Alternative pre-messenger RNA (mRNA) splicing is a key molecular event that allows for protein diversity and plays important roles in development and disease. Alternative pre-mRNA splicing regulations during spermatogenesis and alternative pre-mRNA splicing etiology in testicular tumorigenesis are yet to be characterized. By genome-wide analysis, here we describe alternative splicing features that distinguish distinctive patterns of alternative pre-mRNA splicing among human testis, testicular cancer and mouse testis. Through computationally subtractive analysis, we detected 80 testis-specific transcript candidates in human testis, 175 in human testicular cancer and 262 in mouse testis, which were integrated into a database. Reverse transcription–polymerase chain reaction confirmed that most of these transcript candidates from mouse testis were testis specific. Around 40% of the transcripts were from unknown/hypothetical genes, which were useful for further functional analysis. These transcripts were not overlapped, indicating lack of evolutionary conservation. Further chromosome mapping showed distinct chromosomal preference of alternative pre-mRNA splicing events. Comparison analysis indicated that alternative pre-mRNA splicing in human testicular tumor shared some characters/trends with those in mouse testis. Moreover, human testicular tumor tended to use rare splice sites and there were also distinct sequences adjacent dominant splice sites between normal testis and testicular tumor. These special features of alternative pre-mRNA splicing in human testicular tumor suggested that testicular tumorigenesis was involved in multiple steps/levels of alternative splicing. Using alternative splicing as a potential source for new clinical diagnostic, prognostic and therapeutic strategies for treatment of testicular tumors seems to have a bright prospect.

Introduction

Since the discovery that the coding information of metazoan genes is interrupted by introns (1,2), it still amazes how both intron removal and exon rearrangement are regulated precisely to produce correct proteomes in cell type- or developmental stage-specific manner. Alternative splicing, the process by which the exons of primary transcripts [pre-messenger RNAs (mRNAs)] from genes can be spliced in different arrangements to produce structurally and functionally distinct mRNA and protein variants, is the most widely used mechanism to enhance protein diversity of higher eukaryotic organisms. It was estimated that ~35–70% of all human genes appear to undergo alternative splicing (3–8). Comparable amounts of alternative splicing were also detected in the fly, nematode and other vertebrates (8,9), emphasizing its major role in the biology of all animals.

In certain pathological conditions, aberrantly spliced pre-mRNAs can be generated that go unnoticed by the nonsense-mediated mRNA decay pathway, and are, therefore, translated into aberrant proteins, which are involved in several human diseases (10–12). There have been also multifaceted studies into the relevance of alternative splicing to cancer at the level of individual genes and individual cancers (13). Currently, the analysis of cancer-specific alternative splicing is a promising step forward in potential source for new clinical diagnostic, prognostic and therapeutic strategies.

However, despite the growing interest on the impact of alternative splicing in various aspects of the biological processes, our understanding of alternative splicing is still very primitive and its regulatory mechanisms are mostly unknown (14,15). The unusually high levels of alternative splicing were often observed in the human brain and testis (16,17). Why is alternative splicing so abundant in the tests? Alternative splicing is probably very important throughout germ cell development. In order to advance our understanding of the biological significance of alternative splicing in humans testis and its tumorigenesis, it is essential to identify tissue-specific splicing events. Although tissue-specific alternative splicing patterns was analyzed by computationally determined splice (8,15,16), comparisons between normal and tumor, especially in human testis, combined with experimentally determined analysis, will provide more information on our understanding of human testicular tumorigenesis.

Here, we performed a genome-wide analysis of alternative splicing, combining with computationally subtractive screening and experimental verification, to compare tissue-specific alternative splicing patterns among human normal testis and testicular cancer and mouse testis. Our results distinguish distinctive patterns of alternative splicing and identify pronounced differences in the chromosomal distribution, splice types and splice-/flanking-splice sites among the three kinds of tissues, suggesting multiple and even complex role of alternative splicing in human testicular tumorigenesis.

