[Frontiers in Bioscience 12, 4034-4041, May 1, 2007]

Genomic analysis of RNA alternative splicing in cancers

Yi Xing

Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Alternative splicing and its functional impact
4. Discovery of cancer-specific splice forms from expressed sequences
5. Global profiling of alternative splicing in cancers using microarrays
6. Future Challenges
7. Medical Implications
8. Acknowledgements
9. References

1. ABSTRACT

Alternative splicing is an important and prevalent mechanism of gene regulation in higher eukaryotes. Nearly three quarters of human multi-exon genes are alternatively spliced. There is great interest in discovering alternative splicing events in the transcriptome of cancer cells, and in understanding how alternative splicing contributes to tumorigenesis. In this article, I will review recent advances in global analyses of pre-mRNA alternative splicing, and the applications of these genomic technologies to studies of alternative splicing in human cancers.

2. INTRODUCTION

It was anticipated that human genome had more than 100,000 genes (1). However, after the completion of the Human Genome Project, the analysis of the genome sequence indicated only ~30,000 genes (2,3). Equally surprisingly, it was revealed that a strikingly high percentage of human genes produce multiple transcript and protein product through RNA alternative splicing (4-6). Today, it is widely accepted that alternative splicing plays a ubiquitous role in gene regulation in higher eukaryotes. It affects a wide range of biological processes such as hearing (7), heart development (8), apoptosis (9), and so on. Alternative splicing is a major cause of human diseases (10). An exciting new subject of research is alternative splicing in human cancers. There are well-documented cases of alternative splicing in cancer-related genes (11-13). In this article, I will describe recent advances in genome-wide analyses of alternative splicing, and what we have learned about alternative splicing in human cancers using genomics.

3. ALTERNATIVE SPLICING AND ITS FUNCTIONAL IMPACT

The majority of genes in human and other higher eukaryotes have multiple exons and introns. The pre-

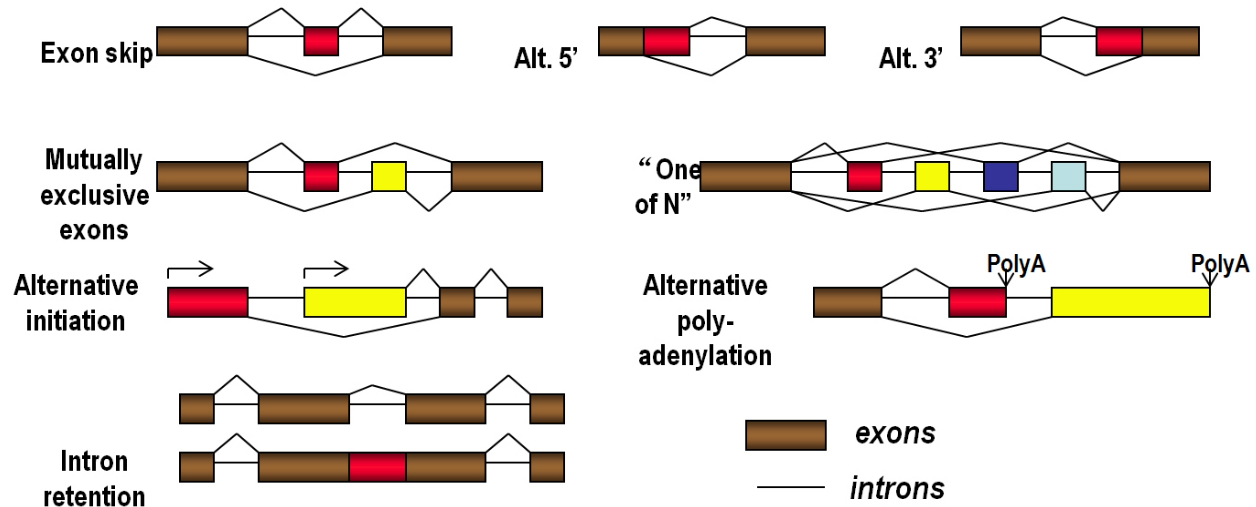

Figure 1. Basic types of alternative splicing events.

RNA of a multi-exon gene undergoes splicing to remove introns before it can be exported out of the nucleus and translated. The control of splicing requires precise recognitions of cis-regulatory elements in exons and their surrounding introns by the splicing machinery. Either mutations of these cis-regulatory elements, or differential expression or activation of trans-acting factors that recognize these elements, can change the default splicing pattern of a gene, leading to alternative splicing (14). Figure 1 illustrates different types of elementary alternative splicing events, such as exon skipping, alternative donor/acceptor sites, mutually exclusive exon usage, etc. In many genes, these elementary alternative splicing events are combined in a complex manner, generating a large number of distinct transcript and protein isoforms. A famous example is the alternative splicing of a Drosophila gene *Dscam*, which has >38,000 distinct protein isoforms (15). Its mosquito homolog, *AgDscam*, has >31,000 protein isoforms (16).

