Molecular analysis of non-culture CD133+ GBM cells revealed different signatures among high grade gliomas.

Juan Luis Garcia¹², Maria Perez-Caro³, Juan Antonio Gomez-Moreta⁴, Francisco Gonzalez³, Javier Ortiz⁵, Oscar Blanco⁵, Magdalena Sancho⁵, Jesús María Hernández-Rivas²⁶, Rogelio Gonzalez-Sarmiento²⁷, and Manuel Sanchez-Martín⁷⁸§.

¹Instituto de Estudios de Salud de Castilla y León (IESCyL). Spain
²IBMCC. Centro de Investigación del Cáncer (USAL/CSIC). Salamanca.
³OncoStem Pharma S.L. Spain.
⁴Hospital Universitario de Salamanca. Department of Neurosurgery. Spain
⁵Hospital Universitario de Salamanca. Department of Pathology. Spain
⁶Hospital Universitario de Salamanca. Department of Haematology. Spain
⁷Department of Medicine. University of Salamanca. Spain
⁸Genetically Engineered Mouse Facility. SEA. University of Salamanca. Spain

§Corresponding author

Email addresses:

JLG: jlgarcia@usal.es
MPC: mpc@oncosp.com
JAGM: gomez_moreta@yahoo.es
FG: fran@oncosp.com
JO: jortiz@usal.es
OB: oscarblancomunez@yahoo.es

MS: MAGDALENASANCHO@terra.es

JMHR: jmhr@usal.es

RGS: gonzalez@usal.es

MSM: adolsan@usal.es
Abstract

Background. Gliomas are the most common type of primary brain tumours, and in this group glioblastomas (GBMs) are the higher-grade gliomas with fast progression and unfortunate prognosis. Two major aspects of glioma biology that contributes to its awful prognosis are the formation of new blood vessels through the process of angiogenesis and the invasion of glioma cells. Despite of advances, two-year survival for GBM patients with optimal therapy is less than 30%. Even among patients with low-grade gliomas that confer a relatively good prognosis, treatment is almost never curative. Recent studies have demonstrated the existence of a small fraction of glioma cells endowed with features of primitive neural progenitor cells, tumour-initiating function, able to growth in vitro forming neurospheres and that can be isolated in vivo using surface markers such as CD133. Methods. Eight fresh, primaries and non culture GBMs were used in order to study the gene expression signatures from its CD133 positive and negative populations isolated by FACs-sorting. In order to establish likely clinical correlations, we completed the genomic study of these tumours by compared genomic hybridization (CGH) arrays, FIS studies (to detect possible alterations in PTEN and EGFR genes) and multiplex ligation-dependent probe amplification (MLPA) to detect the methylation status of the MGMT promoter; Results. Gene expression analysis of CD133+ vs. CD133- cell population from each tumour shown that all glioblastoma stem cells presented common characteristics in all glioblastoma samples such as the up-regulation of genes compromised in angiogenesis, permeability and down-regulation of another cohort of genes implicated in cell assembly, neural cell organization and neurological disorders, according to glioblastoma character. Furthermore, unsupervised gene expression clustering let us to distinguish between two groups of samples: that discriminated by tumour location and, the most important, that
group discriminated by their metastatic potential; **Conclusions.** Primary glioblastomas could be sub-classified accordingly to the properties of their CD133+ cells. The molecular characterization of these stem populations could be critical to found new therapeutic targets and to develop an effective therapy for these tumors with very dismal prognosis.
Background

The cancer relapse and mortality rate suggests that current therapies do not eradicate all malignant cells. In this sense, there is increasing evidence that many types of cancer contain their own stem cells: cancer stem cells (CSCs), which are characterized by their self-renewing capacity and differentiation ability [1]. The study of haematological disorders shed light on the relationship between cancer and stem cell compartments, and mechanisms by which CSCs, capable of forming a tumour at one point in time, might change during the progression of the disease. Evidence, however, for the existence of CSCs in solid tumours has been more difficult to find because of the lack of specific cell surface markers. During last years, different cancer cell subpopulations from selected types of human solid cancer were prospectively identified (breast [2], brain [3-5], colon or colo-rectal [6-8], head and neck [9] and pancreatic cancer [10]). Authors, through the use of cell cultures, FACS and/or magnetic activated cell sorting (MACS) methods, have been able to identify different cells populations within the tumour showing stem cell potential, self-renewal and lineage capacity, through serial transplantation in animal models. Specifically, the investigation of solid tumour stem cells has gained momentum particularly in the area of brain tumours, where gliomas are the most common type. In these group, glioblastomas (GBM) are the highest-grade gliomas [GBM; grade IV] [11] manifested by morphological, genetic and phenotypic heterogeneity [12, 13]. Two major aspects of glioma biology that contributes to its awful prognosis are the formation of new blood vessels through the process of angiogenesis and the invasion of glioma cells, the hallmarks of GBM [14]. In addition, these abnormal blood vessels in gliomas have also been shown to create a vascular niche that houses glioma stem cells [15].

Despite of the recent advances, two-year survival for GBM with optimal therapy is less
than 30%. Even among patients with low-grade gliomas that confer a relatively good prognosis, treatment is almost never curative. Recent studies have demonstrated the existence of a small fraction of glioma cells endowed with features of primitive neural progenitor cells and a tumour-initiating function [1]. In general, this fraction is characterized for forming neurospheres (NS), being endowed with drug resistance properties and that often can be isolated using sorting methods with specific antibodies. Since ten year ago, a nice method vastly applied to isolate neural stem cells (NSCs) is the NS forming assay [16]. This assay has been used to test the presence of NSCs in brain tissues in vitro [17]. NS were one of the key pieces of evidence leading to the now widely accepted view that adult mammalian brain harbours a pool of NSCs responsible for the persistent neurogenesis seen in limited adult brain regions such as the sub-ventricular zone (SVZ), olfactory bulb and hippocampal dentate gyrus [18]. However, the NS assay has also limitations, to isolated NSCs you need to growth the sample in a cell culture with cytokines as EGF and FGF that could change their properties or expression profiles, an important point that you have to keep in mind when you interpreting microarrays expression studies derived from NS cultures.

