Genome Analysis

Insights into the connection between cancer and alternative splicing

Eddo Kim[*], Amir Goren[*] and Gil Ast

Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel

Computational and experimental evidence has revealed that cancerous cells express transcript variants that are abnormally spliced, suggesting that mRNAs are more frequently alternatively spliced in cancerous tissues than in normal ones. We show that cancerous tissues exhibit lower levels of alternative splicing than do normal tissues. Moreover, we found that the distribution of types of alternative splicing differs between cancerous and normal tissues. We further show evidence suggesting that the lower levels of alternative splicing in cancerous tissues might be a result of disruption of splicing regulatory proteins.

Alternative splicing and cancer

Alternative splicing is a mechanism that increases transcriptome and proteome diversity by allowing generation of multiple mRNA products from a single gene. Alternative splicing is the most plausible solution for the paradoxical miscorrelation between the number of genes transcribed by different eukaryotes and the level of phenotypic complexity [1,2]. Defects in mRNA splicing are an important cause of disease. At least 15% of all disease-causing single base-pair mutations affect splicing; hence, the splicing process and its regulation are of great interest [3–5]. The connection between cancer and novel alternative splice forms has been well established. Wang *et al.* [6] found that an alternatively spliced variant of the gene encoding spleen tyrosine kinase is frequently expressed in breast cancer cells but never in matched normal mammary tissues. Other studies have also revealed cancer-specific splice variants that are absent from normal tissues [7,8]. Splice variants found solely in cancerous tissues could be causative of disease or involved in disease progression, because alternative splicing of many genes has a critical impact on all major aspects of cell biology, including cell cycle control, apoptosis and more (reviewed in Refs [9–11]). Recently bioinformatic approaches have been used to identify cancer-associated splice variants. Xu and Lee [12] discovered cancer-specific splice variants in 316 genes. Surprisingly, for a large number of cancer-associated genes, the sequence deposited in GenBank seems to be the cancer-specific form and not the one observed in normal tissues. Roy *et al.* [13] further found that tumor-associated splice variants were twice as likely to be represented in GenBank compared with their normal tissues counterparts. Other

bioinformatic studies have found tumor-associated splice variants, some of which have been validated experimentally [14,15].

Both the experimental and computational data support the notion that alternative splicing is more common in cancerous tissues than in normal ones. However, the overall level of alternative splicing in cancerous tissues compared with normal ones and the distribution of the different types of alternative splicing have not previously been evaluated. Moreover, the reason for the emergence of these cancer-specific variants is not well understood. It is currently unknown whether these variants appeared only after cells have transformed or whether they are one of the forces that propel tumorigenesis. If the former is correct, is aberrant splicing merely a side effect of the severe abnormalities observed in transformed cells or is it induced in tumors to assist in its growth?

Cancer tissues exhibit lower levels of alternative splicing than normal tissues

Cancer cells exhibit aberrant transcript variants not found in normal cells. We set out to determine whether this high level of aberrant splicing in tumor cells is a result of hyperactivity of the splicing machinery, increasing its ability to alternatively splice transcripts, or a result of suppression of the splicing machinery, caused, for example, by improper recognition of splice junctions. We therefore examined human genes that are represented in UniGene clusters [16] to determine whether alternative splicing was more prevalent in tumor tissues than in normal ones. Based on the UniGene annotations, we divided the expressed sequence tags (ESTs) into two subgroups according to their origin: normal tissues and cancerous (neoplastic) tissues. We analyzed the splicing patterns of each exon in each gene represented by a UniGene cluster and determined whether it was constitutively or alternatively spliced (see Supplementary data). Because cancerous tissues were shown to exhibit high levels of aberrant splicing, one would expect the level of alternatively spliced exons and genes (namely, genes consisting of at least one alternatively spliced exon) to be higher in these tissues compared with normal ones. Surprisingly, the prevalence of alternatively spliced exons, as well as alternatively spliced genes, was somewhat higher in normal tissues than in tumor tissues ($P = 2.63e^{-68}$ and $P = 2.01e^{-120}$ for alternatively spliced exons and genes, respectively; Table 1).

Because the coverage of EST sequences in normal tissues is higher than that in cancerous tissues, it is possible these results might be biased. We therefore created comparable

Corresponding author: Ast, G. (gilast@post.tau.ac.il).
  Available online xxxxxx.
[*] These authors contributed equally to this work.