Alternative splicing provides an important mechanism for increasing the regulatory and functional diversity of eukaryotic transcriptomes and proteomes (17). It is common for different protein isoforms of a single gene to have distinct or even antagonistic functions, due to the insertion or deletion of key functional regions. For example, the long isoform of a mammalian serine-threonine kinase gene WNK1 inhibits ROMK1 potassium channel by promoting its endocytosis (18). The kidney-specific alternative splicing of WNK1 disrupts its kinase domain (19). Due to its loss of the kinase activity, this kidney-specific WNK1 isoform reverses the inhibition of ROMK1 by the long WNK1 isoform (18). Another example is the alternative splicing of integral membrane proteins. In many genes encoding single-pass membrane proteins, alternative splicing specifically removes the transmembrane segments, turning the proteins into their soluble forms (20,21). Other studies have investigated the genome-wide impact of alternative splicing on protein domains (22,23), structural elements (24,25), subcellular localization signals (26).

Alternative splicing can impact the function of important cancer-related genes. A classic example is the alternative splicing of Bcl-x. The longer protein isoform of Bcl-x is anti-apoptotic while the shorter protein isoform has pro-apoptotic activity (9). In fact, many genes involved in the regulation of apoptosis and cell proliferation have multiple protein isoforms via alternative splicing (11). Before genomics, however, our knowledge about alternative splicing in cancers was very limited.

4. DISCOVERY OF CANCER-SPECIFIC SPLICE FORMS FROM EXPRESSED SEQUENCES

For almost two decades since the exon-intron structure was discovered (27,28), alternative splicing was regarded as a rare form of gene regulation. Its frequency was considered to be less than 5% in human genes (4). However since mid-1990s, genomics has completely transformed our view on alternative splicing. The first wave of large-scale alternative splicing discovery came from computational analyses of expressed sequences such as full-length cDNAs and expressed sequence tags (ESTs) (4). ESTs are shotgun fragments of full-length mRNA sequences. During the past ten years, EST sequencing projects have generated >6.5 million human ESTs (see http://www.ncbi.nlm.nih.gov/UniGene/). ESTs account for over 95% of human transcript sequences deposited to the NCBI UniGene database. They keep "snapshots" of human transcripts under a variety of conditions, and are particularly useful for discoveries of novel splice variants.

Analyses of expressed sequences suggest that alternative splicing is widespread in human and many other species (4). Figure 2A shows how to detect alternative splicing from expressed sequences. Briefly speaking, mRNA and EST sequences of a gene are aligned to the genome to identify the exon-intron structure. Exons are identified as aligned regions between the expressed sequence and the genomic sequence. Introns are identified as long gaps in the expressed sequence relative to the

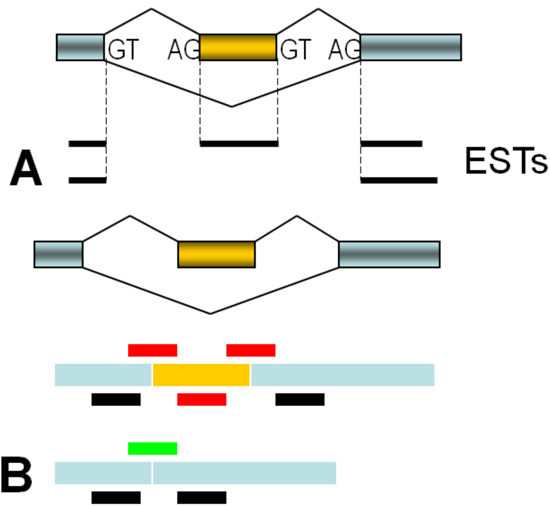

Figure 2. Methods to detect alternative splicing. (A) Detections of alternative splicing from EST data. ESTs are aligned to the genomic sequence. The top EST indicates exon inclusion, while the bottom EST indicates exon skipping. Taken together, they provide evidence for alternative splicing of the middle exon. (B) Profiling alternative splicing using oligonucleotide probes. Probes against exons and exon-exon junctions are used to interrogate alternative splicing of the middle cassette exon. Probes colored in black target constitutive exons and their intensities are measures of overall gene expression. Probes colored in red target the exon-inclusion transcript. Their intensities will be upregulated if exon inclusion becomes predominant. The probe colored in green targets the exon-skipping transcript. Its intensity will be upregulated if exon skipping becomes the predominant splicing pattern.

genomic sequence. The exon-intron boundary can be confirmed by the conserved "GT-AG" or "AT-AC" splice sites, which occupy the first and last two nucleotides of most (>99%) introns in eukaryotes. Next we compare the exon-intron structures of multiple expressed sequences from the same gene for evidence of alternative splicing. For example, in Figure 2A, the middle exon is included in one EST but skipped from another EST, indicating an exon skipping alternative splicing event. Many researchers have used this strategy to detect alternative splicing in animal and plant genomes (29-40). These studies unanimously conclude that alternative splicing is widespread in higher eukaryotes. For example, analyses of human cDNA and EST database suggest that 40-60% of human genes are alternatively spliced. This frequency is much higher than previously expected (4).