In 1999, Sean J. Morrison [19] and finally Irving L. Weissman’s laboratory could identify and isolate the human central nervous system (CNS) stem cell [20] using antibodies against CD133. This five-transmembrane protein, identifies a subset of human fetal brain cells distinct to human haematopoietic stem cells, which are also CD133+ but are also CD34-bright [21]. This subset of human CD133+ fetal brain cells is capable of neurosphere initiation, self-renewal, and multilineage differentiation at the single-cell level. The CD133+ cells can differentiate in vitro to neurons and glial cells, and their transplantation into the lateral ventricles of newborn NOD-SCID mouse brains resulted in specific engraftment in numerous sites of the brain [19, 20, 22].

6
The CD133 marker is a five-transmembrane protein which is expressed in different type of progenitors as human fetal brain cells or human hematopoietic stem cells (HSC)[19-21]. In brain tumours the proportion of these CD133+ cells representing a minority of the tumour cell population and are also capable to initiate tumour formation in vivo. In gliomas, this small fraction of CSC which forming NS, can also be isolated using CD133 marker [4].

We suggest in this study that primary GBM should be further sub-classified according to their CSC properties. The molecular biology and the expression signature of these CSCs that drive and support the tumour growth will shed light to devise novel and specific treatment strategies. To achieve this aim, we have designed a study in which we have analyzed of a short cohort of primary GBMs. We have analyzed thoroughly eight fresh primaries GBMs focused on its CD133 positive and negative cells. Importantly, all tumours were studied before the patient treatment and without previous tumour cell culture. In order to establish likely clinical correlations, we have also analyzed the molecular signature from both CD33+ and CD133- cells using FACs-sorting and expression-arrays assays. Besides, we have completed the genomic study of these tumours by compared genomic hybridization (CGH) arrays, FIS studies (to detect possible alterations in PTEN and EGFR genes) and multiplex ligation-dependent probe amplification (MLPA) to detect the methylation status of the MGMT promoter. Results obtained through these analysis conclude that gene expression signature is able to differentiated between two different GBM subtypes (higher or lower metastatic tumours) according to their CD133+ vs. CD133- characteristics.
Materials and methods

Samples, FACs and sorting assays

Fresh tumors from eight patients affected of primary GBM without any previous treatment were collected (clinic and pathologic features are summarized in Table 1). Patients’ diagnostics were confirmed by the Pathology Facility from the University Hospital of Salamanca, Spain. At the chirurgic extraction moment, a vast proportion of each tumor was processed to isolate the CD133+ and CD133- cells without previous cell culture. Single-cell suspensions were prepared from individual tumors by standard procedures. Briefly, tumors were cut on pieces into Ca²⁺/Mg²⁺ free phosphate-buffered saline. Cell suspension was gone through 0.7µm single-cell filter. All single-cells were used for staining. Cells were immunophenotyped using human CD133/2 (293C3) phycoerythrin conjugated antibody (MACS, Miltenyibiotec). Possible mature red cells were depleted by hypotonic lysis solution (0.38% ammonium chloride for 15 minutes on ice) before staining. Cells, suspended in Ca²⁺/Mg²⁺ free phosphate-buffered saline (PBS) supplemented with 1% FBS, were labeled with this antibody (approximately 1 µg/10⁶ cells) for 30 minutes on ice. Cell fluorescence was analyzed and sorted using the FACS Aria sorter (Becton Dickinson, New Jersey, USA). CD133 antibody was tested previously in human bone marrow (BM) cells in which CD133 positive cells were described before (Figure 1). BM cells were incubated with CD133 and CD34 antibodies (Pharmingen), sustained in studies that demonstrate that antibodies against CD133 also identified a subset of CD34bright BM hematopoietic stem cells [21]. Cell viability was assessed by propidium iodide (5 µg/mL; Sigma) exclusion using flow citometry.

Expression arrays
We studied a dataset generated with Affymetrix U133 Plus 2 arrays (Affymetrix, Santa Clara, CA, USA) in 8 gliomas. Results from this expression analysis have been deposited at GEO [23] with accession number GSE18015.

Isolated cells (CD133+ and CD133- from each tumour) using sorting methods were collected in separated vials containing RNA Later (Qiagen, Chatsworth, CA, USA). Total RNA was extracted from CD133+ and CD133- sorted cells using Trizol (Invitrogen, Carlsbad, CA) making a total of 16 samples (8 positives and 8 negatives). The integrity of the RNA was confirmed with the Agilent Bioanalyzer 2100 using the RNA 6000 Pico kit (Agilent). We used the GeneChip® Expression 3’ Amplification Two-Cycle Target Labeling kit (Affymetrix, Santa Clara, CA, USA) to label the RNA following the manufacturer protocol. The cRNA was hybridized to Affymetrix Human U133 Plus 2 arrays according to the manufacturer protocol. Briefly, double-stranded cDNA was synthesized routinely from less than 1 microgram of total RNA primed with a poly-(dT) -T7 oligonucleotide. The cDNA was used in an in vitro transcription reaction (IVT) in the presence of T7 RNA polymerase and biotin-labelled modified nucleotides during 16 hours at 37°C. Biotinylated cRNA was purified and then fragmented (35–200 nucleotides), together with hybridization controls and hybridized to the microarrays for 16 h at 45°C. Using the GeneChip Fluidics Station 450 (Affymetrix), the biotin-labelled cRNA was revealed by successive reactions with streptavidin R-phycocerythrin conjugate, biotinylated anti-streptavidin antibody and streptavidin R-phycocerythrin conjugate. The arrays were finally scanned in an Affymetrix GeneChip Scanner 7G Plus.

