Identification of *COL6A1* as a differentially expressed gene in human astrocytomas

A. Fujita[1,2], J.R. Sato[2], F. Festa[1], L.R. Gomes[1], S.M. Oba-Shinjo[3],
S.K.N. Marie[3], C.E. Ferreira[2] and M.C. Sogayar[1]

[1]Instituto de Química, Universidade de São Paulo, São Paulo, SP, Brasil
[2]Instituto de Matemática e Estatística, Universidade de São Paulo,
São Paulo, SP, Brasil
[3]Departamento de Neurologia, Faculdade de Medicina,
Universidade de São Paulo, São Paulo, SP, Brasil

Corresponding author: M.C. Sogayar
E-mail: mcsoga@iq.usp.br

ABSTRACT. Diffuse infiltrating gliomas are the most common tumors of the central nervous system. Gliomas are classified by the WHO according to their histopathological and clinical characteristics into four classes: grade I (pilocytic astrocytoma), grade II (diffuse astrocytoma), grade III (anaplastic astrocytoma), and grade IV (glioblastoma multiforme). Several genes have already been correlated with astrocytomas, but many others are yet to be uncovered. By analyzing the public SAGE data from 21 patients, comprising low malignant grade astrocytomas and glioblastomas, we found *COL6A1* to be differentially expressed, confirming this finding by real time RT-PCR in 66 surgical samples. To the best of our knowledge, *COL6A1* has never been described in gliomas.

The expression of this gene has significantly different means when normal glia is compared with low-grade astrocytomas (grades I and II) and high-grade astrocytomas (grades III and IV), with a tendency to be greater in higher grade samples, thus rendering it a powerful tumor marker.

Key words: *COL6A1* gene; Astrocytomas; Differential gene expression; Tumor marker; SAGE

INTRODUCTION

Gliomas are the most common primary central nervous system (CNS) tumors (Akbasak and Sunar-Akbasak, 1992). The incidence of gliomas in the general population is between 5 and 10 per one hundred and thousand people, but for unknown reasons, the number of cases of gliomas is increasing in elderly people (Behin et al., 2003). The World Health Organization (WHO) classifies these tumors according to their histological features and the number of anaplastic characteristics into four grades, namely: grade I (pilocytic astrocytoma), grade II (diffuse astrocytoma), grade III (anaplastic astrocytoma), and grade IV (glioblastoma multiforme) (Louis et al., 2007). Grades III and IV are the most aggressive, highly infiltrating and neurologically destructive, being considered the most lethal of the human tumors. In the case of grade IV (glioblastoma multiforme), the median patient survival outcome is about 1 year. These high-grade tumors can arise from the progression of other classes (Louis et al., 2002). Grade II astrocytoma has an intrinsic tendency for progression to grades III and IV, as a consequence of sequential acquisition of genetic alterations.

Because of their diffusely infiltrating behavior, grade II astrocytomas cannot be completely resected and will usually recur. Progression to anaplastic astrocytoma (WHO grade III) or glioblastoma multiforme, the most malignant variety of gliomas, usually happens. Therefore, it is very important to find genes related to cell migration and to identify infiltrating cells.

In addition, situations arise in which the WHO classification system becomes problematic, since pathological diagnosis remains subjective and intratumor histological variability is common. In addition, high-grade gliomas display poor cellular differentiation, lacking defined histological features (Nutt et al., 2003).

Behin et al. (2003) has already demonstrated that molecular and genetic analysis of gliomas could help in their classification and in the design of treatment protocols.

Results of gene expression analysis, obtained by subtractive hybridization (Vedoy and Sogayar, 2002; van den Boom et al., 2006; Colin et al., 2006) and DNA microarrays (Sallien et al., 2000; Rickman et al., 2001; Tanwar et al., 2002; Sasaki et al., 2003; van den Boom et al., 2003; Nutt et al., 2003; Persson et al., 2007; Liang et al., 2008; Marie et al., 2008), have led to the identification of new genes involved in glioma tumorigenesis. Since glioma development is a complex process in which a large number of genes are involved, we believe it is worthwhile to search for new genes involved in the formation of this tumor. To this end, we employed a statistical analysis of public SAGE data.

