Stem Cell Reports
Report
OPEN ACCESS

# Precursor States of Brain Tumor Initiating Cell Lines Are Predictive of Survival in Xenografts and Associated with Glioblastoma Subtypes

Carlo Cusulin,[1] Charles Chesnelong,[1,4] Pinaki Bose,[2] Misha Bilenky,[2] Karen Kopciuk,[3,4] Jennifer A. Chan,[4,5,6] J. Gregory Cairncross,[1,4,6] Steven J. Jones,[2] Marco A. Marra,[2] H. Artee Luchman,[1,4] and Samuel Weiss[1,4,*]
[1]Hotchkiss Brain Institute, Department of Cell Biology and Anatomy, University of Calgary, Calgary, AB T2N 4N1, Canada
[2]British Columbia Cancer Agency Genome Sciences Centre, Vancouver, BC V5Z 4S6, Canada
[3]Department of Mathematics and Statistics
[4]Southern Alberta Cancer Research Institute
[5]Department of Pathology and Laboratory Medicine
[6]Department of Clinical Neurosciences
University of Calgary, Calgary, AB T2N 4N1, Canada
*Correspondence: weiss@ucalgary.ca
http://dx.doi.org/10.1016/j.stemcr.2015.05.010


## SUMMARY

In glioblastoma multiforme (GBM), brain-tumor-initiating cells (BTICs) with cancer stem cell characteristics have been identified and proposed as primordial cells responsible for disease initiation, recurrence, and therapeutic resistance. However, the extent to which individual, patient-derived BTIC lines reflect the heterogeneity of GBM remains poorly understood. Here we applied a stem cell biology approach and compared self-renewal, marker expression, label retention, and asymmetric cell division in 20 BTIC lines. Through cluster analysis, we identified two subgroups of BTIC lines with distinct *precursor states*, stem- or progenitor-like, predictive of survival after xenograft. Moreover, stem and progenitor transcriptomic signatures were identified, which showed a strong association with the proneural and mesenchymal subtypes, respectively, in the TCGA cohort. This study proposes a different framework for the study and use of BTIC lines and provides precursor biology insights into GBM.

## INTRODUCTION

Glioblastoma multiforme (GBM) is the most common adult primary brain tumor, with an extremely dismal prognosis and high rate of recurrence following standard therapy (Wen and Kesari, 2008). Populations of brain tumor initiating cells (BTICs), which express many of the properties of cancer stem cells (CSCs), have been identified in GBM (Singh et al., 2004). BTICs display CSC characteristics of long-term self-renewal, multilineage differentiation, and tumorigenicity, which may be collectively referred to as cancer "stemness" features (reviewed in Clevers, 2011). BTIC lines have become a valuable tool for modeling GBM (Westphal and Lamszus, 2011) and for the development of experimental therapeutics (Luchman et al., 2014).

In a recent review, Stopschinski and colleagues (Stopschinski et al., 2013) argued that a consensus standardization of the BTIC model will be essential to harness its full cell biologic and experimental therapeutic potential in a heterogeneous disease such as GBM. Initial studies aimed at examining single stemness characteristics, such as dye retention (Deleyrolle et al., 2011) and self-renewal (Campos et al., 2014), have provided insights as to the relationship between BTIC lines and tumor heterogeneity. Nevertheless, a comprehensive precursor cell analysis of multiple BTIC lines has yet to be performed.

It has been suggested (Binda et al., 2014; Vescovi et al., 2006) that understanding BTICs may best be achieved by examining them in relation to normal neural stem cells (NSCs). NSCs are organized in a hierarchical manner, with a quiescent cell (Codega et al., 2014) at the apex. Through asymmetric cell division (Shen et al., 2002; Sun et al., 2005), the NSC self-renews, giving rise to an NSC and transit-amplifying progenitor that proliferates rapidly and produces the bulk of the short-term neurogenic population (Doetsch et al., 1999a). Moreover, BTICs are most often cultured with techniques originally developed for NSCs (Reynolds and Weiss, 1992) and share many stemness characteristics with NSCs, including the expression of markers such as NESTIN, CD133, and SOX2 (Gangemi et al., 2009; Singh et al., 2004), label retention (Deleyrolle et al., 2011), asymmetric cell division (Lathia et al., 2011), and self-renewal (Campos et al., 2014). However, a single study employing multiple stemness characteristics to examine a large set of heterogeneous BTIC lines has yet to be reported.

Hypothesizing that heterogeneity among BTIC lines is related to differences in stemness features, we used a comprehensive and integrated stem cell biology approach to characterize 20 lines. We report the identification of two groups, with distinct *precursor states*, predictive of survival after xenografts and exhibiting strong gene expression profile associations with relevant subtypes of the Cancer Genome Atlas (TCGA).

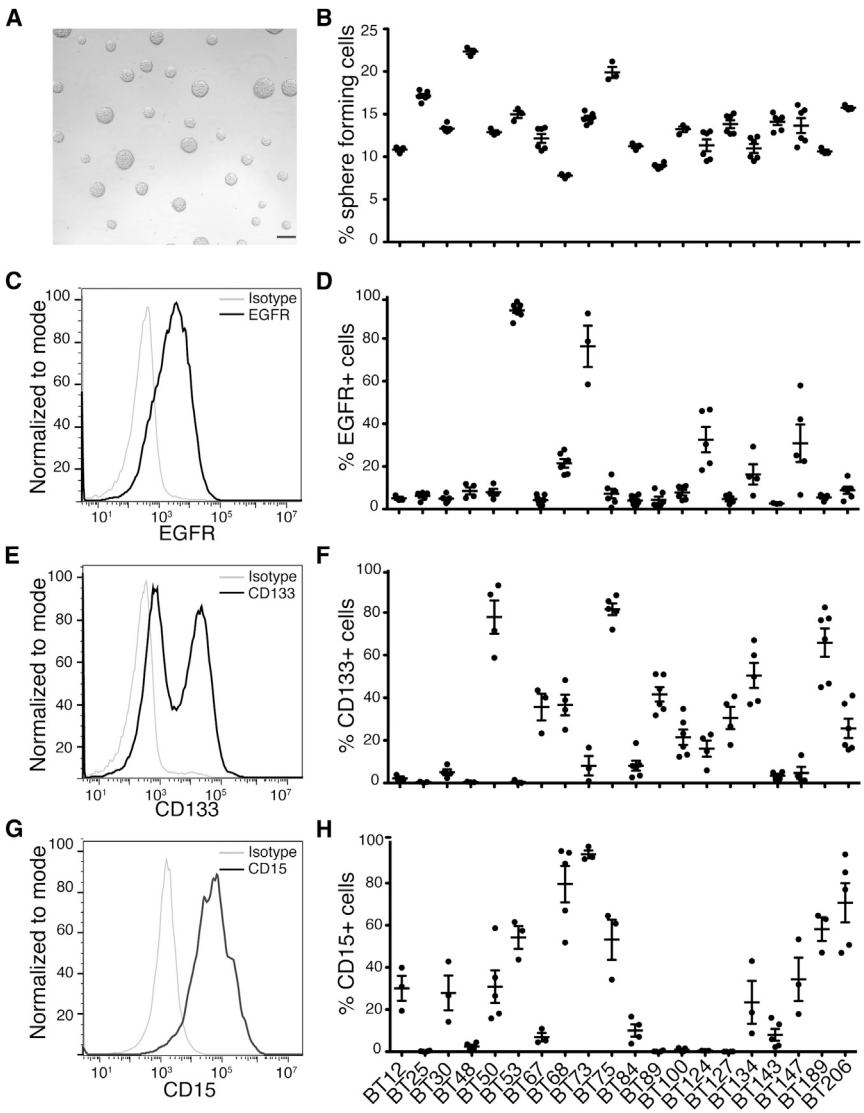


**Figure 1. Complex Self-Renewal and Marker Expression Patterns in BTICs**

(A) Representative phase-contrast micrograph shows BTIC spheres in self-renewal experiments.

(B) Quantification of sphere formation in lines is shown.

(C–H) Expression of cell surface markers common to BTICs and NSCs, EGFR (C and D), CD133 (E and F), and CD15 (G and H), was analyzed by FACS and quantified (D, F, and H). (C), (E), and (G) show examples of the FACS plots for BT147, BT134, and BT189. Scale bar, 100 μm.

Error bars represent SEM (B, D, F, and H). Each dot represents a single experiment. See also Table S2.

## RESULTS

### Heterogeneous Self-Renewal Ability and Marker Expression within BTIC Lines

A comprehensive phenotypic characterization of 20 GBM patient-derived BTIC lines, focusing on features of NSC biology, was performed. The BTIC lines express, although to variable degrees, NSC markers (NESTIN, SOX2, MUSASHI1, VIMENTIN, CD15, and CD133) (Figure S1); differentiate to express neuronal, astrocytic, and oligodendrocytic markers (Figure S1); and are tumorigenic, suggesting that they contain true CSCs. Furthermore, they exhibit mutations/deletions typically found in GBM (*EGFR*, *TP53*, *PTEN*, *NF1*, and *CDKN2A*) (Table S1).

BTIC self-renewal capacity was assessed by calculating the percentage of cells capable of giving rise to a new sphere after dissociation. A high degree of variability was observed across the lines, with sphere-forming cells ranging from 7.8% ± 0.2% (mean ± SEM) to 22.3% ± 0.3% (Figures 1A and 1B; inter-line differences were significant, $p < 0.0001$, one-way ANOVA). Expression of NSC and BTIC markers EGFR, CD133, and CD15 was analyzed by fluorescence-activated cell sorting (FACS) (Figures 1C–1H), and a remarkable inter-line diversity was again observed, with the percentage of positive cells ranging from 2.6% ± 0.2% to 93.3% ± 1.1% for EGFR, from 0.4% ± 0.2% to 81.5% ± 2.3% for CD133, and from 0.2% ± 0.1% to 93.7% ± 1.9% for CD15 (for all, $p < 0.0001$, one-way ANOVA).