Preliminary data analysis was conducted using the software of the Affymetrix Expression Console from AGCC suite (Affymetrix GeneChip Command Console, version 1.1) following the statistical procedures described in the Affymetrix: Expression
Console User Guide, selecting the 3’ Expression Analysis guidelines for MAS5 and PLIER algorithms in two independent steps. MAS5 calculated the present call index for each of the 54,675 probe sets on the chip (settings used were standard for the HG U133 Plus 2 array: alpha1 = 0.04, alpha2 = 0.06, Tau = 0.015, TGT = 500). This present call index was used to select 245 probe sets having Presence index through the 16 analyzed samples. PLIER algorithm was used to calculate the normalized expression values of the probe sets (using quantile normalization and PM-MM background correction methods). Statistical analysis and post-processing were performed using TIBCO Spotfire 9.1 (TIBCO Software Inc., Palo Alto CA, USA).

Network analysis was performed mapping the results on the IPA 8 knowledge database (Ingenuity Pathway Analysis).

**CGH array**

DNA from each fresh-frozen sample was extracted with the standard phenol-chloroform method and normal DNA was prepared from human placenta of healthy donors. All DNAs were quantities using the Nanodrop spectrophotometer. DNA quality was assessed by the 260:280 ratio and its integrity by agarose gel ethidium bromide visualization. Genomic-wide analysis of DNA copy number changes of CD133 negative-patient samples were performed using array-based CGH. Due to the little proportion of CD133 positive cells in each tumor sample, CGH arrays only was performed using genomic DNA from de bulk tumor. Slides containing 3296 BACs were produced in “Centro de Investigación del Cáncer” (Salamanca, Spain). The particular bacterial artificial chromosome (BAC) and P-1 derived artificial chromosome (PAC) set used to produce this array is distributed to academic institutions by the Welcome Trust Sanger Institute (Cambridge, United Kingdom) and contains targets spaced at ≈ 1 Mb density over full genome, a set of subtelomeric sequences for each chromosome arm,
and a few hundred probes selected for their involvement in oncogenesis. The clone content is available in the “Cytoview” windows of the Sanger Institute mapping database site, Ensembl (http://www.ensembl.org/). According to this database, clones were ordered along the chromosomes. For the array, 10 simultaneous hybridizations of normal male versus normal female and placenta: placenta (DNA reference) was performed to define the normal variation for the log₂ ratio. Cy5/Cy3 intensity ratios of every spot were converted into log₂ ratios. The log₂ ratio of each clone was normalized to the median log₂ ratio of the ten control hybridizations, after which the median of triplicate spots was calculated. Data from two-color hybridizations for both DNA were normalized using the corresponding GEPAS module DNMAD. Regions of copy number gain and loss for the BAC array-CGH data were identified by creating sample specific thresholds (see supplementary data- CGH methods). All data sets were carefully revised for frequently affected chromosomal sites of physiologic copy number polymorphisms (CNP). Therefore every clone on the array was compared with ‘Database of Genomic Variants’ (http://www.project.tcag.ca/cariation) [24, 25]. For unsupervised clustering analysis, we converted the relative ratio value for each BAC clone to a score of 1 (gain/amplified), 0 (no change), or -1 (loss) based data obtain by the binary segmentation method and analyzed data with Cluster and TreeView (Multi Experiment Viewer 4.0) based on the average linkage method with the Pearson Uncentered metric correlation. Statistical evaluation was carried out using the SPSS 15.0 statistical software. All P-values reported were two-sided and statistical significance was defined as P-values < 0.05.

Fluorescence in situ hybridization

Dual-probe fluorescence in situ (FISH) analyses were performed with locus-specific
probes for centromere 7/EGFR gene and centromere 10/PTEN gene (Vysis, Downers Grove, IL). FISH studies were carried out following well-established methods [26]. Polysomies (chromosomal gains) were defined as more than 10% of nuclei containing three or more CEP signals. Specimens were considered amplified for EFGR when more than 10% of CD133 negative tumour cells exhibited an EGFR/CEP7 ratio >2 or inestimable tight clusters of signals of the locus probe.

**Real-Time PCR**

CD133+ and CD133- RNA samples were reverse-transcribed to cDNA. PCR reactions were performed using equal amounts of cDNA as template. SYBR Green PCR Master mix (Applied Biosystems, USA) was used for template amplification using standard protocol with specific primers for each of the transcripts examined. Incorporation of the SYBR Green dye into PCR products was monitored in real time with an ABI PRISM 7000 sequence detection system (Applied Biosystems). SDS system software was used to convert the fluorescent data into threshold cycle ($C_t$) at which exponential amplification of products begins. The differences in the $C_t$ values ($dC_t$) between the transcript of interest and endogenous control (GAPDH) were used to determine the relative expression of the gene in each.

RT-PCR was performed using specific primers to corroborate expression array result for several genes of those 245 presented probes (data not shown).

**MS-MLPA analysis.**

MLPA analysis was performed using SALSA MLPA Kit ME011 and performed as described by the manufacturer with minor modifications in order to detect MGMT promoter methylation. Briefly, 5µl of each tumor DNA was denatured and subsequently 12
cooled down to 25ºC. After adding the probe mix, the sample was denatured and the probes were allowed to hybridize. Subsequently, half of the sample was ligated, whereas for the other part of samples, ligation was combined with a HhaI digestion resulting in ligation of the sequences only. PCR was performed with double PCR amounts were used (50 µl PCR volume containing 10µl of the ligation reaction). Agarose gel electrophoresis was used to check MLPA efficiency. Subsequently, PCR reaction fragments were separated by capillary gel electrophoresis (ABI 3739, Applied Biosystems) and quantified using the Genemapper software (Applied Biosystems).MS-MLPA processing was performed using Coffalyser analysis tool developed at MRC-Holland (www.mlpa.com) and performed as described by Jeuken et al [27].

