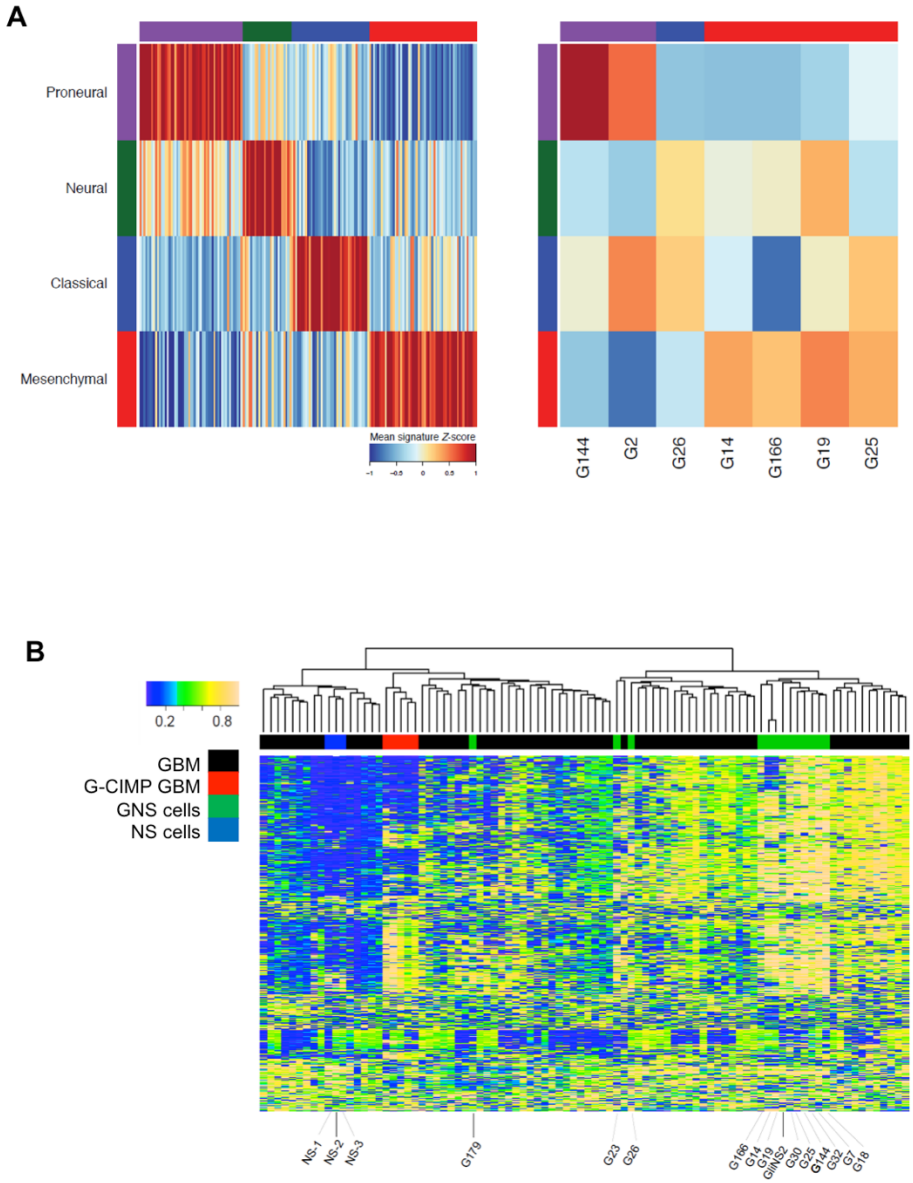

Figure S1, related to Figure 1 | Classification of GNS cell lines used in this study. (A) Subtype classification of GBM tumours (TCGA Project) and GNS cell lines by subtype signature gene expression (Verhaak et al., 2010). Mean Z-score values of signature genes overexpressed in distinct subtypes were used to reclassify GBM data previously described (left) and GNS lines used in this study (right). GNS cells were classified as predominantly Mesenchymal, Proneural or Classical**. (B)** DNA methylation profiles for thirteen different GNS cells were obtained and compared to primary tumour data (TCGA datasets) and normal NS cell lines (NS-1, -2 and -3). Clustering was performed using the top 800 most variable CpG sites identified in primary GBM samples, including the G-CIMP subtype of GBM.

