

Splicing factors are differentially expressed in tumors

Natanja Kirschbaum-Slager¹, Graziela M.P. Lopes¹,
Pedro A.F. Galante^{1,2}, Gregory J. Riggins³ and Sandro J. de Souza¹

¹Ludwig Institute for Cancer Research, São Paulo Branch,
São Paulo, SP, Brazil

²Ph.D. Program, Departamento de Bioquímica, Instituto de Química,
Universidade de São Paulo, São Paulo, SP, Brazil

³John Hopkins University, School of Medicine, Baltimore, MD, USA

Corresponding author: S.J. de Souza

E-mail: sandro@compbio.ludwig.org.br

Genet. Mol. Res. 3 (4): 512-520 (2004)

Received October 4, 2004

Accepted December 16, 2004

Published December 30, 2004

ABSTRACT. Although alternative splicing of many genes has been found associated with different stages of tumorigenesis and splicing variants have been characterized as tumor markers, it is still not known whether these examples are sporadic or whether there is a broader association between the two phenomena. In this report we evaluated, through a bioinformatics approach, the expression of splicing factors in both normal and tumor tissues. This was possible by integrating data produced by proteomics, serial analysis of gene expression (SAGE) and microarray experiments. We observed a significant shift in the expression of splicing factors in tumors in both SAGE and microarray data, resulting from a large amount of experiments. We discuss that this supports the notion of a broader association between alternative splicing and cell transformation, and that splicing factors may be involved in oncogenic pathways.

Key words: Alternative splicing, Tumorigenesis, SAGE

INTRODUCTION

Most human genes undergo a post-transcriptional process, called splicing, that is responsible for the excision of introns from unprocessed messages (Berget et al., 1977; Chow et al., 1977). In many cases, more than one mature mRNA is produced by the alternative choice of splicing sites by the splicing machinery (including the spliceosome and various serine-arginine (SR) proteins). The biological significance of alternative splicing is well represented in many examples, including sex determination in *Drosophila* and sound recognition in complex vertebrates (see Modrek and Lee, 2002; Black, 2003). It is quite likely that most human genes undergo alternative splicing with consequences for almost all aspects of cellular physiology.

Recent reports have suggested a widespread occurrence of new splicing variants in tumors (Wang et al., 2003; Xu and Lee, 2003). Splicing variants from such genes, including *CD44*, *WT1*, *cd79b*, *bin1*, and *Syk* have also been shown to be associated with different aspects of tumorigenesis (Ge et al., 1999; Naor et al., 2002; Baudry et al., 2002; Cragg et al., 2002; Wang et al., 2003). Therefore, there is a growing body of evidence suggesting that splicing variants can be used as tumor markers and that alternative splicing can accompany the process of tumorigenesis (Lee and Feinberg, 1997; Caballero et al., 2001; Adams et al., 2002). The cell surface protein *CD44*, for example, exhibits a large number of splicing variants that seem to be associated with the progression of certain tumors to an invasive state (see Naor et al., 2002). Splicing variants of the gene *WT1* seem to be associated as well with different aspects of tumorigenesis (Menke et al., 1997). It has been suggested that a balance between two variants, with and without *WT1* exon 5, affects several aspects of cell physiology in Wilm's tumor (Baudry et al., 2002). It is still unknown, however, whether these examples are sporadic or if there is a broader and more significant association between alternative splicing and tumorigenesis.

We investigated this possible association between alternative splicing regulation and tumorigenesis by performing a large-scale *in silico* analysis of the expression profiles of 145 genes encoding proteins shown to be involved in the splicing process. Expression data from tumor and normal samples derived from colon, brain, breast, and prostate tissues were used. This was possible by making use of collections of data produced by proteomics, serial analysis of gene expression (SAGE) and microarray analyses. We observed a significant shift in the expression of splicing factors in tumors. This supports the notion that there is a broader association between alternative splicing and cell transformation, and that splicing factors may be involved in oncogenic pathways.

MATERIAL AND METHODS

Gene accession number assignment

The supplementary data from Zhou et al., 2002 were downloaded, and for each splicing factor a corresponding cDNA sequence was manually retrieved from GenBank. The division into 17 functional ontology groups was maintained.