EST databases record the origins of many EST libraries (19,41), such as their tissue sources, developmental states, etc. We can use such information to detect splice forms that are differentially spliced under various conditions. Several groups classified EST libraries into tumor or normal libraries, and developed automatic procedures to detect splice forms enriched in tumors or normal human tissues (42-45). Wang and colleagues analyzed 11,014 RefSeq sequences and >3.4 million EST sequences to identify tumor-associated splice forms. Their genome-wide screen identified 845 isoforms that are enriched in tumor samples (43). Among the 67 genes chosen for RT-PCR validation, alternative splicing was confirmed in 55 (82%) genes, and the enrichment of splice forms in tumor samples was validated in 45 (67%) genes. Some of the tumor-enriched splice forms were restricted to a particular type of cancers, while some others were prevalent in all the tumor samples surveyed in the study. Xu and colleagues developed a rigorous Bayesian approach to identify 316 human genes differentially spliced between tumors and normal tissues (42). These 316 genes were strongly enriched for tumor suppressors and genes involved in cell proliferation and cell cycle control, suggesting that alternative splicing of these genes might play a role in tumorigenesis. Surprisingly, in many of these genes, the tumor-specific splice forms were well-studied, but the splice forms specific to normal tissues were unknown splice forms. For example, estrogen-related receptor α (ERRα) is an orphan nuclear receptor transcription factor implicated in breast cancer. It is constitutively active regardless of the presence of hormone. Xu *et al.* identified a novel splice variant of ERRα which was only present in ESTs from normal tissues. The novel splice form removes an exon from the transcript product, causing a 57aa in-frame deletion. The resulting protein isoform lacks ERRα's transcriptional coactivator binding site but still retains its dimerization region. Xu *et al.* suggested that the novel isoform of ERRα could act as a dominate negative form to repress ERRα activity in normal tissues (42). A follow-up experimental study confirmed the existence of many "novel normal" splice forms (46). These results suggest that many cancer-related genes, which were initially cloned from tumor libraries, have uncharacterized transcripts in normal tissues. Studying the function and regulation of these "normal" transcripts might shed new light on the mechanism of tumorigenesis. In addition, some of the cancer-specific splice variants are potential cancer biomarkers (47) or therapeutic targets (48).

However, these EST-based analyses of cancer-specific alternative splicing have many limitations. First, ESTs are not high-quality sequence data. They contain various types of artifacts and errors (49). Second, due to issues such as the normalization of EST libraries, EST count doesn't accurately reflect the actual abundance of a particular splice form in the cell (41). Third, ESTs are biased toward highly expressed genes. Fourth, EST sequences are usually biased towards 3' end of the transcripts. Alternative splicing events in the 5' end and the middle of the transcripts are underrepresented in the EST data. Finally, EST databases simply don't cover all the tissues and cellular states that are interesting to individual investigators.

Fortunately, microarrays provide an attractive solution to these problems (50). The recent advance in splicing microarray technology offers a powerful tool to profile alternative splicing in the entire transcriptome, to which I will devote the remainder of this review.

5. GLOBAL PROFILING OF ALTERNATIVE SPLICING IN CANCERS USING MICROARRAYS

Although expression microarrays are traditionally used for measuring overall mRNA abundance (51), it can also be used for high throughput profiling of RNA alternative splicing. This new application requires a fundamental change in the design of microarray probes. Traditionally, alternative splicing has been largely ignored throughout the probe design and data analysis of expression microarrays. For example, the conventional Affymetrix GeneChips use 11 perfect-match (PM) probes and 11 mismatch (MM) probes against the 3' end of the mRNA sequence. The signals from multiple probes are summarized into a single value representing the overall expression index (52). On the other hand, in order to obtain quantitative measurements of alternative splicing, we need oligonucleotide probes that specifically target different mRNA isoforms of the same gene. Figure 2B demonstrates how to detect an exon skipping event using microarrays. We use probes complementary to individual exons or exon-exon junctions. The relative intensities of various probes indicate the splicing pattern of an exon (see Figure 2B).