Table 1. Comparison of alternative splicing levels between normal and cancerous tissues

Evidence	Tissue	CE	AE	CG	AG
All ESTs	Normal	102 646	27 044	10 909	8 941
	Cancer	85 739	18 786	13 180	6 670
MGC + eVOC 5 rand ESTs	Normal	40 527	1 889	5 148	1 382
	Cancer	43 068	1 620	5 319	1 211
MGC + eVOC 10 rand ESTs	Normal	34 209	2 594	2 785	1 654
	Cancer	35 954	2 344	2 930	1 509
MGC + eVOC 15 rand ESTs	Normal	26 829	2 638	1 751	1 513
	Cancer	27 678	2 444	1 844	1 420
MGC + eVOC 20 rand ESTs	Normal	20 788	2 466	1 179	1 286
	Cancer	21 215	2 273	1 254	1 211

The numbers of constitutive exons (CE), alternative exons (AE), constitutively spliced genes (CG), and alternatively spliced genes (AG) are shown for normal and cancerous tissues using either all available expressed sequence tags (ESTs) (All ESTs), or randomized Mammalian Gene Collection (MGC) ESTs with similar UniGene and controlled vocabulary for unifying gene expression data (eVOC) annotations for five, ten, 15, and 20 supporting ESTs (MGC + eVOC X rand ESTs).

datasets of human ESTs for normal and cancerous tissues (see Supplementary data). Briefly, we normalized the number of ESTs, used ESTs derived from the Mammalian Gene Collection [17] and used the controlled vocabulary for unifying gene expression data (eVOC) [18] annotations to further support the UniGene annotations. Examining the comparable datasets of high-quality ESTs further supported that the level of alternatively spliced exons and genes is moderately, yet statistically significant, lower in cancer tissues compared with normal ones ($P = 2.91e^{-7}$ and $P = 1.31e^{-3}$ for alternatively spliced exons and genes, respectively; Table 1).

We were also concerned that the use of UniGene clusters of only ten ESTs from normal tissues and ten ESTs from cancerous tissues might introduce a bias to the results. We repeated the above analyses using randomly selected UniGene clusters of five, 15 or 20 sequences from normal and cancerous tissues and revealed similar results, minimizing such a possibility (see Table 1 and Supplementary data).

The distribution of alternative splicing types differs between cancerous and normal tissues

Exons undergo four main types of alternative splicing: exon skipping, alternative acceptor and donor (3′ splice site and 5′ splice site, respectively) splice signal selection, and intron retention [19]. Because we found that cancerous tissues exhibited lower levels of alternative splicing compared with normal tissues, we examined whether this difference was associated with differences in the prevalence of types of alternative splicing. We classified the alternatively spliced events in our EST database into the four major subgroups and determined the prevalence of the different types of alternative splicing in normal and cancerous tissues. There was a significant difference in the distribution of alternative splicing types between normal and cancerous tissues ($P = 6.22e^{-33}$; Figure 1). Examination of the ESTs from cancerous tissues revealed that exon skipping was less frequent compared with normal tissues ($P = 2.62e^{-18}$), whereas alternative 3′ splice site (3′ss) and 5′ splice site (5′ss) selection and intron retention were observed more often in cancerous ESTs ($P = 8.99e^{-3}$, $P = 3.87e^{-3}$ and $P = 4.24e^{-15}$, respectively). Exon skipping is the most prevalent form of alternative splicing in most eukaryotes, and its prevalence exhibits a gradual increase along the eukaryotic tree [20,21]. This suggests that the exon skipping mechanism has been critical in shaping the evolution of higher eukaryotes, where exon definition

is the dominant splicing mechanism [19,22]. By contrast, intron retention was found to be the rarest type of alternative splicing, suggesting that its contribution to the phenotypic complexity of higher eukaryotes is modest. Our findings reveal a combination of reduced levels of exon skipping and higher levels of intron retention in cancerous tissues, presumably indicating aberrant expression or activity of splicing regulatory factors [23–27], which would affect a broad spectrum of cellular activities.

De novo exonizations in cancer are associated with mRNA processing genes

Although the overall level of alternative splicing is lower in cancerous compared with normal tissues, we sought to identify cancer-specific exons. These are exons present in transcripts expressed in cancer cells that are intronic sequences in normal tissues; inclusion of intronic sequence within an mRNA is termed exonization. Overall, we found 325 exons in ESTs from cancerous tissues that were absent from all ESTs derived from normal tissues (at least 10 ESTs supported each normal cluster; see Supplementary data for EST accessions). Novel cancerous exonization events could alter the protein coding sequence or introduce a premature termination codon (PTC), thus lowering the level of the functional protein in the cell [26,27]. Indeed, 202 of the 325 exons are not divisible by three (62.15%) and
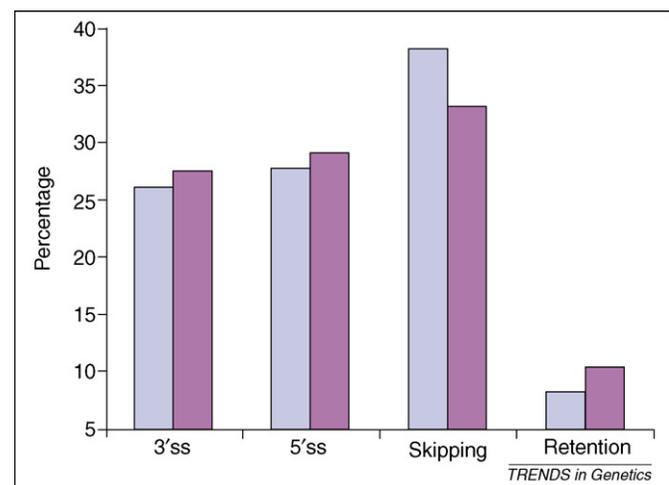
Figure 1. Distribution of alternative splicing types. The percentage (*y*-axis) of each of the four alternative splicing types (*x*-axis), alternative 3′ splice site selection (3′ss), alternative 5′ splice site selection (5′ss), exon skipping (skipping), and intron retention (retention), is shown for normal (blue) and cancerous (purple) tissues.