We applied this technology to identify differentially expressed genes in different grades of glioma, identifying one potential biomarker, namely, the *COL6A1* gene, to be differ-

entially expressed between normal glia and the different grades of gliomas. We validated this finding by quantitative real time reverse transcription-polymerase chain reaction (RT-PCR) analysis using 66 samples of different malignant grades of astrocytic tumors and non-tumor brain tissues, confirming its increasing expression as the tumor grade increases.

The *COL6A1* gene codes for type VI collagen, alpha 1 chain, a component of microfibrillar structures in several tissues (Engel et al., 1985). These microfibrils are located in cells, nerves and blood vessels and are considered to have an anchoring function. In accordance with this anchoring function are the reports that type VI collagen binds cells and, moreover, its fusion protein binds type I collagen. This binding activity implies that type VI collagen may be involved in cell migration, differentiation and embryonic development.

The identification of *COL6A1* as a differentially expressed gene in astrocytomas is important for better understanding of the molecular basis of this type of cancer, and, also, to identify prognostic markers for tumor evolution and/or new targets for alternative therapeutic intervention.

MATERIAL AND METHODS

SAGE analysis

In order to identify possible glioma tumor markers, we analyzed the public SAGE data. The SAGE database of glia/gliomas is composed of 21 libraries (total of 1,714,361 tags) of which one is from normal glia, one is from grade I, four are from grade II, four are from grade III, and 11 are from grade IV. They were downloaded from http://cgap.nci.nih.gov/SAGE/. Since SAGE consists of sequencing tags in a time interval, the number of tags is proportional to the sequencing time. Consequently, for large intervals, we expect more tags than for small intervals. Therefore, we normalized the expression relative to one million tags. In addition, genes which do not appear in at least one library are excluded, since it is not known whether this gene was not expressed due to the low number of tags sequenced or because this gene is not expressed in this sample. Thus, we obtained a set of 1575 genes. In order to obtain genes related to the progression of the tumor, we analyzed the 50 highest t-values in the module between glioblastomas and grades I, II and III and normal glia, and then, we performed the algorithm shown previously (Fujita A, Sato JR, Sogayar MC and Ferreira CE, unpublished results) with parameter $p = 1$ (where p is the number of genes used). This algorithm describes a method in which, given a set of gene expression data, it is possible to infer the sample classification into normal versus tumor classes. The method is based on canonical correlation analysis and Kernel non-parametric regression.

Tissue samples

Fresh surgical samples of astrocytomas of different grades and of non-tumor tissue of the CNS (temporal lobectomy from epilepsy surgeries) were macrodisected and immediately snap-frozen in liquid nitrogen upon surgical removal. Astrocytoma specimens were categorized according to the WHO grading system (Louis et al., 2007). Before RNA extraction, a 4-μm thick section of each sample was obtained at -25°C in a cryostat for histological assessment under light microscopy after hematoxylin-eosin staining. Necrotic and non-neoplastic areas were removed from the frozen block and the tumor tissue was microdisected prior to the RNA extraction procedure. For normal tissue, only white matter was used to avoid neuron-related genes.

For validation of the findings concerning differentially expressed genes by means of quantitative real time RT-PCR, we used a large number of normal and tumor samples: 8 samples of non-tumor samples, 9 pilocytic astrocytomas, 11 low-grade astrocytomas, 12 anaplastic astrocytomas, and 26 glioblastomas.

This study was approved by the University of São Paulo School of Medicine Hospital Ethics Committee, and informed consent was obtained from all patients.

Total RNA isolation and cDNA synthesis

Total RNA was extracted from normal and tumor tissues using guanidine isothiocyanate lysis and ultracentrifugation on a cesium chloride cushion (Chirgwin et al., 1979). RNA quantification and quality assessment were carried out by measuring the absorbance ratio for 260 and 280 nm, taking values between 1.8 and 2.0 as a criterion for purity. Denaturing agarose gel electrophoresis was also used to assess the quality of the samples.