SAGE tag assignment

A virtual SAGE tag corresponds to the 10-bp sequence immediately downstream of the

3' most *Nla*III site of a given transcript (Boon et al., 2002). One representative cDNA sequence was selected for each splicing factor. A full insert mRNA was chosen that showed either a poly A signal and/or a poly A tail. These sequences were then assigned a virtual SAGE tag. The genes that did not fit one of the above criteria were excluded from further SAGE.

SAGE expression analysis

The “virtual” tags were used to query all SAGE libraries from brain, breast, colon, and prostate. Throughout this report we refer to all libraries generated from either patient tumor samples or from tumor cell lines as “tumor”. Frequencies of all tags were normalized (1 tag per 200,000) to take into account differences in library sizes. Genes were defined as over-expressed in either tumor or normal tissue when the difference in tag count was at least three-fold.

Simulation data sets

For the *in silico* simulation, 1000 lists of 145 tags were randomly generated from 17,795 different full insert mRNAs, as described above. A virtual SAGE tag was assigned to all genes in all lists. The number of differentially expressed genes was calculated in each of the four tissues mentioned above for each list of 145 tags.

Microarray expression analysis

The Oncomine database (www.oncomine.org) was manually queried for all 145 genes using their gene names. Gene aliases were used when needed. Only studies comparing normal against tumor tissues were queried and those genes not appearing in these categories were defined as ‘not informative’. The logarithmic average of the log scores of all values in both normal and tumor tissues was calculated for those genes for which more than one experiment was performed in the same tissue. Throughout this report we refer to all libraries generated from either patient tumor samples or from tumor cell lines as “tumor”. Genes were considered differentially expressed when the average value of all normal or tumor tissue experiments was at least three-fold its counterpart.

RESULTS

The observation that different splicing variants are associated with distinct features in different types of tumors led us to investigate the expression pattern of splicing factors. Differential expression of splicing factors in tumors could cause changes in the expression of new variants, when compared to their normal tissue counterparts.

We tested the hypothesis that variations in the expression of components of the splicing machinery are associated with tumorigenesis by using three datasets made publicly available to the scientific community. First, a list of 145 different splicing factors was derived from a proteomics analysis of functional purified human spliceosomes (Zhou et al., 2002). The original division of the 145 genes into 17 functional ontology groups was maintained.

A representative cDNA sequence was retrieved from GenBank for each splicing factor. Then a “virtual” SAGE tag for each cDNA sequence was obtained by scanning the se-

quence for a poly A signal and/or tail. This strategy was recently used by our groups to build SAGE Genie (Boon et al., 2002), and it has generated a reliable set of gene-to-tag and tag-to-gene assignments. Ten genes were excluded from further SAGE analysis, as they did not fit all criteria for tag assignment. The collection of “virtual” SAGE tags was then searched against all publicly available SAGE libraries derived from brain, breast, colon, and prostate. Figure 1 is a graphical view of the tumor-normal ratio of differential expression for a subset of the splicing factors (the whole set of data is available in Supplementary Table 1).

