Microarray-based detection of genomic expression signatures related to tumour recurrence in Glioblastoma patients

Álvaro Abella-Bascarán[*,1], **Eloi Casals-Puig**[*] **and Samuel Miravet-Verde**[*]

[*]Pompeu Fabra University, Barcelona (Spain)

ABSTRACT

Glioblastomas are notorious for resistance to therapy, and consequently for their frequent recurrence. In order to deal with recurrent tumours more complex treatments are needed, and still the survival is very low. A possible approach to this problem is the use of a method to predict which tumours are prone to become recurrent, in order to apply the treatment as soon as possible and give better expectancies to the patients.

In this work, the quality of the data provided by Murat *et al.* (2008) has been validated, and the glioblastoma samples have been analysed to find expression signatures related to recurrent glioblastoma. The results show that recurrence is far from the consequence of a few altered genes, and it must be seen as the result of a combination of multiple over and underexpressed molecular routes. The results also highlight differences in the expression of genes related to the ability to escape immune response and apoptosis, along with an upregulation of DNA damage repair, cell cycle and division processes, probably favoured by a higher metabolic rate.

KEYWORDS Microarray analysis; tumour recurrence; glioblastoma; temozolomide.

Introduction

Glioblastoma multiforme, involving glial cells, is the most frequent and aggressive brain tumour in humans, with an incidence of 2–3 cases per 100,000 person life-years in Europe and North America (Bleeker *et al.* 2012) and its treatment can involve chemotherapy, radiation and surgery. The median survival presented with standard-of-care radiation and chemotherapy with the alkylating agent temozolomide is only 15 months (Johnson *et al.* 2012) while the median survival without treatment is 4 and a half months.

Regretfully glioblastomas are well-known for its resistance to therapy (and consequently for their recurrence). The treatment for this cases requires a more aggressive combination of drugs, including nitrosoureas, temozolomide and bevacizumab, in addition to radiotherapy and surgery (Weller *et al.* 2013). This approach implies several risks for the patient and, therefore, it is only applied after the relapse of the patient.

Biologically, the resistance to treatment has been initially attributed to DNA-repair proficiency, cell proliferation and, more recently, to the particular biologic behavior of tumour stem-like cells, as it is exposed in the work of Anastasia Murat (Murat *et al.* 2008). In that case, HOX and EGFR related genes were identified as the most differentially expressed, using clustering techniques and rigid statistical methods. Hence, we propose in this study a more general analysis to determine the molecular profiles specific to recurrent glioblastoma tumours.

Materials and Methods

Tumour Samples and Patient Characteristics

We analyzed data from 80 frozen glioblastoma samples provided by Murat *et al.* (2008). The data comprised 70 tumours from initial surgery and 10 samples resected after recurrence. All patients were treated within a phase II or a randomized phase III trial (Stupp *et al.* 2002, 2005). The study includes 21 females and 55 males, with a median age of 52 (range, 26 to 70 years). Out of the 76 patients, 28 received radiotherapy treatment only, and 48 received TMZ/radiotherapy treatment.

Gene Expression Profiling

The microarray data with gene expression profiling was obtained from the Gene Expression Omnibus (GEO) database at http://www.ncbi.nlm.nih.gov/geo/ (accession-number GSE7696). The data had been created from 54675 probes prepared with the Enzo BioArray-High Yield Kit (Enzo Life Sciences, Farmingdale, NY) for double amplification and were hybridized to Affymetrix HG-133Plus2.0 GeneChips (Affymetrix, Santa Clara, CA). The data used had been normalized to the expression of the EIF2C3, DNAJA4, and B2M genes that exhibited little variation in the data set (Murat *et al.* 2008).

Data Analysis and Statistical Methods

Analyses were carried out in R, a free software environment available at http://www.r-project.org/. The full list of packages and modules used can be found in the session information part at the end of the supplementary document.

