Published in final edited form as:

Cancer Res. 2018 May 15; 78(10): 2463–2474. doi:10.1158/0008-5472.CAN-17-1724.

Bivalent chromatin domains in glioblastoma reveal a subtype-specific signature of glioma stem cells

Amelia Weber Hall¹, Anna M. Battenhouse¹, Haridha Shivram¹, Adam R. Morris¹, Matthew C. Cowperthwaite², Max Shpak^{2,3}, and Vishwanath R. Iyer^{1,4}

¹Department of Molecular Biosciences, Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas

²St David's Medical Center, Austin, Texas

³Sarah Cannon Research Institute, Nashville, Tennessee

⁴Livestrong Cancer Institutes, Dell Medical School, University of Texas at Austin, Austin, Texas

Abstract

Glioblastoma multiforme (GBM) can be clustered by gene expression into four main subtypes associated with prognosis and survival, but enhancers and other gene regulatory elements have not yet been identified in primary tumors. Here, we profiled six histone modifications and *CTCF* binding as well as gene expression in primary gliomas, and identified chromatin states that define distinct regulatory elements across the tumor genome. Enhancers in mesenchymal and classical tumor subtypes drove gene expression associated with cell migration and invasion, while enhancers in proneural tumors controlled genes associated with a less aggressive phenotype in GBM. We identified bivalent domains marked by activating and repressive chromatin modifications. Interestingly, the gene interaction network from common (subtype-independent) bivalent domains was highly enriched for homeobox genes and transcription factors, and dominated by *SHH* and Wnt signaling pathways. This subtype-independent signature of early neural development may be indicative of poised de-differentiation capacity in glioblastoma, and could provide potential targets for therapy.

Keywords

glioblastoma; bivalent; enhancer; epigenetic; histone modification

Introduction

Glioblastoma multiforme (GBM) is an aggressive primary brain tumor that accounts for 52% of all malignant primary brain neoplasias. The median time of survival with treatment is 14.6 months and only 5% of diagnosed individuals survive five years from diagnosis (1). Given the dismal prognosis of GBM, many studies have focused on analysis of whole-

*Corresponding author: Vishwanath R. Iyer, The University of Texas at Austin, Department of Molecular Biosciences, 100 East 24th St. Stop A5000, Austin, TX 78712-1639, USA, 512-232-7833, vishy@utexas.edu.

The authors declare no potential conflicts of interest.

genome/exome sequencing and gene expression data from primary GBM tumors to identify common gene mutations and expression profiles. These studies identified 4 molecular subtypes of GBM – classical, mesenchymal, neural, and proneural. These data have been invaluable in identifying genes and gene pathways that drive the development of GBM, and subtypes predict some aspects of patient prognosis and response to treatment (2). However, the underlying chromatin context that regulates gene expression programs in primary GBM tumors is largely unknown. As GBM lesions are developmentally plastic, and can change certain aspects of their cellular identity, understanding how they vary with regard to chromatin structure will enable identification of key genes and regulatory motifs controlling differentiation capacity in GBM.

While several studies have quantified single histone modifications in GBM-derived cell lines, none of these studies have been performed in uncultured primary tissue, and few have looked at patterns derived from multiple histone modifications in the same cell line or tumor (3). These studies established the general trend that repressive modifications (particularly polycomb silencing) are globally reduced, and active modifications (such as H3K4me3, H3K9ac, H3K27ac) are generally increased across the GBM genome. This lack of data hinders efforts to conclusively identify and characterize the initiating cell type for glioblastoma tumors. Indeed, a recent chromatin profiling study to identify enhancers revealed cell-type of origin in medulloblastoma, but no comparable dataset currently exists for GBM (4).

In this study, we sought to categorize regulatory regions of the genome in primary GBM tumors by profiling six post-translational modifications of histone H3, and the multifunctional insulator binding protein *CTCF*, in conjunction with gene expression profiling of the same tumors. We used a HMM-based approach (5) and identified combinations of chromatin marks that defined distinct regulatory elements across the genome (Fig. 1A). The resulting model encompassed 21 chromatin states that identified known regulatory elements such as enhancers and promoters, and also identified bivalent regions in tumors for the first time. We were able to annotate any state in this model with matched expression data, generating a context-dependent view of gene expression which identified regulatory regions that may control gene expression indirectly.

Materials and Methods

Chromatin immunoprecipitation in solid tumors and cell lines

All patients provided written informed consent. This study was approved by the Institutional Review Boards of St. David's Medical Center and of the University of Texas at Austin and studies were conducted in accordance with the ethical guidelines of the Belmont Report. Cell lines were obtained from ATCC. Cell lines were not subsequently authenticated or tested for mycoplasma except for the T98G cell line which was verified to be mycoplasma-free by a PCR assay. Tumors were collected during planned surgical resections, and only excess tissue that was not used for pathological analysis was used in this study. Thirteen samples were collected: two meningiomas, two grade 3 anaplastic astrocytomas (AA1 and AA2) and nine grade 4 glioblastoma multiforme tumors (GBM1-GBM9). Tumor samples were flash frozen in liquid nitrogen and homogenized in a liquid nitrogen cooled

Biopulverizer mortar and pestle (BioSpec Products, Bartlesville, OK) until particles were sub-millimeter size. Homogenized tumor tissue was aliquoted by weight into 15 ml conical tubes, and suspended in PBS + 10 µg/mL PMSF in isopropyl alcohol, with 1% formaldehyde for cross-linking. Samples were cross-linked for 15 minutes, rocking at room temperature, washed twice with PBS + PMSF, flash frozen in liquid nitrogen, and stored at -80°C until processing for ChIP-seq.