**Results and discussion**

**Patients with higher number of CD133+ cells could present resistance to the treatment.**

In order to check the functionality of the CD133 antibody, we tested it in a human bone marrow sample. In normal conditions CD133 antibody also identifies a subset of bone marrow stem cells, CD34 positive also. **Figure 1** shows how CD133 antibody identifies a pool of well-defined human CD133+/CD34+ cells. Using the same antibody we sorted the CD133 positive population from each fresh GBM sample without previous cell culture (Figure 1). **Table 1** shows the absolute number of CD133+ cells obtained from each sample and clinical-biological parameters of each patient. It was interested that only those two patients with higher number of CD133+ cells (higher than 1%, more than 10,000 cells) were not able to response to treatment, something not extraordinary if we take into account that some groups have publish the relationship between the high CD133+ cell number and their resistance to therapy in patients [28, 29].
**EGFR, PTEN and MGMT genes are altered in the GBM primary samples.**

Implications of different kinds of genetic alterations in primary glioblastoma prognosis such as *EGFR* amplification, *PTEN* deletion or *MGMT* promoter methylation have been previously described. As *EGFR, PTEN* and *MGMT* genes are usually altered in primary GBM, we decide to corroborate the primary nature of these samples checking the existence of these alterations in the bulk tumour cell population of GBMs by FISH analysis and MLPA assay. *EGFR* amplifications were detected in 3 of 8 samples and *PTEN* deletions in 6 of 8 samples. We have not found correlations with the biological features from patients. Additionally, we also detected *MGMT* promoter methylation but none of these results shown correlations with biological features of patients (see Table 1 and Figure 2). It is important to remark that we have studied a short number of cases.

**Common genomic imbalances identify the patients with higher number of CD133+ cells without treatment response.**

In addition to the expression studies we analyzed the genomic profile by CGH in the bulk tumours. In more than 50% of cases the gains affect to chromosomal on 1q31-q42, 3q25, 4p15, 6p21-6q23, 7p21, 7q21, 9p21, 9q22, 11q22, 12q21 and 18q12. The losses on 1p36, 1p13, 2p23, 5p15, 10q24.16, 12q13, 13q14, and 17p were also affected on more than 50% of cases (see supplementary table -Table CGH data). Two mayor genetic groups emerge from unsupervised clustering of CGH data (Figure 3). Significantly, the only two cases with higher number of CD133+ cells without treatment response, samples G4 and G11, grouped together. This cluster was characterized with genomic gains on 3p21.31, 6p21, 6q25, 7p14.2, 9q22, 15q11, 20q13 and 22q13 chromosomes.
CD133+ vs. CD133- gene expression analysis divided GBM in 2 different groups.

Although the number of samples is short, some novel characteristics in their processing make this study unique in the GBM field. The main feature that distinguishes this work from the rest is the non culture study of GBM cells. Glioma cells and glioma stem cells studies have been previously performed using in vitro culture patient cells and, although these culture cells present the capacity to form neurospheres (NS), we cannot forget that we are modifying essential pathways in this tumour development (cell-cell adhesion, cell-niche junctions,…). Using direct sorting of CD133+ cells without previous cell culture let us to obtained more authentic pool of GSCs. Even when these cells represent a low percentage of total tumour cells, as in a normal tissue, we were able to isolate and amplified their RNA in order to study their gene expression signature in comparison to CD133- population from the same GBM tumour.

mRNA from the each CD133+ sample and its partner CD133- were isolated using standard techniques. Due to the low number of CD133+ cells obtain of each non-culture sample and to obtain enough mRNA to achieve the hybridization of the expression arrays, all samples were subjected to two rounds of mRNA amplification.

Data normalized from the preliminary analysis was used to calculate only those probe sets being Present (as described in the MAS5 algorithm) in all the samples (16 arrays from 8 GBM; hybridization per cell population). Results from this expression analysis have been deposited at GEO [23] with accession number GSE18015.

A final list of 245 probe sets was obtained according to these parameters. Initially, unsupervised clustering of 245 gene list using GEPAS Release v3.1 software (http://gepas3.bioinfo.cipf.es/) allowed us to examine the first classification of these GBMs (Figure 3). Importantly, GBM samples were ordered in two main groups. Only
samples G9 and G11 were far detected from the rest. However, the only one biological
classification able to distinguish them from the rest was the tumour location. G9 and
G11 presented a parietal location whereas the rest presented temporal or local locations
(Table 1). Gene expression profile of both CD33+ and CD133- cells using microarrays
was arranged to establish possible answers between clinical correlations and gene
expression. However, no relevant connections were found between these 2 parameters
(gene expression and biological-clinical correlations). Nevertheless, it is important to
remark that although there is no evidence about the relationship between GBM location
and its prognosis or molecular characteristics, several papers made references about the
importance of microenvironment in stem cell and cancer stem cell maintenance
(perivascular niche recently describe in astrocytomas is possibly the nearest
example)[30]. Probably, in a near future, more evidences teach us about this
association.