Materials and methods

Data sources and filtration

Expressed sequence tag (EST) data of human and mouse were drawn from Cancer Genome Anatomy Project (CGAP) project (http://cgap.nci.nih.gov/Tissues/LibraryFinder). CGAP project collects EST libraries from all over the world and gives good tissue information. We downloaded all available EST libraries of human and mouse from CGAP libraries, Mammalian Gene Collection libraries and Open Reading frame EST Sequencing libraries. We sought to avoid mixing of multiple tissues. Among these libraries that signated ‘pooled’, ‘normalized’ or ‘subtracted’ were excluded because these procedures will affect tissue classification. All EST data in different tissues used were listed in Table I. Human testis libraries were mainly from adults (eight), five were unknown age and one from embryo. Testicular cancers were mainly seminoma and embryonal carcinoma. Mouse testis libraries were from adult, 15 day old and unknown age. All other libraries of the tissues were used regardless of age.

All collection data were then dealt with three procedures: repeat sequence masking (program, repeatmasker; repeat database, repbase; ginsr server: www.girinst.org) to remove simple repeat in the dataset, vector and contamination masking (program, crossmatch; vector database, UniVec_Core; National Center for Biotechnology Information ftp server: ftp://ftp.ncbi.nih.gov/) to clean the vector sequences and a final cleaning of short and redundant sequences (program seqclean from egassembler server: http://egassembler.hgc.jp). Alu repeats were included in following analysis. The filtered ESTs were available for the following analysis.

Computational procedures to identify alternative splicing

A local Basic Local Alignment Search Tool database was constructed for ESTs of each tissue. Testis-specific transcripts were identified based on a subtractive model showed in Figure 1a. Briefly, there are following three steps: (i) Testis EST dataset was blasted to itself. The e-value was set as 1e-30. Gaps (insertion and deletion) of ESTs were identified after alignments. Parameters to identify alternative splicing: the gap length, >10 bp, nucleotide identity, >95%. (ii) Testis ESTs were blasted to the ESTs from the other tissues. Parameters were the same as step (i). (iii) Subtractive ESTs were identified as testis-specific ESTs by insertion/deletion comparisons after the Basic Local Alignment Search Tool. Computer programs were written using Perl language.

EST/genomic sequence alignments, chromosome mapping and splice site analysis

To decrease errors in EST alignments and determine the chromosomal loci of each gene, we localized ESTs to genomic sequences using BLAST-Like
We used the Alignment Tool (http://genome.ucsc.edu). We used the default parameters and selected the best score results. The exon positions of each transcript on the chromosome were recorded and used to determine splice sites and gene structure. Splice sites of both 5' and 3' exon/intron boundaries were aligned online via http://weblogo.berkeley.edu/logo.cgi. We allowed an error of 10 bp in exon/intron boundary. Based on comparisons of EST/genomic alignment, two possible errors can be checked out: (i) candidate EST in the same gene was not on the same chromosome; (ii) candidate EST in the same gene was not in the same locus on the chromosome. The reasons that caused these errors mainly included EST sequencing error, pseudogene and multiple copy genes. The two cases were excluded as false positive in final database.

Reverse transcription–polymerase chain reaction
Reverse transcription–polymerase chain reaction (RT–PCR) was used to amplify transcripts. Total RNAs from mouse and human tissues were prepared by the RNeasy Mini kit (Qiagen, Shanghai, China) according to the manufacturer’s instructions. Human testicular cancer sample (seminoma) was provided by the Clinic of Wuhan University Zhongnan Hospital. All the RNAs were digested by RNase-free DNase I and purified. About 3 µg RNAs were used as template for RT using 0.5 µg poly(T)20 primer and 200 U Moloney Murine Leukemia Virus reverse transcriptase (Promega, Madison, WI, USA). PCR was performed in a 20 µl reaction mix containing 10 mM Tris–HCl, pH 8.3, 1.5 mM MgCl2, 50 mM KCl, 150 mM deoxynucleoside triphosphate, 0.2 µM each primer and 1 U Taq DNA polymerase. Amplification conditions are 94°C, 30 s; 60–68°C (supplementary Table S1 is available at Carcinogenesis Online),