Work flow and scripts : in order to ensure the reproducibility of results, all the analysis steps performed can be found in the supplementary material in *HTML* format. This file was processed directly from R using knitr and markdown. The file contains detailed explanations about the methodology, statistical interpretation and all the graphics generated during the process.

Quality assessment and Normalization: the quality of the microarray data was assessed by a variety of quality checks. Raw chip images were visually inspected to ensure the absence of artifacts. The intensity distributions of the samples showed a similar poisson distribution, without any artifactual distribution. We used the linear probe level model (PLM) (Bolstad 2004; Brettschneider *et al.* 2007) to verify the absence of artifacts in the chip pseudoimages created from the weights and residuals of the sample PLM's. We used Normalized Unscaled Standard Errors (NUSE) (Bolstad 2004) to evalute the deviation of the chip probsets, from here, samples with NUSE median value higher than 1.05 were removed. Using the same models, we also evaluated the Relative Log Expression (RLE) values (Bolstad 2004; Brettschneider *et al.* 2007) to determine technical biases on particular chips resulting in no particular deviation of the median or the interquartile range found on any sample.

The expression intensities for all probe sets from Affymetrix CEL-files were estimated using robust multiarray average with probe-level quantile normalization followed by median polish summarization (Irizarry *et al.* 2003) as implemented in the BioConductor software (http://www.bioconductor.org/). The inspection of MA plots (available at supplementary material) ensured the absence of fluorescent intensity dependent biases (Bolstad 2004). After the quality assurance process, no samples were discarded from the analysis.

Analysis of batch effect and confounding variables: the samples were checked for batch effect, first using hierarchical clustering after measuring the Spearman correlation among samples, and then verifying the absence of batch effects by multidimensional scaling. The presence of confounding variables was assessed by means of principal component analysis, and different sources of heterogeneity were analyzed by surrogate variable analysis (Leek and Storey 2007). We decided adjust for age and methylation status (Qiu *et al.* 2014) excluding the batch as confounding variable.

Differential Expression : the list of analyzed probes was reduced using a non-specific filtering (Bourgon *et al.* 2010), eliminating features with little variation (IQR cutoff = 0.5), consistently low signal across samples, or insufficient annotation. With this process we reduced the list of genes of interest from 54675 genes to 10195.

The differential expression analysis of microarray data was performed with an empirical Bayes method (Smyth 2004) implemented in the Bioconductor package limma. The Benjamini-Hochberg procedure was applied for multiple testing correction (false-discovery rates) (Benjamini and Hochberg 1995). We called Differentially Expressed (DE) genes at 10% FDR resulting in 255 DE genes.

Functional Enrichment : we analysed the enrichment of gene ontology terms (http://www.geneontology.org) for biological processes (GO BP) using a one-tailed Fisher's exact test (Fisher 1922). The significance of the GO terms is computed conditionally to the significance of its child terms (Alexa *et al.* 2006). GO terms formed by less than five genes or with less than five DE genes were removed to improve the reliability of the results. This analysis was repeated three times, one considering DE genes in general and another two with the subsets of overexpressed and underexpressed genes in the recurrent patients.

Simple GSEA : Gene Set Enrichment Analysis (Subramanian *et al.* 2005) was carried out using the simple GSEA algorithm (Irizarry *et al.* 2009), as provided in the GSEABase package from the Bioconductor project. The gene set collection used for the analysis is available in the package GSVAdata, with the name 'c2BroadSets'. We restricted the collection to those gene sets belonging to pathways from KEGG, REACTOME and BIOCARTA. In order to add some robustness to the calculation of the Z-scores, the evaluation was limited to gene sets having 5 or more genes. Finally, only gene sets with an FDR of less than 10% were called differentially expressed.