Soon after the prevalence of RNA alternative splicing was recognized, several studies successfully demonstrated the use of oligonucleotide probes to profile alternative splicing in individual genes (53-56). A genome-wide splicing microarray platform was developed to detect intron-retention events in *S. cerevisiae* (57). Clark and colleagues designed oligonucleotide probes targeting introns, exons and exon-exon junctions of intron-containing genes in yeast. Microarray data indicated that the inactivation of several important splicing regulatory proteins led to global repression of mRNA splicing and accumulation of intron-containing transcripts in yeast. Johnson and colleagues reported a genome-wide survey of mRNA alternative splicing in 52 human tissues using exon junction arrays (58). They designed ~125,000 36-mer probes against the junctions of adjacent exons in ~10000 human multi-exon genes. If two consecutive exon junctions give very low signals, it suggests that the middle exon is skipped. This microarray platform detected hundreds of novel exon-skipping events. Combined with EST data, the microarray data by Johnson and colleagues suggest that 74% of multi-exon genes in human have multiple splice forms (58). Pan and colleagues designed exon probes and exon junction probes targeting 3126 cassette exons in mouse (59). They used a Bayesian model, GenASAP (60), to infer the exon inclusion levels of 3126 cassette exons in ten mouse tissues. This study provided a large set of tissue-specific alternative exons in the mouse transcriptome. Although most of the genome-wide splicing microarray platforms were designed to detect exon skipping events, Le and colleagues showed that microarrays can also be used to detect more subtle splicing changes, such as alternative donor/acceptor sites splicing (61). Taken together, these studies (also see (62)) demonstrate that splicing microarrays provide a powerful tool for genome-wide analysis of alternative splicing on any samples of interest. Moreover, unlike EST-based studies, microarray profiling of alternative splicing is not biased towards highly expressed genes, as long as we have probes for medium and low expressers. Researchers have used splicing microarrays to identify targets of important splicing regulators (63,64), to detect brain-specific and muscle-specific exons in mouse (65), to study the relationship of alternative splicing and mRNA nonsense mediated decay (66), etc.

Besides customized arrays designed by individual laboratories, there are several commercially available splicing microarray platforms from Affymetrix (http://www.affymetrix.com/products/arrays/exon_application.affx), Jivan (http://www.jivanbio.com/products.html) and Exonhit (http://www.exonhit.com/). A major difference of these commercial platforms is their probe design. The splicing microarray platforms from Jivan and Exonhit have both exon probes and junction probes for quantitative measurements of known alternative splicing events. They also provide smaller versions of their arrays that target specific sets of genes (e.g. apoptosis-related genes). By contrast, the Affymetrix Exon Arrays only have exon probes. It is of a significantly higher density, with over six million probes targeting all annotated and predicted exons in a genome. This allows accurate measures of overall gene expression and qualitative discoveries of novel alternative splicing events. However, direct experimental comparisons of these platforms haven't been performed.

Splicing microarrays enable us to depict a high-resolution, exon-level picture of the cancer transcriptome. A few studies have used splicing microarrays to detect alternative splicing events in Hodgkin lymphoma cells (67), prostate cancer (68,69), and breast cancer cell lines (70). In one study, an oligonucleotide splicing microarray platform was used to detect differential splicing in two breast cancer cell lines (MCF7, estrogen-receptor positive; MDA-MB-231, estrogen-receptor negative) and normal mammary epithelial cells (HMEC) (70). 15 events of differential splicing were found between MCF7 and HMEC, of which 11 were validated by RT-PCR. Similarly, 14 events of differential splicing were found between MDA-MB-231 and HMEC, of which 11 were validated by RT-PCR (70). Li and colleagues used a fiber-optic microarray platform to measure the abundance of 1532 mRNA isoforms from 364 prostate-cancer related genes in prostate tumor cell lines and tumor cell lines of non-prostate origins (68). They developed a computational algorithm to simultaneously infer changes in transcript abundance and changes in splicing. Based on RT-PCR validation tests for 24 genes, the method had a false discovery rate of ~5% and a false negative rate of ~20% in detecting differential splicing in prostate tumor cell lines. Li *et al.* then applied their isoform profiling approach to clinical samples of prostate tumors and normal prostate tissues. They identified a total of 104 isoforms characteristic of prostate tumors, and showed that these isoforms could classify tumors from normal tissues at ~90% accuracy. A Support Vector Machine (SVM) classifier combining overall transcript abundance and isoform distributions correctly classified 92% of the samples, compared to an accuracy of 87% if the same classifier was used on overall transcript abundance alone (69). These studies demonstrate that splicing microarray