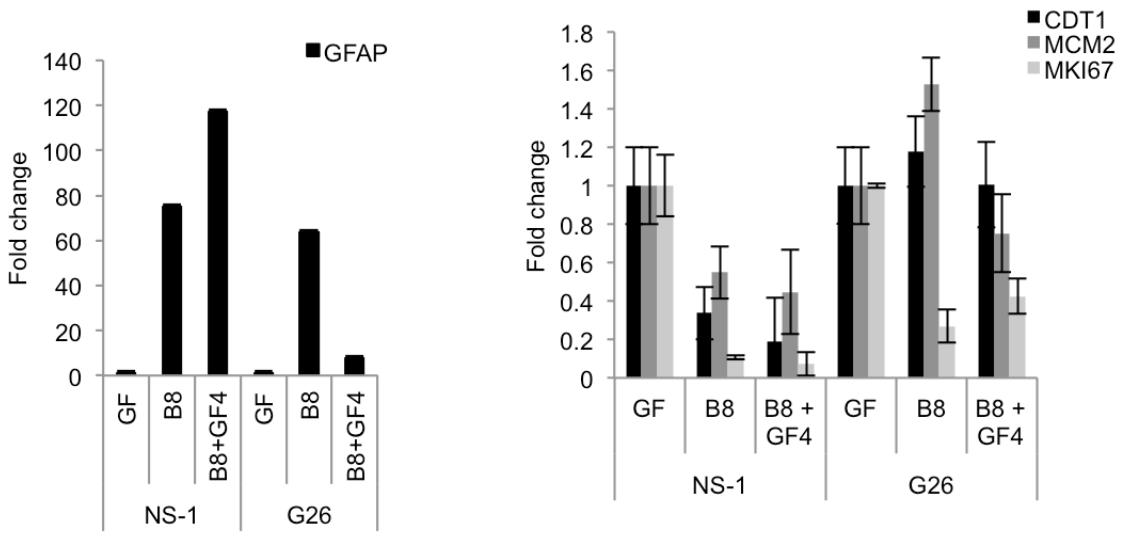

Figure S2, related to Figure 5 | Differentiation and proliferation markers assessed by qRT-PCR following exposure of differentiated GNS cells to growth factors. Independent validation experiment of G26 and NS-1 in proliferating conditions, after 8 days of BMP-treatment and after adding back growth factors for 4 days after the 8 day BMP treatment (no BMP during last 4 days with GFs). Left shows the astrocyte marker GFAP and right the proliferation markers CDT1, MCM2 and MKI67. Error bars denote standard deviation of technical duplicates.

Table S1, related to Figure 6 (see spreadsheet). Transcription factor motif enrichment in the set of loci that fail to be silenced in GNS cells (but not NS cells) in response to BMP.

Supplemental Experimental Procedures (Caren et al.)

DNA methylation data bioinformatics analysis

Bisulfite-modified DNA was hybridized to Infinium HumanMethylation450 BeadChips (Illumina) as previously described (Stricker et al., 2013). Data generated by the BeadStudio software was exported and further analyses were performed in R. The R package ChAMP was used for normalisation and MVP calling (Morris et al., 2014). TCGA datasets were used to generate a 'GBM methylation' set of MVPs by comparing to Pilocytic astrocytoma from GEO, GSE44684, and FACS-sorted human brain orbitofrontal cortex (nonneuronal), GSE50798. $CIMP^+$ cases were excluded since they harbour a specific methylation profile. Differentially methylated sites were identified using limma (Ritchie et al., 2015). Global levels of hydroxymethylation were measured with the Global DNA Hydroxymethylation ELISA Kit (CellBio labs Inc) according to the protocol supplied by the manufacturer. 300ng of each sample was run in duplicates.

