





TECHNICAL ADVANCES

# Chromatin landscapes reveal developmentally encoded transcriptional states that define human glioblastoma

Stephen C. Mack<sup>1\*</sup>, Irtisha Singh<sup>2\*</sup>, Xiuxing Wang<sup>3\*</sup>, Rachel Hirsch<sup>2</sup>, Quilian Wu<sup>3</sup>, Rosie Villagomez<sup>2</sup>, Jean A. Bernatchez<sup>4,5</sup> , Zhe Zhu<sup>3</sup> , Ryan C. Gimple<sup>3,6</sup>, Leo J.Y. Kim<sup>3,6</sup>, Andrew Morton<sup>7</sup>, Sisi Lai<sup>7</sup>, Zhixin Qiu<sup>3</sup>, Briana C. Prager<sup>3,8,9</sup>, Kelsey C. Bertrand<sup>1</sup>, Clarence Mah<sup>10</sup>, Wenchao Zhou<sup>8</sup>, Christine Lee<sup>8,11</sup>, Gene H. Barnett<sup>12</sup>, Michael A. Vogelbaum<sup>12</sup>, Andrew E. Sloan<sup>13</sup>, Lukas Chavez<sup>10</sup>, Shideng Bao<sup>8</sup> , Peter C. Scacheri<sup>7</sup>, Jair L. Siqueira-Neto<sup>4,5</sup>, Charles Y. Lin<sup>2,14</sup> , and Jeremy N. Rich<sup>3</sup>

**Glioblastoma is an incurable brain cancer characterized by high genetic and pathological heterogeneity. Here, we mapped active chromatin landscapes with gene expression, whole exomes, copy number profiles, and DNA methylomes across 44 patient-derived glioblastoma stem cells (GSCs), 50 primary tumors, and 10 neural stem cells (NSCs) to identify essential super-enhancer (SE)-associated genes and the core transcription factors that establish SEs and maintain GSC identity. GSCs segregate into two groups dominated by distinct enhancer profiles and unique developmental core transcription factor regulatory programs. Group-specific transcription factors enforce GSC identity; they exhibit higher activity in glioblastomas versus NSCs, are associated with poor clinical outcomes, and are required for glioblastoma growth in vivo. Although transcription factors are commonly considered undruggable, group-specific enhancer regulation of the MAPK/ERK pathway predicts sensitivity to MEK inhibition. These data demonstrate that transcriptional identity can be leveraged to identify novel dependencies and therapeutic approaches.**

## Introduction

Glioblastoma (GBM; World Health Organization grade IV glioma) is the most prevalent and lethal primary, intrinsic brain tumor (Ostrom et al., 2016). Standard therapy of surgery, radiation, and alkylating chemotherapy offer only palliation, and ultimately almost all patients succumb to this disease (Stupp et al., 2009). Other than conventional therapies, the only US Food and Drug Administration–approved GBM therapies include tumor-treating fields (Stupp et al., 2017) and bevacizumab (Gilbert et al., 2014). Although GBM has been extensively characterized at the molecular level, current clinical management has limited guidance from molecular pathology, beyond mutations in isocitrate dehydrogenase 1/2 (*IDH1/2*) and *MGMT* promoter methylation (Hegi et al., 2005; Cancer Genome Atlas

Research Network, 2008; Parsons et al., 2008; Yan et al., 2009; Brennan et al., 2013; Frattini et al., 2013). Additionally, gliomas have high intratumoral phenotypic heterogeneity, leading to the hypothesis that self-renewing GBM stem cells (GSCs) persist after therapy and are able to regenerate refractory tumors (Singh et al., 2004; Bao et al., 2006; Chen et al., 2012a; Patel et al., 2014; Suvà et al., 2014). GBMs possess remarkable cellular plasticity and are highly responsive to changes in nutrient state, oxygen levels, and the local tumor microenvironment (Li et al., 2009b; Eyler et al., 2011; Flavahan et al., 2013; Zhou et al., 2015; Quail et al., 2016; Miller et al., 2017; Quail and Joyce, 2017). These data support highly dynamic and robust tumor ecosystems that resist standard and genetically defined targeted therapies.

<sup>1</sup>Department of Pediatrics, Division of Pediatric Hematology and Oncology, Baylor College of Medicine, Houston, TX; <sup>2</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; <sup>3</sup>Division of Regenerative Medicine, Department of Medicine, University of California, San Diego, San Diego, CA; <sup>4</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA; <sup>5</sup>Center for Discovery and Innovation in Parasitic Diseases, University of California, San Diego, La Jolla, CA; <sup>6</sup>Department of Pathology, Case Western Reserve University, Cleveland, OH; <sup>7</sup>Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; <sup>8</sup>Department of Stem Cell Biology and Regenerative Medicine, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; <sup>9</sup>Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH; <sup>10</sup>Division of Medical Genetics, Department of Medicine, University of California, San Diego, San Diego, CA; <sup>11</sup>Department of Pharmacology, Case Western Reserve University, Cleveland, OH; <sup>12</sup>Rose Ella Burkhardt Brain Tumor & Neuro-Oncology Center, Cleveland Clinic, Cleveland, OH; <sup>13</sup>Department of Neurological Surgery, Seidman Cancer Center & University Hospitals – Cleveland Medical Center, Cleveland, OH; <sup>14</sup>Therapeutic Innovation Center, Verna and Marrs McClean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX.

\*S.C. Mack, I. Singh, and X. Wang contributed equally to this paper; Correspondence to Jeremy N. Rich: [drjeremyrich@gmail.com](mailto:drjeremyrich@gmail.com); Charles Y. Lin: [charles.y.lin@bcm.edu](mailto:charles.y.lin@bcm.edu); Stephen C. Mack: [scmack@bcm.edu](mailto:scmack@bcm.edu).

© 2019 Mack et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Despite these challenges, transcriptional profiling of GBM lacking *IDH1* or *IDH2* or classic driver mutations has identified molecular subgroups that predominate among patient tumors (Phillips et al., 2006; Verhaak et al., 2010; Wang et al., 2018a). Several transcriptional profiling paradigms have been proposed, but predominantly there have been at least three biologically distinct profiles (Phillips et al., 2006; Verhaak et al., 2010). The first is characterized by a proneural gene expression program, likely defined by the *OLIG1* and *OLIG2* transcription factors (TFs), which are master regulators of oligodendrocyte identity (Lu et al., 2002; Yu et al., 2013). The second is marked by mesenchymal gene expression signatures and, like many mesenchymal tumors, has been associated with increased resistance to radiotherapy (Carro et al., 2010; Kim et al., 2016). The third, referred to as classical or proliferative, is typically driven by recurrent genomic alterations found in *EGFR*, *PTEN*, and *CDKN2A*. Although bulk tumors tend to fall into one category or the other, single-cell RNA sequencing (RNA-seq) suggests that cells of distinct states exist within the same tumor (Patel et al., 2014). Emerging evidence suggests that tumors can transition between these states, designated as proneural-to-mesenchymal transition, suggesting that this plasticity of cellular identity is a hallmark of the disease (Bhat et al., 2013; Mao et al., 2013; Segerman et al., 2016).

At its core, cellular identity is defined and governed by a small number of master TFs that are tissue specific and establish the cell's chromatin and epigenetic landscape (Takahashi and Yamanaka, 2006). Overexpression of master TFs reprogram or transdifferentiate cells (Weintraub et al., 1989; Graf and Enver, 2009). As a rule, master TFs are regulated by large adjacent enhancer elements (often termed super-enhancers [SEs]), and master TFs bind to their own enhancers and those of other master TFs, thereby forming an autoregulatory TF core regulatory circuitry. Cancers often maintain portions of the core regulatory circuitry from their cells of origin, and at the same time, oncogenic deregulation of SEs and core regulatory circuitry are common themes across cancers (Mansour et al., 2014; Lin et al., 2016; van Groningen et al., 2017; Wang et al., 2018b; Zeid et al., 2018). Recently, we and others demonstrated in medulloblastoma, ependymoma, and neuroblastoma that molecularly defined tumor subgroups each possess distinct core regulatory circuitries and that defining TFs were often faithful tracers of cell-of-origin and critically shown to be nonmutated tumor dependencies (Lin et al., 2016; Garancher et al., 2018; Wang et al., 2018b; Zeid et al., 2018).

## Results

### GBMs exhibit distinct chromatin and epigenetic profiles distinguished from other brain tumors

To define the TF core regulatory circuitries of GBM, we integrated active enhancer landscapes defined by histone H3 lysine acetylated chromatin (H3K27ac) with gene expression, DNA methylomes, copy number variations (CNVs), and whole exomes across a large cohort of GSCs ( $n = 44$ ), primary GBM surgical specimens ( $n = 50$ ), and neural stem cells (NSCs;  $n = 10$ ; Fig. 1 and Fig. S1). 101 samples were profiled by at least one

genomic measurement, and within this dataset, 79 samples were matched for H3K27ac, gene expression, and methylation. In sum, this integrated genomic and epigenomics dataset represents the largest integrated cohort of chromatin and epigenetic profiling compiled for GBMs. All datasets were subjected to rigorous quality control based on Encyclopedia of DNA elements (ENCODE) best practices, and only datasets passing stringent quality control metrics are presented in this study (Fig. S1).

Overall, we find strong concordance between gene expression, DNA methylation, and active enhancers. For instance, at the *SOX2* locus we see consistent hypomethylation of promoter DNA, the presence of active euchromatin modifications (H3K27ac) at both the promoter and proximal enhancer, and evidence of abundant mRNA (Fig. 1, a and b). Globally, there is a positive correlation between H3K27ac at gene promoters and proximal enhancer with mRNA levels (Fig. S1 a) and anti-correlation between DNA methylation and mRNA levels (Fig. S1 a). These data suggest that our combined profiling efforts faithfully capture the epigenetic, chromatin, and transcriptional state of GBM.

Previously, we and others have shown that reference chromatin and epigenetic landscapes of brain tumors can be used to accurately to characterize tumor identity, classify tumor subgroups, and faithfully recapitulate molecular features of disease (Capper et al., 2018). Using t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction, we mapped chromatin and epigenetic profiles of our samples onto a comparative landscape of additional tumor types and tissues. By active enhancer H3K27ac profiles, GSCs and GBM surgical samples form clusters distinct from other brain tumors (e.g., medulloblastoma and ependymoma), as well as normal brain and other tissues drawn from the ENCODE and Roadmap Epigenomics Consortium datasets (Fig. 1 c). Integrating DNA methylomes into the German Cancer Research Center brain tumor DNA methylation atlas (<https://www.moleculareuropathology.org/mnp>; Capper et al., 2018), GSCs and GBM surgical samples cluster closely together with reference non-*IDH1*-mutated GBMs and were distinct from other brain tumors, as well as genetically defined GBM subtypes (e.g., *IDH1/2* mutant; Fig. 1 d). Supporting this finding, the most frequent somatic alterations detected in GSC models included *TP53* mutations, *PTEN* mutations, *EGFR* amplifications, and *CDKN2A/p16* deletions, further reinforcing that our models recapitulate genetic features of primary tumors while capturing molecular features of tumor heterogeneity (Fig. S1 b). Within GBMs, classically defined mutational spectra (e.g., *EGFR* mutations) fail to cluster either by active enhancer or DNA methylation landscapes, suggesting an additional contribution of chromatin and epigenetic state to GBM identity (Fig. 1).