### Quiescence and Asymmetry in BTICs

Quiescence, the reversible exit from the cell cycle, is a process common to NSCs and BTICs. Due to slower cellular

turnover, quiescent cells can be detected by the retention of a dye or label (Cheung and Rando, 2013). The abundance of label-retaining cells was assayed with carboxyfluorescein diacetate succinimidylester (CFSE). While a large number of cells were CFSE negative, a substantial population remained clearly labeled after growth in standard conditions (Figures 2A and 2C). Cells that divided only once or twice, as determined by the intensity of fluorescence, were defined as $CFSE^{HIGH}$. These cells were functionally quiescent, as confirmed by the ability of both $CFSE^{HIGH}$- and $CFSE^{LOW}$-sorted cells to form new spheres (Figure S2). Analysis of the percentage of label-retaining cells in all BTIC lines showed different frequencies (Figure 2E), ranging from 0.4% ± 0.1% to 5.2% ± 1.1% ($p < 0.0001$, one-way ANOVA). The presence of label-retaining cells in each line showed a weak inverse correlation to sphere formation and a positive correlation with the percentage of CD133+ cells, but not with the other stemness parameters (Table S2).

Asymmetric cell division is a key step in the homeostatic self-renewal of stem cells and has been previously described in BTICs (Lathia et al., 2011). Using key markers of NSC biology, we analyzed the frequency of asymmetry in BTIC lines by assaying patterned asymmetric distribution of NUMB, EGFR, NESTIN, and GFAP, after cell division. Single cells were plated at low density, synchronized, and analyzed 18–20 hr after the cell division block was removed. Clear examples of asymmetric cell division (Figure 2F) could be found for all four markers. Similar results were observed using additional NSC/BTIC markers (CD133, SOX2, MUSASHI1, and CD15; data not shown). Quantification showed that asymmetric cell division, while very common in some BTIC lines (up to 46.4% ± 5.4% of couples), was virtually non-existent in others (Figure 2G). The inter-line differences were significant for all markers ($p < 0.0001$, one-way ANOVA). Interestingly, the asymmetric division rate did not correlate with the expression of each individual marker, except for *GFAP* (Figure S3).

## Cluster Analysis Reveals Two Distinct Precursor State Groups

To understand parameters relevant to the biology of BTIC lines, we first attempted to find direct correlations between all pairs of assays performed (Table S2). Although some associations were found, no unifying pattern was identified. Instead, using hierarchical clustering with Manhattan distance and Ward's agglomeration method, we identified two clusters (Figure 3A). K-means clustering with Manhattan distance confirmed a two-cluster solution, and the same membership in these two clusters was found (data not shown). We next looked at which variables defined the clustering of BTIC lines in each of these groups. Cluster A was characterized ($Z$ scores > 1.96) by stem cell features, such as higher levels of asymmetry, label-retaining cells, and CD133-expressing cells, suggesting that these BTICs may be more akin to classically defined stem cells (stem-like: SL). Cluster B was defined by higher sphere formation rate, resembling transit-amplifying progenitors found in normal neurogenesis (progenitor-like: PL). Given the resemblance of SL and PL cells to the NSC biology counterparts, we defined these features of the two BTIC groups as a difference in precursor state. Interestingly, neither of the two clusters associated with specific molecular alterations in any of the genes analyzed (Fisher's exact test, $p > 0.05$ for each mutation).

## Precursor States Associate with Survival in Xenografts

We next examined whether SL or PL features of BTICs play a defining role in tumor formation, by implanting 15 lines in immunocompromised mice. All BTIC lines were tumorigenic (Figures 3B and 3C), but, notably, animals xenografted with SL lines survived significantly longer than those implanted with PL lines (average median survival SL = 183.7 ± 24.5 versus PL = 67.4 ± 11.4 days, $p < 0.0001$, log-rank test; SL n = 42, PL n = 59) (Figure 3D; detailed survival times in Table S3).

To test whether the survival difference was due to variations in BTIC proliferation rates, we measured the growth kinetics of 16 lines in vitro (eight for each group). Although cells in SL lines divided slower (doubling time 4.58 ± 0.36 in SL versus 3.57 ± 0.19 days in PL, $p < 0.0001$), survival in vivo was not correlated with the mean doubling time observed in culture (Figure 3E; $p = 0.41$, $R = 0.25$), indicating that shorter survival was not solely due to a difference in proliferation rate. In contrast, median survival showed an inverse correlation with the abundance of sphere-forming cells (Figure 3F; $p = 0.03$, $R = -0.56$).

## Transcriptome Analysis Identifies an Association between Precursor States and GBM Subtypes

RNA sequencing (RNA-seq) was performed on seven BTIC lines from each group. Unsupervised clustering based on the GBM subtype transcriptomic signatures (Verhaak et al., 2010) did not distinguish proneural, mesenchymal, classical, or neural BTICs. We then performed differential expression analysis and found that 1,110 genes were significantly upregulated in SL-BTICs and 269 genes were upregulated in PL-BTICs (Figure 4A). We then derived a signal-to-noise measure ($D_{SN}$) to eliminate genes with high SD within either BTIC group, and ultimately we selected the top tenth percentile of differentially expressed genes based on $D_{SN}$, generating a signature of 136 genes (Figure S4).

To further understand the relevance of this signature in the disease context, we used the publicly available GBM transcriptome dataset (Cancer Genome Atlas Research

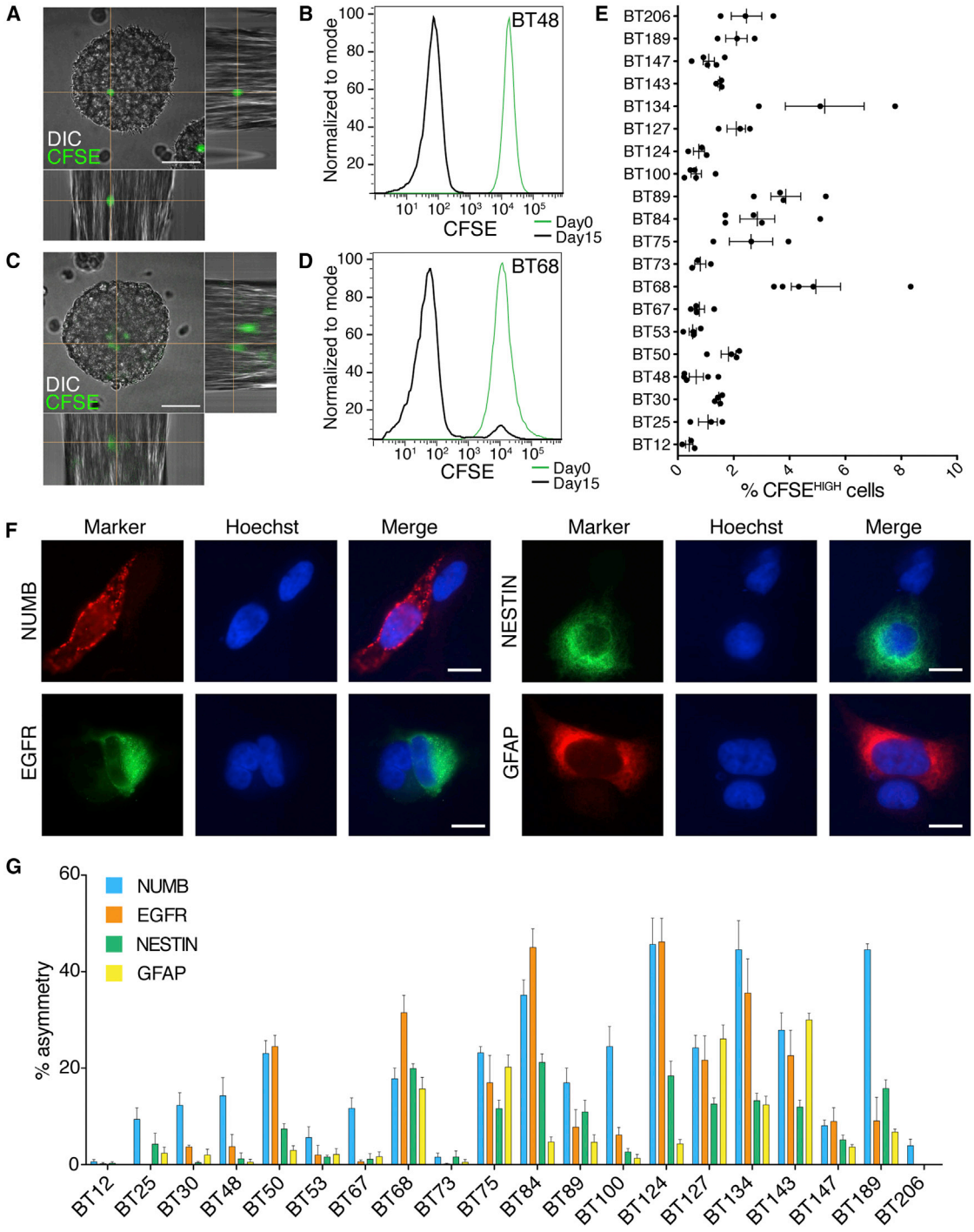


**Figure 2. Heterogeneous Representation of Quiescent and Asymmetrically Dividing Cells in BTIC Lines**

(A–E) CFSE-retaining cells (green in A and C) and by FACS (B and D; day 0, freshly labeled cells). Percentage of label-retaining cells is shown in (E); individual dots represent independent experiments; error bars represent SEM. DIC, differential interference contrast.

(F) Examples of asymmetric distribution of the markers EGFR, NUMB, NESTIN, and GFAP are shown.

(G) Quantification of asymmetric cell division across the 20 BTIC lines. Each bar is the average of at least four independent experiments. Errors bars represent SEM.

Scale bars, 50 μm (A and C) and 10 μm (F). See also Figure S2 and Table S2.