Briefly, samples were lysed by douncing in hypotonic buffer, followed by centrifugation and resuspension in RIPA buffer with protease inhibitors. Lysate was sonicated in an ice bath, 30 seconds on, 60 seconds off, high intensity (Bioruptor, Diagenode, Denville, NJ). Sonication continued for 4 ten-minute cycles, then samples were centrifuged to pellet insoluble particles. We used protein A beads to “pre-clear” the sample before overnight antibody incubation. Samples plus beads were rocked for 30 minutes (4°C), and supernatant was transferred to a new tube. An input sample was removed and antibodies were added for overnight incubation. We performed the following IPs for each tumor: CTCF (EMD Millipore, Billerica, MA, USA, 07-729), H3K4me3 (EMD Millipore, 07-473), H3K4me1 (EMD Millipore, 07-436), H3K27me3 (EMD Millipore, 07-449), H3K9ac (EMD Millipore, 07-352), H3K9me3 (Abcam, Cambridge, MA, USA, ab8898), H3K27ac (Abcam, ab4729), using 10 µg of antibody per IP.

After the overnight incubation, we added 30 µl beads to each IP, samples plus beads rocked for an hour (4°C), followed by salt/detergent washes (4°C) to reduce nonspecific binding. We eluted the antibody/DNA complexes from the beads twice, using 1% SDS and 0.1 M NaHCO₃ buffer. To reverse cross-linking, we added 5 M NaCl to each sample, then incubated at 65°C for >4 hours. For DNA extraction, samples were sequentially digested with RNase A and Proteinase K, then extracted with phenol-chloroform and precipitated with ethanol, DNA was resuspended in sterile water.

Library preparation and sequencing

After assessing enrichment at positive control sites using qPCR, we prepared libraries using the New England Biolabs NEBNext library prep kit (E6240L, New England Biolabs, Ipswich, MA, USA), with some changes. We performed adapter ligation before size selection, used Bioo adapters (514103, Bioo Scientific, Austin, TX, USA), and Ampure XP beads (A63881, Beckman-Coulter, Brea, CA, USA) instead of columns for all purifications and size selections. Sequencing was performed on a HiSeq 2500 in either the Genome Sequencing Facility at MD Anderson Cancer Center at Science Park (Smithville, TX) or the UT Austin Genome Sequencing and Analysis Facility (GSAF).

Alignment and peak calling in ChIP-seq data

Sequencing of input libraries was performed for each tumor. Available datasets and their sequencing details are summarized in the “Overview” and “ChIP-seq” sheets of Supplementary File S1. Alignment of all ChIP-seq and Input sequence datasets was performed with BWA (v0.7.12-r1039)(6) aln and sampe functions after hard trimming all reads to 50 bases. The reference used was the primary GRCh38 (hg38) assembly (GENCODE v24). Duplicate sequences were flagged using the Picard MarkDuplicates tool

(v1.123) (<http://broadinstitute.github.io/picard>). Statistics for these datasets are summarized in the “ChIP-seq” sheet of Supplementary File S1, which groups datasets by mark.

We performed initial identification of enriched regions (peaks) using the MACS2 (v2.1.1.20160309) (7) callpeak function specifying a relaxed *P*-value threshold of 0.01, allowing one duplicate read per locus (keep-dup=2), pairing each ChIP-seq sample with its input control. We removed data aligning to genomic regions with aberrantly high signal due to copy number differences (8), and we defined low complexity regions using the “Duke Excluded Regions” and “DAC Blacklisted regions” tracks at the UCSC Genome Browser, which were lifted over to hg38.

Statistics were gathered for each peak set, including peak counts at a wide range of MACS2-reported *Q*-value and fold enrichment (FE) levels and total base coverage for selected levels. Because *P*- and *Q*-values are strongly affected by sequencing depth, FE provided the most consistent measure for thresholds across chromatin mark sets. Utilizing FE also insulates these peak calls from the effects of copy number alterations in the tumor genome. We could detect large copy number alterations containing amplifications and deletions in the ChIP input libraries (Supplementary Fig. S1A and B). However, since FE was calculated relative to input, this provides an inherent normalization of any amplified regions, preventing them from dominating called peaks with the subsequent use of those peaks in identifying chromatin states. From examination of peak counts and base coverage at various FE levels, we selected three significance levels: high (FE 6+), target (FE 4.5+), and low (FE 3.5 or 4). MACS2-generated narrowPeak files were converted to a custom BED9+ format, including fields for *P*- and *Q*-values, FE, rank, and a significance level designator. All reported results are based on peaks at the target level. In preparation for model-building, target-level peaks were extracted and, for each mark, peaks from all tumors were merged using bedtools (v2.25.0) (9) in such a way as to preserve sample identity.

Chromatin states: systematic identification of histone co-localization

From the above merged consensus peaks for each experiment, we used ChromHMM v1.12 (5) to build a 21-state model, using only tumor ChIP-seq data. To map the distribution of enhancers across elements in the genome, annotation coordinates were defined based on UCSC RefFlat annotations from January 2015. Any genomic region outside of the above annotations was defined as intergenic. Enhancer coordinates were intersected with these elements using bedtools. To understand which genes are regulated by which states, we used the GENCODE (10) annotation, version 24, and used bedtools closest to identify the distance between chromatin states and the closest protein coding gene(s). For enhancers, we analyzed the closest gene even if it was >1 Mb away, but for bivalent regions, the closest gene (and associated bivalent domain) was not analyzed unless the distance was <500 bp.

Assessing gene enrichments for common enhancer and bivalent states using DAVID

In order to assess enriched pathways and terms for our common enhancer and bivalent domains, we utilized DAVID 6.8 (11), and identified enriched terms using functional annotation clustering. The DAVID terms in Fig. 5A and Fig. 6A were derived as follows. For each cluster, the term encompassing the largest number of genes while still having a

significant P -value (FDR corrected) was selected; an additional term could be chosen from any cluster if the terms did not describe redundant features. Annotation clusters were skipped if they appeared completely redundant to any previous cluster, or non-significant by FDR adjusted P -value of <0.05 . This process was continued until at most 10 significant terms were identified.