### Recurrent SE genes define GSC identity and uncover novel dependencies

To delineate genes important for defining and maintaining the GSC state, we mapped SE loci across a large panel of patient-derived GSC cultures. Our analysis revealed 8,533 distinct SE loci detected in at least one GSC sample (Fig. 2 a; Materials and

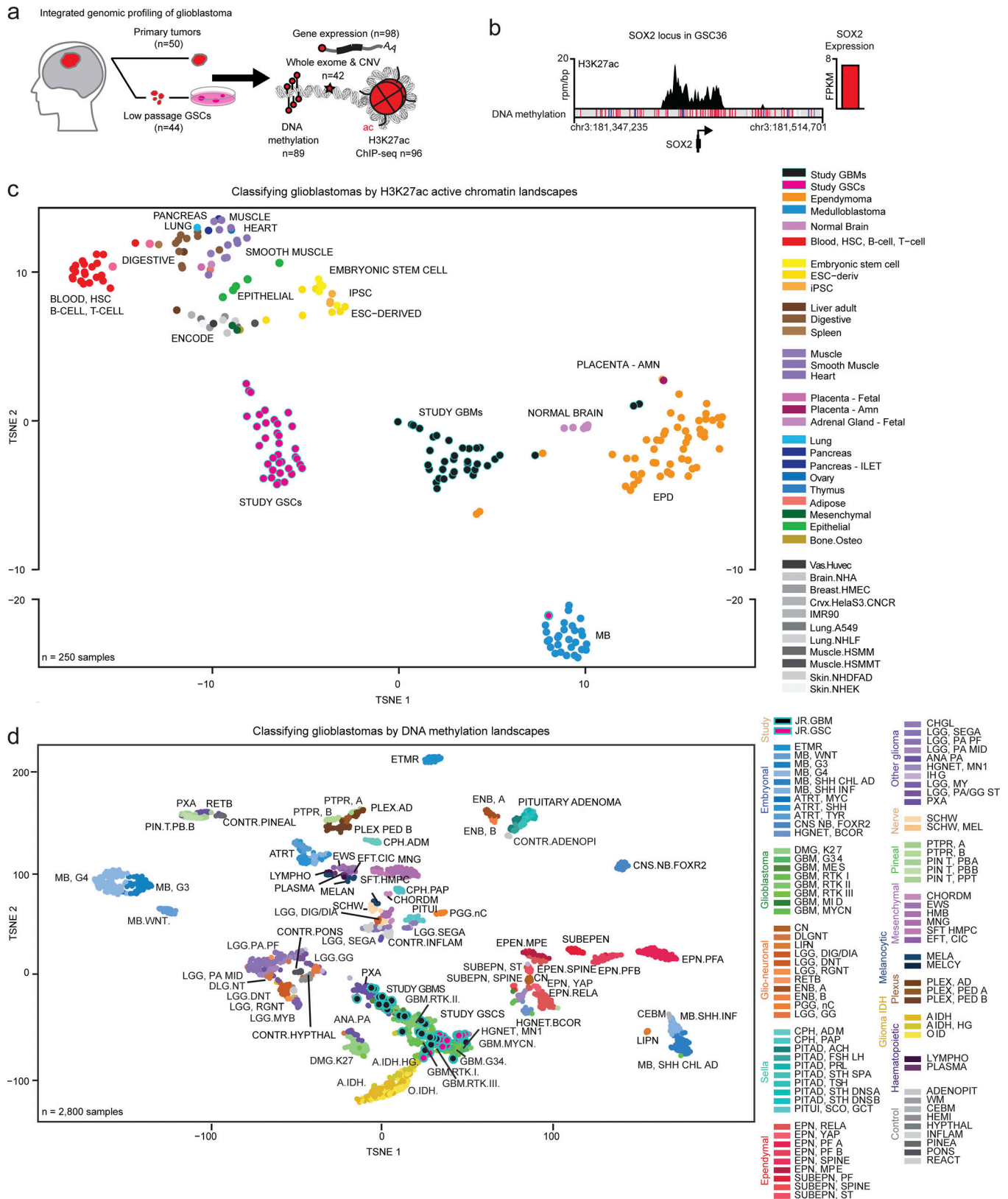


Figure 1. **Active chromatin landscape of GBM and GSCs.** (a) ChIP-seq H3K27 acetylation profile surrounding the SOX2 locus across GBM. (b) DNA methylation levels (Illumina EPIC arrays) surrounding the SOX2 locus. (c) t-SNE analysis of the top 10,000 variant H3K27ac loci of GBM primary tumors, GSCs, medulloblastoma primary tumors, ependymoma primary tumors, Roadmap consortium samples, and ENCODE consortium samples (Kundaje et al., 2015).  $n = 250$  samples. (d) t-SNE analysis of the top 10,000 variant CpG loci across study GBM, GSCs, and DNA methylation brain tumor classification reference samples (Capper et al., 2018). HSC, hematopoietic stem cell; iPSC, induced pluripotent stem cell; AM, placental amnion; EPD, ependymoma; MB, medulloblastoma.

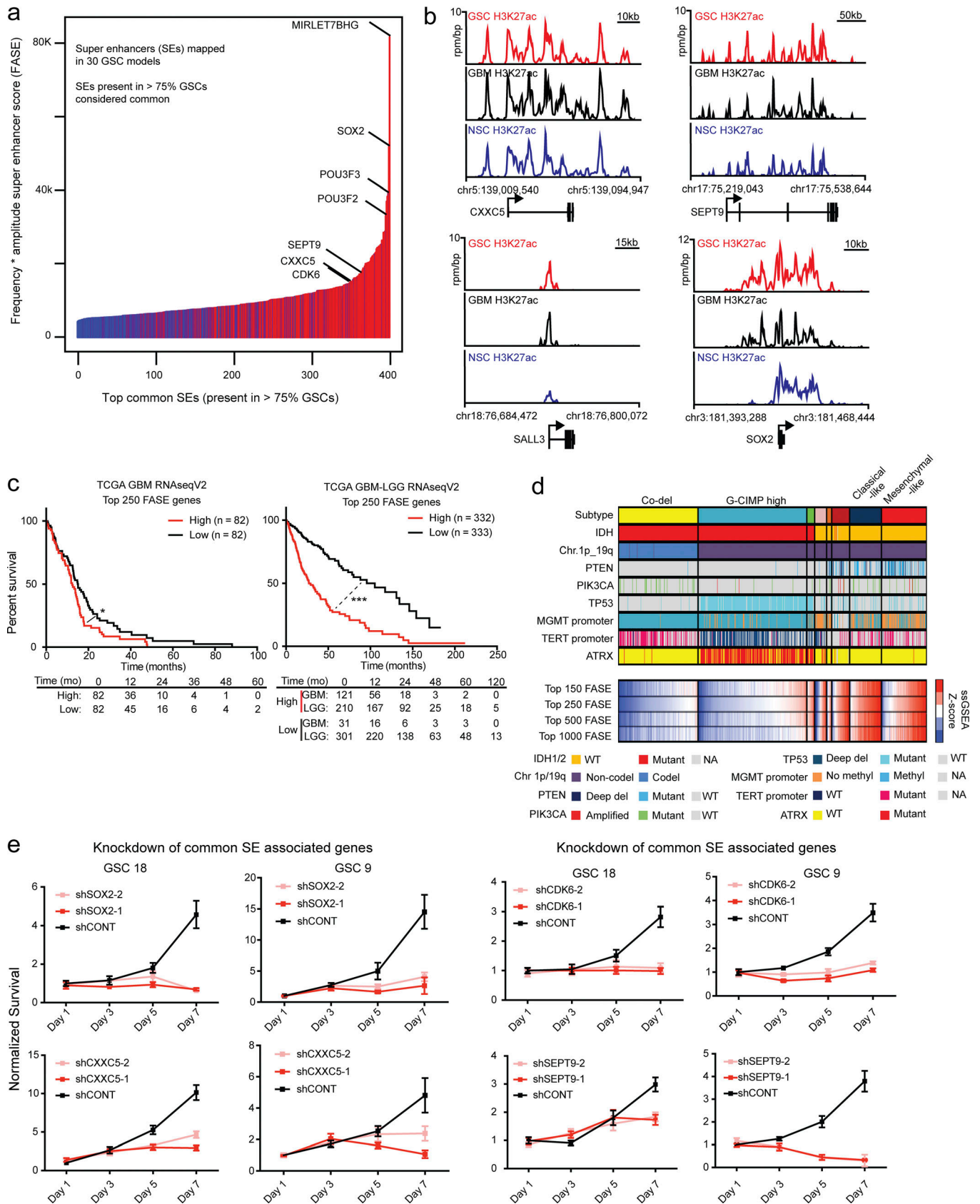


Figure 2. **Core GSC SE genes delineate cellular dependencies and define a negative prognostic signature.** (a) Bar plot indicating recurrent GSC SE genes detected across GSC models with a frequency >75%. Color code indicates frequency of SE presence. (b) Examples of SE loci regulating CXXC5, SEPT9, SALL3, and SOX2 in GSC, GBM, and NSC samples. (c) Kaplan–Meier survival analysis of the top 250 FASE gene signature scores in TCGA GBM and GBM-low-grade

glioma (LGG) RNaseqV2 datasets, respectively. A log-rank test was used to calculate significance of survival differences with P values indicated as \*,  $P \leq 0.05$  and \*\*\*,  $P \leq 0.0001$ . Each patient cohort was stratified as high versus low groups at the corresponding median single-sample gene set enrichment analysis z-score. The number of surviving or uncensored patients in each analysis group at noted time points was also displayed. **(d)** Distribution of top FASE gene signature scores in TCGA GBM low-grade glioma RNaseqV2 dataset among TCGA molecular subtypes and relevant genotypes and epigenotypes (*IDH1* mutation; Chr. 1p/19q codeletion; *PTEN* deletion; *PIK3CA* mutation; *TP53* mutation; *MGMT* promoter methylation status; *TERT* promoter methylation status; *ATRX* mutation status). **(e)** shRNA knockdown of core SE genes *SOX2*, *SEPT9*, *CXXC5*, *CDK6*, *SALL3*, and *EGFR* in GSC models using two distinct hairpins compared with a nontargeting control. Cell proliferation experiments were performed in biological replicates with more than four technical replicates per time point. Error bars indicated as standard deviation of four technical replicates.

methods). Across all samples, we identified a subset of SE loci and genes shared by a high proportion of GSCs (>75% of the samples) that represent potential key mediators of GSC identity (Fig. 2, a and b). These recurrent SEs were ordered by the product of SE frequency and amplitude (H3K27ac levels), termed FASE score. Many of these genes serve functional roles in GSC maintenance, including *CDK6*, *SOX2*, *EGFR*, *BRD4*, *POU3F2*, and *SALL3* (Suvà et al., 2014; Lathia et al., 2015; Liu et al., 2015). Novel core GSC identity transcripts were also identified, such as *MIR4316* and *CXXC5* (Fig. 2 a and Fig. S2). SE-associated genes exhibited elevated H3K27ac in GSCs, primary GBM tumor tissues, and in some cases, NSCs (Fig. 2 b). Globally, proximity of an SE to a gene correlated with increased expression in GSCs and GBM surgical samples (Fig. S2 a). To delineate the impact of core SE transcription on clinical outcome, we derived a gene signature composed of the top 250 SE-regulated genes identified in GSCs, which predicted patient survival in an independent set of transcriptional profiles defined by The Cancer Genome Atlas (TCGA; Verhaak et al., 2010; Brennan et al., 2013; Wang et al., 2018a). Patients with GBM that harbors enrichment of the core GSC SE signature have inferior survival compared with those with reduced enrichment (Fig. 2 c). Patients with high expression of this core GSC SE signature have higher-grade tumor histology and are less likely to have tumors with *IDH1* mutations and a glioma CpG-island methylator phenotype (Noushmehr et al., 2010; Fig. 2 d). Our data support a paradigm in which many genes that define GSC identity are highly expressed, delineated, and regulated by the presence of SEs. To determine if core GSC identity genes were required for tumor cell proliferation, we targeted several top SE genes—*SOX2*, *SEPT9*, *CXXC5*, *CDK6*, *SALL3*, and *EGFR*—using shRNAs in a panel of patient-derived GSC models (Fig. 2 e). Core GSC SE-associated identity genes were essential for tumor cell proliferation over a 7-d time course, suggesting that GSCs depend on expression of core SE genes to facilitate cell growth. Collectively, our data uncovered a set of GSC identity genes that are regulated by SEs and are required for GSC maintenance. Tumors that harbor enrichment of this core GSC SE signature exhibit features of increased tumor malignancy and correlate with a worse prognosis. Core GSC SE genes, therefore, represent a lead set of non-mutationally classified dependency genes as promising targets for therapeutic testing.

