Mutation Research 740 (2012) 6–12

Contents lists available at SciVerse ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

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

Features of missense/nonsense mutations in exonic splicing enhancer sequences from cancer-related human genes

Ping Jin[a], Rongrong Cai[b], Xiaodan Zhou[a], Jesse Li-Ling[c,d], Fei Ma[a,*]

[a] Laboratory for Comparative Genomics and Bioinformatics & Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Science, Nanjing Normal University, Nanjing 210046, China
[b] College of Life Science, Liaoning Normal University, Dalian 116029, China
[c] Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110003, China
[d] Institute of Medical Genetics, School of Life Science & Key Laboratory for Bio-resources and Eco-environment of the Ministry of Education, Sichuan University, Chengdu 610064, China

ARTICLE INFO

Article history:
Received 16 October 2011
Received in revised form 31 August 2012
Accepted 19 October 2012
Available online 1 November 2012

Keywords:
Exonic splicing enhancer
Cancer
Human gene
Missense mutation
Nonsense mutation

ABSTRACT

Missense/nonsense mutations, which are related to pathogenic conditions, are regarded as pathogenic mutations. The features of pathogenic mutations in gene coding regions are still unclear. To explore the pathogenic mutation features of human cancer-related genes, 1227 missense/nonsense mutations from 99 human cancer-related genes were analyzed. We found that the mutability in exonic splicing enhancers (ESEs) is less than that outside ESEs. CpG sites are more enriched in ESEs than outside ESEs. Decrease of mutability in ESEs is much larger than that outside ESEs upon removal of CpG mutations since CpG is more mutable. In addition, the bases in ESEs are prone to undergo C→T/G→A mutations. What is more, mutations in ESEs were preferentially located within 50 nt flanking the short exons (≤250 nt), and tend to be of conservative type with minimum effect on the protein structure. Finally, nonsense mutation located in ESEs might be related to Nonsense Mediated Decay (NMD) pathway. In conclusion, this study explored the features of pathogenic mutations of human cancer-related genes.

1. Introduction

Pathogenesis of cancers may be attributed to DNA mutations which can impact on cell growth and migration [1]. Cancer-related mutations include single base substitutions, insertions, deletions, and copy number aberrations [2], while single nucleotide mutations, which mainly comprise missense and nonsense mutations, are most common [3–5]. Recently, an effort has been made to distinguish cancer-associated missense mutations from common polymorphisms [6]. Studying of the features of such mutations not only can provide insights into cancer biology, but may also help to elucidate the occurrence of new mutations.

Missense mutations usually exert their effect on primary amino acid sequences and protein functions [7]. In contrast, nonsense mutations often result in protein truncations or nonfunctional products [8]. The deleterious effects of point mutations may be splicing-related if they are located in functional exonic splicing enhancers (ESEs) [9–13], which are discrete and degenerate motifs of 6–8 nucleotides located in exons [14,15]. Most exons contain at least one functional ESE site [10,15,16]. As essential splicing factors, ESEs are the target sequences for the serine- and arginine-rich (SR) proteins [16–18]. Nucleotide substitutions in ESEs could abolish the binding between SR proteins and their target ESEs, which results in exon skipping by spliceosome machinery [10–12,19]. Therefore, missense/nonsense mutations in splicing enhancer sequences not only may alter the protein coding sequences, but also lead to aberrant splicing.

Pathogenic mutations are mutations related to pathogenic conditions. However, the studies on the features of pathogenic mutations in gene coding regions are limited. To explore the pathogenic mutation features of human cancer genes, we systematically studied 1227 missense/nonsense mutations from 99 human genes retrieved from the Human Gene Mutation Database (HGMD) [20] and uncovered the features of pathogenic missense/nonsense mutations of human cancer-related genes via investigating their distributions especially inside ESEs of human cancer-related gene.

2. Materials and methods

2.1. Missense/nonsense mutations of human cancer-related genes

A total of 1227 missense/nonsense mutations, together with gene symbols, coding sequences and related information of the 99 human genes were retrieved from the HGMD (http://www.hgmd.cf.ac.uk/) [20] (Table S1).

* Corresponding author. Tel.: +86 25 85891852; fax: +86 25 85891852.
E-mail addresses: feima1968@gmail.com, mafei01@tsinghua.org.cn (F. Ma).

0027-5107/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mrfmmm.2012.10.001

Table 1
The putative ESE motifs.

Type of ESE[a]	Point mutations		Missense mutations		Nonsense mutations	
	Number of ESEs[b] (%)	Average score[c]	Number of ESEs (%)	Average score	Number of ESEs (%)	Average score
SC35	315 (27)	3.28	234 (28)	3.28	81 (23)	3.28
SF2/ASF	316 (27)	3.08	211 (26)	3.04	105 (30)	3.15
SRp40	332 (28)	3.5	214 (26)	3.49	118 (34)	3.52
SRp55	209 (18)	3.61	165 (20)	3.61	44 (13)	3.62

[a] The specific SR proteins binding to the ESEs.
[b] Number of ESEs in Point mutations column is classified to be that as missense mutations and nonsense mutations in the third and fourth columns.
[c] Average score in each sub-column is reported as the mean score of hits above the thresholds predicted by ESEfinder for each SR proteins binding at all sequences.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.mrfmmm.2012.10.001.