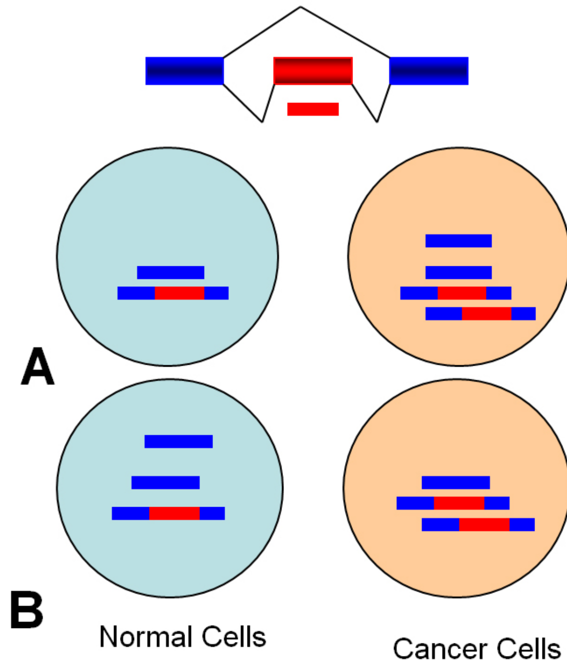

Figure 3. Distinguishing changes in overall transcript abundance and changes in splicing. Consider a probe against the middle cassette exon (in red). (A) There is a two-fold upregulation in the intensity of the probe in cancer cells compared to normal cells. It is caused by a two-fold upregulation in transcription without any change in splicing.(B) There is a two-fold upregulation in the intensity of the probe in cancer cells compared to normal cells. It is caused by an increased percentage of exon-inclusion transcripts in cancer cells (33% to 66%) without any change in the overall mRNA abundance.

technology is an efficient approach to identify isoform biomarkers for particular types of tumors. The information about isoform distributions provides a more refined "molecular signature" for better classifications of clinical tumor samples.

6. FUTURE CHALLENGES

Despite these exciting successes, there are still many unsolved problems and challenges. Compared to conventional gene expression microarrays, splicing microarrays pose a far more complex computational problem. Probe intensities on a splicing microarray are affected by both changes in transcription and changes in splicing (see Figure 3). Moreover, as each exon typically has a small number of probes (for example, the Affymetrix Exon Array platform usually has 4 probes for each exon), variations in probe affinities and cross-hybridization can further complicate data analysis and interpretation. It is a challenging task to design robust algorithms to deconvolute overall gene expression and proportions of individual splice forms. Another key computational problem is how to predict the likely biological function of detected splice variants for prioritization of experimental studies. Since microarrays only generate information about alternative splicing of individual exons, it is important to have algorithms that reconstruct full-length isoforms from splicing microarray data, similar to those developed for EST-based analyses of alternative splicing (71).

It is also critical to understand the biological roles of cancer-specific alternative splicing events discovered by genomic analyses. Functional studies are needed to establish causal relationships between alternative splicing of individual genes and the initiation and progression of certain types of tumors. While it is possible that many cancer-specific splice forms are simply by-products of abnormal splicing regulation in cancer cells, there are well-documented alternative splicing events in cancer-related genes, which might play a direct role in tumorigenesis (see (12,13)). For example, disruption of *BRCA1* is a common and important event in breast and ovarian cancers. A number of mutations in *BRCA1* affect splicing and produce non-functional proteins (72-74). In fact, recent data suggest that as high as 60% of human disease-causing mutations are splicing mutations (75). Combining large-scale mutation screening and whole-transcriptome splicing profiling in cancer cells might shed new light on the biology and clinical treatment of cancers.

7. MEDICAL IMPLICATIONS

The findings from genome-wide analyses of cancer-specific alternative splicing have important medical implications. The discovery of cancer-specific splice forms provides potential cancer biomarkers. Clinical tests using RT-PCR, microarray or antibody-based detection methods can be developed for diagnosis of certain cancers based on observed splicing patterns in the clinical samples (47), as shown successfully by the study of Li and colleagues (68). Alternative splicing events of important cancer-related genes are promising drug targets. For example, alternative splicing of Bcl-x has been the target of several oligonucleotide-based studies which aim to increase the ratio of the proapoptotic isoform (Bcl-xS) over the anti-apoptotic isoform (Bcl-xL) (76,77). There are a variety of therapeutic strategies developed for targeting disease-causing alternative splicing, such as protein isoform specific inhibitors, isoform-specific RNAi, oligonucleotide mediated "re-programming" of aberrant splicing, etc (for details, see the review by Garcia-Blanco and colleagues (78)). These methods might be used for clinical treatment of cancers.

8. ACKNOWLEDGEMENTS

I wish to thank Drs. Christopher Lee, Meenakshi Roy and Qiang Xu for discussions.