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Human</th>
<th>Human cancers</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone marrow</td>
<td>38783</td>
<td>101 216</td>
<td>167 788</td>
</tr>
<tr>
<td>Brain</td>
<td>823 498</td>
<td>220 417</td>
<td>735 123</td>
</tr>
<tr>
<td>Breast</td>
<td>59 128</td>
<td>144 325</td>
<td>96 353</td>
</tr>
<tr>
<td>Colon</td>
<td>30 811</td>
<td>178 032</td>
<td>76 926</td>
</tr>
<tr>
<td>Eye</td>
<td>168 511</td>
<td>51 699</td>
<td>177 157</td>
</tr>
<tr>
<td>Heart</td>
<td>87 371</td>
<td>—</td>
<td>62 084</td>
</tr>
<tr>
<td>Kidney</td>
<td>102 375</td>
<td>101 280</td>
<td>133 060</td>
</tr>
<tr>
<td>Liver</td>
<td>139 417</td>
<td>103 212</td>
<td>112 250</td>
</tr>
<tr>
<td>Lung</td>
<td>138 121</td>
<td>194 825</td>
<td>37 815</td>
</tr>
<tr>
<td>Muscle</td>
<td>92 184</td>
<td>31 839</td>
<td>38 633</td>
</tr>
<tr>
<td>Ovary</td>
<td>12 917</td>
<td>99 916</td>
<td>20 881</td>
</tr>
<tr>
<td>Pancreas</td>
<td>114 179</td>
<td>117 678</td>
<td>123 059</td>
</tr>
<tr>
<td>Placenta</td>
<td>240 162</td>
<td>43 565</td>
<td>40 883</td>
</tr>
<tr>
<td>Prostate</td>
<td>94 512</td>
<td>221 552</td>
<td>27 135</td>
</tr>
<tr>
<td>Skin</td>
<td>62 193</td>
<td>138 279</td>
<td>67 143</td>
</tr>
<tr>
<td>Spleen</td>
<td>56 738</td>
<td>—</td>
<td>90 320</td>
</tr>
<tr>
<td>Stomach</td>
<td>30 152</td>
<td>69 612</td>
<td>27 067</td>
</tr>
<tr>
<td>Uterus</td>
<td>40 646</td>
<td>144 512</td>
<td>7998</td>
</tr>
<tr>
<td>Testis</td>
<td>18 729</td>
<td>40 915</td>
<td>25 721</td>
</tr>
<tr>
<td>Sum</td>
<td>2 350 427</td>
<td>2 002 874</td>
<td>2 070 996</td>
</tr>
</tbody>
</table>

Fig. 1. (a) Prediction model of tissue-specific alternative splicing. There are transcripts 0, 1, 2, 3... in testis, and transcripts 0, 1, 2... in the other tissues. Through alignments and subtraction of ESTs, testis-specific EST (e.g. No. 3) was detected. (b) Testis-specific transcript numbers by gene among three testis samples. Percentages of different transcripts by gene among human testis, human testicular cancer and mouse testis were shown. Distributions of predicted transcripts specific in human testis and testicular cancer (c) and mouse testis (d) by chromosomes. Chromosome mapping information is based on alignments of ESTs and genomic sequences through BLAST-Like Alignment Tool server.
To facilitate future study and reference of alternatively spliced genes of mammals, we constructed a database based on our analysis, which contained 80 tissue-specific transcripts in human testis, 175 in human testicular cancer and 262 in mouse testis, of which, 40% were unknown/hypothetical genes. It gave an overview of testis-specific alternative splicing in human and mouse and was especially useful for further study on testis oncogenesis. The main information of this database includes gene ID, gene alternative splicing patterns, gene structures, locations in chromosomes, sequences, DNA and protein sequences linked with the National Center for Biotechnology Information website. An example page showed the details of gene A1507611 in mouse (Figure 2). The database can be accessed at http://202.114.72.2:8080/input.jsp (password: 12345321).

To validate alternative splicing detected by the computational approaches, we designed RT–PCR primers to amplify testis-specific transcripts of mouse whose function information was not clear in public databases. Of 62 genes amplified, 26 transcripts were testis specific (Figure 3a) and 23 were dominantly expressed in testis (Figure 3b). The two groups accounted for 79% of all cases. Another eight genes (16%) were ubiquitously expressed in the tested tissues (Figure 3c). The remaining five genes cannot be amplified in our validation. Reasons for this may be related to EST sequencing errors, splicing errors or improper primers. Predicted transcripts specific for testicular cancer were verified by RT–PCR, which showed that some were testis cancer specific (Figure 3e1) and some were expressed at equal or lower level in testis cancer (Figure 3e2).