RNA processing and microarray hybridisation

RNA was extracted using Trizol (Invitrogen) followed by treatment with TURBO DNase (Ambion). RNA quality was assessed on the Agilent 2100 Bioanalyzer, and samples were processed for microarray hybridisation according to the GeneChip whole-transcript sense target labeling assay (Affymetrix). Affymetrix Exon Array 1.0 ST arrays were hybridized for 16 h at 45°C, washed, stained with streptavidin-phycoerythrin (SAPE) conjugate on a FS450 automated fluidics station, and imaged on a GCS3000 7G scanner (Affymetrix). Feature extraction was performed using Command Console 3.2.3, and hybridization quality was assessed with Expression Console 1.1.2 (Affymetrix).

Gene expression and subtyping analysis

Stem Cell Reports, Volume *5*

Supplemental Information

Glioblastoma Stem Cells Respond to Differentiation Cues but Fail to Undergo Commitment and Terminal Cell-Cycle Arrest

Helena Carén, Stefan H. Stricker, Harry Bulstrode, Sladjana Gagrica, Ewan Johnstone, Thomas E. Bartlett, Andrew Feber, Gareth Wilson, Andrew E. Teschendorff, Paul Bertone, Stephan Beck, and Steven M. Pollard

Exon arrays were processed using v1.28 of the xps Bioconductor package and processed using the Robust Multi-chip Average (RMA) method (Irizarry et al., 2003), and probe sets were summarised by median polish in xps. Where a gene was represented by multiple splice variants, the transcript model having the maximal value was taken as the dominant isoform and expression levels from replicate arrays were averaged. Sample log2-transformed expression values for the signature centroid genes were produced by taking the mean expression across sample replicates. Centroid genes that could not be assigned to annotated genes were omitted from further analysis. Subtype scores per sample were computed from mean Z-score transformed levels of overexpressed centroid genes for each subtype. Samples were then classified as belonging to the subtype associated with the highest mean Z-score.

ATAC-seq

ATAC library preparation was undertaken according to published protocols (Buenrostro et al., 2013). 8 libraries were derived from cell lines NS-1 and G26, in EGF/FGF and BMP, in biological replicates. These were multiplexed based on quality control and quantification data from the Qubit and 2200 TapeStation (Agilent), then sequenced on the Illumina HiSeq 2500 in 50bp paired-end format.

Paired end reads were trimmed for sequencing primers and aligned to hg19 using Bowtie2 (Langmead and Salzberg, 2012) and a maximum fragment length (-X) of 3000bp. Aligned reads were filtered for duplicates and split into two 27bp intervals to represent the transposase binding site. The 27bp intervals were created by extending 18bp into the read and 9bp beyond the read to represent the transposase binding site and 9bp replicated region. Regions of the genome enriched for transposase loci were identified

using F-seq (Boyle et al., 2008) using both broad and narrow parameters (Narrow: -l 600 -f 28, -t 8. Broad: -l 2000, -f 28, -t 3.). Broad and narrow peaks were called separately in each library and merged using samtools merge into combined broad and combined narrow peak sets and concatenated together into one final set of intervals. Transposase access site counts for the final peak loci were generated using bedtools intersect. Loci counts were normalised for GC bias using conditional quantile normalization (Hansen et al., 2012) and the CQN offsets were passed on to DESeq2 (Love et al., 2014) as the normalization factors for differential testing. Loci with adjusted p-values below 0.05 were called as significantly differentially accessible. Plots were generated using CQN normalized data. Annotation enrichment for genome loci was analysed using the GREAT tool (McLean et al., 2010) and proximal genes were tested for GO term enrichment using DAVID. Motif analysis was completed using the MEME suite 4.10.1 AME motif enrichment tool (Multi-organism motif database (Bailey et al., 2009)) using the 268 'failed silenced' loci sequences as input and a random sample of 1000 loci identified as accessible chromatin in the peak calling analysis as control sequences.

References

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Boyle, A. P., Guinney, J., Crawford, G. E., and Furey, T. S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics *24*, 2537-2538.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics *13*, 204-216.

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

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.
Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.
Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., and Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics *30*, 428-430.
Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.
Stricker, S. H., Feber, A., Engstrom, P. G., Carén, H., Kurian, K. M., Takashima, Y., Watts, C., Way, M., Dirks, P., Bertone, P.*, et al.* (2013). Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. Genes Dev *27*, 654-669.