**Common genes up-regulates and down-regulates in all CD133+ GBMs are
compromised in angiogenesis and assembly functions, respectively.**

Following the initial classification proposed by SOTAarray of GEPAS, we were able to
discriminate a minor group of genes commonly up (19 genes) and down-regulated (23
genes) in all samples (CD133+/CD133-) (Tables 2 and 3), and a mayor group of genes
with a differential gene expression pattern in G4 and G7 samples opposed to the rest.
It was interested to observe how all these up-regulated genes in CD133+ vs. CD133-
cells, were essentially related to connective tissue functions and developmental
disorders. Up-regulation of genes such as COL1A1, COL1A, PGF [31] or TGFB1 [32]
suggested an important role in blood vessel formation, angiogenesis, permeability and
proliferation, essential functions in tumour progression [31, 33, 34]. Significantly, most
of the up-regulated genes codified to secreted proteins to extracellular space. TGFB1,
for example, induces the dissociation of VE-cadherin junctions between endothelial cells which can favour migration of mature tumour or GBM cells [32]. Up-regulation in glioma stem cells of these genes could represent an increase in mobility of glioma stem cells (GSCs) through the brain, consistent with the high metastatic characteristics of these tumours. TMEFF2, also up-regulated in the CD133+ compartment, however, could be reflecting the CD133- compartment state. TMEFF2 is a tumour suppressor gene described in different kind of tumours that appear to be distributed principally in neurons and glial cells of the central nervous system. TMEFF2 plays multiple roles in cell growth, maturation and adhesion and its inactivation may serve as an early event in the initiation of neoplastic progression [35] just what is seems to happen in the CD133- compartment of GBM samples.

Other gene with higher expression level in CD133+ compartment vs. CD133- was LRRFIP1, a transcriptional repressor of EGFR (commonly altered and amplified prognosis factor in GBM)[36]. Over-expression of this gene in the stem cell compartment would implied a lower level of this gene in the mature, tumour bulk of CD133- cells, allowing EGFR gene expression and function in brain tumour development. Negative deregulation of LRRFIP1, HIF-1 (hypoxia inducible factor activated in stress conditions)[37] or OPHN1 (involved in normal development of cerebellum)[38] in CD133- cells, confirms their implication in tumour and neural developmental disorders proposed by Ingenuity Pathway Analysis (Figure 4).

The second minority groups of genes, commonly down-regulated in all CD133+ vs. CD133- cell from human non culture GBMs (that means, over-expressed in the CD133- compartment) were amazingly related to cell assembly, neural cell organization and neurological disorders. That is the case of GNB2L1 (important in adhesion and migration of human glioma cells) [39], DPYSL2 [40], TUBA1A [41] or CFL1 [42, 43].
all of them important pieces in cell migration, morphology and actin polymerization, in resume, motility of neural differentiated cells (Figure 5).

We can conclude that data obtained by gene expression analysis of commonly regulated genes in all samples show that GSCs population plays a really important role in enlarge angiogenesis and migration capacity through the secretion of proteins implicated in these functions. Additionally, differential expression of determined genes in CD133- vs. CD133+ population (mature glioma cell population), shown alterations in several genes involved in neural development disorders and previously described by other authors. High level of $HIF-1$, low levels of $LRRFIP1$ ($EGFR$ transcriptional repressor), $TMEFF2$ and other genes responsible for tissue or cell assembly, provide evidences of GBM as a highly metastatic tumour due down-regulation of cell to cell adhesion junction codifying genes.

Two different GBM groups can be functional defined according to the expression pattern of 40 genes.

In relation to the unsupervised GEPAS software classification, we found out a group of genes clearly over-expressed in two different CD133+ vs. CD133- GBM samples (G4 and G7). Once common up and down-regulated genes were detected, we visualized a group of GBMs whose CD133+/CDD133- gene expression levels vary around 40 well-defined genes. Classification of these 40 genes according to their function revealed their implication in DNA replication, recombination and, definitively, cell growth and movement. Positive deregulation of genes such as $VIM$ (implicated in organization of proteins involve in migration and cell signalling) [44, 45], $GLUL$ (main source of Glutamine)[46], $PLK1$ (mitotic cell control gene)[47], $NEDD6/RPL10A$ (down-regulated in neural precursor cells during development), $RPL22$ (important in adherence) or $RPS4X$ (important role in proliferation)[48] in G4 and G7 CD133+ vs.
CD133- compartments let us conjecture a higher proliferative rate and cellular compromise in GSCs of 2/8 GBM samples (Figure 6). In contrast, the negative deregulation of these same genes in the second GBM group (G2, G5, G6, G8, G9 and G11) shed light on the possible different proliferative potential between similarly characterized GBM tumours. We hypothesize that these up-regulated gene levels could differentiate 2 diverse kind of GBM, one less proliferative and the other one endowed with metastatic characteristics (G4 and G7) respect to the others.

Conclusions

CD133+ vs. CD133- gene expression profile from primary, non-in vitro culture GBM samples demonstrated that GSCs express common, deregulated genes in all samples. Between these genes, those concerned in blood vessel formation, angiogenesis or permeability could be playing a significant position in migration and metastasis. These genes could be a potential targets to develop new therapies. However, the most exciting data has been obtained from the analysis of the opposed expression pattern of a group of 40 genes in GBM samples. This expression profile suggests that primary GBM can be sub-classified accordingly to the properties of their CD133+ cells. Differences between both groups can be provided by the proliferative potential of either GSCs or mature glioma cells. The molecular characterization of these stem populations could be critical to found new therapeutic targets and to develop an effective therapy for these tumors with very dismal prognosis.

Competing interests

The author(s) declare that they have no competing interests
Authors' contributions

JLG participated in the design of the study, carried out CGH, FISH and methylation studies. MPC carried out RT-PCR assays, analysis and interpretation of array data and has been involved in drafting the manuscript. JAGM provided GBM samples and clinical data from patients. FGR carried out gene expression assays and performed gene expression analysis. JO, OB and MS contributed to anatomy-pathological diagnosis of GBM samples. JMHR and RGS have been involved in revising it critically for important intellectual content. MSM conceived of the study, its design, coordination and has been involved in drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

I would like to thank Dr. A.M. Pendas and Dr. I. Sanchez-Garcia for their unconditional help and support. I thank Dr. M. Luz Sanchez and Dr. J.M. Martin Fernandez for continuous and generous help with the human tumour cell sorting and technical assistance. Research in my group is supported by Ramon y Cajal Scientific Spanish Program, Junta de Castilla y León (SAN673/SA09/08) (SA093A08), and Department of Medicine from University of Salamanca. Dr. JL Garcia is supported by IESCyL foundation.
References


17. Marshall GP, 2nd, Reynolds BA, Laywell ED: Using the neurosphere assay to


34. Ramirez F, Di Liberto M: Complex and diversified regulatory programs control the expression of vertebrate collagen genes. In., vol. 4; 1990: 1616-1623.