RNA sequencing in solid tumors and cell lines

RNA was isolated from homogenized tumor tissue using TRIzol (15596-026, Thermo Fisher Scientific, Waltham MA, USA). The Ribo-Zero rRNA removal kit (MRZH116, Illumina, San Diego, CA, USA) was used to remove ribosomal RNAs, and the resulting RNA was used to prepare single-end or paired-end libraries with the NEBNext small RNA kit for Illumina (E7300S, New England Biolabs, Ipswich, MA, USA). The resulting libraries were run on an Illumina HiSeq 2500 as above.

We used cutadapt to remove 3' Illumina RNA-seq adapters from fastq reads (v1.10)(<https://github.com/marcelm/cutadapt>), discarding sequences <36 bases after trimming. We filtered reads from rRNA and tRNA by aligning to a reference composed of human rRNA and tRNA sequences, retaining sequences that did not align. Finally, transcriptome-aware alignment was performed with TopHat2 (v2.1.0) (12). We used the primary GRCh38 (hg38) assembly from GENCODE release 24, and the corresponding GENCODE comprehensive gene annotation set for the transcriptome. Duplicate sequences were flagged using Picard as described above. Alignment statistics for all RNA-seq datasets are summarized in the “RNA-seq” sheet of Supplementary File S1. We used the Tuxedo suite to perform gene FPKM quantification on RNA-seq data (13). The cuffnorm table of tumor-only FPKM and cuffdiff (see below) output, merged based on the unique ENSEMBL identifier is provided in Supplementary File S2.

Assignment of TCGA subtypes to tumors

ENSEMBL Gene IDs from cuffnorm-derived FPKM tables were extracted and matched against 841 gene names from the TCGA subtyping study (14); resulting in 743 direct matches. Another 92 matches were identified using HGNC aliases, with 836 TCGA genes matched. FPKM values for the genes were extracted in TCGA-gene order. The resulting matrices had missing or 0 values replaced by the smallest R double precision value, counts $\log_2$ transformed; rows median centered; complete Spearman pairwise correlation of columns performed to construct a distance matrix; finally, average-linkage hierarchical clustering performed on columns represented by the distance matrix. Assignment of TCGA mesenchymal, classical or proneural subtypes to the tumors was performed based on the three main branches in the resulting dendrogram. We used these subtypes to perform differential gene expression analysis across the proneural and the mesenchymal/classical group. Group assignments were as follows: mesenchymal/classical: AA1, GBM3, GBM5, GBM7, GBM8, GBM9; proneural: AA2, GBM1, GBM2, GBM4, GBM6. Genes described as significantly differentially expressed between the two groups are those marked as significant at Q -value ≤ 0.05 by cuffdiff, included in Supplementary File S2.

Statistical analyses

We used R version 3.3.1 (The R Project for Statistical Computing, <http://www.r-project.org/>) software for calculations. Statistical tests, sample number, and data representation are indicated in the main text or figure legends.

Data access

The raw fastq files from the GBM datasets generated and analyzed in this study are available in the dbGaP repository: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001389.v1.p1

Results

Gene expression programs in tumors recapitulate clinically distinct GBM subtypes

GBM tumors are highly heterogeneous (15), and tumor samples were homogenized before processing, so the data are representative of bulk tumor, as opposed to any specific population of cells. We performed RNA-seq from all primary tumors, as well as several commonly used GBM-derived cell lines and two independent lines of primary normal human astrocytes. After alignment, gene expression was quantified over protein coding genes (10) (Materials and Methods). We plotted gene expression data for previously identified GBM subtype classifier genes (14), and found expression patterns similar to TCGA (Fig. 1B). Thus, the tumors that we used for chromatin profiling represent authentic and clinically relevant GBM subtypes. We did not identify a neural subtype among our tumors, and there is some uncertainty regarding whether this subtype is a distinct molecular subtype of GBM (2). *IDH1* is known to be frequently mutated in lower grade gliomas and in the G-CIMP subclass of the proneural subtype of GBM (16,17). In accordance, we found that 2 of the 5 proneural tumors in our sample set showed the characteristic R132H mutation of *IDH1*. GBM cell lines differed widely from the tumors in their gene expression patterns, and thus are not an accurate model of primary tumor lesions for genome-wide profiling studies (Supplementary Fig. S2A). When we clustered the tumors on the basis of all protein coding genes, three groups were evident: tumors, normal human astrocytes, and GBM-derived cell lines (Supplementary Fig. S2B).

Global profiling of histone modifications reveals biologically relevant patterns

To profile regulatory chromatin states in GBM, we developed a protocol to perform ChIP-seq in primary GBM tumor samples (Materials and Methods) and concurrently sequenced the RNA derived from these samples. We profiled four active histone modifications (H3K4me1, H3K4me3, H3K9ac, H3K27ac), two repressive histone modifications (H3K27me3, H3K9me3), and the multifunctional insulator binding protein *CTCF*. Transcribed genes such as calmodulin (*CALM1*) displayed active histone modifications (Fig. 2A), while transcriptionally silent genes such as keratin 72 (*KRT72*) showed broad repressive marks (Fig. 2B). Genome-wide, samples showing a given mark generally clustered together, with active marks, repressive marks, and *CTCF* binding forming distinct groups (Fig. 2C and D). Thus, the biological state of the chromatin rather than tumor

identity determined the clustering of datasets, indicating that the profiling data was reflective of the underlying chromatin state.

