

Mining Colon Cancer Specific Alternative Splicing in EST Database

Tien-Hsiung Ku M.D., Fang Rong Hsu Ph.D.*

Department of Anesthesia, Changhua Christian Hospital, Department of Information Engineering and Computer Science, Feng Chia University*, Taiwan

Abstract:

Among 75218 splicing sites, 137 colon cancer specific alternative splicing isoforms were found by mining EST database. Alternative splicing database were first constructed by aligning EST to genomic sequence. Numbers of ESTs from normal or cancer colon tissue supporting splicing isoform at each splicing site were then queried and analyzed with Fisher exact test. There were 53 3' splicing, 42 5'splicing, 40 exon skipping and 2 mutual exclusive cancer specific splicing isoforms.

INTRODUCTION

RNA alternative splicing is important for protein production in eukaryotes. Aberration in splicing may produce protein that could be relative to disease. Alternative isoform of SRF is expressed in colon cancer¹. Apobec-1 isoform is also over-expressed in colon cancer cell line. Putative alternative splicing had been derived from EST databases using computational methods. We demonstrate here the colon cancer specific alternative splicing isoforms found by mining human expressed sequence tags (EST) database.

METHODS

EST, genomic sequence and EST library information were obtained from NCBI dbEST, genbank, and uniLib ftp site. There were over 5 million EST entities with information about sequence source. Different diseases were coded by a medical professional manually on the library data file according to disease information. Tissues from colon cancer or normal patients were coded. Since each EST sequence had annotation information about which tissue library it came from, we could obtain the disease status code at EST level via this annotation link. All alternative splicing sites and patterns were found by aligning EST fragment to the genome sequence using multi-layer unique marker alignment method². Splicing isoform patterns were calculated using graph model. Alternative splicing sites and patterns were then clustered if each

alternative splicing site had more than two isoforms.

All EST that support a alternative splicing site were queried. Colon cancer specific or normal tissue alternative splicing sites were found by querying the EST number supporting the specific alternative splicing isoform by the disease state of the EST. Hypothesis is tested that if the frequency between isoform A and isoform B for normal and schizophrenia tissue was the same. Significance was tested by the Fisher exact test. $P < 0.05$ for the two-tailed test was considered significant.

RESULTS

36468 alternative 3' splicing, 24499 alternative 5' splicing, 12906 exon skipping and 1345 mutual exclusion splicing sites were obtained. Among them, there were 53 3' splicing, 42 5'splicing, 40 exon skipping and 2 mutual exclusion colon cancer specific splicing isoforms. Part of these alternative splicing sites are listed in table 1.

Table 1. Information about some of these colon cancer specific alternative splicing sites

chr: chromosome. 3s: 3' splicing. 5s: 5' splicing. cs: exon skippin. me: mutual exclusion.

AS	gene	Chr	contig	position
3s	RPL18	19	29800594	21388871
3s	RPLP0	12	37544143	11155285
5s	TACSTD1	2	37547123	26418375
5s	UQCRH	1	37547124	8339063
cs	LTA4H	12	29802923	19879034
cs	RPLP0	12	37544143	11154008
me	PKM2	15	37540936	43283316

REFERENCES

- [1] Patten LC, Belaguli NS, Baek MJ, Fagan SP, Awad SS, Berger DH. Serum response factor is alternatively spliced in human colon cancer. *J Surg Res.* 2004 Sep;121(1):92-100.
- [2] Hsu FR, Chen JF. Aligning ESTs to genome using multi-layer unique markers. *Proc. Of the IEEE Computational Systems Bioinformatics Conference (CSB2003).* San Francisco, USA. 2003 Aug. p. 564-566.