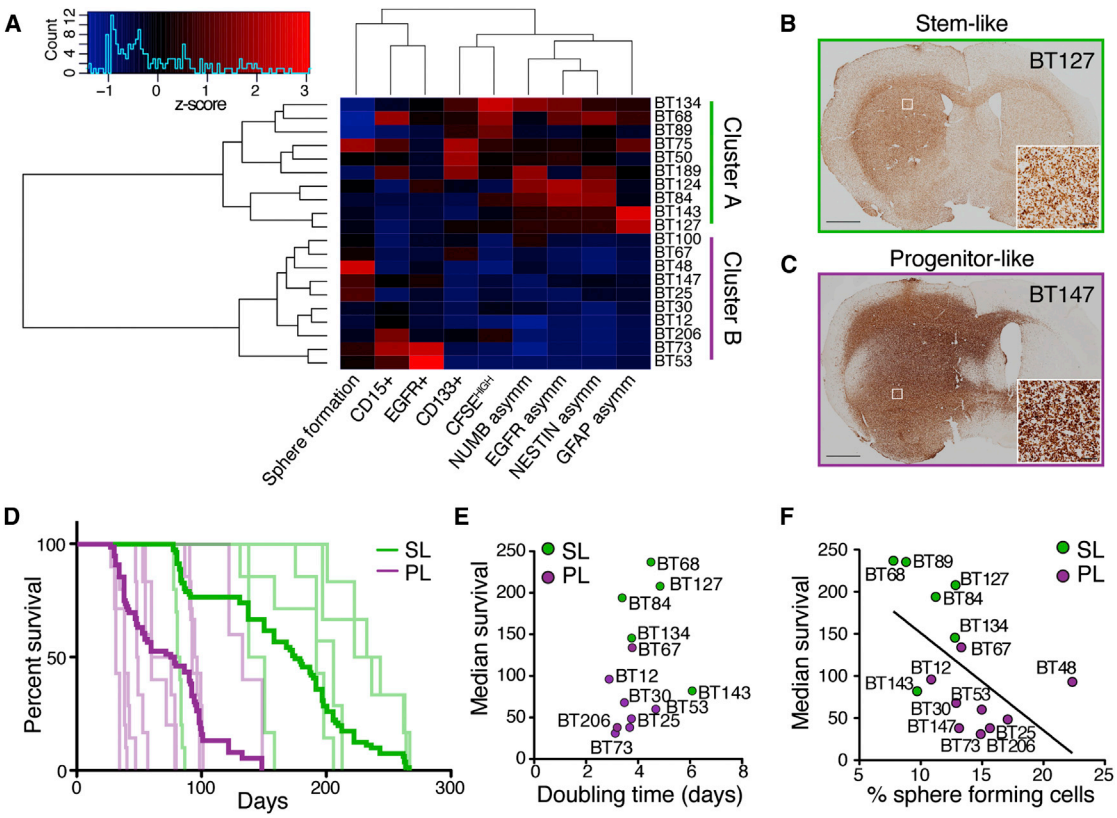


**Figure 3. Cluster Analysis Defines Precursor States of BTICs, Associated with Survival In Vivo**

(A) Heatmap representing unsupervised clustering of 20 BTIC lines based on the parameters studied above. The lines within the two clusters were found to be SL or PL.

(B and C) Representative images of tumor formation after xenograft. BTICs were visualized by human-specific nucleolin staining.

(D) Kaplan-Meier survival curves represent the individual BTICs (light lines) and the combined for each group (bold lines) (n = 5–10 mice per cell line; total SL, n = 42; total PL, n = 59; $p < 0.0001$ in log-rank test statistic; see also Table S3 for specific survival and number of animals per line).

(E) The growth rate of 12 BTIC lines in vitro was not correlated to median survival (Pearson's correlation, $p = 0.44$, $R = 0.23$; for each line, n = 3 independent experiments).

(F) The frequency of sphere-forming cells showed an inverse correlation to the median survival in 15 BTIC lines (Pearson's correlation, $p = 0.03$, $R = -0.56$; sphere formation data from experiments reported in Figure 1).

Scale bars, 1 mm (B and C) and 50 μm (insets).

Network, 2008). To match the precursor state profiles with samples in the TCGA dataset, we calculated $Z$ scores for genes overexpressed in SL- and PL-BTIC lines (referred to as SL- and PL-genes, respectively) for each GBM patient in the dataset. A combined average $Z$ score, hereafter referred to as $Z'$, was calculated by subtracting the average PL-gene $Z$ score from the average SL-gene $Z$ score, for each TCGA patient. $Z'$ was used as a summary measure of similarity to either the SL or the PL group. Notably, in the 152 TCGA samples that were already classified into the four GBM subtypes described previously (Verhaak et al., 2010), there was a significant difference in $Z'$ between proneural and mesenchymal tumors ($p < 0.0001$, one-way ANOVA with Bonferroni post hoc analysis) (Figure 4B).

The patient samples were then separated into three groups, based on $Z'$ values ($Z'$ cutoff 0.1 and −0.1; groups I, II, and III, n = 44, 40, and 68, respectively; Figure 4C). Therefore, for the 136-gene profile, group I ($Z' > 0.1$) had a pattern of expression similar to SL lines and group III ($Z' < -0.1$) similar to PL lines. Group I was predominantly composed of proneural samples (45.5%) and group III had mostly mesenchymal (51.5%) ($\chi^2 = 30.20$, df = 6, $p < 0.0001$); no specific subtype was over-represented in group II (Figure 4D).

To further confirm the associations between groups defined by our 136-gene signature and GBM subtypes, we performed gene set enrichment analysis (GSEA) of genes that define the four GBM subtypes (Verhaak et al., 2010) within the TCGA dataset (Figure 4E). As expected, we

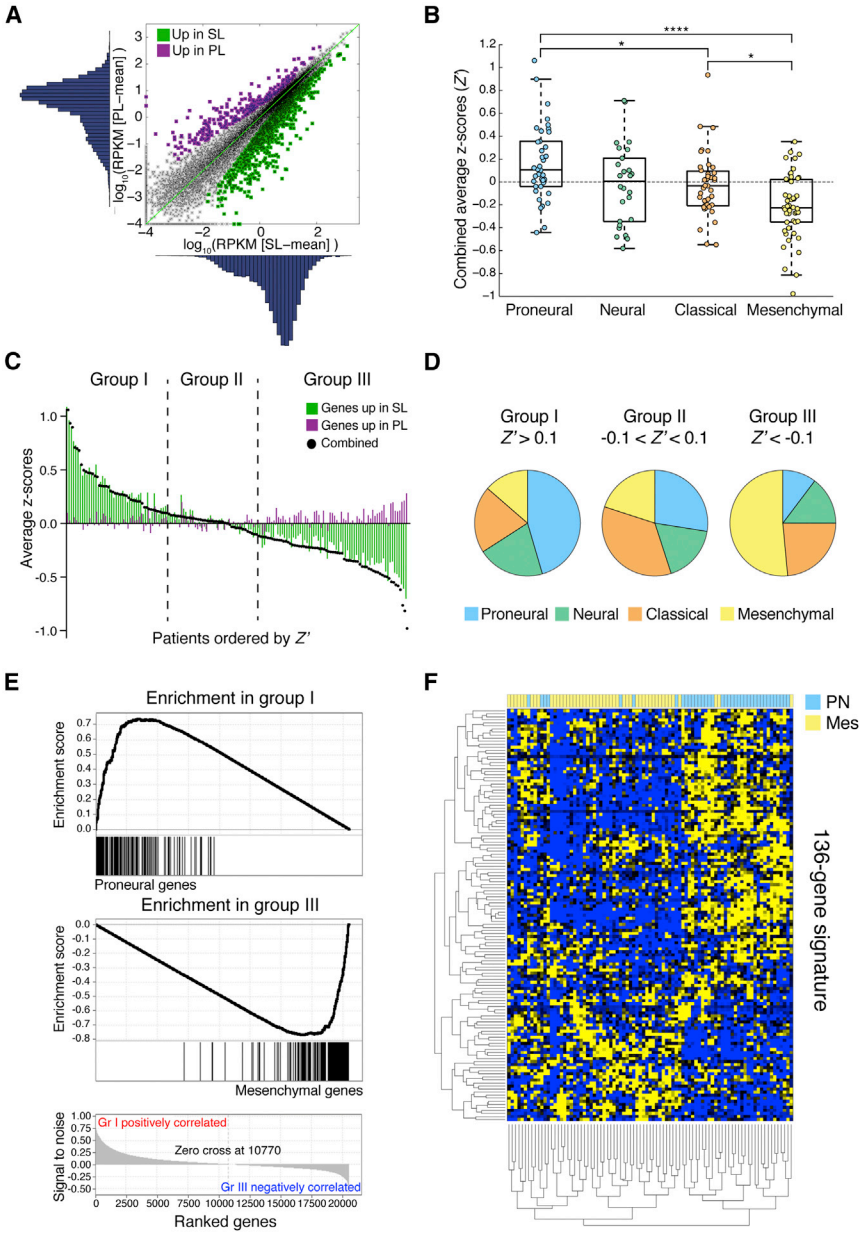


**Figure 4. Transcriptome Analysis Reveals an Association between Precursor States and Proneural and Mesenchymal Subtypes**

(A) Initial selection of genes differentially expressed in our two groups of seven lines each is shown. RPKM, reads per kilobase per million.

(B) Combined average *Z* scores (*Z'*) of TCGA patient samples divided by GBM subtype are shown. *p < 0.05, ****p < 0.0001.

(C) TCGA samples were ranked based on *Z'* and divided into three groups ($Z' > 0.1$, $-0.1 < Z' < 0.1$, and $Z' < -0.1$).

(D) Pie charts represent the distribution of GBM subtypes within the three groups.

(E) Representative gene set enrichment analysis comparing groups I and III for the enrichment of GBM subtype gene sets. All genes are ranked based on the expression in the dataset (bottom; subtype-specific genes are correlated to the ranking, resulting in the enrichment score. See also Table S4 for complete analysis of the three groups.

(F) Clustergram shows the expression of the 136-gene signature in proneural and mesenchymal samples.

TCGA data (B–F) were from 152 samples. See also Figure S4.

observed a strong enrichment of proneural-signature genes in group I ($p < 0.05$, false discovery rate [FDR] < 0.10). Conversely, the mesenchymal gene set was enriched in group III ($p < 0.05$, FDR < 0.05). A complete list of associations is reported in Table S4. Patient survival was not significantly different across the three groups ($p = 0.25$, log-rank; data not shown).