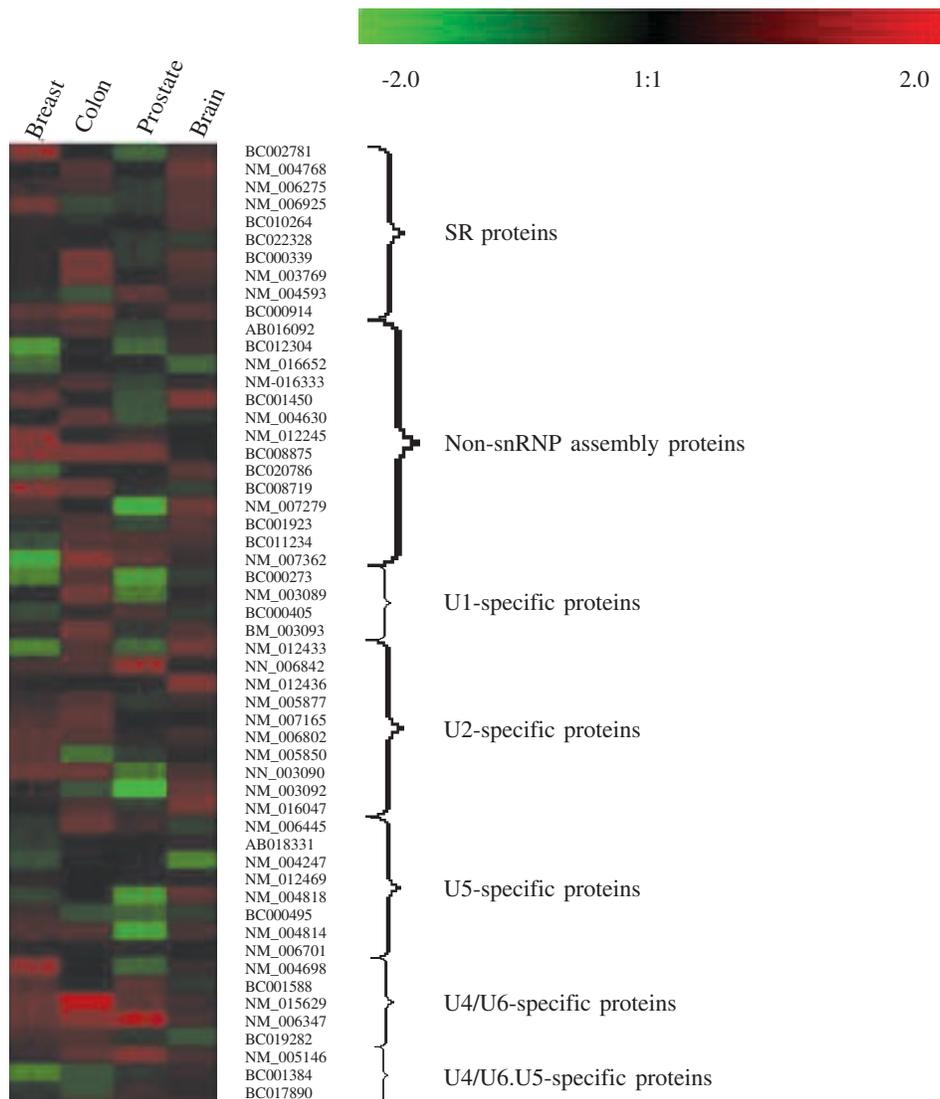


Figure 1. The tumor/normal ratio derived from serial analysis of gene expression (SAGE) for a sample of splicing factors. Each rectangle represents the tumor/normal (T/N) ratio for one gene in one tissue. Virtual SAGE tags were extracted from the corresponding cDNAs retrieved from GenBank according to Boon et al. (2002) (see Material and Methods). Frequencies of all tags were normalized (1 tag per 200,000) to take into account differences in library size. Columns correspond to the different tissues and rows correspond to genes classified according to Zhou et al. (2002). SR = serine/arginine proteins; snRNP = small nuclear ribonucleoproteins.

Supplementary Table 1 is available at <http://www.funpecrp.com.br/gmr/year2004/vol4-3/pdf/icob07st01.pdf>.

A significant fraction of all splicing factors showed a different level of expression (>3-fold difference) between normal tissues and their tumor counterparts. The numbers of differentially expressed genes for breast, prostate, brain, and colon were 46 (32%), 43 (30%), 48 (33%), and 58 (40%), respectively (Table 1). This level of differential expression was significant for all tissues ($P < 0.05$) as evaluated by 1000 simulations of 145 randomly taken genes (32 ± 4.9 , 32 ± 4.9 , 27 ± 4.6 and 42 ± 5.4 random genes differentially expressed in breast, prostate, brain, and colon, respectively, Table 1).

Table 1. The number of differentially expressed and tumor over-expressed splicing factors in each of the tissues as evaluated by 1000 random simulations and as indicated by virtual serial analysis of gene expression (SAGE) tag counts and microarray data obtained from Oncomine (Rhodes et al., 2004).

