

Discovery of Tumor-Specific Alternative Splicing Sites*

Fang Rong Hsu¹, Chia Yang Cheng²

Keywords: alternative splicing, tumor-specific

1 Introduction.

The splicing mechanism recognizes the boundaries between exons and introns. Alternative splicing can be a mechanism to generate more than one mRNA and protein from a single gene. Alternative RNA splicing functions in two ways. One is turn on/off control of gene expression; the other is in the formation of multiple protein isoforms [1]. It has been estimated that 35%-59% of human genes have more than one RNA isoforms [2].

Alternative splicing can display a family of different protein in different tissues [1]. In the other hand, cancer associated splice variants have been reported for many genes [3]. ~15% of diseases-causing mutations in human genes involve misregulation of alternative splicing and errors in mRNA processing have been associated with cancer and other human diseases. [4]

Currently, many alternative splicing databases were developed [5, 7], and some detected tissue-specific and tumor-specific alternative splicing sites [2, 3]. However, after being sifted by clustering and aligning to genomic sequence, few EST sequence could be used to find tumor-specific sites for each tissue.

AVATAR, a value added transcriptome data base, which was developed by aligning ESTs to genomic sequence directly [7, 8], offer more rich resource to analysis EST express in many tissues at each alternative splicing sites. In this study, we present a method to detect the tumor-specific alternative splicing isoforms from certain tissue.

2 Materials and Methods

Three steps were present to detect tumor-specific alternative splicing sites. First, 14,099 human alternative splicing sites were collected from AVATAR, which exons, introns and alternative splicing sites were identified by aligning five million ESTs [9] to human genomic sequence (Build 31).

Second, five million ESTs from 8,431 libraries were categorized into 45 tissues and three types of histology, normal, tumor and unknown [6, 10]. ESTs were divided into four pools by tissue and histology (isoform 1 and tumor, isoform 2 and tumor, isoform 1 and normal, and isoform 2 and normal) at each alternative splicing site.

Third, we calculated theses data by Fisher's exact test, divided left tail of P-value by right tail of P-value as confidence C. The splicing sites with certain tissue, which Log odd ratio of C was greater than 2, were suggested as tumor-specific alternative splicing sites.

3 Result

We found 20 genes which's LOD greater than 2. For each gene, most ESTs dates in the four pools were from more than one library. 40% of tumor-specific isoforms were happened in Brain, and 50% were 5' type alternative splicing (see table 1). Only three alternative splicing sites were also

¹ Department of Bioinformatics, Taichung Healthcare and Management University, Taichung, Taiwan. E-mail: frshu@thmu.edu.tw

² Department of Bioinformatics, Taichung Healthcare and Management University, Taichung, Taiwan. E-mail: std91242006@ms1.thmu.edu.tw

* support in part by NSC 92-3112-B-468-001 and NSC 91-2662-E-468-001-CC3, Taiwan

tissues-specific sites. GNAS complex locus (GNAS), LOD score is 2.09, located on chromosome 20, is one of the three genes which's alternative splicing isoforms were published in literatures.

type	No. of tumor specific sites	Tissue	No. of tumor specific sites
cassette	3	Brain	8
3'	7	Lung	4
5'	10	Skin	3

Table 1: Tissue and types distribution of human tumor-specific alternative splicing sites

GNAS gene structure has been well investigated [11]; *Gsa* is encoded by exons 1-13 of GNAS. Exon 2-13 are also included in two additional overlapping transcripts, *XL α S* and *NESP55*, each with a distinct first exon. (Fig. 1) In brain, exon 3 skipping isoform (S) and exon 3 including isoform (S') were both expressed in tumor and normal EST pools. In S, tumor ESTs (7) are expressed more than normal ones (4), but oppositely tumor ESTs (5) were less than normal ESTs (12) in S'. (Fig. 1) We could ratiocinate that GNAS exon 3 might associated with cancer, though, reversed from which one of the three mRNAs the ESTs were unknown. We have good ground for thinking that it is worth to validate these 20 tumor-specific alternative splicing sites by traditional experiment.

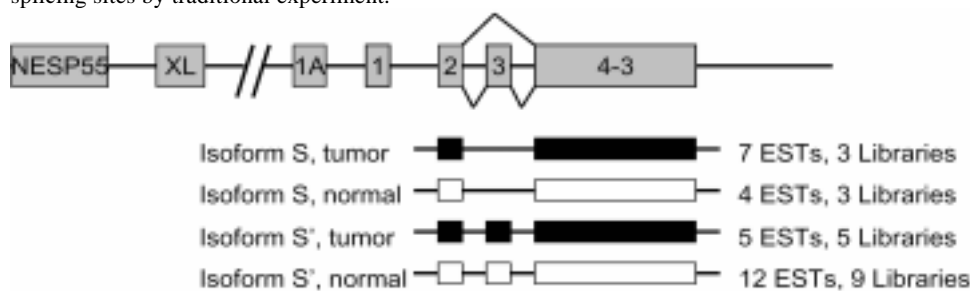


Figure 1: GNAS gene structure and EST expression in brain.

References

- [1] A. R. Krainer 1997. Eukaryotic mRNA Processing, Oxford Univ., pp.242-279
- [10] Cancer Genome Anatomy Project, <http://www.ncbi.nlm.nih.gov/ncicgap/>
- [11] CancerGene Card GNAS1, <http://caroll.vjf.cnrs.fr/cancergene/CG46.html>
- [7] F. R. Hsu, Ying Tsong Chen, Hwan-You Chang, Yaw-Lin Lin, Yin-Te Tsai, Hui-Ling Peng, Min Yao Shih, Chia-Hung Liu, and Jer Feng Chen, 2003, Proc. Of the 6th Conference on Engineering Technology and Chinese/Western Medicine Applications, Sep. 13, Taichung Taiwan, pp. 196-201.
- [8] F. R. Hsu and J. F. Chen 2003, Proc. Of the IEEE Computational Systems Bioinformatics Conference (CSB2003), Aug., 11-14, San Francisco, USA, pp. 564-566.
- [5] Lee, C., Atanelov, L., Modrek, B., Xing, Y., 2003, Nucleic Acids Res. 31: 101-5.
- [6] Library Browser. <http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=9606>
- [9] NCBI dbEST, <http://www.ncbi.nlm.nih.gov/dbEST>
- [4] Philips, A.V. and Cooper, T. A., 2000, Cell. Mol. Life Sci., 57, 235-249
- [2] Xu, Q., Modrek, B., Lee, C., 2002, Nucleic Acids Res. 30: 3754-66
- [3] Xu, Q., Lee, C., 2003, Nucleic Acids Res. 31: 5635-5643