To systematically identify distinct chromatin states in the tumor genome, we called ChIP-seq peaks for each sample, then used ChromHMM (5) to build a 21-state model of histone modification co-localizations across the genome (Materials and Methods). Based on known associations of histone marks and *CTCF* binding with regulatory activities, we identified several functionally distinct chromatin states in the tumor genome (18), including an active enhancer, polycomb, and heterochromatin silenced state (Supplementary Table S1). Interestingly, the 21-state model revealed the existence of a bivalent state marked by active H3K4me3 and repressive H3K27me3 modifications. Such bivalent states were first identified in embryonic stem cells (ESCs) (19), and have been identified in glioblastoma-derived cell lines (20) but to our knowledge, this is the first time they have been seen to exist in primary GBM tumors. A view of the ChIP-seq signal surrounding a given chromatin state in a tumor showed that globally, the identified states faithfully reflected the underlying combinations of histone marks (Fig. 3A). At individual loci, the identified states captured the appropriate combination of marks corresponding to different types of functional elements such as promoters, enhancers, and silenced heterochromatin regions (Fig. 3B). Although our analysis was not geared towards identifying super-enhancers, there was good overlap between our identified enhancers in GBM and super-enhancers reported in medulloblastoma and in the U87MG GBM-derived cell line based on ChIP-seq for *MED1*, *BRD4* and *EP300* (4,21) (Fig. 3C).

Although expression was variable across states, any state with a repressive mark was associated with a statistically significant reduction in expression compared with other states (Fig. 4A; Supplementary Fig. S3A–D). We used whole genome bisulfite sequencing (WGBS) data from TCGA GBM tumors to examine DNA methylation levels corresponding to each state. The polycomb and heterochromatin silenced states were highly methylated, reflecting their transcriptional inactivity. Interestingly, although the genes nearest enhancers were highly expressed, the enhancers themselves were highly methylated (Fig. 4B, Supplementary Fig. S4A–D). While methylation is generally considered to be silencing (22), WGBS is unable to distinguish between methylation and 5-hydroxymethylation (5hmC) of cytosine residues. Enhancers and gene bodies are specifically marked by 5hmC in ES cells (23), and enhancers were recently reported to be targeted by 5hmC in GBM (24). By comparing our enhancer annotations to 5hmC data from GBM tumors (24), we found that GBM enhancers were enriched for 5hmC compared to promoters ($P = 6.212 \times 10^{-7}$, *t*-test), bivalent regions ($P = 4.08 \times 10^{-9}$), polycomb silenced regions ($P = 1.866 \times 10^{-9}$), and background 5hmC levels in the genome ($P = 2.778 \times 10^{-12}$) (Supplementary Fig. S5A). Genes corresponding to bivalent states were not expressed, and bivalent loci showed lower levels of methylation than polycomb and heterochromatin silenced regions. *CTCF* binding sites were associated with high levels of methylation, consistent with previous reports (25).

GBM enhancers are located predominantly within introns and intergenic regions, and contain degenerate STAT/Klf/SP family motifs

Active enhancers (state 12) varied widely in number among tumors, with a median number of 3,640 (Supplementary Table S2). Generally, enhancers localized within or upstream of gene bodies. The vast majority of enhancers in any given tumor were located in introns. The genomic distribution of enhancers that we identified in GBM showed small differences, particularly for introns and intergenic regions, compared to enhancers identified in clinical medulloblastoma tumors (4) or in cell lines (26) (Supplementary Table S3). Although the difference is small, the reduced representation of GBM enhancers in intergenic regions, and increased representation in intronic regions was consistent across tumor samples (Supplementary Fig. S5B). Though expression levels for genes associated with enhancers varied across tumors, there were no statistically significant changes in expression across tumors or subtypes, or in the average distance to a gene (Supplementary Fig. S5C and D).

We used MEME-ChIP (27) to identify *de novo* motifs overrepresented in enhancers in each tumor. These motifs were largely degenerate, and bore resemblance to motifs for several families of transcription factors, such as KLF and STAT-like binding motifs (Supplementary Fig. S6). Some motifs from 6–8 nucleotides in length strongly resembled primary and secondary binding sites for the *AP2* transcription factor family, as well as *TCF3*, *ASCL2*, and *TCF5*. From a functional perspective, the observed motifs indicate that enhancers are enriched for binding of transcription factors controlling cellular proliferation and immune response.

Enhancers are subtype-specific and control genes involved in cell-cell contacts

The enhancer state in our model is defined by co-localization of H3K27ac and H3K4me1 (Fig. 3A, Supplementary Table S2). Some transcriptional activity originated from these regions, but in a less defined manner than from promoter regions (Fig. 3A). 1,817 enhancers, which covered 1,227 genes, were present in at least 3 out of 11 tumors, and we defined this set as “common enhancers”. 307 of these genes were strongly enriched for pathways that mediate cell-cell interactions such as cell adhesion and cell-cell adherens junctions (Fig. 5A, Supplementary File S3).

These 307 genes with a DAVID (11) annotation from Fig. 5A were split between proneural (PN) and mesenchymal/classical (MES/CL) tumors, with genes expressed in PN tumors being minimally expressed in the MES/CL tumors and vice versa (Fig. 5B). Of these 307 genes, 33 were on the TCGA list of subtype genes (2). This includes five genes significantly upregulated in the PN tumors (*EPHB1*, *MAPT*, *NCAM1*, *KIF21B*, *STMN1*), and two genes that were significantly upregulated in the MES/CL tumors (*EGFR*, *OSBPL3*). In total, we identified 274 novel genes associated with GBM that are controlled by an enhancer in at least 3 tumors (Supplementary File S3). The subtype specificity of enhancers was often visually evident, as with the gene *PODXL*, which showed strong enhancer signals and expression in the MES/CL tumors, but much weaker enhancer signals and expression in PN tumors (Fig. 5C).