Tissue	1000 random simulations		Virtual SAGE tags		Microarray data	
	Differentially expressed genes	Tumor over-expressed genes	Differentially expressed genes	Tumor over-expressed genes	Differentially expressed genes	Tumor over-expressed genes
Brain	27 ± 4.6	12 ± 3.4	48 (33%)	41 (85%)	48 (41%)	42 (88%)
Breast	32 ± 4.9	16 ± 3.8	46 (32%)	30 (65%)	70 (60%)	20 (29%)
Colon	42 ± 5.4	19 ± 4.2	58 (40%)	49 (84%)	58 (50%)	43 (74%)
Prostate	32 ± 4.9	15 ± 3.7	43 (30%)	25 (58%)	16 (14%)	9 (56%)

Most of the differentially expressed factors in brain and colon were up-regulated in tumors, 41(85%) and 49 (84%), respectively. However, this trend was not as pronounced in breast and prostate, in which 30 (65%) and 25 (58%) of all differentially expressed splicing factors were up-regulated in tumors, respectively. The simulation showed 16 ± 3.8 , 15 ± 3.7 , 12 ± 3.4 and 19 ± 4.2 , genes to be over-expressed in tumor in breast, prostate, brain, and colon (Table 1).

The differential expression of factors was not evenly distributed among the 17 ontology classes. For example, 11 of 14 examples within the category “non-snRNP assembly proteins” were down-regulated in either breast or prostate. On the other hand, most transcripts belonging to the “SR” class of splicing factors were up-regulated in all tumors (Figure 1).

We evaluated the expression pattern of all splicing factors in the same tissues by using the Oncomine microarray database (version 1.0; Rhodes et al., 2004) as a means to validate the above observation performed with SAGE data. In the Oncomine database, only 117 splicing factors were informative (see Material and Methods). The numbers of microarray experiments, per splicing factor, for prostate, brain, breast, and colon were on average 5.5, 4.3, 2.2, and 1.9, respectively (see Supplementary Table 2 for the complete list of experiments per gene). We found that 101 splicing factors (86%) showed differential expression (equal or higher than 3-fold difference) in at least one tissue (see Material and Methods). Most of the differentially expressed genes (73%) showed an average over-expression in tumors for at least one of the

tissues. Fifty-eight percent of the genes showed tumor under-expression in at least one of the tissues. The expression pattern of all genes in the different tissues can be found in Supplementary Table 2. Similar to the SAGE, the microarray experiments showed that most of the differentially expressed factors in brain and colon were up-regulated in tumor (42 of 48 - 88% and 43 of 58 - 74%, respectively). In breast and prostate the percentage of tumor over-expressed genes was 20 of 70 (29%) and 9 of 16 (56%), respectively (Table 1).

Supplementary Table 2 is available at <http://www.funpecrp.com.br/gmr/year2004/vol4-3/pdf/icob07st02.pdf>.

Analyzing the microarray expression pattern of the genes according to their functional spliceosome ontology groups (Table 2), one can observe that 6 of 8 down-regulated genes from the non-snRNP assembly proteins occurred in breast. In brain, 10 of 17 ontology groups showed only over-expressed genes in tumor. Three other ontology groups showed more genes over-expressed in tumors than in normal brain. In colon, a similar, albeit weaker signal, was observed: six ontology groups showed only genes that were over-expressed in tumors, while five groups showed more genes being over-expressed in tumor than in normal colon.

Most (17 of 24 cases) of the splicing factors belonging to the Sm/Lm core proteins category were over-expressed in tumor in at least one of the tissues. Among the seven factors that were under-expressed in the same family, five appeared in breast, one in prostate and one in brain. Furthermore, in 13 tumor samples factors from the "SR" category were over-expressed, while 4 of 5 down-regulated cases were found in breast.

DISCUSSION

Although there are several reports on splicing factors differentially expressed in tumors (Ghigna et al., 1998; Scorilas et al., 2001; Maeda and Furukawa, 2001), this is the first large-scale computational analysis approaching this issue. In general, we observed differential expression for a significant fraction of splicing factors. Over-expression was a general trend, although this was more pronounced for colon and brain. Using our approach it cannot be determined whether tumor over-expression of splicing factors acts as a causative factor for tumorigenesis or whether it is rather a consequence of the oncogenic process.

It would be interesting to investigate a possible correlation between this expression pattern of splicing factors in tumors and the general splicing pattern of all expressed sequences. Such analysis would show whether the over-expression of the spliceosome factors in fact does cause an increase in the rate of alternative splicing, giving rise to increased expression of new transcripts, or whether it enhances the expression level of the constitutively spliced transcripts. In addition, it could shed some light on functional aspects of the splicing process. For example, we show here that proteins associated with the U2 snRNP seem to be positively associated with cell transformation in brain and breast. According to our SAGE analysis, there are 10 cases of genes being over-expressed in brain and breast and only 3 genes down-regulated in colon. According to the microarray data, three splicing factors associated with the U2 snRNP were over-expressed in brain and/or colon tumors while only one was down-regulated in breast tu-

Table 2. The number of splicing factors in each of the tissues according to their different spliceosome ontology groups as evaluated by microarray data available from the OncoPrint database (Rhodes et al., 2004).