Patterns of the identified alternative splicing in testis

After comparisons between EST and genomic sequences on chromosomes, we detected five types of alternative splicing patterns: cassette alternative exon, alternative 5’ splice site, alternative 3’ splice site, intron retention and mutually exclusive alternative exons (Figure 4a). It was intriguing that the cassette alternative exon variants formed the most frequent category (50% of all the types). For example, by subtractive analysis, we detected that loss of the exon III by cassette alternative exon of mouse Tcte3 (t-complex-associated testis expressed 3) gene resulted in a testis-specific transcript 2 which deleted Tcte-1 domain (Figure 4c). Similarly, deletion of the exon VI by cassette alternative exon of human TMBIM4 (trans-membrane BAX inhibitor motif containing 4) gene generated a testis-specific short form of transcript 2, which lost the 3’ part of the domain UP000005 (Figure 4d). Alternative splicing may cause testis cancers. In support of this, we detected a testis cancer-specific transcript 1 of human PRDX3 (peroxiredoxin 3) gene, which was produced by the alternative 5’ and 3’ splice sites (Figure 4e). Indeed, we detected >110 of human testis cancer-specific alternatively splicing transcripts.

Comparisons of ratios of these types of alternative splicing patterns among human testis, testicular cancer and mouse testis showed that both alternative 3’ splice site and intron retention had a higher proportion in cancer, however ratio of both cassette alternative exon and alternative 5’ splice site decreased in cancer. In addition, in mouse testis, the type alternative 5’ splice site was highest, whereas the cassette alternative exon was lowest among the three samples (Figure 4b). These results showed that there was significant preference in use of these types of alternative splicing in testicular cancer and mouse tests.

Diversification of the identified alternative splice sites

To see preference/diversification of alternative splice sites in these testis-expressed genes, we analyzed 1747 splice sites in human testis, 5159 in human testicular cancer and 7002 in mouse testis by comparison of each EST with its genomic sequence and mapping it onto chromosome. We detected four basic donor–acceptor splice sites: GT–AG, GC–AG, GG–AG and GT–GG, of which GT–AG was the most dominant site (90–93%; Figure 5a). Furthermore, several interesting points we found were as follows: (i) GG–AG was detected in both normal (0.38–0.47%) and cancer tests (0.74%). (ii) GT–GG was only detected in cancer (0.45%) and mouse tests (0.19%), not in normal human tests. (iii) Besides the four basic donor–acceptor
splice sites, there were quite a number of rare sites especially in testicular cancer (6.3%). (iv) In all, 27% of the rare sites were shared by human testicular cancer and mouse testis, 21% of them were human testis cancer specific, whereas only a few of them were normal human testis specific (4%; Figure 5b). These results suggested that the rare splice sites may be involved in testis tumorigenesis; moreover, most of these rare splice sites were mouse derived.

We further compared sequences adjacent dominant splice sites among human testis cancer, testicular cancer and mouse testis. In 5’ splice site region, downstream dominant splice site GT, though a consensus AAGT at position 3–6, less A at the position 3 (60%) and more AG at the position 4 (74%) and 5 (78%) were observed in testicular cancer compared with relevant positions in normal human testis (70%, 65% and 70%, respectively; Figure 6a). In the upstream of the 3’ splice site AG, no preference of nucleotide at the −2 position and a C/T preference at the −1 position were observed in the three kinds of samples (Figure 6b). However, obvious T preference from position −3 to −9 was detected in testicular cancer, especially at the position −9 (64%), while in normal human testis the positions with the T preference was wider upstream position −3, especially less T (57%) at the position −9 and more T at the position 10 (64%; cancer, 42%) and 11 (63%; cancer, 50%), and less obvious T preference in mouse testis was observed.

Discussion

We have developed a method for precisely delineating alternative splice patterns by alignment EST/cDNA sequences first, and comparing the ESTs/cDNAs with their genomic sequences, and then mapping them onto chromosomes. These procedures eliminated many EST errors, pseudogene and multiple copy/repeat gene problems when data were from diverse EST databases. Considering the fact that genome-wide studies found transposable elements (mainly Alus) in nearly 4% of human protein-coding regions (18), the Alu repeats were not excluded in our subtractive analysis. The alternatively spliced transcripts were subject to subtractive screening of testis versus all the other tissues, and these analyses finally yielded testis-specific transcripts. Interestingly, ~40% of the transcripts were from unknown/hypothetical genes, which were useful for further functional analysis. RT–PCR analysis further confirmed most of our computationally determined results. Nevertheless, there were a small number of transcripts from computationally subtractive results, which were ubiquitously expressed in the tested tissues. This may be caused by EST coverage, their fragmentary character and sequencing errors escaped from our analysis. In any case, we believe our database of 80 testis-specific transcript candidates in human testis, 175 in human testicular cancer and 262 in mouse testis can be a rich source of discovery for our studying spermatogenesis and especially testicular tumorigenesis.