9. REFERENCES

1. Liang, F., I. Holt, G. Pertea, S. Karamycheva, S. L. Salzberg & J. Quackenbush: Gene Index analysis of the human genome estimates approximately 120,000 genes. *Nature Genet.* 25, 239-240 (2000)

2. Consortium., I. H. G. S.: Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001)
3. Lee, C.: The incredible shrinking human genome. *Trends in Genetics* 17, 187-188 (2001)
4. Modrek, B. & C. Lee: A genomic view of alternative splicing. *Nature Genet.* 30, 13-9 (2002)
5. Graveley, B. R.: Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17, 100-7 (2001)
6. Lareau, L. F., R. E. Green, R. S. Bhatnagar & S. E. Brenner: The evolving roles of alternative splicing. *Curr Opin Struct Biol* 14, 273-82 (2004)
7. Fettiplace, R. & P. A. Fuchs: Mechanisms of hair cell tuning. *Annu Rev Physiol* 61, 809-34 (1999)
8. Xu, X., D. Yang, J. H. Ding, W. Wang, P. H. Chu, N. D. Dalton, H. Y. Wang, J. R. Bermingham, Jr., Z. Ye, F. Liu, M. G. Rosenfeld, J. L. Manley, J. Ross, Jr., J. Chen, R. P. Xiao, H. Cheng & X. D. Fu: ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* 120, 59-72 (2005)
9. Boise, L. H., M. González-García, C. E. Postema, L. Ding, T. Lindsten, L. A. Turka, X. Mao, G. Nuñez & C. B. Thompson: bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* 74, 597-608 (1993)
10. Faustino, N. A. & T. A. Cooper: Pre-mRNA splicing and human disease. *Genes Dev* 17, 419-37 (2003)
11. Bracco, L. & J. Kearsey: The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol* 21, 346-53 (2003)
12. Venables, J. P.: Unbalanced alternative splicing and its significance in cancer. *Bioessays* 28, 378-86 (2006)
13. Srebrow, A. & A. R. Kornblihtt: The connection between splicing and cancer. *J Cell Sci* 119, 2635-41 (2006)
14. Black, D. L.: Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291-336 (2003)
15. Schmucker, D., J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon & S. L. Zipursky: *Drosophila*Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671-684 (2000)
16. Dong, Y., H. E. Taylor & G. Dimopoulos: AgDscam, a Hypervariable Immunoglobulin Domain-Containing Receptor of theAnopheles gambiae Innate Immune System. *PLoS Biol* 4, e229 (2006)
17. Black, D. L.: Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367-70 (2000)
18. Lazrak, A., Z. Liu & C. L. Huang: Antagonistic regulation of ROMK by long and kidney-specific WNK1 isoforms. *Proc Natl Acad Sci U S A* 103, 1615-20 (2006)
19. Xu, Q., B. Modrek & C. Lee: Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 30, 3754-66 (2002)
20. Xing, Y., Q. Xu & C. Lee: Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett* 555, 572-8 (2003)
21. Cline, M. S., R. Shigeta, R. L. Wheeler, M. A. Siani-Rose, D. Kulp & A. E. Loraine: The effects of alternative splicing on transmembrane proteins in the mouse genome. *Pac Symp Biocomput* 17-28 (2004)
22. Liu, S. & R. B. Altman: Large scale study of protein domain distribution in the context of alternative splicing. *Nucleic Acids Res* 31, 4828-35 (2003)
23. Resch, A., Y. Xing, B. Modrek, M. Gorlick, R. Riley & C. Lee: Assessing the impact of alternative splicing on domain interactions in the human proteome. *J. Proteome Res.* 3, 76-83 (2004)
24. Romero, P. R., S. Zaidi, Y. Y. Fang, V. N. Uversky, P. Radivojac, C. J. Oldfield, M. S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic & A. K. Dunker: Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A* 103, 8390-5 (2006)
25. Wang, P., B. Yan, J. T. Guo, C. Hicks & Y. Xu: Structural genomics analysis of alternative splicing and application to isoform structure modeling. *Proc Natl Acad Sci U S A* 102, 18920-5 (2005)
26. Nakao, M., R. A. Barrero, Y. Mukai, C. Motono, M. Suwa & K. Nakai: Large-scale analysis of human alternative protein isoforms: pattern classification and correlation with subcellular localization signals. *Nucleic Acids Res* 33, 2355-63 (2005)
27. Berget, S. M. & P. A. Sharp: A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven Symp Biol* 332-44 (1977)
28. Chow, L. T., R. E. Gelinas, T. R. Broker & R. J. Roberts: An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8 (1977)
29. Mironov, A. A., J. W. Fickett & M. S. Gelfand: Frequent alternative splicing of human genes. *Genome Res.* 9, 1288-1293 (1999)
30. Kan, Z., E. C. Rouchka, W. R. Gish & D. J. States: Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.* 11, 889-900 (2001)
31. Kan, Z., D. States & W. Gish: Selecting for Functional Alternative Splices in ESTs. *Genome Res* 12, 1837-45 (2002)
32. Clark, F. & T. A. Thanaraj: Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum Mol Genet* 11, 451-64 (2002)
33. Brett, D., J. Hanke, G. Lehmann, S. Haase, S. Delbruck, S. Krueger, J. Reich & P. Bork: EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Letters* 474, 83-86 (2000)
34. Modrek, B., A. Resch, C. Grasso & C. Lee: Genome-wide analysis of alternative splicing using human expressed sequence data. *Nucleic Acids Res.* 29, 2850-9 (2001)
35. Wang, B. B. & V. Brendel: The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol* 5, R102 (2004)
36. Lee, B. T., T. W. Tan & S. Ranganathan: DEDB: a database of Drosophila melanogaster exons in splicing graph form. *BMC Bioinformatics* 5, 189 (2004)
37. Sugnet, C. W., W. J. Kent, M. Ares, Jr. & D. Haussler: Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput* 66-77 (2004)