#### GSCs are composed of at least two distinct SE states

Given that core GSC SE genes define a shared axis of GSC identity genes (Singh et al., 2017), we next sought to investigate SE landscape heterogeneity in a large series of GSC models. Using unsupervised hierarchical clustering of H3K27ac activity, we found that two highly distinct SE states predominated among

GSCs (Fig. 3 a; Materials and methods). These subgroups, annotated as group 1 and group 2, comprised the majority (24/30) of GSCs profiled and exhibited high similarity within groups and low similarity between groups. Differential analysis of SEs between group 1 and group 2 revealed 597 group 1 and 651 group 2 SEs that drive group-specific differences (Fig. 3 b; Materials and methods). These regions show opposing patterns of H3K27ac at exemplary loci and globally (Fig. 3 c). Differential H3K27ac patterns often appear binary—most regions have greater than fivefold change in H3K27ac (Fig. 3 d)—indicative of de novo chromatin acetylation in each subgroup as opposed to a more quantitative increase/decrease in acetylation. Almost all genes proximal to and associated with GSC group-specific SEs are differentially expressed between groups consistent with enhancer acetylation correlating to target gene expression (Fig. 3 d). Using single-sample gene set enrichment analysis, TCGA-defined molecular signatures from bulk primary GBM samples profiles were used to classify group 1 and group 2 samples. While group 2 was highly associated with “mesenchymal” features, group 1 samples contained both “proneural” and “classical” or “proliferative” features (Phillips et al., 2006; Verhaak et al., 2010; Mao et al., 2013; Wang et al., 2018a), suggesting potentially convergent pathways at the level of active chromatin regulation. We hypothesized that group-specific SE-associated genes are also likely to be enriched for dependencies but in a group-specific manner. Genetic knockdown of group-specific genes, such as *BCAN* and *HMG2A*, in GSCs expressing concordant SEs attenuated tumor growth both in vitro and in vivo (Fig. 3, e and f), reinforcing the relationship between SE presence and dependency. Overall, these data suggest that GBM and GSCs are composed of at least two distinct SE states and that maintenance of SEs that regulate group-specific identity is required for tumor growth/proliferation.

#### Group-specific TFs define group identity and are required for glioma growth

Cellular identity in both tumors and normal tissues is defined by a small number of TFs that establish cell-type SEs, are often regulated by large SEs, and regulate one another in what has been termed a transcriptional core regulatory circuitry (Boyer et al., 2005). Reverse analysis of active enhancer landscapes maps putative TF core regulatory circuitry (Lin et al., 2016; Mack et al., 2018). Here, we strengthened TF enhancer-target gene regulatory predictions by implementing a ridge regression model, which incorporates gene expression to identify TFs where gene expression of the TF and target correlate or anti-correlate significantly (Fig. 4 a). Using this approach, we identified cohorts of SE-associated TFs that form highly interconnected

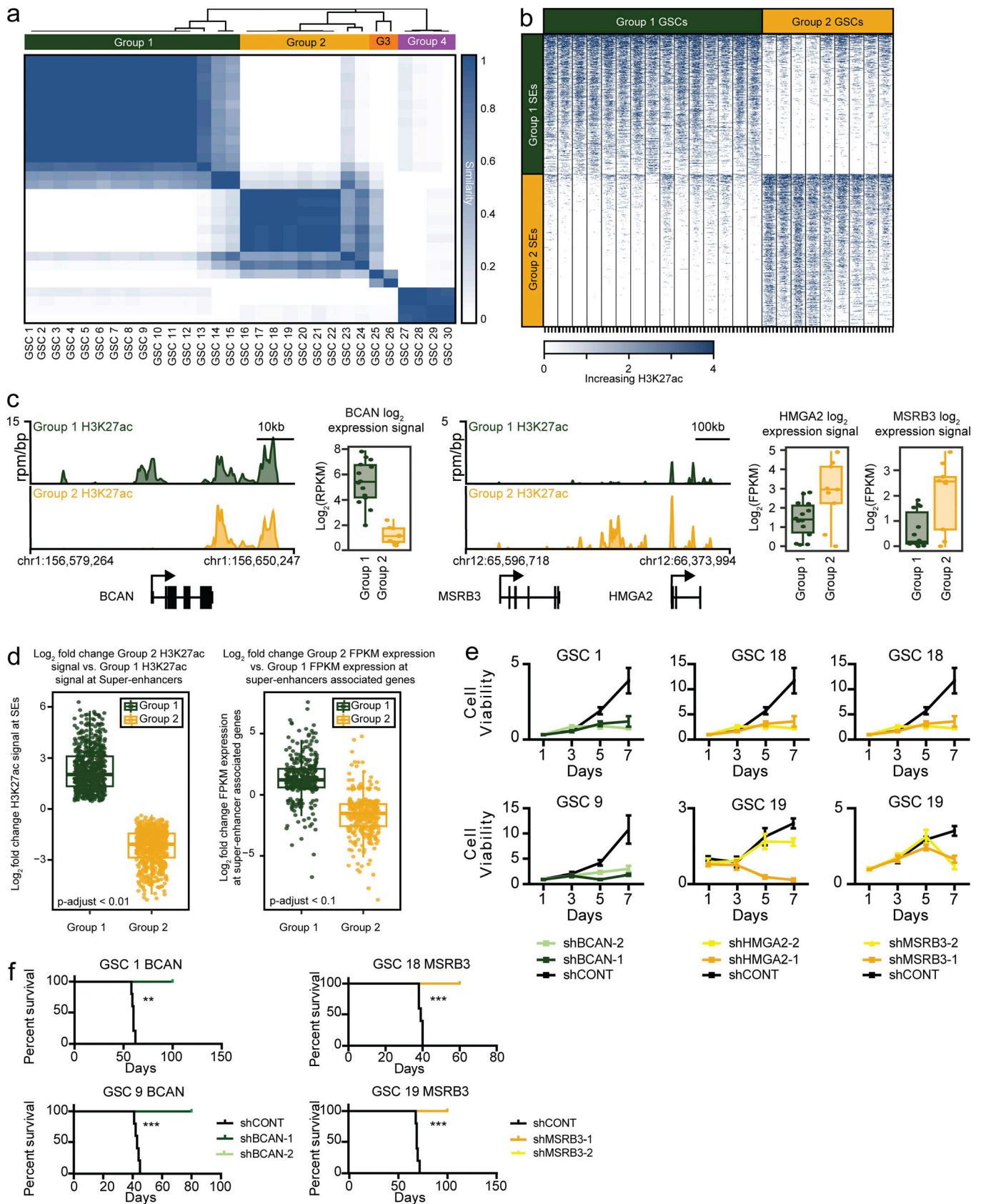


Figure 3. **GBMs display at least two distinct SE states.** (a) Similarity of 30 GSC samples by consensus clustering of 10% most variable regions by H3K27ac signal ( $n = 14,621$ ). The GSCs clustered into four groups, group 1 GSCs ( $n = 15$ ), group 2 GSCs ( $n = 9$ ), group 3 GSCs ( $n = 2$ ), and group 4 GSCs ( $n = 4$ ). (b) H3K27ac occupancy around the significantly different SE (FDR-adjusted  $P < 0.01$ ) between group 1 GSCs and group 2 GSCs (group 1 SEs,  $n = 597$ ; group 2 SEs,  $n = 651$ ).

(c) H3K27ac activity at the *BCAN*, *HMGGA2*, and *MSRB3* SEs for group 1 GSCs ( $n = 15$ ) and group 2 GSCs ( $n = 9$ ). Gene expression from RNA-seq (FPKM,  $\log_2$ FPKM) of *BCAN*, *HMGGA2*, and *MSRB3* shown on the right (group 1 GSCs,  $n = 14$ ; group 2 GSCs,  $n = 9$ ). (d) Change in H3K27ac activity at group 1 and group 2 SEs (FDR-adjusted  $P < 0.01$ , left) and change in expression of SE-associated genes (FDR-adjusted  $P < 0.1$ , right). (e) Effect of in vitro shRNA knockdown of *BCAN*, *HMGGA2*, and *MSRB3* on growth of group 1 and group 2 GSCs. Cell proliferation experiments were performed in biological replicates with more than four technical replicates per time point. Error bars indicated as standard deviation of four technical replicates. (f) Effect of in vivo shRNA knockdown of *BCAN* and *HMGGA2* on survival of mice. A log-rank test was used to calculate significance of survival differences with  $P$  values indicated as \*\*,  $P \leq 0.001$  and \*\*\*,  $P \leq 0.0001$ .

circuitries and are enriched at the SEs of each respective group (Fig. 4 b; Materials and methods). Group-specific TFs were also shown to have group-specific SE association, expression, and exhibited highly restricted expression patterns across human tissues (Fig. 4 c). Functional analysis of high-confidence target gene sets identified thematic pathways consistent with features of oligodendrocyte and neuronal differentiation for group 1 and JAK/STAT signaling for group 2 (Fig. S4 f). shRNA knockdown of *RUNX2*, a group 2-specific TF with no prior association to GBM, was lethal in vitro and in vivo (Fig. 4, d and e) but not in group 1 GSCs. *OLIG2*, a group 1-specific TF that has been implicated in oligodendrocyte identity, showed a reciprocal pattern of dependency with its knockdown highly sensitive in group 1 but not group 2 (Fig. 4 d). These data suggest that GSC subgroups are defined by and driven by highly tumor-specific TFs identified within core regulatory programs and that group-specific TFs are themselves highly specific dependencies.

#### Group-specific TF core regulatory circuitries are recapitulated in primary tumors

Primary GBMs exhibit high intratumoral heterogeneity and are highly dynamic to signaling queues from the tumor microenvironment (Hu et al., 2016; Venkatesh et al., 2017; Griveau et al., 2018; Wang et al., 2018b). As TF core regulatory circuitries were defined by using in vitro GSCs, we next sought to determine whether primary GBMs contained similar group-specific transcriptional regulation both across surgical tumor specimens and in the individual cells comprising tumors. Using group-specific TFs and their high-confidence target genes, we found that by RNA-seq and H3K27ac profiling GBM biopsy specimens strongly associated with one group or the other with a high overall degree of similarity (Fig. 5, a-c; Materials and methods). These data are consistent with prior multi-omic profiling efforts on primary tumors that originally elucidated GBM subgroups (Phillips et al., 2006; Verhaak et al., 2010). Moreover, they suggest that tumor identity as defined by GSC TF core regulatory circuitry is stable across GBM independent of cell culture conditions.