Conventional reverse transcription was performed to obtain single-strand cDNA for real time RT-PCR. First-strand cDNA was synthesized from 1 μg total RNA, which was previously treated with 1 unit DNase I (FPLC-pure, Amersham Biosciences, Uppsala, Sweden), using random and oligo(dT) primers (Invitrogen Life Technologies, Maryland, USA), RNase inhibitor and SuperScript III reverse transcriptase following manufacturer (Invitrogen) recommendations. The resulting cDNA was then treated with 1 unit RNase H (Amersham Biosciences) and diluted with TE buffer.

Quantitative real-time RT-PCR

The levels of expression of the selected gene were determined by real time PCR analysis. Primers were designed to amplify 101-bp length amplicons. Primer sequences were as follows (5' to 3'): F: TCAGAATAGTGATGTGTTCGACGTT, R: AGCAACATGGATATGGTTCAGAAA. All primers were synthesized by IDT (Coralville, USA).

The minimum concentration of primers was determined by the lowest threshold cycle and maximum amplification efficiency while minimizing non-specific amplification. Analysis of DNA melting curves demonstrated a single peak for the whole set of primers. SYBR Green I amplification mixtures (25 μL) contained 6.25 μL cDNA, 2X SYBR Green I Master Mix (Applied Biosystems), and forward and reverse primers at a final concentration of 800 nM. Reactions were run on an ABI Prism 5700 sequence detector (Applied Biosystems). The cycle conditions comprised 10 min of polymerase activation at 95°C, and 40 cycles at 95°C for 15 s and 60°C for 1 min. DNA melting curve analysis showed a single peak for all five genes analyzed. Quantitative data was normalized relative to the internal housekeeping control glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene, whose primers were as follows (5' to 3'): GAPDH F: CTGCACCACCAACTGCTTA and GAPDH R: CATGACGGCAGGTCAGGTC. The $2^{-\Delta\Delta Ct}$ equation was applied to calculate the relative expression of the genes in tumor samples versus the median of normal CNS tissues, where $\Delta Ct = Ct$ gene - Ct GAPDH, and $\Delta\Delta Ct = \Delta Ct$ tumor - mean ΔCt normal tissues (Livak and Schmittgen, 2001).

Data analysis

Statistical analysis was applied to gene expression data obtained from both astrocytomas of different grades and normal CNS tissues. To identify the outliers, we considered the following criterion: let J_1, J_2 and J_3 be the first, second and third quartiles, respectively, and set $d_J = J_3 - J_1$. The outliers were defined as the samples with gene expression levels lower than $J_1 - 1.5d_J$ and greater than $J_3 + 1.5d_J$ (Tukey, 1977). Thus, we excluded three samples from grade I, two samples from grade II and six samples from grade IV.

For pairwise class comparisons, a standard statistical test was used, namely, the two-sample *t*-test (Student, 1908).

RESULTS

We studied 66 astrocytic gliomas using real time RT-PCR as illustrated in Table 1. It is interesting to note that normal samples and lower malignant-grade astrocytomas display low *COL6A1* expression, whereas higher malignant grade tumors display high expression of this gene.

Table 1. mRNA expression levels in increasing order of *COL6A1* expression in samples of normal glia and of different grade astrocytomas obtained from a set of 66 patients (8 normal glia, 9 pilocytic astrocytomas, 11 low-grade astrocytomas, 12 anaplastic astrocytomas, and 26 glioblastoma multiforme) analyzed by real time PCR.

		Grades		
Normal	I	II	III	IV
0.35386	0.561069	2.753468	2.743942	1.339087
0.451016	3.184904	3.29722	2.763028	2.659675
0.658041	4.031314	4.320655	6.898344	3.867094
1.198517	6.803372	4.335655	6.922293	5.300947
1.227948	8.006934	6.110329	9.358323	6.369808
1.371971	10.60189	6.640307	9.857693	7.092282
1.496145	25.83491	7.266446	10.67564	8.346955
3.151962	199.639	8.346955	15.68431	10.45593
	439.9665	9.588134	16.23741	10.7127
		24.18851	21.2774	11.48159
		56.1514	36.15818	13.84459
			37.43329	15.90325
				16.52124
				17.40283
				17.46325
				20.34003
				20.91187
				20.98447
				24.35675
				29.77955
				75.1265
				100.1661
				101.9169
				122.4671
				137.7826
				157.7231

The expression values were obtained relative to GAPDH expression levels and normal brain values.