Bivalent regions are enriched for hedgehog signaling and embryonic development genes

The model identified a bivalent state (state 16, Supplementary Table S4) predominantly defined by co-localization of H3K4me3 and H3K27me3 (Fig. 3A). Transcription from this state was slightly higher than a polycomb silenced state, but lower than genes near active promoters or enhancers (Fig. 4A). There were 1,435 frequently bivalent regions present in at least 5 tumors, and these regions covered 1,511 distinct genes (Materials and Methods). 840 of these genes showed strong enrichment for pattern specification, Wnt signaling, embryonic development, transcription factor DNA binding domains (Fig. 6A, Supplementary File S4), and revealed a PN versus MES/CL division in gene expression, similar to enhancers. Frequently bivalent regions were divided in expression by subtype and fell into 2 groups: Group 1 contained genes expressed in MES/CL tumors that were often bivalent in PN tumors, and Group 2 exhibited largely the reciprocal pattern (Fig. 6B, Supplementary Fig. S7A–D). For example, Group 1 contained many genes within the *HOXB* locus that were bivalent in PN tumors but expressed in MES/CL tumors (Fig. 6C). When compared with existing data from adult brain (28), the GBM samples showed distinct epigenetic profiles (Supplementary Fig. S8).

To identify regions that were bivalent in a subtype-independent manner in GBM, we examined bivalent regions common to at least 8 out of 11 tumors, and identified 349 regions, which covered 381 unique genes (Supplementary File S5). 68 of these 381 genes (17.8%) were homeobox genes, a highly significant enrichment as only 1.25% of all genes are homeobox genes ($P < 2.2 \times 10^{-16}$, Fisher's Exact Test). Moreover, there were 127 transcription factors (33.3%) among the 381 commonly bivalent genes, an equally significant enrichment ($P < 2.2 \times 10^{-16}$, Fisher's Exact Test) compared to the 10% of all genes that are transcription factors. The occurrence of homeobox genes and transcription factors is thus likely to represent a functional attribute of bivalent chromatin domains in GBM. The bivalent domains were marked by punctate H3K4me3, with H3K27me3 more broadly distributed across the region (Fig. 6C).

Commonly bivalent genes were highly interconnected, with 192 of the genes connected through StringDb (29). 30 genes were not connected to the main network, so the primary network comprises 162 nodes. Using HumanNet, 176 of the nodes were connected (AUC = 0.612; $P = 1.54 \times 10^{-12}$) (30). The resulting network is dominated by *SHH* and *IHH*, with *WNT1*, GATA-family transcription factors, and the growth factor *FGF10* forming additional hubs (Fig. 6D). The presence of bivalent chromatin domains in cancer may indicate de-differentiation towards a more stem cell-like phenotype.

Discussion

Although gene expression profiling of primary tumors suggests 4 molecular subtypes – classical, mesenchymal, neural and proneural – detailed phenotypic and molecular characterization of glioma stem cells (GSCs), which are thought to be the tumor initiating cells in glioblastoma, reveal two distinct subtypes of GSCs, corresponding to the mesenchymal and proneural types (31). Strikingly, the genes targeted by the enhancers and bivalent chromatin states that we identified in primary tumors also separate them into two groups corresponding to the GSC-based classification of GBM. Moreover, genes targeted by

enhancers appear to regulate pathways that are differentially active in the two GSC subtypes. Many enhancer-associated genes that were significantly upregulated in MES/CL tumors promote cellular invasion and angiogenesis, a hallmark of mesenchymal GSCs (32,33). For example, *PODXL* (Podocalyxin-like) promotes cell migration, and its overexpression in GSCs is associated with a poor outcome (34); *MMP11* (Matrix metalloprotease 11) cleaves ECM, promotes tumorigenesis, and cellular invasion (35); *S100A16* (a Ca^{++} binding protein) promotes the epithelial-to-mesenchymal transition (EMT) in breast cancer (36); the protein kinase *FAM20C* is a marker of mesenchymal GSCs and promotes proliferation in triple-negative breast cancer (37); *LMO2* (LIM Domain Only 2) promotes angiogenesis (38), and is a marker of GSCs (32,33). Conversely, many enhancer-associated genes significantly upregulated in PN tumors were associated with increased survival. This set included, for example, *AKT3* and *DNM1* (dynamin-1) which are associated with survival in GBM (39), and *NCAM1* (Neural cell adhesion molecule 1) which is involved in neuron-neuron interactions and is negatively correlated with invasion (40). Many of these genes have not been previously associated with GBM.

Genes adjacent to frequently bivalent regions could be placed into two groups showing the same reciprocal relationship in expression between MES/CL and PN tumors as observed with enhancer-associated genes (Fig. 5B, Fig. 6B, Supplementary Fig. S7A–D). Thus, Group 1 genes active in MES/CL tumors, such as *COL6A2*, *SMOC2*, *ITGB2*, *FOXC2* and *HOXB3* have been associated with angiogenesis, cellular migration and invasive growth (41–43). Some Group 2 genes active in PN tumors were protective. For example, *ICAM5* is an intracellular adhesion molecule that regulates interactions between neurons and microglia, and it is often repressed in colon cancer (44). Another Group 2 gene, *SLIT2*, provides axon guidance in the developing forebrain, and patients with *SLIT2* positive gliomas show better survival (45). However, other Group 2 genes highly expressed in PN tumors were strongly suggestive of GSCs. Notable among these was *OLIG2*, a lineage specific transcription factor for oligodendrocytes that is required for proliferation of GSCs (46). *OLIG2* regulates *PDGFRA* (47), which was also a Group 2 bivalent gene highly expressed in PN samples, and while it is required for gliomagenesis (48), its expression is associated with an improved prognosis (49).

Genes marked by bivalent chromatin in 70% of GBM (8/11 tumors) were highly interconnected and formed a network dominated by Wnt (*WNT1*, *WNT2B*, *WNT6*), and hedgehog (*SHH*, *IHH*) signaling, as well as *HOX* and homeobox genes, and transcription factors. This bears strong similarity to signatures of bivalent chromatin both in embryonic stem cells, where the opposing active and polycomb repressed marks poise genes for developmental expression, as well as in cancer stem cells (CSCs) (19,50). *WNT5B* was recently identified as vital for differentiation and cell growth in GSCs (51). Bivalent loci marked by H3K4me3 and H3K27me3 have not been identified previously in primary GBM tumors; unexpectedly, the bivalent signature associated with GSCs, which comprise a small fraction of the overall tumor, was instead observable in the bulk tumor.