25


48. Watanabe M, Zinn AR, Page DC, Nishimoto T: Functional equivalence of

**Figure legends**

**Figure 1. FACS sorting of glioblastoma cells using CD133 antibody.**
A) Control samples from human bone marrows incubated with CD133 antibody. 1: Total cellularity; 2: Gate CD34 without CD133-PE; 3: Gate CD34 with CD133-PE.
B) Sorting of tumor sample cells CD133+. C) RNA isolation from CD133+ / CD133- tumor cells.

**Figure 2. CGH array and MGMT promoter methylation assays in GBM samples.**
A) Unsupervised cluster analysis of CGH data from 8 primaries GBMs. Each column represents one case and each row represents one BAC clone. We assigned values of 1, 0 and -1 for gain, no change and loss, respectively. Losses are in green and gains in red. **B) Ideogram showing MGMT promoter methylation.**

**Figure 3. Unsupervised clustering of CD133+ cells vs. CD133- cell gene expression signature from each tumour sample show 2 main GBM groups.**
To molecularly characterize glioblastoma stem cells of GBM tumours, we compared the gene expression profiles of purified CD133+ cells from GBM patients versus CD133- cells from control mice. Each gene (identified at right) is represented by a single row of coloured boxes; each patient is represented by one single column. Data are displayed by a colour code where red indicates over-expression in CD133+ fraction versus CD133- cells. A group of genes over-expressed for almost all samples are grouped in the bottom. SOTArray tool from GEPAS Release v3.1, let us to classify CD133+ vs.
CD133- cells from each tumour in 2 mainly groups: G9, G11 and the rest. The only one biological correlation between this genetic cluster and GBM biological characteristics resided in the tumour location.

**Figure 4. Commonly glioblastoma stem cell up-regulated genes participate in angiogenesis and in tumour and neural developmental disorders.** Ingenuity representation and classification by functions of those common up-regulated genes in all CD133+ vs. CD133- cell GBM samples. Red colour genes are the most positive deregulated and grey one those with a lower over-expression levels in this group. The first cluster of genes (COL1A1, COL1A2, TGFB1…) has been described largely in angiogenesis and permeability whereas the second cluster (LRRFIP1 and OPHN1) participates in developmental disorders. Changing transcription pattern of all of them favour tumour development.

**Figure 5. Common glioblastoma stem cell down-regulated genes are involved in cell assembly organization and cancer.** Ingenuity representation and classification by functions of those commonly down-regulated genes in all CD133+ vs. CD133- cell GBM samples. Green colour represents those genes differentially regulated in CD133+ vs. CD133- that participates in cell assembly, migration and cancer pathways.

**Figure 6. Forty differential genes in G4 and G7 samples discriminate between high or low proliferative tumours.** Unsupervised clustering and ingenuity pathways representation of 40 differentially expressed genes. A) **Unsupervised clustering of this 40 gene list** let us to distinguish 2 well defined and opposite groups. Ingenuity principal represented pathways include B) **recombination and repair pathways** and C) **cancer and cell compromise.** Those GBMs with a positive pattern CD133+/CD133- for this
gene expression signature, could present a higher proliferative potential of their tumour stem cells or, by the opposite, a lower proliferative potential of the mature glioma cells.
### Tables

**Table1. Clinical characteristics, MGMT promoter methylation and FISH analysis of eight primary GBMs.**

<table>
<thead>
<tr>
<th>GBM sample</th>
<th>Sex</th>
<th>Age years</th>
<th>Number CD133</th>
<th>Stage</th>
<th>Tumour Location</th>
<th>Resection</th>
<th>DFP months</th>
<th>Survival days</th>
<th>Response</th>
<th>MGMT Methylation</th>
<th>EFGFR (% cells)</th>
<th>PTEN (% cells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>M</td>
<td>73</td>
<td>2300</td>
<td>Death</td>
<td>Temporal</td>
<td>Total</td>
<td>9</td>
<td>398</td>
<td>CR</td>
<td>No met</td>
<td>Amplification (74%)</td>
<td>del (46%)</td>
</tr>
<tr>
<td>G4</td>
<td>M</td>
<td>70</td>
<td>37000</td>
<td>Death</td>
<td>Temporal</td>
<td>Total</td>
<td>5</td>
<td>479</td>
<td>NR</td>
<td>Met</td>
<td>Polysomy (61%)</td>
<td>del (86%)</td>
</tr>
<tr>
<td>G5</td>
<td>F</td>
<td>68</td>
<td>6000</td>
<td>Alive</td>
<td>Frontal</td>
<td>subtotal</td>
<td>8</td>
<td>630</td>
<td>PR</td>
<td>Met</td>
<td>Polysomy (55%)</td>
<td>del (64%)</td>
</tr>
<tr>
<td>G6</td>
<td>M</td>
<td>65</td>
<td>6000</td>
<td>Death</td>
<td>Multifoci</td>
<td>subtotal</td>
<td>5,5</td>
<td>410</td>
<td>CR</td>
<td>No Met</td>
<td>Polysomy (47%)</td>
<td>Normal</td>
</tr>
<tr>
<td>G7</td>
<td>M</td>
<td>59</td>
<td>4800</td>
<td>Death</td>
<td>Frontal</td>
<td>Total</td>
<td>3</td>
<td>192</td>
<td>P</td>
<td>No met</td>
<td>Amplification (54%)</td>
<td>del (40%)</td>
</tr>
<tr>
<td>G8</td>
<td>M</td>
<td>65</td>
<td>6000</td>
<td>Alive</td>
<td>Temporal</td>
<td>Total</td>
<td>4,5</td>
<td>166</td>
<td>P</td>
<td>Met</td>
<td>Normal</td>
<td>del(70%)</td>
</tr>
<tr>
<td>G9</td>
<td>M</td>
<td>69</td>
<td>2900</td>
<td>Death</td>
<td>Parietal</td>
<td>subtotal</td>
<td>2</td>
<td>108</td>
<td>P</td>
<td>nd</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>G11</td>
<td>F</td>
<td>76</td>
<td>13000</td>
<td>Death</td>
<td>Tempo-parietal</td>
<td>Total</td>
<td>6</td>
<td>377</td>
<td>NR</td>
<td>No met</td>
<td>Amplification (72%)</td>
<td>del (79%)</td>
</tr>
</tbody>
</table>