Within tumors, single-cell RNA-seq profiling has revealed varying degrees of heterogeneity with some tumor cell populations more homogeneously aligning with a specific GBM subgroup and others containing more mixed populations and traits of multiple subgroups (Patel et al., 2014). To investigate the TF core regulatory circuitry of single cells within GBM tumors, we analyzed these single-cell expression data from five GBM tumors with 50–100 cells each. The expression profile of each cell was correlated against signatures composed of group-specific TFs and their high-confidence target genes (Fig. 5 d; Materials and methods). Considering significant correlations to TFs, two of the tumors contained mostly cells assigned to group

1 (MGH26 and MGH30). Cells for tumors MGH28 and MGH29 showed a strong correlation to group 2 and more moderate correlation to group 1. For MGH31, only a small number of cells significantly correlated with a specific subgroup, and no group-specific bias was observed. Our findings demonstrate the high sample-sample variance in group 1 and group 2 identities and potential utilization of other distinct GSC states found to be less abundant across GBM (Fig. 3 a). Given our delineation of group 1- and group 2-specific genetic dependencies, there may be utility in estimating GSC group/state-specific compositions within GBM samples as a basis for molecular targeting.

#### Group-specific aberrant activation of developmental TFs in GBM

The presence of at least two GSC subgroups suggests either multiple cells of origin for GBMs or multiple transforming events that fundamentally alter cell state and transcriptional identity. To further investigate this question, we interrogated a cohort of NSCs, commonly considered to be useful comparators for GSCs and upstream or primordial to the GBM developmental cell of origin. By H3K27ac enhancer landscapes and TF activity, NSCs subdivided and clustered with either group 1 or group 2 GSCs (Fig. 6, a-d). Further analysis revealed that although most group-specific TFs are consistent between group-specific GSCs and their associated NSCs, a small number of TFs exhibited aberrant enhancer activity and expression in GSCs relative to NSCs (Fig. 6, e and f). For group 1, GSC-specific TFs included *OLIG1/2*, *NR2E1*, *NKX2-2*, *EN1*, and *MEOX2*. *OLIG1* and *OLIG2* are master regulators of oligodendrocyte identity and GSC drivers (Lu et al., 2002; Yu et al., 2013; Tachon et al., 2019; Fig. 6 e). *NR2E1* (*TLX*) is an orphan nuclear hormone receptor previously implicated in both GSC and NSC renewal (Zhu et al., 2014). *NKX2-2*, *EN1*, and *MEOX2* are developmental TFs with neural-specific expression whose roles in glioma are heretofore unknown. For group 2, *HOXA* and *HOXB* developmental TFs were up-regulated compared with NSCs. Additionally, a cohort of TFs with known roles in mesenchymal identity and inflammatory signal transduction were up-regulated, including *RUNX2*, *STAT3/5*, *TFAP2A*, and *IRX3* (Fig. 6 e). As data above support *OLIG1/2* and *RUNX2* as GSC-specific dependencies, we hypothesized that GSC-specific TFs define a highly oncogenic gene expression program. For each group, we defined high-confidence target gene sets for GSC-specific and GSC/NSC shared TFs. Thus, when compared with NSCs, GSCs activate and integrate highly lethal oncogenic TFs into their core regulatory circuitry.

#### Core cell identity genes uncover therapeutic targets in GBM

Core GSC SE genes were integrated with the Washington University Drug Gene Interaction Database (Cotto et al., 2018) and

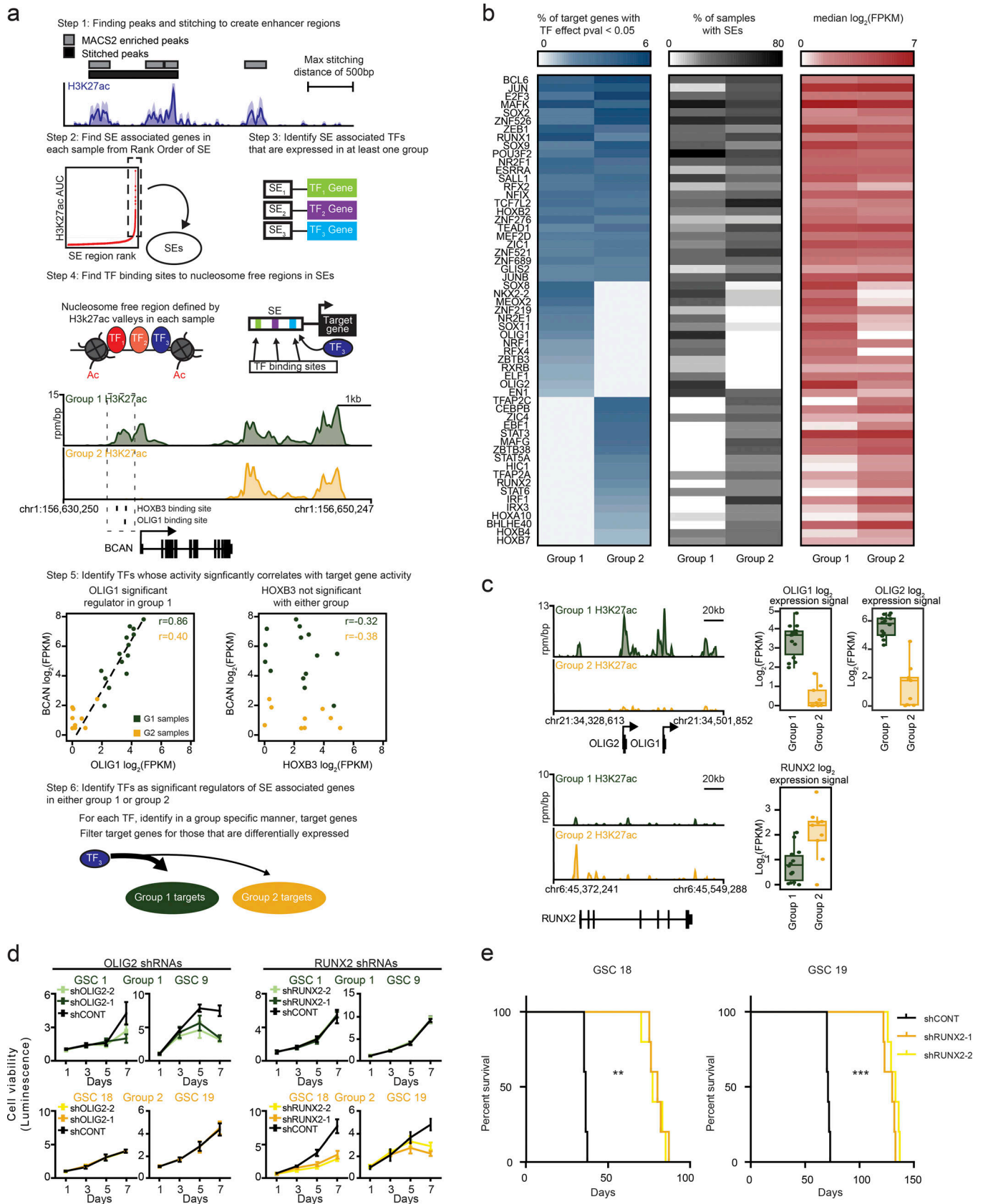


Figure 4. **Group-specific TFs define group identity.** (a) Schematic representation of the methodology used for identification of group-specific TFs. (b) Consensus TFs and group-specific TFs ( $P < 0.05$ ), with fraction-of-target genes for which the TF potentially plays an important role in regulation (left), fraction of group 1 and group 2 GSC samples that had an SE associated with the TF (middle), and median expression of TFs for both the groups (right). (c) Comparative



traces for H3K27ac ChIP-seq and mRNA levels from GSCs in groups 1 and 2 are displayed (two-sided Wilcoxon rank-sum test,  $P < 6.37 \times 10^{-05}$ ,  $P < 3.05 \times 10^{-06}$ ,  $P < 5.5405 \times 10^{-03}$  for OLIG1, OLIG2, and RUNX2, respectively). **(d)** GSCs from group 1 and group 2 were transduced with either a control shRNA sequence that does not target a sequence in the mammalian genome (shCONT) or one of two shRNAs targeting OLIG2 and RUNX2 (shOLIG2 and shRUNX2). Cellular proliferation was measured over a time course using CellTiter-Glo. Cell proliferation experiments were performed in biological replicates with more than four technical replicates per time point. Error bars indicated as standard deviation of four technical replicates. **(e)** GSCs from group 2 were transduced with either a control shRNA sequence that does not target a sequence in the mammalian genome (shCONT) or one of two shRNAs targeting RUNX2 (shRUNX2) then xenotransplanted into the right frontal lobes of immunocompromised NSG mice. Survival of tumor-bearing mice is displayed by the Kaplan–Meier method. A log-rank test was used to calculate significance of survival differences with P values indicated as \*\*,  $P \leq 0.001$  and \*\*\*,  $P \leq 0.0001$ .

annotated 718 SE associated genes as targets with drug interactions with 309 genes nonannotated (Figs. 2 a and 7 a). Many of these candidates were divided into targets defined as “clinically actionable” or classified as “druggable” proteins, kinase targets, or histone-associated proteins (Fig. 7 a). Four GSC models driven by predicted clinically actionable SE genes *EGFR* and *CDK6* were treated with a combination of EGFR and CDK4/6 inhibitors, lapatinib, and palbociclib, revealing cellular growth inhibition by treatments (Fig. 7, b and c). GBM models were further inhibited by combinatory treatments against EGFR and CDK6 in mice bearing orthotopic xenograft tumors with relevant GSC SEs (Fig. 7 d), validating this approach as potentially informative of preclinical efficacy for precision-based therapy selection. Given our findings that GBM identity (and group-specific specification) are regulated by SEs and distinct TF programs, and that these represent potential cellular dependency genes, we asked whether this list of potential oncogenic drivers could be explored for therapeutic targeting by small-molecule inhibitors. High-confidence group-specific target genes and pathways were mined with publicly available gene–drug interaction databases (Cotto et al., 2018) to pinpoint 16 small-molecule inhibitors of interest targeting group 1 or group 2 SE genes. A high-throughput screen was performed on 12 GSC models ( $n = 7$  for group 1 and  $n = 5$  for group 2), testing 16 drugs, at 10 concentrations, in eight technical replicates, and in two biological replicates (Fig. 7 e). This compound screen identified MAP3K1 (MEKK1) as a target with almost exclusive group 1 enhancer regulation (Fig. 7 f). MAP3K1 is upstream of canonical MAPK/ERK and JNK activity, and we hypothesized that group 1 GSCs may harbor selective dependency to MAPK/ERK inhibition. Using AZD8330, a potent and selective inhibitor to the MAPK/ERK MEK family of kinases, we observed selective effects on GSC group 1 viability at low micromolar doses with little effect in group 2 GSCs (Fig. 7, e and g). Our findings, therefore, demonstrate that SE programs that regulate GSC identity (both core and group specific) can be effectively mapped and used for therapeutic selection as a basis for discovery of non-mutationally defined oncogenic dependencies.

