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

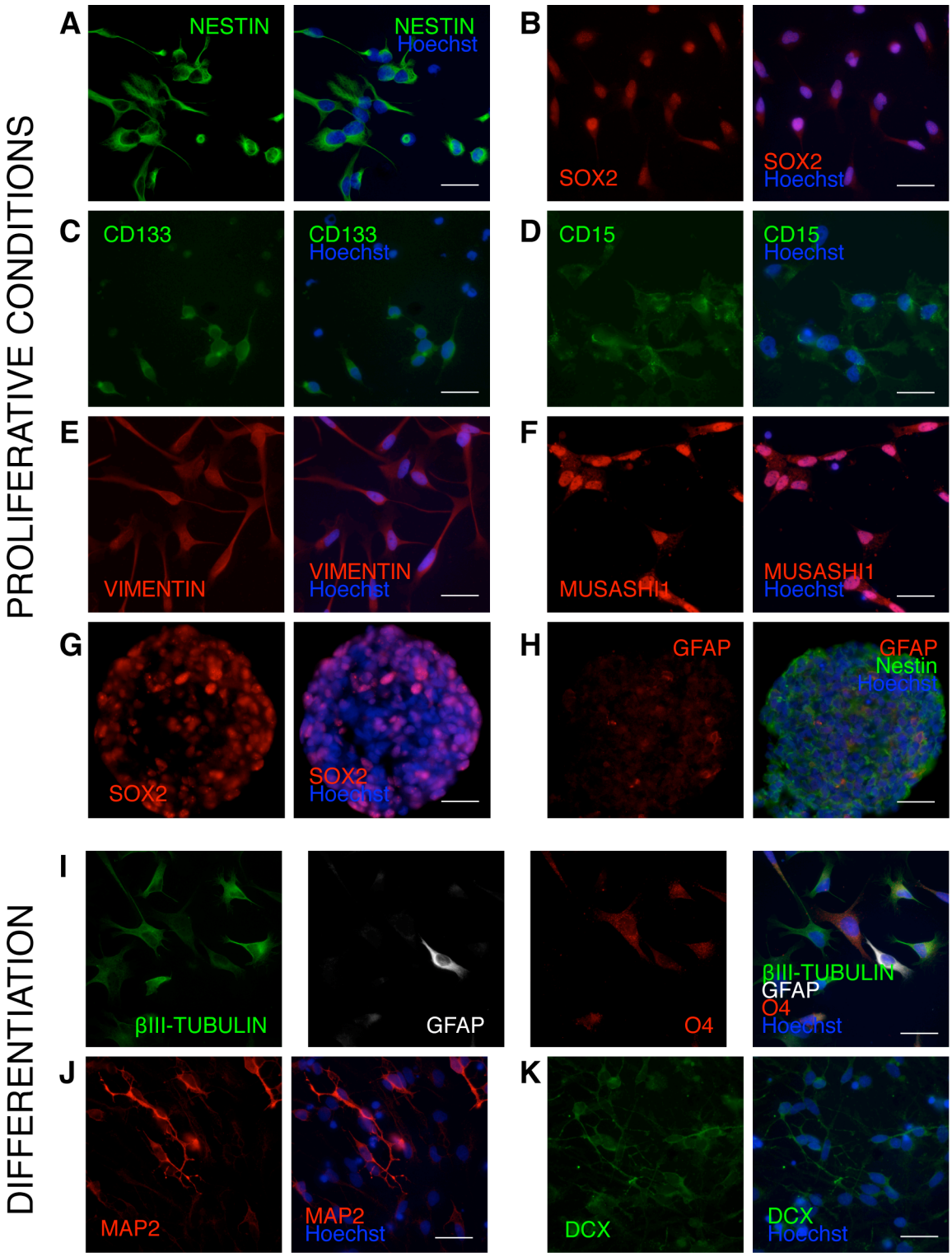

Figure S1, related to Experimental procedures. Characterization of BTICs

(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

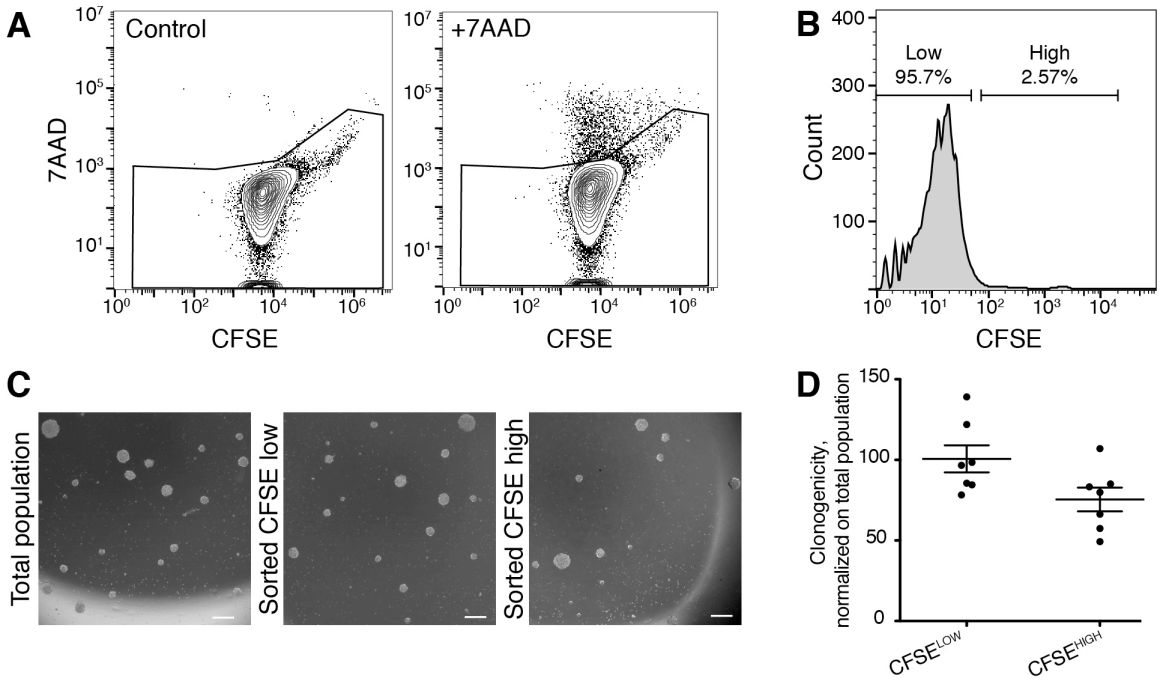