Finally, we asked whether our 136-gene precursor state signature could effectively stratify the proneural (n = 38) and mesenchymal (n = 49) patients in the TCGA dataset. In unsupervised clustering (Figure 4F), our signature segregated proneural and mesenchymal GBM samples, further confirming the relationship between these subtypes and BTIC precursor states. SL genes significantly overlapped with the cluster of genes overexpressed in proneural patients (hypergeometric p = 0.001). Similarly, PL genes overlapped significantly with those overexpressed in mesenchymal patients (hypergeometric p = 0.001).

## DISCUSSION

By applying cluster analysis to the study of stemness properties of BTICs, we report the emergence of two distinct

groups of patient-derived BTIC lines. Utilizing concepts taken from NSC biology, we defined these two groups as SL and PL. Mice xenografted with PL lines showed strikingly shorter median survival, which was not simply due to their higher proliferation rate. Given that SL- and PL-BTICs both contain CSCs, yet show remarkable differences in vitro and in vivo, we propose that they differ in precursor state. Our model provides a conceptual link between BTICs and NSCs, which may be critical in developing a better understanding of GBM. This is highlighted by the fact that the gene expression signatures derived from SL and PL lines were found to associate with GBM subtypes, further underscoring the differences between the two groups of BTIC lines.

The CSC hypothesis hinges on the presence of a hierarchical organization within a tumor (Kreso and Dick, 2014), reminiscent of normal stem cell biology, which results in a heterogeneous cell population. Furthermore, diverse genetic and mutational backgrounds, as well as differences in disease evolution, can result in heterogeneity among CSCs (Meacham and Morrison, 2013). BTICs are a widely used model for GBM but show great inter-line heterogeneity (Stopschinski et al., 2013), mirroring the inherent heterogeneity of GBM, recently observed even at the single-cell level within and between patient tumors (Patel et al., 2014).

Others previously have used in vitro growth characteristics in attempts to identify subgroups among BTIC lines. Günther and colleagues (Günther et al., 2008) classified BTICs based on free-floating or adherent growth, finding that they had different gene expression and in vivo behavior. Clonogenicity, measured in a colony formation assay, was the basis of the classification proposed by Campos et al. (2014). Interestingly, this group found that highly clonogenic lines had fewer label-retaining cells and were more aggressive, a pattern resembling our findings. Recently, Mao and colleagues (Mao et al., 2013) identified and characterized two mutually exclusive glioma stem cell subtypes. Transcriptomic analysis distinguished proneural and mesenchymal groups, and the latter displayed more aggressive phenotypes both in vitro and in vivo as well as increased resistance to radiation compared with the proneural group. These studies offer interesting insights in BTIC biology and help strengthen our conclusions. Importantly, by taking a multi-dimensional stem cell biology approach, our study also allowed the identification of two BTIC subgroups that appear highly relevant to the subtypes in the TCGA dataset.

Expression profiles from tumor tissue and derived cell lines are extremely different, thus the current transcriptomic signature for GBM subtyping may not apply in vitro. Using our precursor state signature, we were able to identify an association with the proneural and mesenchymal subtypes. Interestingly, it has been proposed that, upon recurrence, GBMs tend to shift toward the mesenchymal subtype (Phillips et al., 2006), suggesting that the latter might be downstream in hierarchy. At the same time, proneural GBMs show a worse response to standard therapeutic treatment (Verhaak et al., 2010), as would be expected for stem-like cells. Moreover, Ozawa et al. (2014) recently showed that most GBMs might derive from a common proneural-like precursor. These reports suggest that the mesenchymal subtype may be downstream of the proneural. Our study provides a possible interpretation of these data, in the context of stem cell biology. SL-BTICs would be less dysregulated, maintaining more normal stem cell features; this would be reflected in longer survival of xenografted mice and in the fact that they are associated with the proneural subtype. In addition, the SL lines may be more CSC-like and, hence, possibly have a larger fraction of tumor-initiating cells; while displaying less aggressive growth, they may be more tumorigenic in vivo. This premise needs be confirmed by xenografting smaller numbers of cells in tumorigenicity studies. Furthermore, our data suggest that the precursor state model will be essential to test hypotheses in cancer cell lineage and tumor evolution, currently assessed with mouse models (Ozawa et al., 2014).

The precursor state model offers the conceptual basis for developing more targeted GBM cell biology and experimental therapeutic studies. SL lines may be more useful to study tumor recurrence and hierarchy, while PL-BTICs might be best employed to investigate the pathways involved in rapid and aggressive growth. It is conceivable that the precursor state also could influence response to cytotoxic treatment in BTICs, as it does in NSCs. For example, neural progenitors in adult mouse brain are ablated easily by treatment with the chemotherapy agent cytarabine, while the quiescent stem cells are able to survive and repopulate the subventricular zone (Doetsch et al., 1999b).

Failure of several targeted therapies, both in experimental and clinical settings (Ohka et al., 2012), is a grim reminder of the complexity and heterogeneity of GBM. Here we provide a fresh classification of BTICs using a classic stem cell biology approach combined with cluster analysis. Improved characterization of BTICs, as a model of GBM, will allow for a better understanding of their response to treatment, both in vitro and in vivo, and ultimately open the way for findings of high relevance to GBM therapeutics.

## EXPERIMENTAL PROCEDURES

### BTIC Cultures and Growth Assays

BTIC lines (n = 20) were isolated, established, and maintained as described previously (Kelly et al., 2009). Lines were used within

25–30 passages of establishment from primary cells. For sphere formation assays, 1,000 cells per well were plated (n = 6) and spheres were counted when they reached a diameter of 200–250 μm. For growth curves, cells were seeded at specific densities and alamarBlue (Life Technologies) conversion was measured every day over the course of 2 weeks.

### Flow Cytometry

Single-cell suspensions were labeled with appropriately conjugated antibodies and analyzed with an ATTUNE flowcytometer (Life Technologies). For label retention studies, single cells were stained with 1 μM CFSE (Life Technologies) and grown in standard conditions.

### Asymmetry Studies

Asymmetry was analyzed by plating cells at low density, synchronizing them with nocodazole or thymidine (Sigma); the block was released and the cells were fixed and stained after 18–20 hr. Asymmetric distribution of markers was counted in at least 80 couples of daughter cells.

### In Vivo Studies

Cells (100,000) were implanted in the right striatum of 6- to 8-week-old female C17/SCID mice (n = 6–10 per line). Animals were euthanized upon demonstration of overt disease symptoms.

### RNA-Seq and Analyses

RNA-seq was performed using HiSeq2000 (Illumina). The list of differentially expressed genes was derived using DEfine v.0.9.2 (FDR cutoff of 0.01). Ranking of the TCGA samples was performed by subtracting the average $Z$ scores of SL genes by the average $Z$ scores of PL genes ($Z'$), then dividing the dataset in three groups based on $Z'$ (TCGA data Ver.2014-08-28).

### Statistical Analyses

All data reported for in vitro experiments are representative of at least three independent replicates and are illustrated in scatterplots or bar graphs, including mean ± SEM. Statistical analyses and graphing were performed with Prism 6.0 (GraphPad), SPSS (IBM), and R (version 3.0.1).

## ACCESSION NUMBERS

The accession number for the RNA-seq data reported in this paper is NCBI Sequence Read Archive (SRA): SRP057855.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.05.010.

## AUTHOR CONTRIBUTIONS

C. Cusulin performed the experiments. C. Cusulin and C. Chesnelong designed the study, analyzed the data, and wrote the paper. P.B., M.B., S.J.J., and M.A.M. performed the RNA-seq and analyzed the data. K.K. performed the cluster analysis. J.A.C. and J.G.C. contributed to the acquisition and analysis of samples. H.A.L. and S.W. supervised the study and wrote the paper.

## ACKNOWLEDGMENTS

We thank Rozina Hassam, Dorothea Livingstone, and Orsolya Cseh for technical assistance; the Calgary Brain Tumour and Tissue Bank for providing patient samples; and the sequencing and bioinformatics teams at the Michael Smith Genome Science Center for data processing. Funding to S.W. and H.A.L. was from the Stem Cell Network of Canada and to S.W., H.A.L., J.A.C., J.G.C., S.J., and M.A.M. was from the Terry Fox Research Institute.

Received: February 13, 2015
Revised: May 14, 2015
Accepted: May 18, 2015
Published: June 18, 2015

## REFERENCES

Binda, E., Reynolds, B.A., and Vescovi, A.L. (2014). Glioma stem cells: turpis omen in nomen? (The evil in the name?). J. Intern. Med. *276*, 25–40.

Campos, B., Gal, Z., Baader, A., Schneider, T., Sliwinski, C., Gassel, K., Bageritz, J., Grabe, N., von Deimling, A., Beckhove, P., et al. (2014). Aberrant self-renewal and quiescence contribute to the aggressiveness of glioblastoma. J. Pathol. *234*, 23–33.

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061–1068.

Cheung, T.H., and Rando, T.A. (2013). Molecular regulation of stem cell quiescence. Nat. Rev. Mol. Cell Biol. *14*, 329–340.

Clevers, H. (2011). The cancer stem cell: premises, promises and challenges. Nat. Med. *17*, 313–319.

Codega, P., Silva-Vargas, V., Paul, A., Maldonado-Soto, A.R., Deleo, A.M., Pastrana, E., and Doetsch, F. (2014). Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. Neuron *82*, 545–559.

Deleyrolle, L.P., Harding, A., Cato, K., Siebzehnrubl, F.A., Rahman, M., Azari, H., Olson, S., Gabrielli, B., Osborne, G., Vescovi, A., and Reynolds, B.A. (2011). Evidence for label-retaining tumour-initiating cells in human glioblastoma. Brain *134*, 1331–1343.

Doetsch, F., Caillé, I., Lim, D.A., García-Verdugo, J.M., and Alvarez-Buylla, A. (1999a). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. Cell *97*, 703–716.

Doetsch, F., García-Verdugo, J.M., and Alvarez-Buylla, A. (1999b). Regeneration of a germinal layer in the adult mammalian brain. Proc. Natl. Acad. Sci. USA *96*, 11619–11624.

Gangemi, R.M., Griffero, F., Marubbi, D., Perera, M., Capra, M.C., Malatesta, P., Ravetti, G.L., Zona, G.L., Daga, A., and Corte, G. (2009). SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. Stem Cells *27*, 40–48.