38. Leipzig, J., P. Pevzner & S. Heber: The Alternative Splicing Gallery (ASG): bridging the gap between genome and transcriptome. *Nucleic Acids Res* 32, 3977-83 (2004)
39. Eyras, E., M. Caccamo, V. Curwen & M. Clamp: ESTGenes: alternative splicing from ESTs in Ensembl. *Genome Res* 14, 976-87 (2004)
40. Croft, L., S. Schandorff, F. Clark, K. Burrage, P. Arctander & J. S. Mattick: ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nature Genet.* 24, 340-1 (2000)
41. Gupta, S., D. Zink, B. Korn, M. Vingron & S. A. Haas: Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. *BMC Genomics* 5, 72 (2004)
42. Xu, Q. & C. Lee: Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res* 31, 5635-5643 (2003)
43. Wang, Z., H. S. Lo, H. Yang, S. Gere, Y. Hu, K. H. Buetow & M. P. Lee: Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res* 63, 655-7 (2003)
44. Hui, L., X. Zhang, X. Wu, Z. Lin, Q. Wang, Y. Li & G. Hu: Identification of alternatively spliced mRNA variants related to cancers by genome-wide ESTs alignment. *Oncogene* 23, 3013-23 (2004)
45. Kirschbaum-Slager, N., R. B. Parmigiani, A. A. Camargo & S. J. de Souza: Identification of human exons overexpressed in tumors through the use of genome and expressed sequence data. *Physiol Genomics* 21, 423-32 (2005)
46. Roy, M., Q. Xu & C. Lee: 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, 5026-33 (2005)
47. Brinkman, B. M.: Splice variants as cancer biomarkers. *Clin Biochem* 37, 584-94 (2004)
48. Atanelov, L., Q. Xu, R. Rad & C. Lee: Targeting alternatively spliced sequence features for cancer diagnosis and therapeutics. *J Gastroenterol* 40 Suppl 16, 14-20 (2005)
49. Sorek, R., O. Basechess & H. M. Safer: Expressed sequence tags: clean before using. Correspondence re: Z. Wang et al., computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. Cancer Res., 63: 655-657, 2003. *Cancer Res* 63, 6996; author reply 6996-7 (2003)
50. Lee, C. & M. Roy: Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol* 5, 231 (2004)
51. Hoheisel, J. D.: Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 7, 200-10 (2006)
52. Li, C. & W. H. Wong: Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98, 31-6 (2001)
53. Wang, H., E. Hubbell, J. S. Hu, G. Mei, M. Cline, G. Lu, T. Clark, M. A. Siani-Rose, M. Ares, D. C. Kulp & D. Haussler: Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* 19 Suppl 1, i315-22 (2003)
54. Castle, J., P. Garrett-Engele, C. D. Armour, S. J. Duenwald, P. M. Loerch, M. R. Meyer, E. E. Schadt, R. Stoughton, M. L. Parrish, D. D. Shoemaker & J. M. Johnson: Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol* 4, R66 (2003)
55. Yeakley, J. M., J. B. Fan, D. Doucet, L. Luo, E. Wickham, Z. Ye, M. S. Chee & X. D. Fu: Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol* 20, 353-8 (2002)
56. Neves, G., J. Zucker, M. Daly & A. Chess: Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* 36, 240-6 (2004)
57. Clark, T. A., C. W. Sugnet & M. J. Ares: Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907-910 (2002)
58. Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton & D. D. Shoemaker: Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141-4 (2003)
59. Pan, Q., O. Shai, C. Misquitta, W. Zhang, A. L. Saltzman, N. Mohammad, T. Babak, H. Siu, T. R. Hughes, Q. D. Morris, B. J. Frey & B. J. Blencowe: Revealing global regulatory features of Mammalian alternative splicing using a quantitative microarray platform. *Mol Cell* 16, 929-41 (2004)
60. Shai, O., Q. D. Morris, B. J. Blencowe & B. J. Frey: Inferring global levels of alternative splicing isoforms using a generative model of microarray data. *Bioinformatics* 22, 606-13 (2006)
61. Le, K., K. Mitsouras, M. Roy, Q. Wang, Q. Xu, S. F. Nelson & C. Lee: Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. *Nucleic Acids Res* 32, e180 (2004)
62. Fehlbaum, P., C. Guihal, L. Bracco & O. Cochet: A microarray configuration to quantify expression levels and relative abundance of splice variants. *Nucleic Acids Res* 33, e47 (2005)
63. Ule, J., A. Ule, J. Spencer, A. Williams, J. S. Hu, M. Cline, H. Wang, T. Clark, C. Fraser, M. Ruggiu, B. R. Zeeberg, D. Kane, J. N. Weinstein, J. Blume & R. B. Darnell: Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* 37, 844-52 (2005)
64. Blanchette, M., R. E. Green, S. E. Brenner & D. C. Rio: Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* 19, 1306-14 (2005)
65. Sugnet, C. W., K. Srinivasan, T. A. Clark, G. O'Brien, M. S. Cline, H. Wang, A. Williams, D. Kulp, J. E. Blume, D. Haussler & M. Ares: Unusual Intron Conservation near Tissue-Regulated Exons Found by Splicing Microarrays. *PLoS Comput Biol* 2, e4 (2006)
66. Pan, Q., A. L. Saltzman, Y. K. Kim, C. Misquitta, O. Shai, L. E. Maquat, B. J. Frey & B. J. Blencowe: Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev* 20, 153-8 (2006)
67. Relogio, A., C. Ben-Dov, M. Baum, M. Ruggiu, C. Gemund, V. Benes, R. B. Darnell & J. Valcarcel: Alternative splicing microarrays reveal functional