Figure S2, related to Figure 2. LRCs are viable, functionally quiescent and capable of reactivation

(**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.

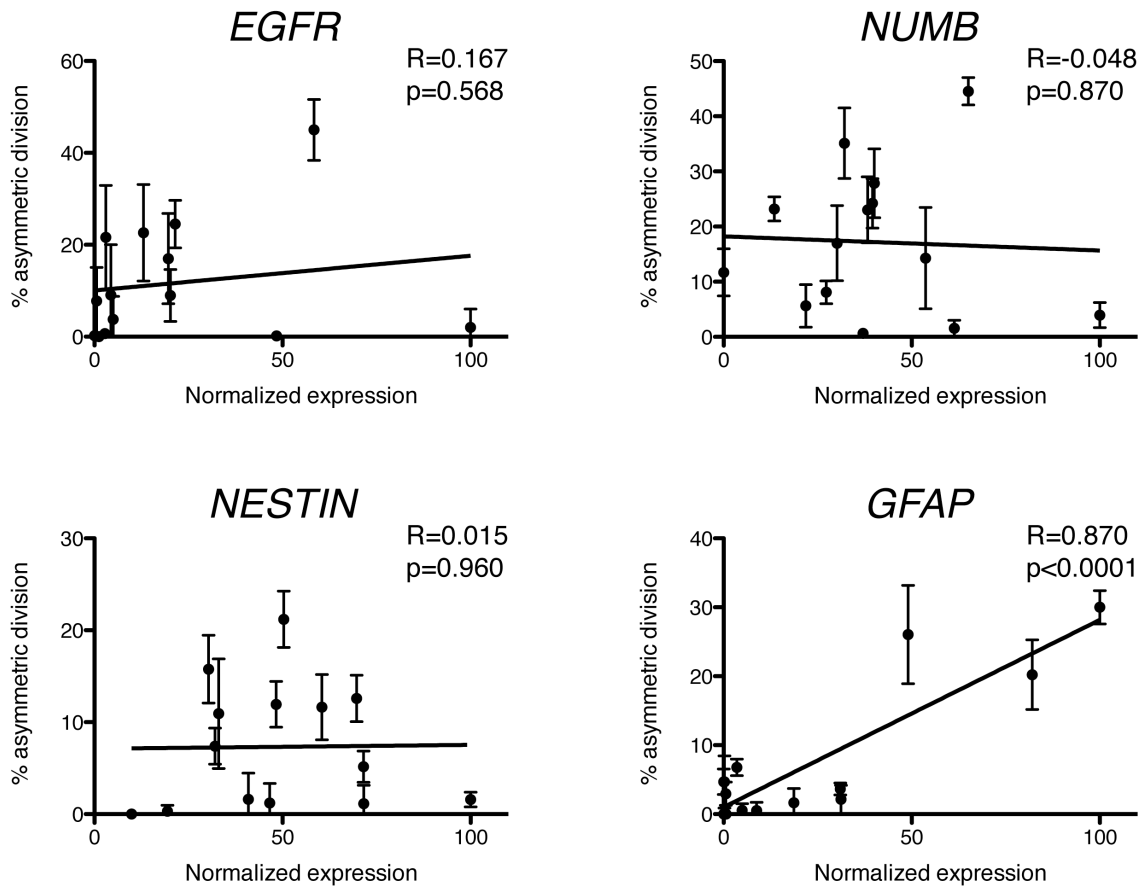

Figure S3, related to Figure 2. Correlation between asymmetric division and marker expression in BTIC lines.