Results and Discussion

The analysis of differential expression showed 255 genes differentially expressed (figure 1) (corrected p-value < 0.1) across the two conditions (recurrent and non-recurrent GBM). From this set of genes we performed a Gene Ontology (GO) enrichment analysis, first using all differentially expressed genes, and then separating by under and overexpression in recurrent tumours. Additionally, a simple GSEA was carried out to check similarities and possible new differentially expressed pathways. The results from this analysis are presented below, including some biological inferences. The GO enrichment analysis are exposed in three different subsections, and the results from the GSEA can be found at the end of the section. A general overview of the results obtained can be found in the table 1.

General Enrichment : the GO enrichment revealed 43 enriched terms (the corresponding table can be found here) when considering together under and overexpressed genes. In total, 161 genes significantly differentially expressed were included in this group.

The enrichment include terms related to metabolic processes (eg. glucose metabolic process and hexose biosynthetic process), immune response (eg. regulation of T-cell activation and regulation of lymphocyte proliferation), and cell proliferation (regulation of cell cycle G1/S phase transition and cell proliferation).

Table 1 Resume of results

Condition	Number of terms[a]	n° genes	Important functions included
General GO Enrichment	43	161	Metabolic processes, Regulation of differentiation, Immune response, Intracellular transport, Extracellular transport & Homeostasis, negative regulation of cell cycle G1/S phase transition, Cell adhesion, Anatomical cell structure, Protein modification.
Underexpression GO Enrichment	15	66	Cell surface, Negative regulation of cell cycle, proliferation & apoptisis, Immune response, Cell adhesion, RNA processing, Negative regulation of protein transport.
Overexpression GO Enrichment	34	116	Negative regulation of cell death & apoptosis, Metabolic processes, Cell structure organization, Transport and Homeostasis, DNA & Protein modification, Immune response & Inflammation.
Gene Set Enrichment Analysis	304	NA[b]	Metabolism regulation, Cell cycle regulation, Damage response, Negative regulation of Apoptosis, Immune system.

[a] Gene Ontology terms or Pathway terms
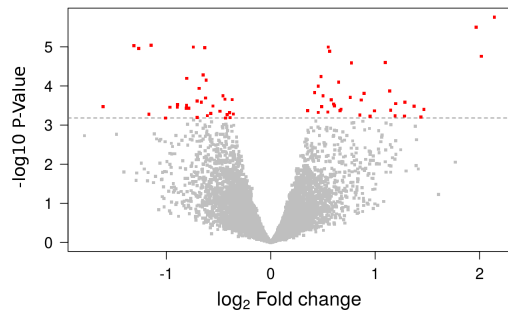[b] No concrete genes were considered as the results come directly from the Expression Set

Figure 1 Volcano plot showing DE genes (red), with negative and positive fold changes indicating underexpressed and overexpressed genes, respectively, in recurrent tumours.

This profile suggests that glioblastoma recurrence is due to heterogeneous causes, possibly combining a higher ability to escape immune response with an even more dysregulated cell cycle, supported by a higher glucose metabolism.

Underexpression Enrichment: the enrichment with underexpressed DE genes revealed 66 different genes grouped in 15 GO terms (table A1 in the appendix, full results here), among which we can highlight the 'apoptotic signaling pathway'. It is known that defects in apoptosis signalling contribute to tumour resistance (Debatin 2004), and we can therefore include it as a feature contributing to recurrence in glioblastoma. This situation of resistance and uncontrolled growth is also favored by the downregulation of pathways negatively related to the proliferation and cell cycle control.

In addition, it is interesting to remark the bigger set of genes, 29, present in the 'Cell surface receptors pathways'. This could be related to the fact that cancerous cells inhibit the exposure on their surface of cell death receptors, in addition to other possible antigens that could be detected by the immune system (Ozören and El-Deiry 2003).

Overexpression Enrichment: the enrichment done with overexpressed DE genes in recurrent patients revealed 34 different GO terms (table A2 in the appendix, full results here) including a total of 116 different genes. An interesting result is the presence of the enriched terms 'exocytosis' and 'transport proteins' (both included in the class 'Transport and Homeostasis').