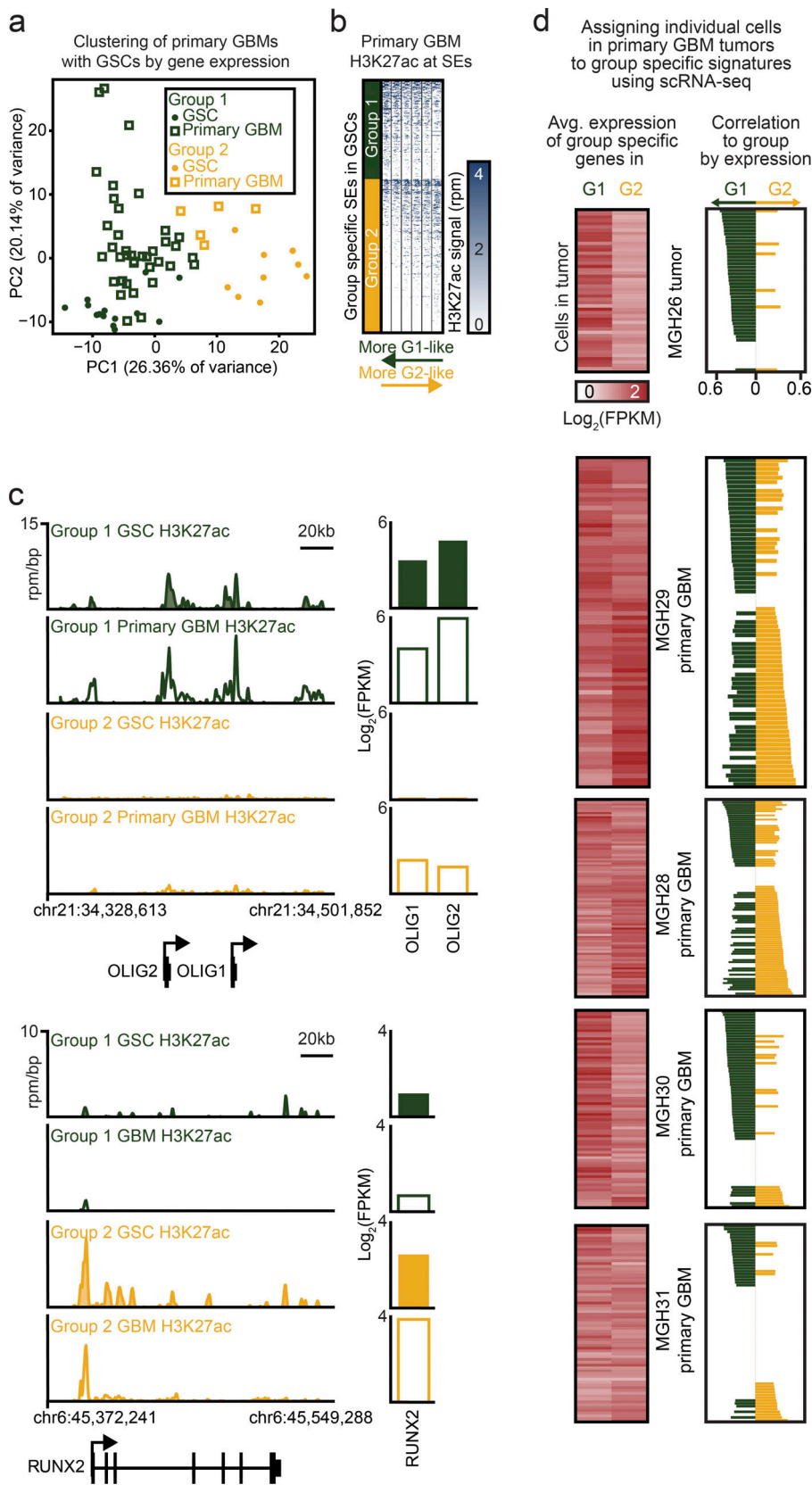
## Discussion

Numerous studies have sought to define groups of GBMs that share transcriptional profiles. Classification schemes based on these transcriptional signatures have ranged in the number of groups widely, but none has reliably informed clinical management beyond the identification of the IDH1 or IDH2 mutant tumors, which are tumors with a distinct molecular basis. Transcriptional profiles appear dynamic, displaying both spatial

and temporal transitions. We recently showed that defined radiographical features were associated with selected chromatin regulation and transcriptional profiles: contrast enhancement with an EZH2-associated vascular (proneural) transcriptional signature and necrosis with a BMI1-associated hypoxic (mesenchymal signature; Jin et al., 2017). GBMs display spatial diversity of genetic events, which have not, to date, been associated with specific regional variation, but genetic analyses have suggested that tumors may accumulate mutations selectively based on heterochromatin and euchromatin states, supporting crosstalk between epigenetic and genetic tumor regulation (Chen et al., 1998; Zheng et al., 2014; Makova and Hardison, 2015).

Further, GBMs may transition in the dominant signature upon recurrence after radiation and chemotherapy (Phillips et al., 2006; Verhaak et al., 2010). Despite these complexities or perhaps because of them, the identification of core, reinforcing TF circuitry may provide relatively preserved molecular regulation over time and space. We demonstrated across a large panel of GSCs maintenance of such consistent chromatin patterns as defined by presence and elevated expression of core SE genes. Our results are in line with previous studies that use comparable approaches to characterize the active chromatin landscapes of GSCs in an effort to delineate GSC identity (Suvà et al., 2014; Singh et al., 2017). However, owing to the large and diverse panel of GSC models that we characterized in our study, we were able to 1) robustly quantify the frequency of core SEs detected, and 2) dissect the molecular subgroups of GSCs as defined distinct SE landscapes. Elevated expression of the top 250 most frequent GSC SE-linked genes stratified patients with poor survival. Furthermore, our characterization of a large number of GSC models from patients revealed consistent patterns of at least four GSC subgroups harboring unique SE signatures. The two predominant groups, which we assigned group 1 and group 2, have some similarities in terms of their regulation by key TFs (i.e., OLIG1/2 and RUNX2) that define proneural and mesenchymal GBM subgroups, respectively. However, at the transcriptional level, group 1 and group 2 do not significantly match proneural and mesenchymal subgroup signature genes, thus supporting that the GSC groups we identify are restricted to specific tumor cell populations, highly consistent between distinct sets of patients, and defined by distinct transcriptional programs.

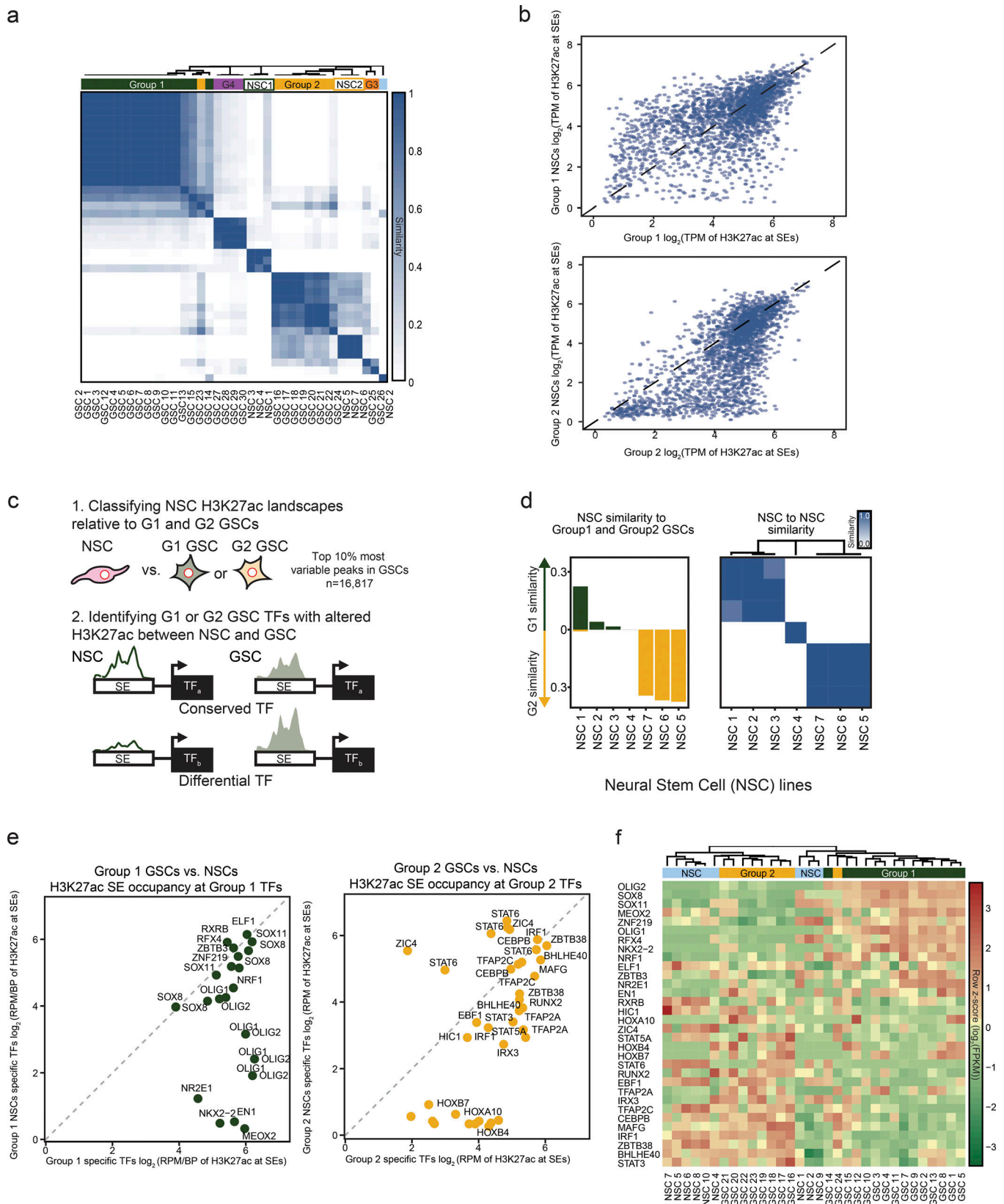
A proneural–mesenchymal transition (Bhat et al., 2013; Mao et al., 2013; Barnes et al., 2018) has been described in numerous studies on GBM as a cell identity switch in response to micro-environmental and/or treatment-associated factors (i.e., irradiation). As nearly all of our specimens were derived from newly



**Figure 5. Group-specific TF core regulatory circuitries are recapitulated in primary tumors and coopted in NSCs.** (a) Assigning the primary GBM samples by H3K27ac signal at the SEs of group 1 gene signature (group 1-specific TFs,  $n = 14$ ; target genes,  $n = 125$ ) and group 2 gene signature (group 2-specific TFs,  $n = 18$ ; target genes,  $n = 171$ ) by *k*-means. Assignment was reconfirmed by using gene expression of the two gene signatures (group 1-like primary GBMs = 38; group 2-like primary GBMs = 5). The figure shows the principal components of the GSCs and GBMs for log<sub>2</sub> (FPKM) of the above set of genes along with groups to which they were assigned. (b) The same as in Fig. 4 b but for primary GBMs (group 1-like primary GBMs = 6; group 2-like primary GBMs = 1). (c) The same as in Fig. 4 c with gene expression of the sample on the right. (d) Similarity of single cells of five primary GBMs to group 1 and group 2 GSCs using the gene signatures used for panel a. Left: The average gene expression of the group 1 gene signature and group 2 gene signature in every single sample. Right: The median Spearman rank-correlation with  $P < 0.01 \geq 75\%$  of group 1 and group 2 GSCs.