M: Male; F: Female; CR: complete response; DFP: disease free period; PR: partial response; P: progression; NR: no response; nd: no date.
Table 2. Common up-regulated in CD133+ vs. CD133- GBM cell patients.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Probe set</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL1A1</td>
<td>Collagen, type I, alpha 1</td>
<td>1556499_s_at</td>
</tr>
<tr>
<td>TGFB1</td>
<td>Transforming growth factor, beta-induced, 68kDa</td>
<td>201506_at</td>
</tr>
<tr>
<td>C10orf104</td>
<td>Chromosome 10 open reading frame 104</td>
<td>224667_x_at</td>
</tr>
<tr>
<td>UACA</td>
<td>Uveal autoantigen with coiled-coil domains and ankyrin repeats</td>
<td>236715_x_at</td>
</tr>
<tr>
<td>C14orf153</td>
<td>Chromosome 14 open reading frame 153</td>
<td>232814_x_at</td>
</tr>
<tr>
<td>OCIAD1</td>
<td>OCIA domain containing 1</td>
<td>239748_x_at</td>
</tr>
<tr>
<td>CMBL</td>
<td>Carboxymethylenebutenolidase homolog (Pseudomonas)</td>
<td>234981_x_at</td>
</tr>
<tr>
<td>PDE4C</td>
<td>Phosphodiesterase 4C, cAMP-specific (phosphodiesterase E1 dunce homolog, Drosophila)</td>
<td>206792_x_at</td>
</tr>
<tr>
<td>DBT</td>
<td>Dihydrolipoamide branched chain transacylase E2</td>
<td>205370_x_at</td>
</tr>
<tr>
<td>LRRFIP1</td>
<td>Leucine rich repeat (in FLII) interacting protein 1</td>
<td>211452_x_at</td>
</tr>
<tr>
<td>PRR11</td>
<td>Proline rich 11</td>
<td>219392_x_at</td>
</tr>
<tr>
<td>PGF</td>
<td>Placental growth factor, vascular endothelial growth factor-related protein</td>
<td>215179_x_at</td>
</tr>
<tr>
<td>ZNF160</td>
<td>Zinc finger protein 160</td>
<td>214715_x_at</td>
</tr>
<tr>
<td>SEPP1</td>
<td>Selenoprotein P, plasma, 1</td>
<td>237475_x_at</td>
</tr>
<tr>
<td>OPHN1</td>
<td>Oligophrenin 1</td>
<td>206323_x_at</td>
</tr>
<tr>
<td>COL1A2</td>
<td>Collagen, type I, alpha 2</td>
<td>202403_s_at</td>
</tr>
<tr>
<td>FLJ45803</td>
<td>FLJ45803 protein</td>
<td>238701_x_at</td>
</tr>
<tr>
<td>TMEFF2</td>
<td>Transmembrane protein with EGF-like and two follistatin-like domains 2</td>
<td>224321_at</td>
</tr>
<tr>
<td>SRGN</td>
<td>Serglycin</td>
<td>201859_at</td>
</tr>
</tbody>
</table>
Table 3. Common down-regulated in CD133+ vs CD133- GBM cell patients.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Probe set</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNRPA3</td>
<td>Heterogeneous nuclear ribonucleoprotein A3</td>
<td>1555653_at</td>
</tr>
<tr>
<td>ATP13A5</td>
<td>ATPase type 13A5</td>
<td>1553567_s_at</td>
</tr>
<tr>
<td>IQWD1</td>
<td>IQ motif and WD repeats 1</td>
<td>224373_s_at; 224372_at</td>
</tr>
<tr>
<td>TUBA1A</td>
<td>Tubulin, alpha 1a</td>
<td>209118_s_at</td>
</tr>
<tr>
<td>DPYS1L2</td>
<td>Dihydropyrimidinase-like 2</td>
<td>200762_at</td>
</tr>
<tr>
<td>RAB13</td>
<td>RAB13, member RAS oncogene family</td>
<td>202252_at</td>
</tr>
<tr>
<td>MATR3</td>
<td>Matrin 3</td>
<td>214363_s_at</td>
</tr>
<tr>
<td>DSTN</td>
<td>Destrin (actin depolymerizing factor)</td>
<td>201022_s_at</td>
</tr>
<tr>
<td>LDHB</td>
<td>Lactate dehydrogenase B</td>
<td>213564_x_at</td>
</tr>
<tr>
<td>UBC</td>
<td>Ubiquitin C</td>
<td>211296_x_at</td>
</tr>
<tr>
<td>CFL1</td>
<td>Cofilin 1 (non-muscle)</td>
<td>200021_at</td>
</tr>
<tr>
<td>LOC729548</td>
<td>Similar to ribosomal protein S2</td>
<td>203107_x_at</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>213453_x_at; 212581_x_at; 217398_x_at</td>
</tr>
<tr>
<td>LOC388076</td>
<td>Similar to ribosomal protein S8</td>
<td>200858_s_at</td>
</tr>
<tr>
<td>RPS19</td>
<td>Ribosomal protein S19</td>
<td>202649_x_at; 213414_s_at</td>
</tr>
<tr>
<td>TUT1</td>
<td>Terminal uridylyl transferase 1, U6 snRNA-specific</td>
<td>200689_x_at; 211345_x_at; 211927_x_at</td>
</tr>
<tr>
<td>RPSAP15</td>
<td>Ribosomal protein SA pseudogene 15</td>
<td>213801_x_at</td>
</tr>
<tr>
<td>LOC390860</td>
<td>Similar to 60S acidic ribosomal protein P0 (L10E)</td>
<td>211720_x_at; 208856_x_at; 201033_x_at</td>
</tr>
<tr>
<td>DDX17</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 17</td>
<td>208718_at</td>
</tr>
<tr>
<td>RPL13A</td>
<td>Ribosomal protein L13a</td>
<td>212790_x_at</td>
</tr>
<tr>
<td>GNB2L1</td>
<td>Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1</td>
<td>200651_at</td>
</tr>
<tr>
<td>LOC390861</td>
<td>Similar to cytoplasmic beta-actin</td>
<td>200801_x_at</td>
</tr>
</tbody>
</table>
1. Supplementary data -CGH-methods.