With respect to exocytosis, it is known that invasion by cancer cells is facilitated by the secretion of enzymes that degrade the extracellular matrix, and that glioma invasion can be inhibited through inhibition of exocytosis (Liu *et al.* 2012). Recurrence of the tumour could thus be related to a higher capacity to invade new areas.

On the other hand, about the 'Transport proteins', one of the most common types of cancer resistance consists on the overexpression of drug pumps able to expel the chemotherapeutical drugs from the cell (Borst 2012). The overexpression of proteins with transmembrane transport function could also be related to recurrence, as the tumour cells would be able to avoid the toxic effect of temozolomide by expelling it from the cytosol or just generating gradients in order to avoid the uptake.
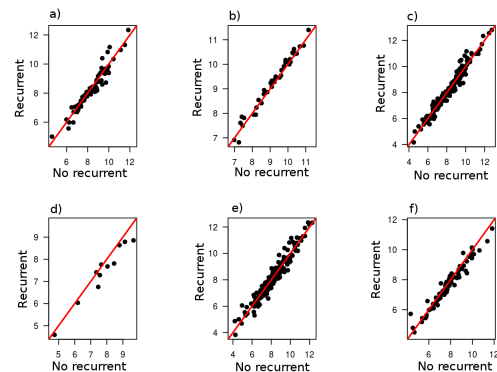

Figure 2 Example of some relevant enriched gene sets: a) Reactome cell surface interactions at the vascular wall. b) Reactome p53 independent DNA damage response. c) Reactome signaling in immune system. d) Reactome tight junction interactions. e) Reactome hemostasis. f) Reactome SLC mediated transmembrane transport.

Gene Set Enrichment Analysis : with this analysis 304 pathways were found to be differentially expressed in the recurrent tumours. Among all the results, it is interesting to remark the alteration of metabolism, cell cycle, immune system and apoptosis. All of this terms were also present along the gene ontology enrichment analysis.

However, this analysis highlights as one of the most important pathways (highest Z-score) the p53-independent DNA damage response from the Reactome database. This result matches with one of the outcomes proposed by Anastasia Murat (Murat *et al.* 2008), who explains that tumour resistance could be due to a higher ability to repair DNA damages like those produced by temozolomide.

In the figure 2 six biologically interesting pathways were selected to provide a visual clue of the differential expression among recurrent and not recurrent tumours. For example, in the case of p53-independent repair, it is possibly to see that in recurrent tumours the genes are slightly but consistently overexpressed. The whole list of pathways altered in recurrent tumours can be found in the supplementary material.

Conclusion

Murat *et al.* (2008) show that the genes related to the survival of the patients were those participating in DNA-repair proficiency and stem cell-like behaviour (HOX and EGFR). Here, we have extended their results studying the genetic profile of recurrent patients of glioblastoma. A general overview of the results shows clearly that recurrence cannot be attributed to a single pathway, and it must be seen as the consequence of a multiple set of features, among which we can highlight a higher ability to escape immune response and apoptosis and a more dysregulated cell cycle and monosacharyde metabolism.

As this study is a comparison of recurrent and non-recurrent glioblastoma, we do not dispose of information to know whether this pathways are also enriched in the case of glioblastoma with respect to controls. This information would be useful to draw conclusions on whether recurrence is a consequence of a qualitatively different expression profile, or a quantitative matter (simply the result of a higher expression of cancer-related pathways).

In conclusion, this article exposes a statistically significant genetic profile related to the relapse of glioblastoma patients due to the recurrence of the tumour. As the glioblastoma is one of the most recurrent types of cancer (Bleeker *et al.* 2012), the definition of a specific recurrence expression profile opens the possibility to detect potential recurrent patients, and to prevent recurrence by personalizing the treatment with the current therapies for recurrent glioblastoma tumours (Weller *et al.* 2013).