Functional ontology group	Informative factors	Any of the four tissues		Over-expression							
				Brain		Breast		Colon		Prostate	
				T	N	T	N	T	N	T	N
SR proteins	10	8	4	4	0	2	4	5	1	2	0
Non-snRNP assembly proteins	15	8	6	8	0	2	6	3	2	2	0
U1-specific proteins	4	3	2	1	0	0	2	3	0	1	0
U2 snRNP specific proteins	3	3	1	2	0	0	1	2	0	0	0
U5-specific proteins	5	3	2	0	0	0	2	2	0	0	0
U4/U6-specific proteins	4	3	3	0	1	0	2	3	0	0	0
U4/U6,U5-specific proteins	1	1	0	1	0	0	0	0	0	0	0
Sm/Lm core proteins	10	7	6	5	1	4	5	6	0	2	1
Catalytic step II	3	1	1	1	0	0	1	0	1	0	0
Spliced mRNP/EJC proteins	5	1	3	1	1	0	2	1	1	0	0
Other splicing proteins	6	4	2	2	0	0	2	2	0	0	0
RRM-containing proteins	10	6	8	3	1	4	4	4	2	1	2
DEXD box proteins	4	3	2	1	0	2	2	0	1	0	0
Proteins with other known motifs	10	5	5	3	2	1	5	3	2	1	2
Proteins with no known motif	9	4	5	1	0	2	4	2	3	0	1
Cyclophilin-like proteins	5	3	1	0	0	2	0	1	1	0	0
Proteins designated as H-complex component	13	11	8	8	0	1	8	5	1	0	1
Total	117	74	59	42	6	20	50	43	15	9	7

T = tumor; N = normal sample; snRNP = small nuclear ribonucleoproteins; EJC = exon junction complex; RRM = RNA-recognition motif.

mors. The U2 ribonucleoproteic particle is associated with the definition of the 3' splice site and it is therefore expected that variations in the expression of these factors will affect the choice of the 3' splice site. Thus, it may be expected that this choice of the 3' splice site will be somehow different between the above tissues.

For two other categories, the Sm/Lm core proteins and the “SR” proteins, a similar trend was observed; the functional impact of over-expression of splicing factors from these categories remains to be investigated.

We used experimental expression data from SAGE libraries and from microarray data available online from the Oncomine database (Rhodes et al., 2004) to evaluate the expression pattern of all splicing factors characterized by a proteomics approach. A correlation between these two data sources has been shown before, especially in genes having high expression levels (Ishii et al., 2000; Evans et al., 2002). The data presented here confirm this trend. However, there are still few experiments per tissue source available for the microarray analysis. When more data become available, more solid statistical analyses should be done to obtain a realistic picture of the variability found in the expression pattern between samples.

There is a growing body of evidence suggesting that several oncoproteins can act as splicing factors (Burns et al., 1999; McGarvey et al., 2000; Meissner et al., 2003). It would be interesting to test whether there is an overlap between oncogenic and splicing pathways. We observed that, according to SAGE, 15 splicing factors were up-regulated in at least three types of tumors. Microarray analysis revealed 11 of such factors. We found 7 of the splicing factors to be known genes related to cancer. Experimental analyses are needed to evaluate whether splicing factors have oncogenic activity *in vivo*.

Our data support the possibility that altered gene expression of splicing factors might be involved in the differential rate of alternative splicing in human tumors, therefore establishing a broader association between cell transformation and alternative splicing.

ACKNOWLEDGMENTS

The authors thank Maria D. Vbranovski for careful reading of the manuscript. N. Kirschbaum-Slager and P.A.F. Galante were supported by Ph.D. fellowships from FAPESP.