One potential issue with identifying bivalent regions in a population of primary tumor cells is that active and repressive marks could be present at the same locus, in distinct subpopulations of cells in heterogeneous tumors. Performing ChIP-reChIP to conclusively

identify active and repressive marks in the same cells could resolve this ambiguity, but the large number of cells required makes such experiments infeasible in clinical tissue samples. To establish if bivalent regions could be discerned in a homogeneous population of GBM-derived cells, we performed ChIP-seq for H3K4me3 and H3K27me3 in the T98G cell line. Interestingly, we identified 1073 loci showing both active and repressive marks in these cells. These loci overlapped significantly ($P < 4.257 \times 10^{-201}$; Bedtools Fisher) with the set of bivalent loci present in at least 8 tumors (Supplementary Fig. S9A and B) and included key genes like *HOXA3-7*, *SHH* and *WNT6*. Additionally, if active and repressive marks occurred in different populations of cells, the expectation is that such loci would show expression levels that are intermediate between the active and repressed state. However, the bivalent regions we identified in tumors showed undetectable expression levels, similar to polycomb repressed regions (Fig 3A, Fig. 4, Supplementary Fig. S2). Thus, although we cannot formally exclude the possibility that the active and repressive states occur in distinct cells in tumors, our data suggest that the bivalent state is not dispersed across active and repressed sub-populations of cells.

Wnt and hedgehog signaling (Hh) regulate EMT, invasion and proliferation in cancers; certain components of these pathways were expressed in the GBM tumors profiled here. The master regulators *IHH*, *SHH*, and *WNT1* were nearly always silent, but poised for expression. The large number of commonly bivalent transcription factors and homeobox genes suggests a rapidly deployable program controlling a Hh and Wnt-mediated transcriptional response with the capacity to drive the production of multipotent stem cells from more differentiated bulk tumor cells. Many genes expected to be specific to GSCs were highly expressed in unsorted tumor in a subtype specific manner, indicating the genetic pathways necessary for GSC programming are present in any GBM tumor cell. These findings corroborate other recent studies in cultured GSCs indicating that stemness in GBM is controlled in part by Wnt signaling (52) and by epigenetic regulation of *HOX* genes (53). Our study raises the possibility that targeting epigenetic states with small molecule modulators, in combination with agents that target Wnt and hedgehog signaling pathways, might be a fruitful approach for exploring subtype-specific therapy in GBM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded in part by grants from the Cancer Prevention Research Institute of Texas (RP120194) and NIH (HG004563, CA130075 and CA198648).

We are grateful to the patients for consenting to donate tissue. This study would not be possible without them. We thank the staff at NeuroTexas Institute at St. David's Medical Center (Austin, TX), the Next Generation Sequencing Core Facility at the University of Texas MD Anderson Cancer Center Science Park, which was supported by CPRIT Core Facility Support Grant RP120348, as well as the Genome Sequencing and Analysis Facility at the University of Texas at Austin for sequencing, and the Texas Advanced Computing Center at The University of Texas at Austin for HPC resources. This work was funded in part by grants from the Cancer Prevention Research Institute of Texas (RP120194) and NIH (HG004563, CA130075 and CA198648) to V.R. Iyer.

Abbreviations

ChIP-seq	Chromatin immunoprecipitation sequencing
GBM	Glioblastoma multiforme
AA	Anaplastic astrocytoma
TCGA	The Cancer Genome Atlas

References

1. Delgado-Lopez PD, Corrales-Garcia EM. Survival in glioblastoma: a review on the impact of treatment modalities. *Clin Transl Oncol*. 2016; 18:1062–71. [PubMed: 26960561]
2. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17:98–110. [PubMed: 20129251]
3. Lucio-Eterovic AK, Cortez MA, Valera ET, Motta FJ, Queiroz RG, Machado HR, et al. Differential expression of 12 histone deacetylase (HDAC) genes in astrocytomas and normal brain tissue: class II and IV are hypoexpressed in glioblastomas. *BMC cancer*. 2008; 8:243. [PubMed: 18713462]
4. Lin CY, Erkek S, Tong Y, Yin L, Federation AJ, Zapatka M, et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*. 2016; 530:57–62. [PubMed: 26814967]
5. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*. 2012; 9:215–6. [PubMed: 22373907]
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–60. [PubMed: 19451168]
7. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology*. 2008; 9:R137. [PubMed: 18798982]
8. Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, et al. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*. 2011; 21:456–64. [PubMed: 21106903]
9. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–2. [PubMed: 20110278]
10. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012; 22:1760–74. [PubMed: 22955987]
11. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
12. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013; 14:R36. [PubMed: 23618408]
13. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7:562–78. [PubMed: 22383036]
14. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061–8. [PubMed: 18772890]
15. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 2013; 110:4009–14. [PubMed: 23412337]
16. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013; 155:462–77. [PubMed: 24120142]
17. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. 2009; 360:765–73. [PubMed: 19228619]

18. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011; 470:279–83. [PubMed: 21160473]
19. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–26. [PubMed: 16630819]
20. Lin B, Lee H, Yoon JG, Madan A, Wayner E, Tonning S, et al. Global analysis of H3K4me3 and H3K27me3 profiles in glioblastoma stem cells and identification of SLC17A7 as a bivalent tumor suppressor gene. *Oncotarget*. 2015; 6:5369–81. [PubMed: 25749033]
21. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013; 153:320–34. [PubMed: 23582323]
22. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462:315–22. [PubMed: 19829295]
23. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome biology*. 2011; 12:R54. [PubMed: 21689397]
24. Johnson KC, Houseman EA, King JE, von Herrmann KM, Fadul CE, Christensen BC. 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nature communications*. 2016; 7:13177.
25. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012; 22:1680–8. [PubMed: 22955980]
26. Ashoor H, Klefogiannis D, Radovanovic A, Bajic VB. DENdb: database of integrated human enhancers. *Database : the journal of biological databases and curation*. 2015; 2015
27. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res*. 2015; 43:W39–49. [PubMed: 25953851]
28. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518:317–30. [PubMed: 25693563]
29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015; 43:D447–52. [PubMed: 25352553]
30. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res*. 2011; 21:1109–21. [PubMed: 21536720]
31. Nakano I. Stem cell signature in glioblastoma: therapeutic development for a moving target. *J Neurosurg*. 2015; 122:324–30. [PubMed: 25397368]
32. Cheng L, Huang Z, Zhou W, Wu Q, Donnola S, Liu JK, et al. Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell*. 2013; 153:139–52. [PubMed: 23540695]
33. Kim SH, Kim EJ, Hitomi M, Oh SY, Jin X, Jeon HM, et al. The LIM-only transcription factor LMO2 determines tumorigenic and angiogenic traits in glioma stem cells. *Cell death and differentiation*. 2015; 22:1517–25. [PubMed: 25721045]
34. Binder ZA, Siu IM, Eberhart CG, Ap Rhys C, Bai RY, Staedtke V, et al. Podocalyxin-like protein is expressed in glioblastoma multiforme stem-like cells and is associated with poor outcome. *PLoS One*. 2013; 8:e75945. [PubMed: 24146797]
35. Kou YB, Zhang SY, Zhao BL, Ding R, Liu H, Li S. Knockdown of MMP11 inhibits proliferation and invasion of gastric cancer cells. *Int J Immunopathol Pharmacol*. 2013; 26:361–70. [PubMed: 23755751]
36. Zhou W, Pan H, Xia T, Xue J, Cheng L, Fan P, et al. Up-regulation of S100A16 expression promotes epithelial-mesenchymal transition via Notch1 pathway in breast cancer. *J Biomed Sci*. 2014; 21:97. [PubMed: 25287362]
37. Chandran UR, Luthra S, Santana-Santos L, Mao P, Kim SH, Minata M, et al. Gene expression profiling distinguishes proneural glioma stem cells from mesenchymal glioma stem cells. *Genom Data*. 2015; 5:333–6. [PubMed: 26251826]

38. Yamada Y, Pannell R, Forster A, Rabbitts TH. The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. *Proc Natl Acad Sci U S A*. 2000; 97:320–4. [PubMed: 10618416]
39. Patel VN, Gokulrangan G, Chowdhury SA, Chen Y, Sloan AE, Koyuturk M, et al. Network signatures of survival in glioblastoma multiforme. *PLoS Comput Biol*. 2013; 9:e1003237. [PubMed: 24068912]
40. Wang Z, Dai X, Chen Y, Sun C, Zhu Q, Zhao H, et al. MiR-30a-5p is induced by Wnt/beta-catenin pathway and promotes glioma cell invasion by repressing NCAM. *Biochemical and biophysical research communications*. 2015; 465:374–80. [PubMed: 26255203]
41. Liu Y, Carson-Walter EB, Cooper A, Winans BN, Johnson MD, Walter KA. Vascular gene expression patterns are conserved in primary and metastatic brain tumors. *J Neurooncol*. 2010; 99:13–24. [PubMed: 20063114]
42. Shvab A, Haase G, Ben-Shmuel A, Gavert N, Brabletz T, Dedhar S, et al. Induction of the intestinal stem cell signature gene SMOC-2 is required for L1-mediated colon cancer progression. *Oncogene*. 2016; 35:549–57. [PubMed: 25915847]
43. Fu H, Fu L, Xie C, Zuo WS, Liu YS, Zheng MZ, et al. miR-375 inhibits cancer stem cell phenotype and tamoxifen resistance by degrading HOXB3 in human ER-positive breast cancer. *Oncol Rep*. 2017; 37:1093–9. [PubMed: 28075453]
44. Ashktorab H, Rahi H, Wansley D, Varma S, Shokrani B, Lee E, et al. Toward a comprehensive and systematic methylome signature in colorectal cancers. *Epigenetics : official journal of the DNA Methylation Society*. 2013; 8:807–15.
45. Liu L, Li W, Geng S, Fang Y, Sun Z, Hu H, et al. Slit2 and Robo1 expression as biomarkers for assessing prognosis in brain glioma patients. *Surg Oncol*. 2016; 25:405–10. [PubMed: 27916173]
46. Kupp R, Shtayer L, Tien AC, Szeto E, Sanai N, Rowitch DH, et al. Lineage-Restricted OLIG2-RTK Signaling Governs the Molecular Subtype of Glioma Stem-like Cells. *Cell reports*. 2016; 16:2838–45. [PubMed: 27626655]
47. Lu F, Chen Y, Zhao C, Wang H, He D, Xu L, et al. Olig2-Dependent Reciprocal Shift in PDGF and EGF Receptor Signaling Regulates Tumor Phenotype and Mitotic Growth in Malignant Glioma. *Cancer Cell*. 2016; 29:669–83. [PubMed: 27165742]
48. Liu KW, Feng H, Bachoo R, Kazlauskas A, Smith EM, Symes K, et al. SHP-2/PTPN11 mediates gliomagenesis driven by PDGFRA and INK4A/ARF aberrations in mice and humans. *J Clin Invest*. 2011; 121:905–17. [PubMed: 21393858]
49. Chen D, Persson A, Sun Y, Salford LG, Nord DG, Englund E, et al. Better prognosis of patients with glioma expressing FGF2-dependent PDGFRA irrespective of morphological diagnosis. *PLoS One*. 2013; 8:e61556. [PubMed: 23630597]
50. Nicolis SK. Cancer stem cells and “stemness” genes in neuro-oncology. *Neurobiol Dis*. 2007; 25:217–29. [PubMed: 17141509]
51. Hu B, Wang Q, Wang YA, Hua S, Sauve CG, Ong D, et al. Epigenetic Activation of WNT5A Drives Glioblastoma Stem Cell Differentiation and Invasive Growth. *Cell*. 2016; 167:1281–95. e18. [PubMed: 27863244]
52. Rheinbay E, Suva ML, Gillespie SM, Wakimoto H, Patel AP, Shahid M, et al. An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell reports*. 2013; 3:1567–79. [PubMed: 23707066]
53. Kurscheid S, Bady P, Sciuscio D, Samarzija I, Shay T, Vassallo I, et al. Chromosome 7 gain and DNA hypermethylation at the HOXA10 locus are associated with expression of a stem cell related HOX-signature in glioblastoma. *Genome biology*. 2015; 16:16. [PubMed: 25622821]