Günther, H.S., Schmidt, N.O., Phillips, H.S., Kemming, D., Kharbanda, S., Soriano, R., Modrusan, Z., Meissner, H., Westphal, M.,

and Lamszus, K. (2008). Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. Oncogene *27*, 2897–2909.

Kelly, J.J., Stechishin, O., Chojnacki, A., Lun, X., Sun, B., Senger, D.L., Forsyth, P., Auer, R.N., Dunn, J.F., Cairncross, J.G., et al. (2009). Proliferation of human glioblastoma stem cells occurs independently of exogenous mitogens. Stem Cells *27*, 1722–1733.

Kreso, A., and Dick, J.E. (2014). Evolution of the cancer stem cell model. Cell Stem Cell *14*, 275–291.

Lathia, J.D., Hitomi, M., Gallagher, J., Gadani, S.P., Adkins, J., Vasanji, A., Liu, L., Eyler, C.E., Heddleston, J.M., Wu, Q., et al. (2011). Distribution of CD133 reveals glioma stem cells self-renew through symmetric and asymmetric cell divisions. Cell Death Dis. *2*, e200.

Luchman, H.A., Stechishin, O.D., Nguyen, S.A., Lun, X.Q., Cairncross, J.G., and Weiss, S. (2014). Dual mTORC1/2 blockade inhibits glioblastoma brain tumor initiating cells in vitro and in vivo and synergizes with temozolomide to increase orthotopic xenograft survival. Clin. Cancer Res. *20*, 5756–5767.

Mao, P., Joshi, K., Li, J., Kim, S.H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U.R., Benos, P.V., Smith, L., et al. (2013). Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. USA *110*, 8644–8649.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Ohka, F., Natsume, A., and Wakabayashi, T. (2012). Current trends in targeted therapies for glioblastoma multiforme. Neurol. Res. Int. *2012*, 878425.

Ozawa, T., Riester, M., Cheng, Y.K., Huse, J.T., Squatrito, M., Helmy, K., Charles, N., Michor, F., and Holland, E.C. (2014). Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell *26*, 288–300.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Phillips, H.S., Kharbanda, S., Chen, R., Forrest, W.F., Soriano, R.H., Wu, T.D., Misra, A., Nigro, J.M., Colman, H., Soroceanu, L., et al. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell *9*, 157–173.

Reynolds, B.A., and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science *255*, 1707–1710.

Shen, Q., Zhong, W., Jan, Y.N., and Temple, S. (2002). Asymmetric Numb distribution is critical for asymmetric cell division of mouse cerebral cortical stem cells and neuroblasts. Development *129*, 4843–4853.

Singh, S.K., Hawkins, C., Clarke, I.D., Squire, J.A., Bayani, J., Hide, T., Henkelman, R.M., Cusimano, M.D., and Dirks, P.B. (2004). Identification of human brain tumour initiating cells. Nature *432*, 396–401.

Stopschinski, B.E., Beier, C.P., and Beier, D. (2013). Glioblastoma cancer stem cells—from concept to clinical application. Cancer Lett. *338*, 32–40.

Sun, Y., Goderie, S.K., and Temple, S. (2005). Asymmetric distribution of EGFR receptor during mitosis generates diverse CNS progenitor cells. Neuron *45*, 873–886.

Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al.; Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110.

Vescovi, A.L., Galli, R., and Reynolds, B.A. (2006). Brain tumour stem cells. Nat. Rev. Cancer *6*, 425–436.

Wen, P.Y., and Kesari, S. (2008). Malignant gliomas in adults. N. Engl. J. Med. *359*, 492–507.

Westphal, M., and Lamszus, K. (2011). The neurobiology of gliomas: from cell biology to the development of therapeutic approaches. Nat. Rev. Neurosci. *12*, 495–508.

**Stem Cell Reports**

**Supplemental Information**

# Precursor States of Brain Tumor Initiating Cell Lines Are Predictive of Survival in Xenografts and Associated with Glioblastoma Subtypes

**Carlo Cusulin, Charles Chesnelong, Pinaki Bose, Misha Bilenky, Karen Kopciuk, Jennifer A. Chan, J. Gregory Cairncross, Steven J. Jones, Marco A. Marra, H. Artee Luchman, and Samuel Weiss**

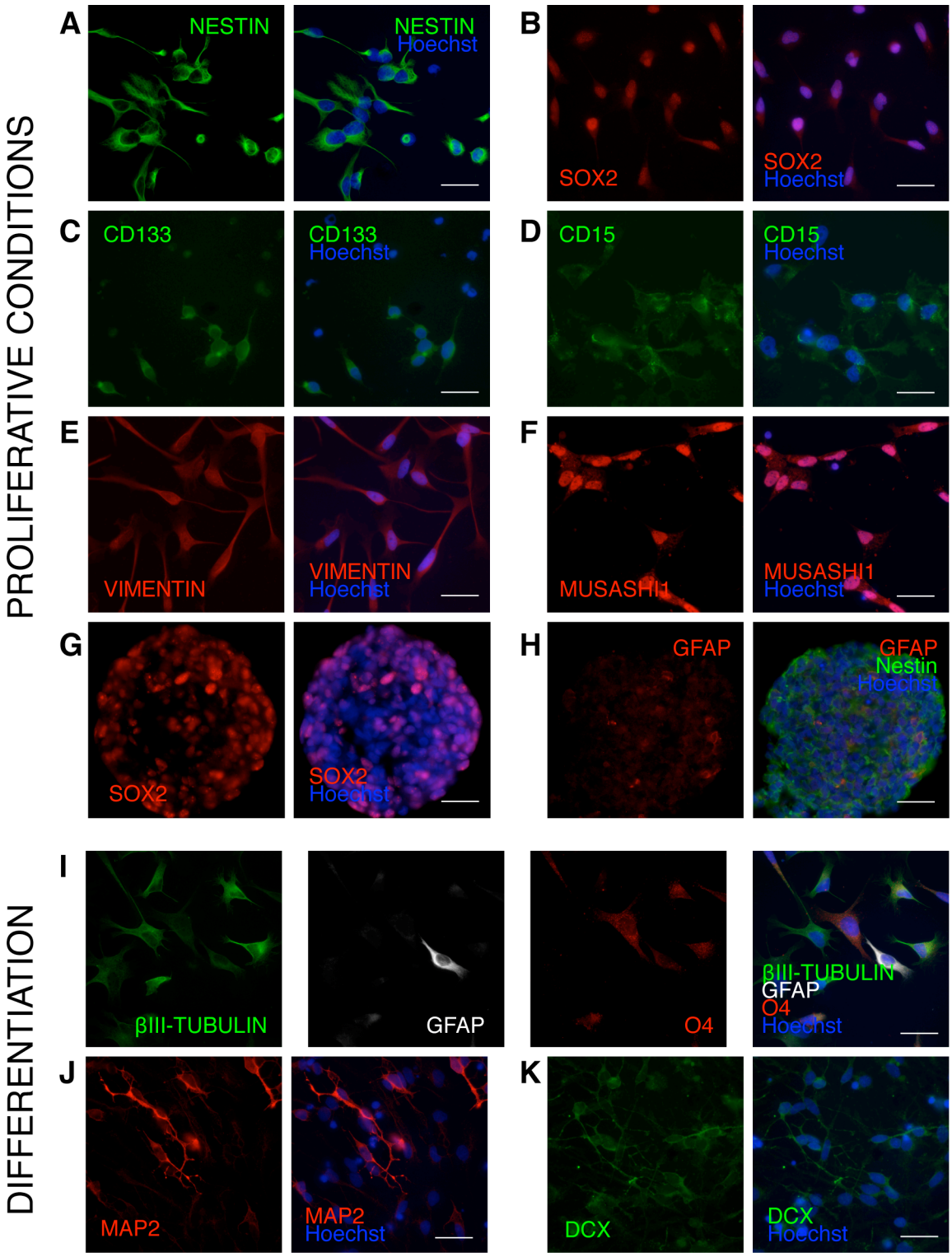


**Figure S1, related to Experimental procedures. Characterization of BTICs**

**(A-F)** Representative pictures of immunostaining for neural stem cell marker expression:

NESTIN, SOX2, CD133, CD15, VIMENTIN and MUSASHI1 in BTICs.

(**G** and **H**) Immunofluorescent staining on sectioned spheres shows the distribution of markers within the sphere. (**I)** Differentiated BTICs express neuronal, astrocytic and oligodendrocytic markers (βIII TUBULIN, GFAP and O4, respectively); not all BTIC lines differentiated with the same efficiency, but were all able to give rise to cells positive for each of the markers. (**J** and **K**) With longer differentiation (7 days), BTICs can be induced to express mature neuronal markers such as MAP2 and DOUBLECORTIN (DCX). Scale bars=50μm