The results presented in Figure 1 show overexpression of *COL6A1* in higher grade gliomas, clearly suggesting that the *COL6A1* gene expression profile allows discrimination between normal glia, low-grade astrocytomas (grades I and II) and high-grade astrocytomas (grades III and IV).
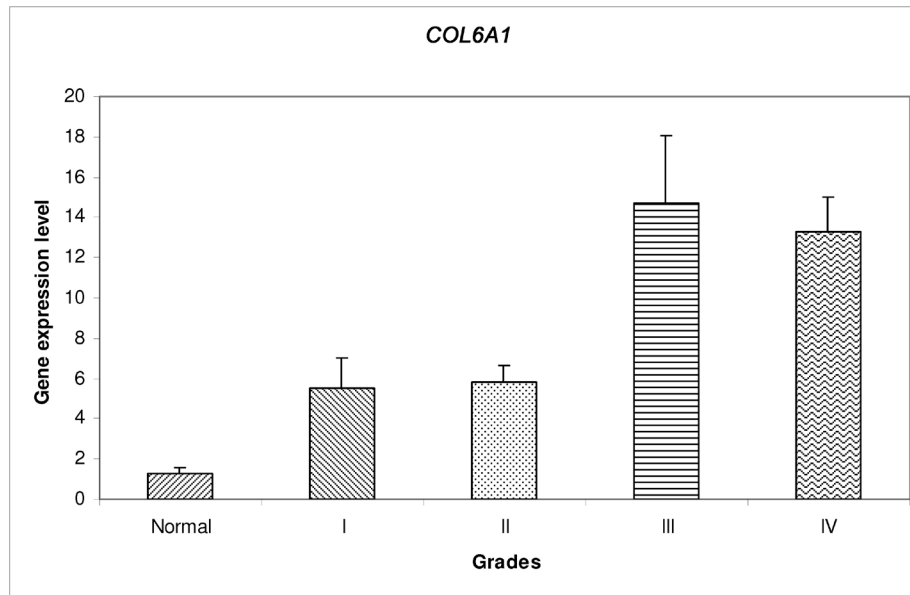

Figure 1. Differences in means between normal glia, grades I and II, grades III and IV showing that the level of *COL6A1* expression is increased in higher grade astrocytoma tumors.

The corresponding P values are shown in Table 2. It is noteworthy that the six grade IV samples with the highest gene expression levels are quite different from the expression pattern of the other samples in this group; therefore, these samples deserve closer future inspection.

Table 2. Pairwise comparisons between normal glia, low-grade astrocytomas (grades I and II) and high-grade astrocytomas (grades III and IV).

	I	II	III	IV
Normal	0.037	0.000	0.002	0.000
I		0.854	0.026	0.003
II			0.026	0.001
III				0.715

No significantly different means were found between grades I and II and between grades III and IV. Therefore, the levels of *COL6A1* expression may discriminate between normal glia, low-grade astrocytomas and high-grade astrocytomas. Fisher discriminant analysis (Fisher, 1936) with cross-validation procedure was performed to verify the predictive property

of this gene. The accuracy was 72%. This rate is not good enough for use of this marker in the clinic, but if associated with other genes, it could represent a potential predictor.

DISCUSSION AND CONCLUSIONS

In order to find novel differentially expressed genes in astrocytomas, we analyzed the available public SAGE data, identifying the *COL6A1* gene and several others as differentially expressed.

The *COL6A1* gene was chosen for validation by real time RT-PCR, since it has yet to be described in gliomas and, also, because this gene codes for type VI collagen, an extracellular matrix protein, which is clearly involved in tumor invasiveness. In addition to a structural role, type VI collagen may be involved in cell migration, thus explaining its higher expression in higher grade astrocytomas, which are more infiltrative.

By analyzing the data presented in Table 1, it is possible to verify that samples displaying low levels of expression of *COL6A1* are found in all malignant grades of tumor; however, what discriminates the normal tissue from tumor samples of different grades is the presence of samples displaying very high expression only in the tumor sample sets.

