue grinders.

2.4. Loci

Identification of appropriate indicator loci for this study was based on multiple criteria. (1) Loci expressed in all vertebrates, tissues, and developmental stages with minimal turn-over; all loci considered were "housekeeping" genes with well defined functions and expression levels. (2) Loci with mRNA sequence characterized in different vertebrates. (3) Loci with non-essential function to avoid selection pressures that might mask errors. (4) Transcript length less than 2500 bp to allow efficient cloning but containing multiple exons. (5) Not reported to generate "productive" alternative transcripts. (6) Loci that are not subject to NMD surveillance. Due to the extensive criteria, selecting appropriate loci presented considerable difficulty particularly because mRNA sequence was frequently not available in most non-model species and thus we had to perform the characterization ourselves. HPRT is involved in the purine salvage pathway. In humans, *HPRT* is 60,639 nucleotides in length with 654 protein coding nucleotides in 9 exons. The non-terminal exons, those involved in splicing, range in size from 18 to 134 nucleotides (Refseq, NCBI reference sequence). POLB is the main polymerase participating in base excision repair. In humans *POLB* is 33,355 nucleotides in length with 1005 protein coding nucleotides in 14 exons. The non-terminal exons range in size from 50 to 140 nucleotides (Refseq). TRPV1 forms a receptor cation channel. In humans TRPV1 is 31,597 nucleotides in length with 2517 protein coding nucleotides in 17 exons. The non-terminal exons range in size from 63 to 322 nucleotides (Refseq). To identify any tissue specific differences, the sample sources included liver, brain, breast tumor, and transformed and primary cell lines.

2.5. Detection and quantification of splice variants by transcript cloning and sequencing

Transcript cloning involved the reverse transcription of poly(A) mRNAs to cDNA, the amplification of transcripts using gene-specific primers, and the cloning of the amplified products into vectors. Oligo(dT)-primed cDNA synthesized from total RNA as described above, was amplified using gene-specific primers. Gene specific primer pairs were designed based on highly conserved regions on the first and last exons of each locus. The PCR product was purified using spin columns (Microcon-PCR; Millipore) and then was ligated into the vector pGEM-T-Easy kit (Promega). Blue-white colony screening was performed according to the vendor's instructions, and each white colony was then picked into 100 μl of distilled water and heated to 97 °C to lyse. The lysed colony was then used as a template (2 μl) for the amplification of *HPRT*, *POLB* or *TRPV1* sequences. PCR products were then analyzed by electrophoresis, restriction digests, and sequencing as previously reported [16]. Typical types of splice variants are illustrated in Fig. 1.

3. Results and discussion

In order to estimate the error rate of constitutive mRNA splicing, we analyzed the frequency of splice variants at three loci in six vertebrate species. SV frequencies are reported as the number of splice variants per 100 characterized transcripts.

3.1. Splice variant output in different species and tissues

Splice variants were detected at all loci investigated with frequency ranging widely from 0.8% to 91.7%. In humans, the average SV frequencies were *HPRT* 1.0%, *POLB* 56.1%, and *TRPV1* 10.5%

Download English Version:

https://daneshyari.com/en/article/2146172

Download Persian Version:

https://daneshyari.com/article/2146172

Daneshyari.com