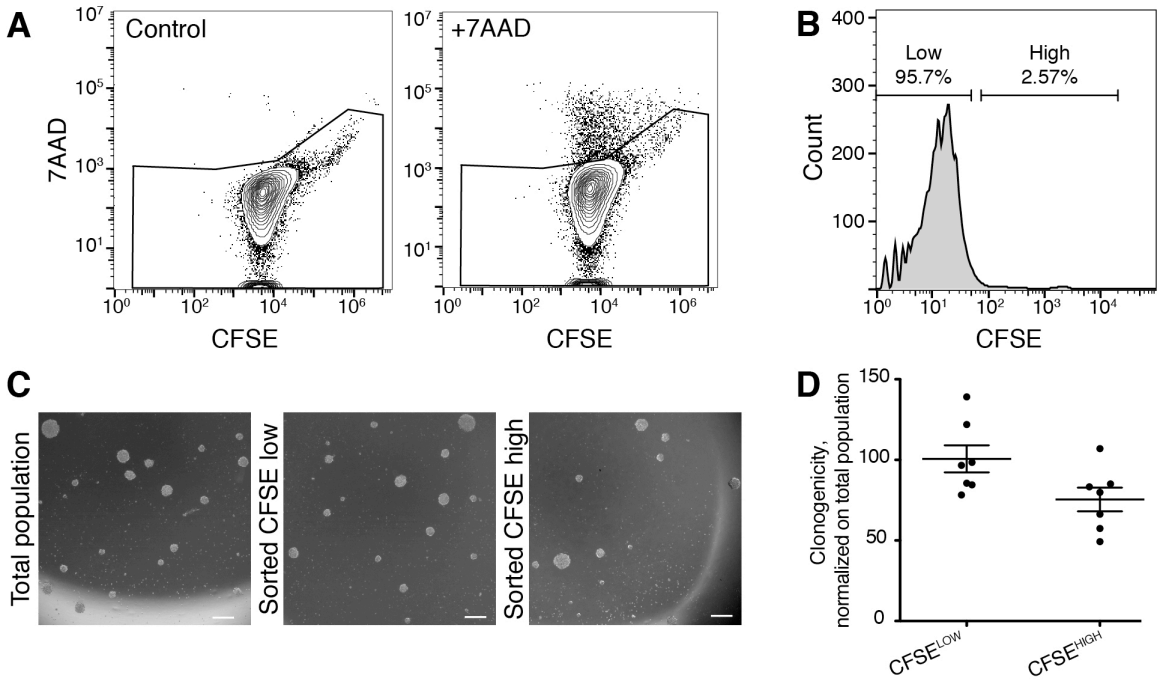


**Figure S2, related to Figure 2. LRCs are viable, functionally quiescent and capable of reactivation**

(**A**) CFSE-retaining cells are 7AAD negative, indicating that they are viable. (**B-D**) Functional quiescence was confirmed by sorting $CFSE^{HIGH}$ and $CFSE^{LOW}$ cells (**B**); both populations are capable of sphere formation, although at different frequencies (**C** and **D**). Each dot represents independent experiments performed at least in triplicate; error bars represent SEM. Scale bars=200μm.

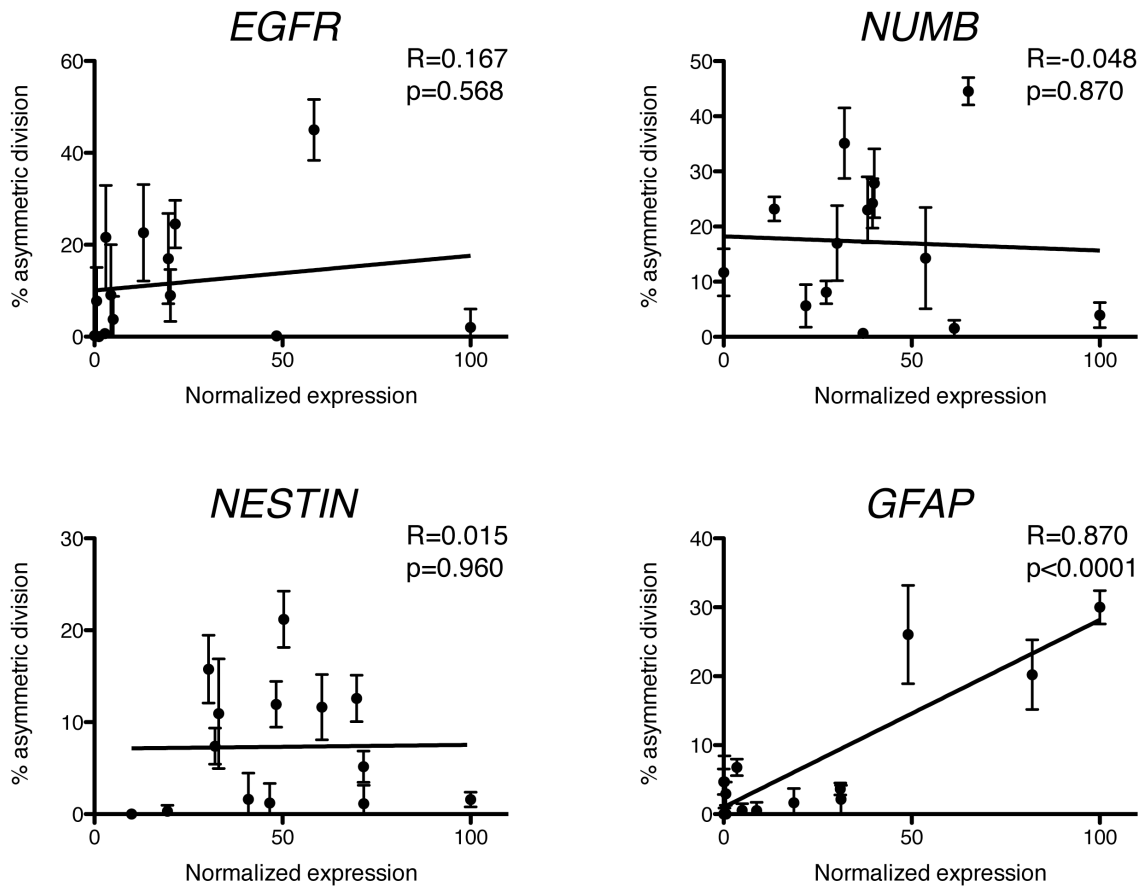


**Figure S3, related to Figure 2. Correlation between asymmetric division and marker expression in BTIC lines.**

Available RNA sequencing data was used to correlate asymmetric division, as reported in Figure 2, and expression of each marker (*EGFR*, *NUMB*, *NESTIN* and *GFAP*). Each dot represents the expression of the marker based on RNA sequencing data and the percentage of asymmetric division in individual BT line (n > 3 independent experiments; errors bars represent SEM). R indicates Pearson's correlation factor.

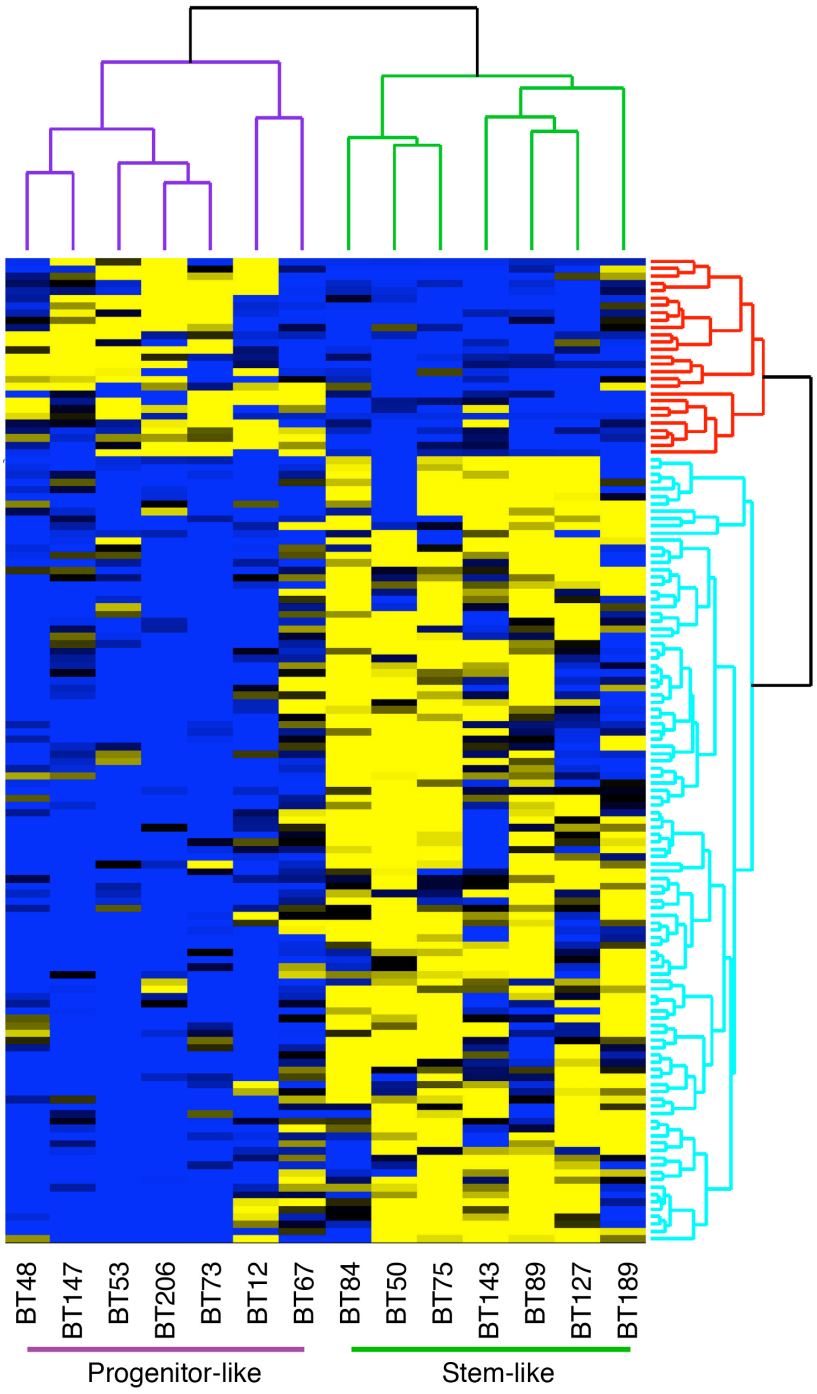


**Figure S4, related to Figure 4. Clustergram of RNA sequencing data for 14 BTIC lines using 136 differentially expressed genes.** Heatmap representing the relative expression of the 136 genes in the signature with the 14 BTIC lines. The lines correctly segregate in the stem- or progenitor-like group.