Literature Cited

Alexa, A., J. Rahnenführer, and T. Lengauer, 2006 Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22**: 1600–1607.

Benjamini, Y. and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**: 289 – 300.

Bleeker, F. E., R. J. Molenaar, and S. Leenstra, 2012 Recent advances in the molecular understanding of glioblastoma **108**: 11–27.

Bolstad, B. M., 2004 *Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization*. Ph.D. thesis.

Borst, P., 2012 Cancer drug pan-resistance: pumps, cancer stem cells, quiescence, epithelial to mesenchymal transition, blocked cell death pathways, persisters or what? **2**: 120066–120066.

Bourgon, R., R. Gentleman, and W. Huber, 2010 Independent filtering increases detection power for high-throughput experiments. Proceedings of the National Academy of Sciences of the United States of America **107**: 9546–9551.

Brettschneider, J., F. Collin, B. M. Bolstad, and T. P. Speed, 2007 Quality assessment for short oligonucleotide microarray data. Access p. 32.

Debatin, K.-M., 2004 Apoptosis pathways in cancer and cancer therapy. Cancer immunology, immunotherapy : CII **53**: 153–9.

Fisher, R., 1922 On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. Journal of the Royal Statistical Society **85**: 87–94.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England) **4**: 249–264.

Irizarry, R. A., C. Wang, Y. Zhou, and T. P. Speed, 2009 Gene set enrichment analysis made simple. Statistical methods in medical research **18**: 565–75.

Johnson, D. R., A. M. Sawyer, C. A. Meyers, B. P. O'Neill, and J. S. Wefel, 2012 Early measures of cognitive function predict survival in patients with newly diagnosed glioblastoma. Neuro-Oncology **14**: 808–816.

Leek, J. T. and J. D. Storey, 2007 Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genetics **3**: 1724–1735.

Liu, Y., Y. Zhou, and K. Zhu, 2012 Inhibition of glioma cell lysosome exocytosis inhibits glioma invasion. PloS one **7**: e45910.

Murat, A., E. Migliavacca, T. Gorlia, W. L. Lambiv, T. Shay, M. F. Hamou, N. De Tribolet, L. Regli, W. Wick, M. C. M. Kouwenhoven, J. A. Hainfellner, F. L. Heppner, P. Y. Dietrich, Y. Zimmer, J. G. Cairncross, R. C. Janzer, E. Domany, M. Delorenzi, R. Stupp, and M. E. Hegi, 2008 Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. Journal of Clinical Oncology **26**: 3015–3024.

Ozören, N. and W. S. El-Deiry, 2003 Cell surface Death Receptor signaling in normal and cancer cells. Seminars in cancer biology **13**: 135–147.

Qiu, Z. K., D. Shen, Y. S. Chen, Q. Y. Yang, C. C. Guo, B. H. Feng, and Z. P. Chen, 2014 Enhanced MGMT expression contributes to temozolomide resistance in glioma stem-like cells. Chinese Journal of Cancer **33**: 115–122.

Smyth, G. K., 2004 Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology **3**: Article3.

Stupp, R., P. Y. Dietrich, S. O. Kraljevic, A. Pica, I. Maillard, P. Maeder, R. Meuli, R. Janzer, G. Pizzolato, R. Miralbell, F. Porchet, L. Regli, N. de Tribolet, R. O. Mirimanoff, and S. Leyvraz, 2002 Promising survival for patients with newly diagnosed glioblastoma multiforme treated with concomitant radiation plus temozolomide followed by adjuvant temozolo-

mide. Journal of Clinical Oncology **20**: 1375–1382.

Stupp, R., W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. B. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer, and R. O. Mirimanoff, 2005 Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England Journal of Medicine **352**: 987–996.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, 2005 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America **102**: 15545–50.

Weller, M., T. Cloughesy, J. R. Perry, and W. Wick, 2013 Standards of care for treatment of recurrent glioblastoma-are we there yet? **15**: 4–27.