2.2. ESEs

At present, there are two main programs for ESE prediction. One is ESEfinder, and the other is RESCUE-ESE [21,22]. We referred to the method that Gorlov et al. analyzed the relationship between the missense mutations of cancer-related genes and exonic splicing enhancers corresponding to several SR proteins [7,23]. Alternatively, an ESEfinder software (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi) was used to identify the putative ESEs from the selected genes [21,24]. The software predicts motifs responsive to human SR proteins SF2/ASF, SC35, SRp40 and SRp55 with different weight matrices. As recommended, the thresholds were set as 1.956 for heptamer SF2/ASF, 2.383 for octamer SC35, 2.67 for heptamer SRp40 and 2.676 for hexamer SRp55 [21,24]. ESEs across exonic boundaries were excluded. The frequencies of such motifs in exons and genes were calculated. All mutations have mapped within or outside the ESEs.

2.3. 'Conservative' and 'radical' mutations

According to their carried electric charges, amino acids can be classified into three classes: positive (R, H, K), negative (D, E) and uncharged (A, N, C, Q, G, I, L, M, F, P, S, T, W, Y, V) [25]. Accordingly, a missense mutation is considered to be 'conservative' should the corresponding amino acid be replaced with one from the same class, or 'radical' if it is replaced by one from a different class [26]. Nonsense mutations were not grouped as they will result in a premature stop codon.

2.4. Statistical tests

Fisher exact tests were carried out using statistics programs GraphPad Prism 5 (GraphPad Software Inc., CA, USA). All other statistical analyses were carried out using a Statistic Analysis System (version 9.1.3, SAS Institute Inc, NC, USA). The associations of mutations inside ESEs and outside ESEs with the ratio of the ESE length/CDS length were analyzed using a generalized linear model by the GENMOD procedure (SAS version 9.1.3). The logit link function with the binomial logistic regression was used to determine the effects of the independent variable to the response variable. Parameter estimates were calculated using the maximum likelihood method.

3. Results

3.1. Features of missense/nonsense mutations from human cancer-related genes

Eight hundred and forty-six missense mutations and 381 nonsense mutations were identified from the 99 human cancer-related genes with 287 mutation-containing exons. These genes are distributed across all human chromosomes except chromosome Y (Table S1). The numbers of mutations on each chromosomes ranged from 0 (Y chromosome) to 378 (chromosome 17). The longest CDS consisted of 13,968 nucleotides (*LRP2*), whilst the shortest (*CDKN2B*) only consisted of 417 nucleotides (mean = 2871 nt, S.D. = 2473 nt). Each gene contains 1–337 mutations. Two breast cancer-related genes, *BRCA1* and *BRCA2*, harbored the greatest numbers of mutations. *MLH1* and *MSH2*, both related to nonpolyposis colon cancer, also contained many mutations (162 and 140, respectively). The exons of human cancer-related genes are heterogeneous with various sizes ranging from 11 nt (exon 8 of *MSH2*, with 4 mutations) to 6574 nt (exon 15 of *APC*, with 3 mutations) (mean = 316 nt, S.D. = 645 nt). Two hundred and twenty-four out of the 287 exons (78%) are less than 250 nt in size (Fig. S1). The length of coding sequence is significantly correlated with the number of mutations ($r = 0.2276$, $P = 0.0235 < 0.05$ by Pearson correlation test) (Table S1).

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.mrfmmm.2012.10.001.

3.2. The missense/nonsense mutation-containing ESEs

In total, 60.31% (740 of 1227) of the mutations were located in the 1172 putative ESEs from these 99 cancer-related genes. Eleven of these cancer-related genes (*CDKN1A*, *CREBBP*, *AURKA*, *LRP2*, *PARP1*, *XPC*, *AKAP9*, *AKAP13*, *ABCB1*, *GSTO1* and *NCOA3*) harbor 12 mutations with no mutation in the putative ESEs. The mutation-containing ESEs account for 80.49% (231 of 287) of the exons. However, when mutations of different types are considered, 61.23% (518/846) of missense mutations and 58.27% (222/381) of nonsense mutations are located within ESEs. Considering base pairs located within ESEs, 740 nt were mutated while 179,777 nt were not. Considering based pairs outside ESEs (but within CDS), 487 were mutated while 103,176 were not. These figures indicate that more pathogenic mutations occurred inside ESEs than outside ESEs (Fisher exact test, Two sided, $P = 0.0207 < 0.05$). This argument was further supported by the GENMOD (Generalized Linear Model) analyses using maximum likelihood parameter estimation (estimate inside ESEs = 0.4690, estimate outside ESEs = 0, Wald Chi-Square = 18.69, $P < 0.0001$) (Table S2). However, mutations inside ESEs have a less mutability (defined as number of mutations per hundred nucleotides: $740/180{,}517 \approx 0.4099\%$) than that outside ESEs ($487/103{,}663 \approx 0.4698\%$). These results indicate that bases within ESEs appear less mutable than those outside ESEs, although pathogenic mutations may be to some extent enriched in number from these cancer related genes.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.mrfmmm.2012.10.001.

The proportions of ESEs recognized by the essential splicing factors SC35, SF2/ASF, SRp40 and SRp55 were 27%, 27%, 28%, and 18%, respectively (Table 1). However, no significant difference was found between the above distributions (ANOVA: $F = 0.30$, $P = 0.8248$, $n = 352$) (Table S3). The number of ESEs recognized by SRp55 was the lowest and the average scores for four SR proteins were all greater than 3.00. In addition, 518 missense mutations and 222 nonsense mutations have been found from 824 and 348 putative ESEs, respectively.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.mrfmmm.2012.10.001.

Single base mutations within the four types of ESEs, namely SC35, SF2/ASF, SRp40 and SRp55, are shown in Table 2. The X/Y indicates from nucleotide X to Y unidirectional mutation, while the X|Y represents X to Y and Y to X bidirectional substitutions.