**Table S1. Mutational status of BTICs, related to experimental procedures.**

| | Line | *EGFR* | *TP53* | *PTEN* | *IDH1* | *NF1* | *CDKN2A* |
|---|---|---|---|---|---|---|---|
| SL | **BT50** | wt | wt | mut | wt | wt | homo del |
| | **BT68** | vIII | wt | wt | wt | NA | NA |
| | **BT75** | wt | wt | wt | wt | wt | wt |
| | **BT84** | wt | wt | mut | wt | wt | homo del |
| | **BT89** | wt | wt | wt | wt | mut | homo del |
| | **BT100** | wt | wt | wt | wt | NA | NA |
| | **BT124** | wt | mut | mut | wt | NA | NA |
| | **BT127** | vIII | wt | wt | wt | wt | homo del |
| | **BT134** | wt | wt | wt | wt | NA | NA |
| | **BT143** | wt | mut | mut | wt | mut | homo del |
| PL | **BT12** | wt | mut | mut | wt | NA | NA |
| | **BT25** | wt | mut | mut | wt | NA | NA |
| | **BT30** | wt | wt | mut | wt | NA | NA |
| | **BT48** | mut | wt | mut | wt | wt | homo del |
| | **BT53** | mut | mut | wt | wt | wt | homo del |
| | **BT67** | wt | wt | mut | wt | mut/het del | homo del |
| | **BT73** | vIII | mut | mut | wt | NA | homo del |
| | **BT147** | vIII | mut | mut | wt | wt | homo del |
| | **BT189** | wt | mut | mut | wt | wt | wt |
| | **BT206** | wt | mut | wt | wt | NA | homo del |

Mutant or wild-type status of genes frequently mutated in GBM for the 20 BTIC lines used in this study. SL – Stem-like, PL- Progenitor-like. vIII indicates *EGFR* variant III, an activating deletion characteristic of GBM, het del indicates a heterozygous deletion, homo del indicates a homozygous deletion, NA – not available.

**Table S2. Assay to assay correlations, related to Figures 1 and 2.**

| | %EGFR+ | | %CD133+ | | %CD15+ | | %CFSE$^{HIGH}$ | | EGFR asymm | | NUMB asymm | | NES asymm | | GFAP asymm | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Self-renewal** | 0.31 | 0.18 | -0.32 | 0.17 | -0.07 | 0.77 | **-0.57** | **0.01** | -0.37 | 0.10 | -0.38 | 0.10 | **-0.44** | **0.05** | -0.31 | 0.18 |
| **%EGFR+** | | | -0.19 | 0.41 | 0.42 | 0.06 | -0.21 | 0.36 | -0.12 | 0.61 | -0.23 | 0.32 | -0.15 | 0.53 | -0.27 | 0.25 |
| **%CD133+** | | | | | 0.09 | 0.71 | **0.66** | **0.00** | 0.43 | 0.06 | **0.53** | **0.02** | 0.38 | 0.10 | 0.44 | 0.05 |
| **%CD15+** | | | | | | | 0.10 | 0.68 | -0.23 | 0.32 | -0.33 | 0.15 | -0.16 | 0.50 | -0.18 | 0.45 |
| **%CFSE$^{HIGH}$** | | | | | | | | | **0.58** | **0.01** | **0.50** | **0.03** | **0.63** | **0.00** | **0.66** | **0.00** |
| **EGFR asymm** | | | | | | | | | | | **0.82** | **0.00** | **0.87** | **0.00** | **0.80** | **0.00** |
| **NUMB asymm** | | | | | | | | | | | | | **0.82** | **0.00** | **0.75** | **0.00** |
| **NES asymm** | | | **ρ** | **p** | | | | | | | | | | | **0.88** | **0.00** |

Spearman's correlation was used to analyze the direct relations between the parameters studied. The ρ correlation coefficient and p-value are represented for each pair. Significant correlations ($p<0.05$, $|\rho|>0.4$) are in bold. Correlations with $0.4<|\rho|<0.6$ were considered weak.

**Table S3. Survival of xenografted animals, related to Figure S3**

| Group | Cell line | N= | Median survival (Days) |
|---|---|---|---|
| Stem-like | BT68 | 6 | 237 |
| | BT84 | 7 | 194 |
| | BT89 | 6 | 235.5 |
| | BT127 | 7 | 208 |
| | BT134 | 6 | 145.5 |
| | BT143 | 10 | 82 |
| | **TOTAL** | **42** | **183.7±24.5** |
| Progenitor-like | BT12 | 6 | 96 |
| | BT25 | 6 | 48.5 |
| | BT30 | 6 | 68 |
| | BT48 | 7 | 93 |
| | BT53 | 5 | 60 |
| | BT67 | 5 | 134 |
| | BT73 | 7 | 31 |
| | BT147 | 10 | 38 |
| | BT206 | 7 | 38 |
| | **TOTAL** | **59** | **67.4±11.4** |

Survival of mice implanted with individual BTIC lines. N indicates the number of animals per line included in the analysis. The mean per group is the average of the medians ± SEM.

**Table S4. GSEA for the three TCGA groups, related to Figure S4**

| | vs Group I | vs Group II | vs Group III | vs REST |
|---|---|---|---|---|
| **Group I** | | Neural *** | Proneural * <br> Neural ** | Proneural * <br> Neural ** |
| **Group II** | n.s. | | Proneural *** | n.s. |
| **Group III** | Mesenchymal * | Mesenchymal * | | Mesenchymal *** |
| **REST** | Mesenchymal ** | n.s. | Proneural ** <br> Neural * | |

Enrichment of GBM subtype genes in the three groups. All combinations were tested; only significant associations are reported. *p<0.05, **p<0.01, ***p<0.0001; for all, FDR<0.10.

## Supplemental Experimental Procedures

Patient-derived material was obtained following informed consent and approved by the University of Calgary Ethics Review Board. All animal procedures received ethical approval by the University of Calgary Animal Care and Use Committee, in accordance with the Canadian Council for Animal Care.

### Cell culture

Cell lines were maintained as spheres in Neurocult Basal Medium (Stem Cell Technologies), complemented with Proliferation Supplement (Stem Cell Technologies), EGF and bFGF (20ng/ml each, Peprotech) and heparan sulfate (2µg/ml, Sigma). Growth factors were omitted for mitogen independent lines (n=5) (Kelly et al., 2009). Spheres were passaged every 10-21 days (determined by line-specific growth characteristics) through dissociation with Accumax (ICT) and plating (100000 cells) in T25 flasks.

Differentiation was performed with either Neurobasal medium (Life Technologies) supplemented with B27 (Life Technologies) or Neurocult Basal Medium with 1% fetal bovine serum (FBS). Cells (50000 cells/$cm^2$) were plated on coverslips and maintained in medium for 3 or 7 days, and then processed for immunofluorescence.

CFSE labeling was performed, as per manufacturer's instructions, with cells grown in standard conditions and medium. When the largest spheres reached an appropriate size (200-250µm in diameter), they were either fixed for sectioning (see below) or dissociated for FACS analysis or replating.

To study asymmetric cell division, single cells were plated at low density, synchronized with thymidine (2mM, Sigma) or nocodazole (100ng/ml, Sigma) for 24h. Cells were fixed 18-20h after removal of this block and subsequently stained. Asymmetric cell division was evaluated by a blinded observer, counting at least 80 cell couples per sample and assessing the asymmetric distribution of the markers.

Sphere formation assays were performed by plating 1000 cells per well (n=6) in 96-well plates, and the spheres were counted when they reached a diameter of 200-250μm.

For growth curves, 1000 cells/well were plated in 96-well plates and grown in standard conditions. Over the course of 2 weeks, alamarBlue (Life Technologies) was added to wells at different time points and the fluorescence measured after 6 hours. Data were normalized to fluorescence levels on day1.

**Flow cytometry and sorting**

FACS analysis experiments were performed with an ATTUNE flowcytometer (Life Technologies); cell sorting was performed on a BD FACSAria II cell sorter. For all experiments 7-actinomycinD (7AAD - BD Biosciences), propidium iodide (Sigma) or Aqua Dead Cell stain (Life Technologies) were used to exclude non-viable cells from analysis.

For CFSE retention experiments, the fluorescence of single cells was measured and compared with the signal from cells fixed at the time of CFSE labeling (Day0). The gate for $CFSE^{HIGH}$ cells was placed, based on the FlowJo's proliferation analysis tool, to include cells that had cycled one or two times after plating.

For marker expression analysis, single cells were resuspended in PBS with 1% bovine serum albumin (BSA) and incubated with primary antibodies for 1h at 4°C. Cells were then washed and resuspended in PBS/BSA for subsequent analysis. Primary antibodies used include: αEGFR-APC or -FITC (R&D, FAB10951A and FAB10951F); αCD133-PE or -APC (Miltenyi, 130-098-826 and 130-098-829); αCD15-FITC or -APC (BD Biosciences, 560997 and 551376). Gates were placed based on isotype controls or, in case of multicolor experiments, on all labels except the one of interest to ensure the accuracy of the gate. No compensation was applied. Analysis of all FACS data was performed with FlowJo (Treestar).

**Immunofluorescence and immunohistochemistry**

Single cells were plated on coverslips coated with poly-L-ornithine -alone (5µg/ml, Sigma) -or with laminin (10 µg/ml, Sigma). Density of plating was 2000 cells/$cm^2$ for asymmetric cell division studies, 25000 cells/$cm^2$ for general characterization, 50000 cells/$cm^2$ for differentiation, or 50000 cells/$cm^2$ for label retaining cell reactivation. In all cases, the cells were fixed with cold 4% paraformaldehyde (PFA) for 20 minutes at room temperature. Immunostaining of CFSE-labeled spheres was performed after cryosectioning of PFA fixed, OCT embedded spheres, followed by subsequent processing of specimens for immunofluorescence.

For histological procedures, xenografted mice were transcardially perfused with phosphate buffer saline (PBS) and 4% PFA. Brains were then cryopreserved, sectioned and processed for immunofluorescence.

In all cases, specimens were washed and non-specific binding was blocked with appropriate normal serum in presence of Triton X100 (omitted when staining for membrane proteins). Antigen retrieval was performed for SOX2 and Ki67 detection in citrate buffer pH6.0. Samples were incubated with primary antibodies (see below), and then the staining developed with appropriate secondary antibody (DyeLight 488-, Cy3-, DyeLight 647- or biotin-conjugated; Cy3- or Dyelight 647-conjugated streptavidin was used for developing biotinylated antibodies; all Jackson Immunolabs). Hoechst 33342 was used as nuclear counterstain. For immunohistochemistry, ABC Elite kit (Vector) and diaminobenzidine (Sigma) were used to develop the staining.