diagnosed primary unirradiated specimens, the proneural-mesenchymal transition process does not explain the group 2 GSC signatures we observed in treatment-naive GSC models

and tumors. This does not exclude the possibility that group 1 and group 2 GSC populations may transition between states as this was not explicitly tested in our study and may be an



**Figure 6. Group-specific TF core regulatory circuitries are co-opted in NSCs. (a)** Similarity of group 1 and group 2 GSC samples along with NSC samples ( $n = 7$ ) by consensus clustering of the 10% most variable regions by H3K27ac signal ( $n = 16,817$ ). **(b)** Median H3K27ac signal (rpm/bp) at SEs in group 1-like NSCs and group 1 GSCs (top). Median H3K27ac signal (rpm/bp) at SEs in group 2-like NSCs and group 2 GSCs. **(c)** Schematic showing (1) the identification of NSCs similar to group 1 and group 2 GSCs by their H3K27ac landscape and (2) the identification of conserved and differential TFs of group 1 and group 2 GSCs compared with NSCs. **(d)** Classifying the NSCs into group 1-like GSCs ( $n = 4$ ) and group 2-like GSCs ( $n = 3$ ) based on their average similarity with the group

1 and group 2 GSC samples (left) and the similarity of NSCs within themselves (right). Similarity measure was obtained by consensus clustering of the 10% most variable regions by H3K27ac signal ( $n = 16,817$ ). **(e)** Median H3K27ac signal (rpm/bp) at SEs of group 1-specific TFs (top) in group 1-like NSCs and group 1 GSCs (top). Median H3K27ac signal (rpm/bp) at SEs of group 2-specific TFs (bottom) in group 2-like NSCs and group 2 GSCs. **(f)** Row-normalized gene expression levels ( $\log_2$ FPKM) of the group-specific TFs for group 1 and group 2 GSCs along with NSCs.

important area of further investigation, particularly in the setting of prospective therapies targeting GSC group identity.

Our data also suggest that GSCs may adopt TF circuitry similar to distinct signatures of neural stem and progenitor cells. These findings may represent maintenance of distinct cells of origin or convergent evolution driven by the collective effects of genetic and epigenetic events. Genetically engineered mouse models of GBM have supported divergent cells of origin (e.g., NSCs, oligodendroglial progenitor cells, or astrocytes; [Chen et al., 2012b](#)), which may reflect divergent precursor cell states adopted by GBMs. Further expansion of normal to tumor comparisons (at the chromatin level) in a large series of adult NSC, oligodendroglial progenitor cell, and immature/mature astrocyte cell populations (cultured under the same growth conditions) would be potentially informative of transcriptional programs that mediate tumor initiation and potential cells of origin.

Targeted therapies have transformed the therapeutic landscape of several cancers, but many cancer types fail to respond to molecularly guided precision medicine approaches that are based solely on tumor genetics or transcriptional signatures. Prominently, GBM ranks among the most deeply characterized cancers, yet current therapies focus on generalized treatment paradigms. Here, we demonstrated that the epigenetic landscapes of GSCs and extension to GBM tumors yields novel insights into cellular dependencies and core transcriptional regulation. While TFs represent essential factors in cellular transformation and tumor maintenance, they have been notoriously challenging to target. We find that core TF circuits are associated with the presence of SEs that identify essential genes that are amenable to therapeutic targeting and may provide the basis of developing novel therapeutic paradigms. These findings are in keeping with growing literature that supports that SEs and associated TF circuitries represent important nonmutated cancer dependencies ([Durbin et al., 2018](#); [Mack et al., 2018](#); [Ott et al., 2018](#)). Such vulnerabilities may be identified from SEs shared across GSCs and those that occur in a GSC subtype-specific fashion. It is important to note that, although our approach points to a novel set of target genes, single-agent approaches are unlikely to be effective against this highly genetically and phenotypically heterogeneous disease. As a key example, EGFR inhibition used in EGFR SE-containing GSC tumor models, although effective *in vitro*, showed limited efficacy as a single agent *in vivo*. Combinatorial approaches against multiple SE genes in addition to mutated drivers may represent a potential approach that may prove efficacious. Potential translation of targets (and combinations of targets) will require comprehensive *in vitro* and *in vivo* functional validation to prioritize and screen candidates for further drug testing. To this end, we recently developed an *in vivo* screening method to evaluate a prioritized set of candidate genes using inducible

loss-of-function technology in mouse orthotopic GBM xenografts ([Miller et al., 2017](#)). This approach could be used to systematically evaluate the targets we identified in this study with expansion to multiple GSC group-specific tumor models. In sum, our data support that targeting of transcriptional identity has the potential for a large therapeutic index as we find elements of core TF circuits and their associated pathways to be not only highly GBM specific but also related to the presence of distinct GSC populations defined by specific chromatin signatures.

## Materials and methods

### Derivation of GSC models and maintenance of xenografts

GBM tissues were acquired from excess surgical resection tissues from patients at the Case Western Reserve University. All specimens were reviewed by a neuropathologist, and appropriate informed consent was obtained from patients in accordance with a Cleveland Clinic Institutional Review Board-approved protocol (090401). All studies involving human patients were conducted in accordance with the Declaration of Helsinki. To limit cell culture-based artifacts, patient-derived xenografts were generated and dissociated as needed to serve as a renewable source of tumor cells for study. Short tandem repeat analyses were performed to authenticate the identity of each tumor model used in this article every 6 mo. Cells were stored at  $-160^{\circ}\text{C}$  when not being actively cultured. All cells were thawed within 1 mo of these experimental procedures. All experiments conformed to the relevant regulatory standards.

### In vivo tumorigenesis

50,000 human-derived GSCs were implanted into the right cerebral cortex of NSG mice (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ; The Jackson Laboratory) at a depth of 3.5 mm to establish intracranial xenograft models. All murine experiments were performed under an animal protocol approved by the Cleveland Clinic and University of California, San Diego, Institutional Animal Care and Use Committee. Healthy, wild-type female mice of NSG background, 4–6 wk old, were randomly selected and used in this study for intracranial injection. None of the mice had experienced any treatment or procedures before the experiments described. Mice were housed together in a controlled environment with 14 h of light and 10 h of dark per day. Animal husbandry staff at the Cleveland Clinic and University of California, San Diego, regularly observed all animals, and no more than five mice were housed in each cage. Animals were monitored until neurological signs were observed, at which point they were sacrificed. Hunched posture, gait changes, lethargy, and weight loss served as neurological signs or signs of morbidity, which indicated an endpoint condition. Brains were harvested and fixed in 4% formaldehyde, cryopreserved in 30%

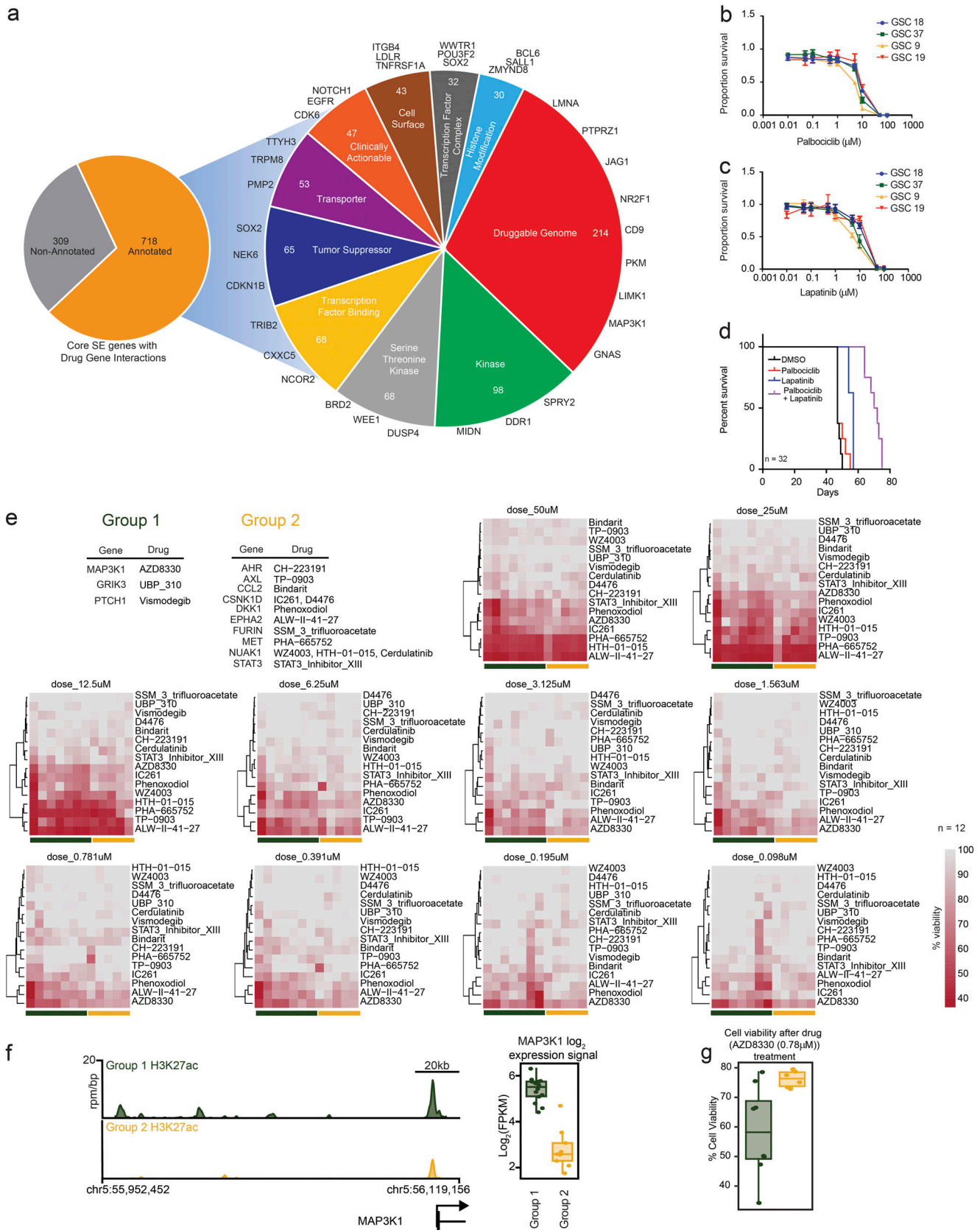


Figure 7. **Core- and group-specific GSC SE genes pinpoint small-molecule inhibitors effective against GBM.** (a) Integration of core GSC SE genes with the Washington University Drug Gene Interaction Database (Cotto et al., 2018) annotating 718 targets with small-molecule interactions (left) that could be classified into several drug classification categories (right). (b) In vitro treatment of GSC models harboring SEs regulating CDK6 with the CDK6 inhibitor

palbociclib. **(c)** In vitro treatment of GSC models harboring SEs regulating *EGFR* with the EGFR inhibitor lapatinib. **(d)** In vivo drug treatment of the CW738 orthotopic xenograft GBM model with single or combinatorial agents palbociclib and lapatinib. **(e)** List and heatmap of drugs predicted to be GSC-group selective through integration with the Washington University Drug Gene Interaction Database (Cotto et al., 2018). **(f)** Depiction of the *MAP3K1* locus harboring group 1 specific enhancer activity (left) and group 1-specific gene expression (right, two-sided Wilcoxon rank-sum test,  $P < 0.029$ ). **(g)** Percentage of cell viability of group 1 and group 2 cells treated with the *MAP3K1* inhibitor, AZD8330 at 0.78  $\mu\text{M}$ . Cell proliferation experiments were performed in biological replicates with more than four technical replicates per time point. Error bars indicated as standard deviation of four technical replicates.

sucrose, and then cryosectioned. H&E staining was performed on sections for histological analysis. In parallel survival experiments, mice were observed until the development of neurological signs as described above. Healthy 4–6-wk-old female NSG mice were randomly selected and used in this study for intracranial injection. For inhibitor treatments, mice were treated with control DMSO vehicle, palbociclib (50 mg/kg every day), lapatinib (50 mg/kg every day), or respective combinations until endpoint criteria were reached. Mice were sacrificed when neurological signs were observed, and tissues were collected for further analysis. Brains were harvested and fixed in 4% formaldehyde, cryopreserved in 30% sucrose, and cryosectioned. H&E staining was performed on brain sections for histological analysis.

