

# The use of expression data to study the association between alternative splicing and cancer

Natanja Kirschbaum - Slager<sup>1</sup>, Raphael Parmigiani<sup>2</sup>, Graziela M. P. Lopes<sup>2</sup>, Pedro A. F. Galante<sup>3</sup>, Gregory J. Riggins<sup>4</sup>, Anamaria A. Camargo<sup>2</sup>, Sandro J. de Souza<sup>5</sup>

**Keywords:** transcriptome, alternative splicing, tumor, splicing factors

Alternative splicing is one of the major sources of the large transcriptional diversity found in human cells [1]. Splicing variants have been shown to be associated with features like spreading and progression in several human tumors [2]. Therefore understanding the association between alternative splicing and tumors may be of great importance for the generation of both diagnostic and therapeutic tools. In this study we analyzed different types of expression data in order to find tumor associated splicing variants and to investigate the expression pattern of splicing factors that might cause the appearance of such tumor associated expression of splicing variants.

We screened the genome for exons over-expressed in tumors of specific tissues by implying strict criteria and statistical analyses of EST data. We then performed a Serial Analysis of Gene Expression (SAGE) analysis excluding exons belonging to genes that are up-regulated in tumors. This allowed us to predict the over-expression of single exons in specific tumors. Our final group of candidates includes 1386 exons belonging to 638 genes. Experimental validation of a few candidates in normal tissue, tumor cell lines and patient samples suggests that most of these candidates are indeed tumor associated exons.

Having shown the existence of *bona fide* tumor associated splicing variants, we then used proteomics, SAGE and microarray data to further explore the association between regulation of alternative splicing and tumorigenesis. We analyzed the expression pattern of splicing factors, shown to be associated with the spliceosome, in both normal and tumor tissues. Virtual SAGE tags were assigned to 145 genes, encoding proteins involved in splicing as shown by a proteomics analysis [3]. The expression pattern of the genes was assessed by analyzing their tag counts in normal and tumor SAGE libraries. The number of differentially expressed (equal or higher than three fold difference) genes for brain, colon breast and prostate, were 48 (33%), 58 (40%), 46 (32%) and 43 (30%), respectively ( $p < 0.05$  as evaluated by 1000 simulations of 145 randomly taken genes). The expression pattern of the same genes was also analyzed using a large data set of microarray experiments available in the Oncomine database [4]. In brain, colon, breast and prostate 48(41%), 58(50%), 70 (60%) and 16 (14%) genes were differentially expressed, respectively. Both the approaches showed that, on average, most of the differentially expressed genes were over-expressed in tumor in at least one of the analyzed tissues. Tumor over-expression of splicing factors

---

<sup>1</sup> Ludwig Institute for Cancer Research, Sao Paulo, Brazil, E-mail: natanja@compbio.ludwig.org.br

<sup>2</sup> Ludwig Institute for Cancer Research, Sao Paulo, Brazil

<sup>3</sup> Ludwig Institute for Cancer Research, Sao Paulo, Brazil. Ph.D Program, Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Sao Paulo, Brazil.

<sup>4</sup> John Hopkins University, School of Medicine, Baltimore, MD, USA.

<sup>5</sup> Ludwig Institute for Cancer Research, Sao Paulo, Brazil, E-mail: sandro@compbio.ludwig.org.br

may support the notion of a broader association between alternative splicing and cell transformation and suggests that splicing factors may be involved in oncogenic pathways.

## References

[2] Caballero, O.L., de Souza, S.J., Brentani, R.R. and Simpson, A.J. 2001. Alternative spliced transcripts as cancer markers. *Dis.Markers*. 17:67-75.

[1] Modrek, B., and Lee, C. 2002. A genomic view of alternative splicing. *Nature Genetics* 30: 13-19.

[4] Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A.M. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 6: 1-6.

[3] Zhou, Z., Licklider, L.J. Gygi, S.P. and Reed, R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature*. 419:182-185.

[supported by FAPESP]