Primary antibodies: mouse α human NESTIN 1:350, (Millipore, MAB5326), mouse α SOX2 1:50 (R&D, MAB2018), mouse α CD133 1:100 (Millipore MAB4301), mouse α CD15 1:100 (BD Biosciences, 555400), goat α VIMENTIN 1:150 (Millipore, AB1620), rat α MUSASHI1 1:200 (eBiosciences, 14-9896-80), rabbit α GFAP 1:400 (Biomedical Technologies, BT-575), mouse α EGFR 1:25 (Millipore, 05-101), rabbit α EGFR 1:500 (Abcam, AB2430), rabbit α NUMB 1:250 (Abcam, AB14140), mouse α βIII-TUBULIN 1:300 (Sigma T8578), mouse IgM α O4 1:40 (R&D, MAB1326), mouse α MAP2 1:500 (Sigma, M9942), goat α DCX 1:400 (SantaCruz, sc8066), mouse α Ki67 1:100 (Novocastra, NCL-L-Ki67-MM1), rabbit α Ki67 1:1000 (Novocastra, NCL-Ki67p), rabbit α cleaved CASPASE 3 1:400 (CST, 9661), mouse α NUCLEOLIN, human specific 1:1000 (Abcam, AB22758).

**Microscopy**

Samples were imaged with a Zeiss Axiovert 40 CFL inverted microscope, a Zeiss Axioplan 2 fluorescence microscope, a Nikon C1si Spectral confocal microscope or an Olympus VS120-5 Slide scanner. Images were processed with NIS elements 4.3 (Nikon), Olyvia (Olympus) or Photoshop CS6 (Adobe).

**BTICs xenografts**

C17/SCID female mice (6-8 weeks old) were implanted with 100000 BTICs in the right striatum (coordinates, mm from Bregma and dura mater surface: AP +0.5, ML -2.0, DV -3.0). Mice were sacrificed upon significant weight loss or presentation of neurologic symptoms necessitating euthanasia as per University of Calgary animal care guidelines. 15 BTIC lines were used in this study and 6 to 10 animals were xenografted per line. Kaplan-Meier curves were constructed based on the survival data and followed by Log-Rank survival analysis. Mice that died for causes other than brain tumor were excluded from the analysis.

**RNA sequencing and transcriptomic analysis**

PolyA+ mRNA was column purified (including on column DNaseI treatment) using the MACS 96 Separation Unit (Miltenyi Biotec) as per the manufacturer's protocol. 2-10 ug of total RNA with a RIN>=7 (Agilent Bioanalyzer) were used as input. The purified polyA+ RNA was used as substrate for double-stranded cDNA synthesis using the Superscript II Double-Stranded cDNA Synthesis kit (Life Technologies) and 200ng random hexamers (Life Technologies). Double stranded cDNA was purified using 2

volumes of Ampure XP beads, fragmented using Covaris E series shearing (20% duty cycle, Intensity 5, 55 seconds), and used for paired-end sequencing library preparation (Illumina). Prior to library amplification uridine digestion was performed at 37$^{o}$C for 30 min following with 10 min at 95$^{o}$C in Qiagen Elution buffer (10mM Tris-Cl, pH 8.5) with 5 units of Uracil-N-Glycosylase (UNG: AmpErase).

The resulting single stranded sequencing library was amplified by PCR (10-13 cycles) to add Illumina P5 and P7 sequences for cluster generation. PCR products were purified on Qiaquick MinElute columns (Qiagen) and assessed and quantified using an Agilent DNA 1000 series II assay and Qubit fluorometer (Life Technologies) respectively. Libraries were sequenced using paired-end 76nt sequencing chemistry on a cBot and Illumina $GA_{iix}$ or HiSeq2000 following manufactures protocols (Illumina).

RNA-seq pair-end reads were aligned to a transcriptome reference consisting of the reference genome extended by the annotated exon-exon junctions (Morin et al., 2008). To generate transcriptome reference we use the JAGuaR v 1.7.6 pipeline (Butterfield et al., 2014), specifically developed to allow the possibility for a single read to span multiple exons. Reads aligned to a custom transcriptome reference (build from NCBI GRCh37-lite reference and Ensembl v69 gene annotations) are then "repositioned" on to genomic coordinates, transforming reads spanning exon-exon junctions into large-gapped alignment. Using repositioned reads, the RPKM (Reads Per Kilobase per Million) metric was calculated (Mortazavi et al., 2008) for every collapsed transcripts gene model that we used in the subsequent analysis. Collapsed transcripts gene model was defined by overlap of all exons of all known isoforms for a given gene.

**Differential expression analysis**

To evaluate coding genes that are differentially expressed between the SL and PL BTIC clusters, mean expression for each gene was calculated across seven SL and seven PL BTIC lines. Mean expression of genes in SL-BTICs was compared to PL-BTICs using an in-house differential expression analysis tool (DEfine v.0.9.2) to detect significantly up- and down-regulated genes (FDR <0.01). To reduce noise, DEfine applies thresholds on total number of reads aligned to a gene (N>30) and uses a normalization-free procedure to derive the differentially-expressed genes as well as corrects for potential biases associated with GC content and gene length.

The list of differentially expressed genes consisted of 1110 down-regulated and 269 up-regulated genes in SL BTICs. Of the down-regulated genes, 36 belonged to the proto-cadherin α and γ gene clusters (15 *PCDHα* and 21 *PCDHγ* genes). We found that the 15 *PCDHα* genes were expressed at similar levels within each BTIC cluster; the same was observed for 21 *PCDHγ* genes (likely an artefact of exon sharing between members). We therefore collapsed the expression of these two gene groups into two meta-genes (*PCDHA* and *PCDHB*) by averaging the expression of 15 *PCDHα* and 21 *PCDHγ* genes, separately. We further reduced the list of differentially-expressed genes by calculating a signal to noise statistic ($D_{SN}$) for each gene using the following formula:

$$D_{SN} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 - \sigma^2}_2}$$

where $\mu_1$ and $\mu_2$ are the mean RPKM values for each gene in SL and PL BTIC groups, respectively; $\sigma_1$ and $\sigma_2$ are the standard deviations.

We included the top $10^{th}$ percentile differentially expressed genes based on $D_{SN}$, eliminating genes with high expression variance within each BTIC cluster from further

analyses. The final list of differentially expressed genes consisted of 109 down-regulated and 27 up-regulated genes in SL BTICs. This list of differentially expressed genes was used for unsupervised clustering of the seven SL and seven PL BTIC lines (Figure S3).

**The Cancer Genome Atlas (TCGA) GBM gene expression analysis**

The publicly available, processed (Level 3) TCGA GBM RNA sequencing data (Ver.2014-08-28) was downloaded from the TCGA Data Portal. Briefly, RNA sequencing was performed by the University of North Carolina TCGA genome characterization center using the Illumina HiSeq 2000 RNA Sequencing platform. Of the 172 samples in the dataset, 8 samples with missing GBM subtype information and 12 recurrent GBM samples were excluded. Further analysis was performed on data from 152 TCGA GBM patients, including 11 patients with glioma-CpG island methylator phenotype (G-CIMP). Raw Data: Dataset ID: TCGA_GBM_exp_HiSeqV2 (https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/gbm/cgcc/unc.edu/illuminahiseq_rnaseqv2/rnaseqv2/)

Patients in the TCGA dataset were ordered based on similarity to the levels of expression of the 136 genes defined in SL and PL cells. Briefly, we calculated z-scores, across the whole dataset, for each of the 136 genes:

$$z = \frac{X - \mu}{\sigma}$$

where X is the level of expression, μ is the average expression across all patients in the dataset and σ the standard deviation, for a given gene. We then calculated an average z-score for SL genes, an average z-score for PL genes, and the difference between the two average z-scores for each patient:

$$Z' = \overline{z_{SL}} - \overline{z_{PL}}$$

Therefore, higher the $Z'$, the more similar is a patient's expression profile to the SL gene profile. The samples in TCGA data set were then ranked based on $Z'$ and divided in three groups with $Z'>0.1$, $-0.1<Z'<0.1$ and $Z'<-0.1$ (n=44, 40 and 68, respectively; see Figure 4C).

The GSEA analysis was performed using the software provided by the Broad Institute (http://www.broadinstitute.org/gsea) (Subramanian et al., 2005). The gene sets that define the four GBM subtypes were downloaded from the original paper (Verhaak et al., 2010) and consisted of 178 proneural genes, 129 neural genes, 162 classical genes and 216 mesenchymal genes.

Briefly, GSEA is a method to determine whether a set of genes defined *a priori* (in this case, the GBM subtype) is concordant with subgroups (groups I-III) of dataset (the TCGA patient samples). All the genes are ranked based on the subgroups and then an enrichment score is calculated for each gene of the defined set within this ranking. This analysis indicates whether a gene set (e.g. proneural) is overrepresented in a set of samples (e.g. group I).

**Statistical analyses**

All data reported for *in vitro* experiments are representative of at least 3 independent replicates and are illustrated in scattered plots or bar graphs, including mean ± SEM. Hierarchical clustering using Manhattan distance metric with Ward's agglomeration method were used to generate the heat map. The medians of all replicates (3-6) for each cell line were centered and scaled prior to clustering. Multivariate analysis was carried

out in the R software program (version 3.0.1 - www.R-project.org) using hclust (stats) and heatmap.2 (gplots) functions. All other statistical analyses and graphing were performed with Prism 5.0f or 6.0 (GraphPad) or SPSS 20 (IBM).

## Supplemental references


Butterfield, Y.S., Kreitzman, M., Thiessen, N., Corbett, R.D., Li, Y., Pang, J., Ma, Y.P., Jones, S.J., and Birol, I. (2014). JAGuaR: junction alignments to genome for RNA-seq reads. PLoS One *9*, e102398.

Kelly, J.J., Stechishin, O., Chojnacki, A., Lun, X., Sun, B., Senger, D.L., Forsyth, P., Auer, R.N., Dunn, J.F., Cairncross, J.G.*, et al.* (2009). Proliferation of human glioblastoma stem cells occurs independently of exogenous mitogens. Stem Cells *27*, 1722-1733.

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., and Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques *45*, 81-94.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621-628.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545-15550.

Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P.*, et al.* (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98-110.