REFERENCES

- Adams, M., Jones, J.L., Walker, R.A., Pringle, J.H. and Bell, S.C. (2002). Changes in tenascin-C isoform expression in invasive and preinvasive breast disease. *Cancer Res.* 62: 3289-3297.
- Baudry, D., Faussillon, M., Cabanis, M.O., Rigolet, M., Zucker, J.M., Patte, C., Sarnacki, S., Boccon-Gibod, L., Junien, C. and Jeanpierre, C. (2002). Changes in WT1 splicing are associated with a specific gene expression profile in Wilms' tumour. *Oncogene* 21: 5566-5573.
- Berget, S.M., Moore, C. and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74: 3171-3175.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72: 291-336.
- Boon, K., Osorio, E.C., Greenhut, S.F., Schaefer, C.F., Shoemaker, J., Polyak, K., Morin, P.J., Buetow, K.H., Strausberg, R.L., De Souza, S.J. and Riggins, G.J. (2002). An anatomy of normal and malignant gene expression. *Proc. Natl. Acad. Sci. USA* 99: 11287-11292.
- Burns, C.G., Ohi, R., Krainer, A.R. and Gould, K.L. (1999). Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* 96: 13789-13794.
- 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.
- Chow, L.T., Gelinias, R.E., Broker, T.R. and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12: 1-8.
- Cragg, M.S., Chan, H.T., Fox, M.D., Tutt, A., Smith, A., Oscier, D.G., Hamblin, T.J. and Glennie, M.J. (2002). The alternative transcript of CD79b is overexpressed in B-CLL and inhibits signaling for apoptosis. *Blood* 100: 3068-3076.

- Evans, S.J., Datson, N.A., Kabbaj, M., Thompson, R.C., Vreugdenhil, E., De Kloet, E.R., Watson, S.J. and Akil, H. (2002). Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. *Serial Analysis of Gene Expression. Eur. J. Neurosci.* 16: 409-413.
- Ge, K., Duhadaway, J., Du, W., Herlyn, M., Rodeck, U. and Prendergast, G.C. (1999). Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl. Acad. Sci. USA* 96: 9689-9694.
- Ghigna, C., Moroni, M., Porta, C., Riva, S. and Biamonti, G. (1998). Altered expression of heterogenous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. *Cancer Res.* 58: 5818-5824.
- Ishii, M., Hashimoto, S., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T. and Aburatani, H. (2000). Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics* 68: 136-143.
- Lee, M.P. and Feinberg, A.P. (1997). Aberrant splicing but not mutations of TSG101 in human breast cancer. *Cancer Res.* 57: 3131-3134.
- Maeda, T. and Furukawa, S. (2001). Transformation-associated changes in gene expression of alternative splicing regulatory factors in mouse fibroblast cells. *Oncol. Rep.* 8: 563-566.
- McGarvey, T., Rosonina, E., McCracken, S., Li, Q., Arnaout, R., Mientjes, E., Nickerson, J.A., Awrey, D., Greenblatt, J., Grosveld, G. and Blencowe, B.J. (2000). The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. *J. Cell Biol.* 150: 309-320.
- Meissner, M., Lopato, S., Gotzmann, J., Sauermann, G. and Barta, A. (2003). Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. *Exp. Cell Res.* 283: 184-195.
- Menke, A.L., Shvarts, A., Riteco, N., van Ham, R.C., van der Eb, A.J. and Jochemsen, A.G. (1997). Wilms' tumor 1-KTS isoforms induce p53-independent apoptosis that can be partially rescued by expression of the epidermal growth factor receptor or the insulin receptor. *Cancer Res.* 57: 1353-1363.
- Modrek, B. and Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.* 30: 13-19.
- Naor, D., Nedvetzki, S., Golan, I., Melnik, L. and Faitelson, Y. (2002). CD44 in cancer. *Crit. Rev. Clin. Lab. Sci.* 39: 527-579.
- 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.
- Scorilas, A., Kyriakopoulou, L., Katsaros, D. and Diamandis, E.P. (2001). Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of pre-mRNA splicing factors: overexpression in aggressive ovarian cancer. *Br. J. Cancer* 85: 190-198.
- Wang, L., Duke, L., Zhang, P.S., Arlinghaus, R.B., Symmans, W.F., Sahin, A., Mendez, R. and Dai, J.L. (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.
- Wang, Z., Lo, H.S., Yang, H., Gere, S., Hu, Y., Buetow, K.H. and Lee, M.P. (2003). Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res.* 63: 655-657.
- 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.
- Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419: 182-185.