could therefore introduce a PTC, resulting in a truncated protein. We examined whether specific types of genes were associated with these exonization events, which might suggest potential effects on cell growth. We extracted the Gene Ontology (GO) annotations [28] for each of the genes that harbored these exons and compared them with the annotations of the entire gene collection in our dataset (see Supplementary data). The genes associated with cancer *de novo* exonizations were significantly associated with five GO annotations, four of which are linked with mRNA processing (Table S1). This result indicates a direct link between cancer and alterations of splicing patterns and suggests that alteration of proteins associated with mRNA processing is associated with cancer. It might be possible that the splicing factors that are altered in cancer are those responsible for alternative rather than constitutive splicing, and therefore, the lower levels of alternative splicing in cancer tissues. This supports the recent finding that SR protein genes, which are involved in regulation of alternative splicing, contain an alternatively spliced PTC-containing 'poison' exon that marks the transcript for degradation by the nonsense-mediated mRNA decay mechanism [26,27]. It will be interesting to study whether these genes are more prone to nonsense-mediated mRNA decay (NMD) degradation in cancerous than in normal tissues. Moreover, we also found several intriguing cases of exonization in genes that are not related to RNA processing (see Supplementary data). These cases imply that exonizations in genes that are involved in antitumor activities impair their function in cancer cells by insertion of PTC-containing exons.

Concluding remarks

In summary, we showed that the level of alternative splicing is surprisingly lower in cancerous tissues than in normal tissues. We found that this difference stems from a difference in the level of exon skipping events. Moreover, we found that cancer-specific *de novo* exonizations often interfere with genes involved in mRNA processing, suggesting that splicing regulatory factors are not expressed at proper levels and/or that their functions are impaired in cancer. Our data suggest that the aberrant splicing patterns observed in cancer are not side effects of cancer development, but rather that alternations in the splicing machinery and its regulation are key to cancer progression.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.10.001.

References

1 Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107
2 Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19
3 Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437
4 Garcia-Blanco, M.A. *et al.* (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.* 22, 535–546
5 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106–110
6 Wang, L. *et al.* (2003) Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. *Cancer Res.* 63, 4724–4730
7 Song, S.W. *et al.* (2005) Inactivation of the invasion inhibitory gene IIp45 by alternative splicing in gliomas. *Cancer Res.* 65, 3562–3567
8 Ermak, G. *et al.* (1996) Novel CD44 messenger RNA isoforms in human thyroid and breast tissues feature unusual sequence rearrangements. *Clin. Cancer Res.* 2, 1251–1254
9 Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell* 126, 37–47
10 Venables, J.P. (2006) Unbalanced alternative splicing and its significance in cancer. *Bioessays* 28, 378–386
11 Watahiki, A. *et al.* (2004) Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. *Nat. Methods* 1, 233–239
12 Xu, Q. and Lee, C. (2003) Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res.* 31, 5635–5643
13 Roy, M. *et al.* (2005) Evidence that public database records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. *Nucleic Acids Res.* 33, 5026–5033
14 Wang, Z. *et al.* (2003) Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res.* 63, 655–657
15 Hui, L. *et al.* (2004) Identification of alternatively spliced mRNA variants related to cancers by genome-wide ESTs alignment. *Oncogene* 23, 3013–3023
16 Wheeler, D.L. *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 33, D39–D45
17 Gerhard, D.S. *et al.* (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* 14, 2121–2127
18 Kelso, J. *et al.* (2003) eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.* 13, 1222–1230
19 Ast, G. (2004) How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773–782
20 Sugnet, C.W. *et al.* (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.* 9, 66–77
21 Kim, E. *et al.* (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35, 125–131
22 Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
23 Kirschbaum-Slager, N. *et al.* (2004) Splicing factors are differentially expressed in tumors. *Genet. Mol. Res.* 3, 512–520
24 Relogio, A. *et al.* (2005) Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.* 280, 4779–4784
25 Karni, R. *et al.* (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* 14, 185–193
26 Ni, J.Z. *et al.* (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* 21, 708–718
27 Lareau, L.F. *et al.* (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929
28 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29