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.

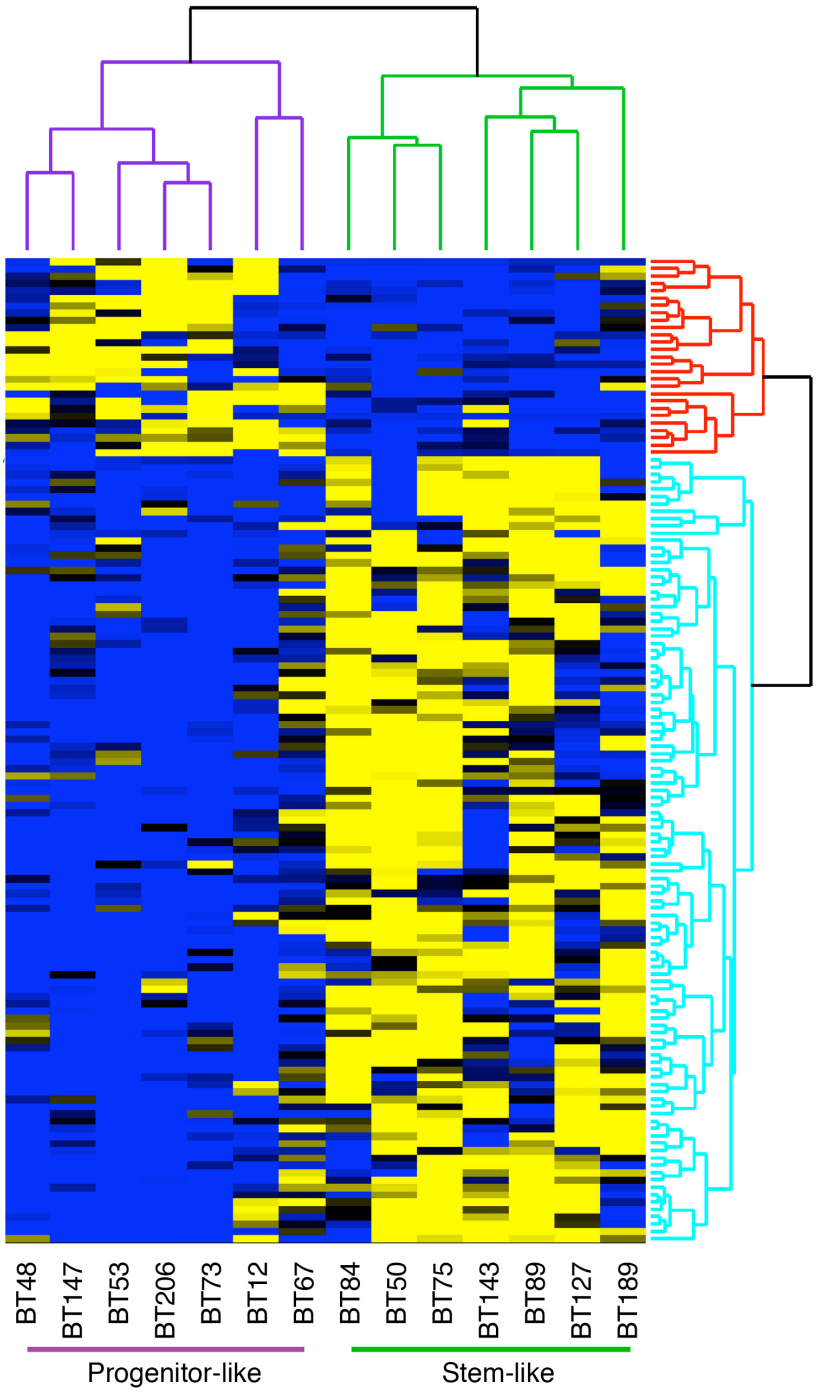

Figure S4, related to Figure 4. Clustergram of RNA sequencing data for 14 BTIC lines using 136 differentially expressed genes. Heatmap representing the relative expression of the 136 genes in the signature with the 14 BTIC lines. The lines correctly segregate in the stem- or progenitor-like group.

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 * Neural **	Proneural * Neural **
Group II	n.s.		Proneural ***	n.s.
Group III	Mesenchymal *	Mesenchymal *		Mesenchymal ***
REST	Mesenchymal **	n.s.	Proneural ** 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 μ_1 and μ_2 are the mean RPKM values for each gene in SL and PL BTIC groups, respectively; σ_1 and σ_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.