We propose that the expression of the *COL6A1* gene is related to the progress of tumor development, with expression increasing as tumor progresses to higher grades, thus explaining the presence of low-expression samples in all tumor grades.

The identification of an overexpressed gene to an extracellular matrix protein related to migration points to an interesting target for drug design for an alternative therapeutic approach to infiltrative astrocytomas. Moreover, our data suggest the use of *COL6A1* in class prediction models, based on defined molecular profiles.

The identification of genes that are overexpressed in higher grade astrocytomas is important in understanding the molecular basis of the development of these tumors and to define possible targets for future therapeutic intervention.

In summary, we showed that much information is still to be mined from high-throughput gene expression technologies, such as SAGE, and that statistical tumor classification models could be very useful in this kind of analysis.

ACKNOWLEDGMENTS

REFERENCES

Akbasak A and Sunar-Akbasak B (1992). Oncogenes: cause or consequence in the development of glial tumors. *J. Neurol. Sci.* 111: 119-133.

Behin A, Hoang-Xuan K, Carpentier AF and Delattre JY (2003). Primary brain tumours in adults. *Lancet* 361: 323-331.

Chirgwin JM, Przybyla AE, MacDonald RJ and Rutter WJ (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18: 5294-5299.

Colin C, Baeza N, Bartoli C, Fina F, et al. (2006). Identification of genes differentially expressed in glioblastoma versus pilocytic astrocytoma using suppression subtractive hybridization. *Oncogene* 25: 2818-2826.

Engel J, Furthmayr H, Odermatt E, von der Mark H, et al. (1985). Structure and macromolecular organization of type VI collagen. *Ann. N. Y. Acad. Sci.* 460: 25-37.

Fisher RA (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* 7: 179-188.

Liang Y, Bollen AW and Gupta N (2008). CC chemokine receptor-2A is frequently overexpressed in glioblastoma. *J. Neurooncol.* 86: 153-163.

Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25: 402-408.

Louis DN, Pomeroy SL and Cairncross JG (2002). Focus on central nervous system neoplasia. *Cancer Cell* 1: 125-128.

Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, et al. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 114: 97-109.

Marie SK, Okamoto OK, Uno M, Hasegawa AP, et al. (2008). Maternal embryonic leucine zipper kinase transcript abundance correlates with malignancy grade in human astrocytomas. *Int. J. Cancer* 122: 807-815.

Nutt CL, Mani DR, Betensky RA, Tamayo P, et al. (2003). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res.* 63: 1602-1607.

Persson O, Krogh M, Saal LH, Englund E, et al. (2007). Microarray analysis of gliomas reveals chromosomal position-associated gene expression patterns and identifies potential immunotherapy targets. *J. Neurooncol.* 85: 11-24.

Rickman DS, Bobek MP, Misek DE, Kuick R, et al. (2001). Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res.* 61: 6885-6891.

Sallinen SL, Sallinen PK, Haapasalo HK, Helin HJ, et al. (2000). Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res.* 60: 6617-6622.

Sasaki T, Arai H, Beppu T and Ogasawara K (2003). Detection of gene amplification and deletion in high-grade gliomas using a genome DNA microarray (GenoSensor Array 300). *Brain Tumor Pathol.* 20: 59-63.

Student (1908). The probable error of a mean student. *Biometrika* 6: 1-25.

Tanwar MK, Gilbert MR and Holland EC (2002). Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma. *Cancer Res.* 62: 4364-4368.

Tukey JW (1977). Exploratory Data Analysis. Addison-Wesley, Reading.

van den Boom J, Wolter M, Kuick R, Misek DE, et al. (2003). Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. *Am. J. Pathol.* 163: 1033-1043.

van den Boom J, Wolter M, Blaschke B, Knobbe CB, et al. (2006). Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. *Int. J. Cancer* 119: 2330-2338.

Vedoy CG and Sogayar MC (2002). Isolation and characterization of genes associated with the anti-tumor activity of glucocorticoids. *Brain Res. Mol. Brain Res.* 106: 57-69.