expression of neuron-specific regulators in Hodgkin lymphoma cells. *J Biol Chem* 280, 4779-84 (2005)
68. Li, H. R., J. Wang-Rodriguez, T. M. Nair, J. M. Yeakley, Y. S. Kwon, M. Bibikova, C. Zheng, L. Zhou, K. Zhang, T. Downs, X. D. Fu & J. B. Fan: Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. *Cancer Res* 66, 4079-88 (2006)
69. Zhang, C., H. R. Li, J. B. Fan, J. Wang-Rodriguez, T. Downs, X. D. Fu & M. Q. Zhang: Profiling alternatively spliced mRNA isoforms for prostate cancer classification. *BMC Bioinformatics* 7, 202 (2006)
70. Li, C., M. Kato, L. Shiue, J. E. Shively, M. Ares, Jr. & R. J. Lin: Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. *Cancer Res* 66, 1990-9 (2006)
71. Lee, C. & Q. Wang: Bioinformatics analysis of alternative splicing. *Brief Bioinform* 6, 23-33 (2005)
72. Mazoyer, S., N. Puget, L. Perrin-Vidoz, H. T. Lynch, O. M. Serova-Sinilnikova & G. M. Lenoir: A BRCA1 nonsense mutation causes exon skipping. *Am J Hum Genet* 62, 713-5 (1998)
73. Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang & A. R. Krainer: ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31, 3568-71 (2003)
74. Pettigrew, C., N. Wayte, P. K. Lovelock, S. V. Tavtigian, G. Chenevix-Trench, A. B. Spurdle & M. A. Brown: Evolutionary conservation analysis increases the colocalization of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms. *Breast Cancer Res* 7, R929-39 (2005)
75. Xing, Y. & C. Lee: Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes. *Nat Rev Genet* 7, 499-509 (2006)
76. Mercatante, D. R., P. Sazani & R. Kole: Modification of alternative splicing by antisense oligonucleotides as a potential chemotherapy for cancer and other diseases. *Curr Cancer Drug Targets* 1, 211-30 (2001)
77. Villemaire, J., I. Dion, S. A. Elela & B. Chabot: Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J Biol Chem* 278, 50031-9 (2003)
78. Garcia-Blanco, M. A., A. P. Baraniak & E. L. Lasda: Alternative splicing in disease and therapy. *Nat Biotechnol* 22, 535-46 (2004)

Key Words: Alternative Splicing, Microarray, Cancer, Genomics, Bioinformatics, Review

Send correspondence to: Dr Yi Xing, Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA, Tel: 310-904-2162, Fax: 319-356-8608, E-mail: yi-xing@uiowa.edu

http://www.bioscience.org/current/vol12.htm