However, its major value is providing previously unknown patterns of alternative splicing in human testicular tumor. (i) Alternative splicing events in testicular tumor have chromosomal preference, a strong preference to chromosome 2, 11, 12, 19 and X. (ii) Alternative splicing in human testicular tumor shares some characters/trends with those in mouse testis. For example, in the usage of types of alternative splicing, both human testicular cancer and mouse testis tend to use types of both alternative 3’ splice site and intron retention. Moreover, a considerable number (27%) of the rare splice sites are shared by human testicular cancer and mouse testis, and genes involved in synthesis/degradation are frequently spliced in these two groups. (iii) Human testicular tumor tends to use rare splice sites. (iv) There are distinct sequences adjacent dominant splice sites in human testicular tumor. These special features of alternative splicing in human testicular tumor suggest that testicular tumorigenesis is involved in multiple steps/levels of alternative splicing events. This view is consistent with results observed in testicular germ cell tumorigenesis, which is involved in multiple aspects including somatic genetic
changes, allelic imbalance, multiple gene mutations and methylation patterns (19). Testicular cancer seems to need multiple genetic events to occur. Taken together, these data suggest a complex etiopathology of testicular tumorigenesis. Our data include discovery of many novel splice forms of testicular cancer-associated genes, alternative splicing patterns of testicular cancer and suggest a significant new direction for testicular cancer research.

Interestingly, these tissue-specific alternative splicing variants are not evolutionarily conserved between human and mouse. Whether would the biological significance of their existence questionable? The same results were obtained to show that vast majority of alternative splicing events represented in the human EST database are not conserved in mouse (20). Main explanation for this issue may be because most of these splicing events may represent tissue-, disease- or allele-specific splicing, or events for which phylogenetic distribution is highly restricted. And alternative splicing could be also the easiest road to diversification of the species. We believe that species-specific alternative splicing represents a fascinating aspect of it that has undoubtedly had a major impact on genome evolution (21). Moreover, analyses of substitution rates in the genomes of different mammalian species indicate that the exons of species-specific splice variants evolve more rapidly than constitutive or conserved alternatively spliced exons (22). Indeed, it will be important in future studies to decipher the interplay among species-specific alternative splicing, protein diversity and pathogenic alternative splicing in cancer.

Main interest of this paper is data mining for testis-specific transcripts by the genome-wide identification and subtractive analysis of alternative splicing variants. Some of the testis cancer-specific splice forms we detected may be medically important. However, this should be further verified in detail by experimental approaches such microarray analysis and functional analysis, so as to determine which are initial triggers of tumorigenesis or just reflections of transformed state, because at present time, some factors such as grade of aggressiveness, aging and stress were not considered owing to limitation of data sources. Indeed these factors affected the alternative splicing events (23–26). Computational analyses have gone some way to providing a global view of tissue-specific splicing, but these studies are limited by inadequate transcript coverage, by difficulties in assigning ESTs to tissue or cell types or sorting different categories.
In conclusion, evidences are accumulating that just as alternative splicing is important for spermatogenesis, so aberrations of alternative splicing patterns are important for testicular cancer. Testicular tumorigenesis needs multiple genetic events to occur; therefore, processes at the top of genetic hierarchies such as alternative splicing could be the cause, and the prospect that we might start to use alternative splicing as a potential source for new clinical diagnostic, prognostic and therapeutic strategies. For example, novel therapeutic strategies could be designed to target the splicing machinery, which can be defective in some testicular tumors. Further efforts into alternative splicing etiology in testicular cancer seem bright.

Supplementary material

Supplementary Table S1 can be found at http://carcin.oxfordjournals.org/

Funding

National Natural Science Foundation of China; the National Key Basic Research project (2006CB102103); the Program for New Century Excellent Talents in University; 111 project (B06018).

Acknowledgements

The authors thank Dr J.Feng for help in writing software.

Conflict of Interest Statement: None declared.

References


Received May 7, 2007; revised July 30, 2007; accepted August 14, 2007