Significance

Enhancers and bivalent domains in glioblastoma are regulated in a subtype-specific manner that resembles gene regulation in glioma stem cells.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Bulk tumors represent the known molecular subtypes in GBM

(A) Overview of our approach. We profiled histone modifications and used ChromHMM (5) to produce a model of chromatin states, and associated distinct chromatin states with gene expression profiles from the same tumors.

(B) The panel on the left displays microarray data used by TCGA to establish molecular subtypes in GBM (14). The TCGA array data was previously clustered by rows, and we retained this order, while hierarchically clustering the samples using Spearman's rho. The right-hand panel displays RNA sequencing data generated in this study. The gene order is the same as on the left and tumors were arranged according to the similarity of their expression profiles with the TCGA data (Materials and Methods).

Figure 2. Epigenetic profiles in GBM tumors

(A) Active chromatin over the promoter region for calmodulin (*CALM1*), which is highly expressed in the tumor GBM7. Region displayed: chr14:90,391,001-90,427,000 on the hg38 assembly.

(B) A polycomb-repressed region, marked by H3K27me3, in the same tumor as A over the gene *KRT72*, which encodes a keratin protein. Region displayed: chr12:52,580,318-52,607,332 on the hg38 assembly.

(C) A clustered correlation heatmap of all chromatin profiles generated in this study. The pairwise correlation coefficients across the 33,188 genomic loci showing measurable signal in at least 15 experiments are shown, with clusters of chromatin marks indicated by the text.

(D) Heatmap of normalized ChIP-seq signals in the same genomic loci shown in C. Both genomic loci (vertical) and tumors (horizontal) were hierarchically clustered.

Figure 3. Chromatin states in the GBM tumor genome

(A) Epigenetic signal in a 20 kb region surrounding four different chromatin states in GBM4. The data were sorted by H3K4me3 signal for promoters, H3K4me1 signal for enhancers, and H3K27me3 signal for bivalent regions, as well as polycomb-repressed regions. See Supplementary Table S1 for details on each chromatin state.

(B) A view of 877 kb on chromosome 1 for the tumor GBM8, showing 7 ChIP-seq tracks, RNA-seq, chromatin states and gene annotations. Below, a close-up view of chromatin states showing, from left, a weak enhancer, a promoter, and a heterochromatin-silenced region over a lncRNA of unknown function. Region displayed on the hg38 assembly:

chr1:18,621,857-19,499,690.

(C) Intersection of super-enhancers identified in medulloblastoma (4) or the GBM-derived cell line U87MG (21), with GBM enhancers identified in this study. A single super-enhancer could overlap with many individual enhancers identified in our study, and therefore two numbers are reported in the intersection. *P*-values for the intersections were calculated using Bedtools Fisher (9) and were approximately zero.

Figure 4. Expression and methylation across chromatin states in the model

(A) Average expression of genes closest to each of the four states displayed in Fig. 3A. Normalized FPKM counts across all tumors were derived from cuffnorm.

(B) Average methylation across each of the states in Fig. 3A. Fractional methylation levels were derived from WGBS data in GBM and intersected with our data using bedtools intersect.

Figure 5. GBM enhancers regulate gene expression in a cell-type specific manner

(A) Functional enrichment for genes associated with enhancers in at least 3 tumors. FDR-adjusted *P*-values provided by DAVID (11) are shown above each column, with the shading proportional to the significance.

(B) Heatmap of gene expression from the 307 enhancer associated genes with a functional annotation in A, demonstrating clustering of PN and MES/CL tumors with regard to gene expression. The genes are sorted based on the ratio between average FPKM values across PN and MES/CL groups (Supplementary File S3).

(C) A 140 kb genome browser view of the region surrounding the gene *PODXL*. This enhancer is subtype-specific, and is much stronger in the MES/CL tumors (indicated by purple bars, with the enhancer region in MES/CL tumors outlined by a red dashed rectangle). The hg38 coordinates of the region shown are: chr7:131,500,691-131,640,269.

Figure 6. Bivalent regions underlie Wnt and SHH signaling in GBM

(A) Functional enrichment for genes associated with bivalent domains in at least 5 tumors. FDR-adjusted P -values provided by DAVID (11) are shown above each column, with the shading proportional to the significance.

(B) Heatmap of gene expression from the 840 genes with DAVID annotations in A, demonstrating clustering of PN and MES/CL tumors with regard to gene expression. The genes are sorted based on the ratio between average FPKM values across PN and MES/CL groups (Supplementary File S4).

(C) A genome browser view of a 90 kb region on chromosome 17, encompassing the HOXB cluster of genes. From top to bottom, tracks show chromatin states, genes, RNA-seq, H3K4me3 binding, and H3K27me3 binding. Proneural tumors (green bars) are bivalent, while 4 out of 6 MES/CL tumors (purple bars) show expression over this region. The hg38 coordinates of the region shown are chr17:48,539,612-48,628,935.

(D) Interconnectivity between genes identified as bivalent in at least 8 tumors. The size of the nodes is scaled by the number of connected edges. The seven most common functional gene classes are colored, with a legend at the bottom of the panel; TF = transcription factor, RTK = receptor tyrosine kinase.