#### Tumor dissociation and GSC culture

Dissociation of xenografted tumors was performed using a papain dissociation system according to the manufacturer's instructions. Cells were then cultured in Neurobasal medium supplemented with 2% B27, 1% L-glutamine, 1% sodium pyruvate, 1% penicillin/streptomycin, 10 ng/ml basic fibroblast growth factor, and 10 ng/ml epidermal growth factor (EGF) for  $\geq 6$  h to recover expression of surface antigens. Because no single-cell surface marker is uniformly informative for marking GSCs, we used functional criteria to define and validate the presence of GSCs. GSCs were isolated immediately following dissociation or following transient xenograft passage in immunocompromised NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice using prospective sorting followed by assays to confirm stem cell marker expression, sphere formation, and secondary tumor initiation. When experiments were performed, we isolated AC133 marker-positive populations using CD133/1 antibody-conjugated magnetic beads. Neurobasal medium was used to wash specimens, which were then acutely dissociated to remove nontumor tissue and subjected to enzymatic dissociation using the Papain dissociation system (Worthington Biomedical Corp; LK003150). The isolated tumor cells were briefly placed in Neurobasal medium with B27 supplement (Life Technologies; 12587010) to permit recovery following enzymatic dissociation. Cells were labeled with CD133/2(293C)-allophycocyanin antibody kit (Miltenyi Biotec; 130098826), and the CD133<sup>+</sup> cells were sorted and analyzed by flow cytometry. The sorted CD133<sup>+</sup> cells were cultured in NBM-B27 medium containing 20 ng/ml of both EGF (R&D Systems; 236-EG-01M) and recombinant human basic fibroblast growth factor (R&D Systems; 4114-TC-01M) for a short period before treatment and analysis (Bao et al., 2006; Flavahan et al., 2013). Expression of stem cell markers (SOX2 and OLIG2), functional assays of self-renewal (serial neurosphere passage), and tumor propagation by using in vivo limiting dilution assays were used to validate GSC phenotypes.

#### Proliferation and neurosphere formation assay

Cell proliferation experiments were conducted by plating cells of interest at a density of 1,000 cells/well in a 96-well plate with six replicates. CellTiter-Glo (Promega) was used to measure cell proliferation. All data were normalized to day 0 and presented as mean  $\pm$  SD. In vitro limiting dilution assays were used to assess neurosphere formation capacity as previously reported (Flavahan et al., 2013). Briefly, decreasing numbers of cells were plated into each well of a 96-well plate. The presence and number of neurospheres in each well were recorded 7 d after plating. All tumorsphere and proliferation experiments were performed at least six times.

#### Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) of 5–10 mg of flash-frozen primary GBM tumors was performed by using 5  $\mu\text{g}$  H3K27ac antibody per ChIP experiment (Active Motif; 39133). In the case of cells, 5 million cells were used for ChIP studies. ChIPs were performed as described previously (Wang et al., 2018b). Enriched DNA was quantified by using PicoGreen (Invitrogen) and ChIP libraries were amplified and barcoded by using the ThruPLEX DNA-seq library preparation kit (Rubicon Genomics) according to the manufacturer's recommendations. Following library amplification, DNA fragments were agarose gel (1.0%) size selected ( $< 1$  kb), assessed by using a Bioanalyzer (Agilent Technologies), and sequenced at The Centre for Applied Genomics (The Hospital for Sick Children) by using Illumina HiSeq 2500 150-bp single-end reads (GeneWiz).

#### RNA-seq library preparation

Total RNA was extracted from cell pellets or liquid nitrogen-pulverized tissue using the miRNeasy mini kit (Qiagen) according to instructions from the manufacturer. Stranded RNA library preparation was performed with ribosomal RNA depletion according to instructions from the manufacturer (Epicentre) to achieve greater coverage of mRNA and other long noncoding transcripts. Paired-end sequencing was performed on the Illumina HiSeq 2500 with  $2 \times 150$ -bp paired-end read configuration.

#### Quantitative RT-PCR

Trizol reagent (Sigma-Aldrich) was used to isolate total cellular RNA from cell pellets. The qScript cDNA Synthesis Kit (Quanta BioSciences) was used for reverse transcription into cDNA. Quantitative real-time PCR was performed by using Applied Biosystems 7900HT cyler using SYBR-Green PCR Master Mix (Thermo Fisher Scientific). Quantitative PCR (qPCR) primers used in this study were human *ASCL1* forward 5'-CCCAAGCAA GTCAAGCGACA-3' and reverse 5'-AAGCCGCTGAAGTTGAGCC-3'; human *BCAN* forward 5'-TGGAAGGAGACAGCTCAGAGG-3' and

reverse 5'-CGGGACAGGAAAGTCCACTTG-3'; human *CIORF61* forward 5'-CTTTTCCAGCTTTGGGAGTCA-3' and reverse 5'-GTGGCTCTATCGTCCACACG-3'; human *CDK6* forward 5'-GCTGACCAGCAGTACGAATG-3' and reverse 5'-GCACACATCAAA CAACCTGACC-3'; human *CXCL8* forward 5'-TTTTGCCAAGGA GTGCTAAAGA-3' and reverse 5'-AACCTCTGCACCCAGTT TTC-3'; human *CXXC5* forward 5'-CCGAGCGTCGGAACAAG-AG3' and reverse 5'-CCACTGCTGCCAAAAGAAGAG-3'; human *EGFR* forward 5'-AGGCACGAGTAACAAGCTCAC-3' and reverse 5'-ATGAGGACATAACCAGCCACC-3'; human *HMG2* forward 5'-ACCCAGGGGAAGACCCAAA-3' and reverse 5'-CCTCTTGGC CGTTTTTCTCCA-3'; human *MSRB3* forward 5'-CGGTTCAAG TTGGCCTTCATT-3' and reverse 5'-GTGCATCCCATAGGAAAA GTCA-3'; human *RUNX2* forward 5'-TGGTTACTGTCATGGCGG GTA-3' and reverse 5'-TCTCAGATCGTTGAACCTTGCTA-3'; human *SALL3* forward 5'-CCAATGTGTGGTTCGAG-3' and reverse 5'-CCGGTAAGGGTTCATCTGG-3'; human *SEPT9* forward 5'-TTCGGCTACGTGGGGATTG-3' and reverse 5'-CTGCCC GACCACCATGATG-3'; human *SOX2* forward 5'-GCCGAGTGG AAATTTTTGTGC-3' and reverse 5'-GGCAGCGTGTACTTATCC TTCT-3'; and 18S RNA forward 5'-ACCCGTTGAACCCATT-3' and reverse 5'-CCATCCAATCGGTAGTAGCG-3'.

#### Plasmids and lentiviral transduction

Lentiviral clones expressing two nonoverlapping shRNAs directed against human *ASCL1* (TRCN0000235657, TRCN0000244309), human *BCAN* (TRCN0000151872, TRCN0000371366), human *CIORF61* (TRCN0000172451, TRCN0000167021), human *CDK6* (TRCN0000196261, TRCN0000196337), human *CXCL8* (TRCN0000058030, TRCN0000232050), human *CXXC5* (TRCN0000139498, TRCN0000142729), *EGFR* (TRCN0000121329, TRCN0000121068), *HMG2* (TRCN0000342671, TRCN0000021966), *MSRB3* (TRCN0000064761, TRCN0000064759), *RUNX2* (TRCN0000013656, TRCN0000013655), *SALL3* (TRCN0000418831, TRCN0000019755), *SEPT9* (TRCN0000119069, TRCN0000445092), *SOX2* (TRCN0000355694, TRCN0000231642), or a nontargeting control shRNA that has no targets in the human genome were obtained from Sigma-Aldrich. Nonoverlapping shRNAs were selected based on knockdown efficiency and were then used for all following experiments. 293FT cells were used to generate lentiviral particles through cotransfection of the packaging vectors pCMV-dR8.2 dvpr and pCI-VSVG by using a standard calcium phosphate transfection method in Neurobasal complete medium.

#### Whole-exome sequencing variant discovery

BWA-MEM version 0.7.12 (Li and Durbin, 2010) was used to align paired-end exome sequencing reads to the hg19 reference genome, and SAMtools (Li et al., 2009a) was used for sorting. PCR duplicates were removed with PicardTools (<http://broadinstitute.github.io/picard/>). Single-nucleotide variants and indels were identified with the Genome Analysis Toolkit (McKenna et al., 2010) version 3.8 in accordance with the Genome Analysis Toolkit best practices with the principal steps of base quality score recalibration, variant genotyping for single-nucleotide variants and indels, and variant hard-filtering with standard recommendations (DePristo et al., 2011; Van der Auwera et al., 2013). Variants with predicted moderate or

high impact on gene function (primarily nonsynonymous variants or those that directly affect splice donors or acceptors) were annotated with SNPeff (Cingolani et al., 2012). Variants were further annotated with allelic frequencies from the Exome Aggregation Consortium release 1 (Lek et al., 2016) or the 1000 Genomes Project (Auton et al., 2015). Potential pathogenic variants were identified in a group of commonly mutated PI3K pathway genes (John Lin et al., 2017) and GBM genes (Brennan et al., 2013) with an allelic frequency of <1/10,000 in both Exome Aggregation Consortium and the 1000 Genomes Project. CopywriteR (Kuilman et al., 2015) was used to identify CNVs. Focal CNVs of commonly altered GBM and PI3K genes were identified with a log<sub>2</sub> (copy number ratio) of >2 or >1 and focal (i.e., <10 megabases and clearly delimited) for amplifications with a threshold of <-2 or <-1 and focal for deletions. A small minority of focal CNVs were also excluded manually based on small size and having limited support in the BAM file.