The DNA from each fresh-frozen sample was extracted with the standard phenol-chloroform method and reference DNA was prepared from human placenta of healthy donors. All DNAs were quantified using the Nanodrop spectrophotometer. DNA quality was assessed by the 260:280 ratio and its integrity by agarose gel ethidium bromide visualization. Genomic-wide analysis of DNA copy number changes of patient samples were performed using array-based CGH. Slides containing 3296 BACs were produced in “Centro de Investigación del Cáncer” (Salamanca, Spain). The particular bacterial artificial chromosome (BAC) and P-1 derived artificial chromosome (PAC) set used to produce this array is distributed to academic institutions by the Welcome Trust Sanger Institute (Cambridge, United Kingdom) and contains targets spaced at ≈ 1 Mb density over full genome, a set of subtelomeric sequences for each chromosome arm, and a few hundred probes selected for their involvement in oncogenesis. The clone content is available in the “Cytoview” windows of the Sanger Institute mapping database site, Ensembl (http://www.ensembl.org/). According to this database, clones were ordered along the chromosomes. These clones were isolated from their bacterial cultures with the relevant antibiotics and the DNA was extracted with the Welcome Trust Sanger Institute standard protocol (Cambridge, UK). 10 ng of DNA (BAC/PAC) was used as a template for three DOP-PCR. These products were ethanol precipitated and dissolved in distilled water. A minimum of three replicates per clone were printed by Microgrid II (Biorobotics) on each slide (Ultragaps Coated Slides, Corning) using the aqueous DMSO buffer as spotting solution. Briefly, for labelling reactions, 3 µg of nonamplified genomic DNA, test (DNA of tumor) and reference (DNA of placenta) were digested
separately with DpnII restriction enzyme (New England Biolabs, Beverly, MA). For microarray hybridization, the digested DNAs were separately labeled using random primers (Bioprimer labelling kit, Invitrogen) and Cy3-dCTP and Cy5-dCTP fluorescent dye to paired hybridization samples (Amersham Biosciences). The incorporated of the label nucleotide was quantified using the Nanodrop spectrophotometer. Labeled test and reference DNAs were mixed equitably, co-precipitated in presence of Cot-1 human DNA (Roche, Indianapolis, IN) with ethanol, washed, and resuspended in hybridization solution (50% Formamide, 10% Dextran sulfate, 2X standard saline citrate, 10 mM Tris pH 7.6, 2.7% sodium dodecyl sulfate and 10 µg/ µl of yeast tRNA). DNA mixtures were cohybridized to the arrays in the GENETAC for 48 hours at 37ºC according to the manufacturer’s recommended protocol. Images and signal intensities were acquired using GenePix4000B (Axon Instruments, Burlingame, CA) dual laser scanner in combination with GenePixPro4.0 (Axon Instruments) imaging software. For the array, 10 simultaneous hybridizations of normal male versus normal female and placenta: placenta (DNA reference) was performed to define the normal variation for the log₂ ratio.

The Cy5/Cy3 intensity ratios of every spot were converted into log₂ ratios. The log₂ ratio of each clone was normalized to the median log₂ ratio of the ten control hybridizations, after which the median of triplicate spots was calculated. Data from two-color hybridizations for both DNA were normalized using the corresponding GEPAS module DNMAD. Regions of copy number gain and loss for the BAC array-CGH data were identified by creating sample specific thresholds. The clones with log2 ratios above or below +/- a control sample’s threshold value (0.4) were considered as gains or losses, respectively. A BAC clone with a log₂ ratio of – 0.4 or less was defined as loss region and log₂ ratio of +0.4 or more was defined as gain region. Moreover, spots with
weak Cy3 or Cy5 below $R^2 < 0.2$ intensity were excluded. Furthermore, clones with standard deviation more than 0.3 of triplicate spots were also excluded. In total, approximately 10% of clones were excluded. Additionally we applied the binary segmentation method described by Olshen et al.

All data sets were carefully revised for frequently affected chromosomal sites of physiologic copy number polymorphisms (CNP). Therefore every clone on the array was compared with ‘Database of Genomic Variants’ (http://www.project.tcag.ca/variation). For unsupervised clustering analysis, we converted the relative ratio value for each BAC clone to a score of 1 (gain/amplified), 0 (no change), or -1 (loss) based data obtain by the binary segmentation method described by Olshen et al. and analyzed data with Cluster and TreeView (Multi Experiment Viewer 4.0) based on the average linkage method with the Pearson Uncentered metric correlation.

Statistical evaluation was carried out using the SPSS 15.0 statistical software. All P-values reported were two-sided and statistical significance was defined as P-values < 0.05.
Figure 4
Figure 6
Additional files provided with this submission:

Additional file 1: Table CGH data.pdf, 3401K