#### H3K27ac ChIP-seq processing

The H3K27ac ChIP-seq was processed following the guidelines of ENCODE (phase-3) by using the AQUAS pipeline ([https://github.com/kundajelab/chipseq\\_pipeline](https://github.com/kundajelab/chipseq_pipeline)). This ChIP-seq processing pipeline (1) mapped the sequenced reads to the human reference genome hg19/GRCh37 using the Burrows-Wheeler Aligner (Li and Durbin, 2009), (2) filtered low-quality, duplicate, multimapping, unmapped reads along with reads mapping to the mitochondrial genome, and (3) performed peak calling using MACS2 with a P value threshold of 10<sup>-5</sup>. The quality control measures such as mapping statistics, library complexity (PCR bottlenecking coefficients, PBC1, and PBC2), cross-correlation scores (normalized strand cross-correlation coefficient and relative strand cross-correlation coefficient) and fraction of reads in the peaks as defined by ENCODE data standards (<https://www.encodeproject.org/chip-seq/histone/>) were also determined using the pipeline. Additionally, we also determined the fraction of reads mapping within 2-kb of an annotated promoter as a quality control measure. The H3K27ac ChIP-seq experiments with PBC1 >0.8 with total mapped reads >10 million and MACS2 peaks (P < 10<sup>-05</sup>) >5,000 were used for further analysis. Experiments detected as outliers based on acetylation signal and potential impurity (n = 4) were further removed to provide a robust set of experiments of GSCs (n = 30), primary tumors (n = 7), and NSCs (n = 7).

#### RNA-seq processing

The sequenced reads of RNA-seq experiments were aligned to human reference genome hg19/GRCh37 using STAR aligner (Dobin et al., 2013). The expression of the genes was determined as fragments per kilobase million (FPKM) by counting the number of uniquely mapping fragments and normalizing it by the length coding sequence and the library size.

Gene annotation was based on UCSC RefSeq. Genes expressed with >1 FPKM in ≥50% of the samples of each group were considered to be expressed genes.

#### Defining GSCs groups

Unsupervised consensus clustering using the ConsensusClusterPlus package from Bioconductor was performed on the top 10% most

variable regions by acetylation signal ( $n = 14,621$ ). This was done by calculating the median absolute deviation of tags per base per million of all the peaks with an acetylation signal. Unsupervised clustering was based on the Euclidean distance between the experiments with maximum clusters  $k = 6$  using 80% of samples and 80% of features for every  $k$ . The consensus clustering results showed that GSCs optimally fell into four clusters, group 1 ( $n = 15$ ), group 2 ( $n = 9$ ), group 3 ( $n = 2$ ), and group 4 ( $n = 4$ ).

### Defining NSCs groups

NSC samples were clustered with the GSC samples using the consensus clustering approach as described above with maximum clusters  $k = 7$  for on top 10% most variable regions by acetylation signal ( $n = 16,817$ ).

### Mapping SEs using H3K27ac enhancer definitions

H3K27ac SEs and typical enhancers in samples were mapped using the ROSE2 software package (Lovén et al., 2013; Whyte et al., 2013; and available at <https://github.com/BradnerLab/pipeline>). Genes were assigned to SEs by proximity.

### Identifying group-specific SEs

The SEs that were present in at least two samples of each of the two GSC groups were compared for differential occupancy ( $n = 2,445$ ). SEs with significantly differential H3K27ac occupancy (false discovery rate [FDR]-adjusted  $P < 0.01$ ) were identified by modeling H3K27ac read counts as negative binomial distribution using a generalized linear model (Anders and Huber, 2010). Bioconductor package DESeq2 was used for this purpose.

### Identifying differentially expressed genes between the GSC groups

Differentially expressed genes (FDR-adjusted  $P < 0.1$ ) between the two GSC groups were identified using the generalized linear model as described above from the RNA-seq gene counts. Genes expressed in either of the GSC groups were used for this.

### Occupancy heatmap of SEs

The occupancy heatmap of significantly differential SEs was generated using the Bioconductor package genomation. The heatmap is centered at the center of the differential SEs with total width as the 75th quartile width of all the differential SEs. The H3K27ac density in the heatmaps is shown in reads per million (rpm).

### Identifying group-specific TFs

SE-associated genes have been shown to be critical for establishing cell identity. Thus, we wanted to identify group-specific SE-associated TFs that are key regulators of SE-associated target genes. We predicted TF binding to SEs by motif and used proximity to assign SEs to genes. The H3K27ac valleys were used as the nucleosome-free regions where the TFs could potentially bind. TF binding sites were obtained from TRANSFAC (Matys et al., 2006) and detected at nucleosome-free regions by using FIMO (Grant et al., 2011). For each group of GSCs, we identified the SE-associated target genes and TFs. We used ridge regression

to predict the target gene activity from the binding TF activity and then used a permutation test to assign significance to ridge regression coefficients. The regularizer in the ridge regression was chosen by 10-fold cross-validation. The target gene activity was randomly permuted followed by model fitting by ridge regression. We performed 10,000 permutations to obtain a null distribution of the regression coefficients. This null distribution was used to assign the significance to the true model coefficients. TFs with  $P < 0.05$  were considered the potential key regulators of the target genes. TFs with significant correlation to  $>2\%$  of target genes in both the groups were considered as consensus TFs ( $n = 24$ ) while the ones that were significantly correlated with  $>1\%$  target genes in the specific groups only were the potential group-specific TFs (group 1 TFs = 14 and group 2 TFs = 18).

### High-confidence group-specific target genes

Genes with group-specific SEs with predicted binding sites for TFs along with significantly different gene expression between the two groups were categorized as group-specific target genes (common target genes = 111, group 1 target genes = 125, and group 2 target genes = 171).

### Gene ontology enrichment analysis

Functional annotation enrichment done for group-specific TFs, common TFs, and target genes was performed by using DAVID with the expressed genes as the background.

### Primary GBM classification

A gene-signature set was created for group 1 by combining the group 1-specific TFs and group 1 target genes. Similarly, the group 2 gene signature was defined by combining group 2-specific TFs and group 2 target genes.  $k$ -means ( $k = 2$ ) was used to classify the GSC samples and primary GBMs into two groups by using the quantile normalized  $\log_2(\text{rpm}/\text{bp})$  acetylation signal at the SEs of these genes with 10,000 iterations with 100 different starting points. This created two clusters, cluster 1 corresponding to group 1 GSCs (GSCs = 15, primary GBMs = 6) and cluster 2 corresponding to group 2 GSCs (GSCs = 9, primary GBM = 1). Primary GBMs that clustered with group 1 GSCs were assigned to be group 1-like GSCs, and the one that clustered with group 2 GSC was assigned to be group 2-like GSC. This methodology was repeated on the gene expression ( $\log_2$  RPKM) of group 1 and group 2 gene signatures yielding consistent assignments of primary GBMs to be like group 1 GSC and group 2 GSC. This created two clusters, cluster 1 corresponding to group 1 GSCs (GSCs = 14/15 and primary GBMs = 38) and cluster 2 corresponding to group 2 GSCs (GSCs = 9/9 and primary GBM = 5).

### Group representation at single-cell level in bulk tumors

Single-cell RNA-seq samples from Gene Expression Omnibus dataset GSE57872 (Patel et al., 2014) were aligned and processed as the other RNA-seq samples as described previously. This study has single-cell RNA-seq from five bulk tumors. Only single-cell samples and genes that had been described to have robust expression in the study were used for further analysis



( $n = 468$ ). We additionally filtered for single-cell samples that had expression for at  $\geq 30\%$  of filtered genes ( $n = 5,516$ ). These overlapped with four group 1-specific TFs, four group 2-specific TFs, 60 group 1 target genes, and 64 group 2 target genes. The gene expression in tags per million of the single-cell samples and RNA-seq of GSCs were  $\log_2$  transformed and quantile normalized together. Spearman rank correlation coefficient for the group-specific target genes and TFs was determined for every single-cell sample of each of the five tumor types with all the GSC samples in group 1 or group 2. To assess the robustness of the correlation, the gene expression values of the GSC sample were randomized 5,000 times, and the correlation coefficient was calculated. Only if the absolute correlation coefficient was  $>99\%$  of the random absolute correlation coefficients then the single-cell sample was considered to have a correlation with the GSC sample. For a group level similarity score, the median correlation coefficient for every single-cell sample with all the samples in the GSC group was determined. This was done only for cells that had significant correlation with  $>75\%$  samples of each group.

### Compound drug screening assay

Compounds and DMSO vehicle controls were transferred to black, clear-bottom 1,536-well plates (Greiner; 789092) using an acoustic transfer system (ATS) Gen4+ instrument (EDC Biosystems). Cells were dispensed in a volume of 10  $\mu\text{l}$ , and a density of 1,000 cells per well by using a Multidrop Combi liquid handler (Thermo Fisher Scientific; 5840300). Plates were incubated at  $37^\circ\text{C}$  and  $5\% \text{CO}_2$  for 72 h, at which point CellTiter-Glo reagent (Promega; G7572) was added to each well using the Multidrop Combi instrument. Plates were shaken at  $25^\circ\text{C}$  for 30 min and then read for luminescence output by using an EnVision plate reader (Perkin Elmer). Luminescence intensity of compound wells was then expressed as a percentage of the average of the DMSO controls for the corresponding plate for each cell line and plotted against the log of the concentration of compound using GraphPad Prism 7 (GraphPad Software Version 7.04). Data were fit to a  $\log_{10}$  (concentration) versus a normalized response curve, and 50% cytotoxic concentration values were obtained for each compound for individual cell lines.

### Data accession statement and information

All raw data files are deposited in the Gene Expression Omnibus repository under accession nos. [GSE119834](#), [GSE119776](#), [GSE119755](#), and [GSE119774](#).

### Online supplemental material

Fig. S1 shows a summary of study DNA methylation, H3K27ac, DNA exome, and copy number data. Fig. S2 shows validation of core GSC SE genes. Fig. S3 shows functional validation of group-specific GSC identity genes. Fig. S4 shows gene expression and pathway analysis of group-specific GSC TFs. Fig. S5 shows examples of group-specific GSC TF enhancer and gene expression profiles.

### Acknowledgments

We thank Mary McGraw for her excellent assistance in sample curation and processing.

This work was supported by National Institutes of Health grants CA217066 (to B.C. Prager), CA217065 (to R.C. Gimple), CA203101 (to L.J.Y. Kim), and CA197718, CA154130, CA169117, CA171652, NS087913, NS089272, and NS103434 (to J.N. Rich). C.Y. Lin is supported by the Cancer Prevention and Research Institute of Texas (RR150093), the National Institutes of Health, and the National Cancer Institute (1R01CA215452-01), and is a Pew-Stewart Scholar for Cancer Research (Alexander and Margaret Stewart Trust). S.C. Mack is supported by Cancer Prevention and Research Institute of Texas scholar award (RR170023), Alex's Lemonade Stand Foundation for Childhood Cancer young investigator award, Rally Foundation research grant, BEAR Necessities Pediatric Cancer Foundation Grant, Children's Cancer Research Fund award, Children's Brain Tumor Foundation Award, and Baylor College of Medicine Junior Faculty Award.

C.Y. Lin is a shareholder and inventor of intellectual property licensed to Syros Pharmaceuticals. The other authors declare no competing financial interests.

Author contributions: S.C. Mack and X. Wang led and performed a majority of the functional studies and were actively involved in study design, data analysis, interpretation, and manuscript preparation. I. Singh led the majority of the bioinformatics analysis and also was actively involved in study design, data analysis, interpretation, and manuscript preparation. R. Hirsch, R.C. Gimple, L.J.Y. Kim, A. Morton, B.C. Prager, K.C. Bertrand, C. Mah, and L. Chavez, contributed to data collection, analysis, and study design. Q. Wu, S. Lai, R. Villagomez, C. Lee, and W. Zhou, assisted with in vitro and in vivo functional studies. J.A. Bernatchez, Z. Zhu, Z. Qiu, and J.L. Siqueira-Neto led the drug validation studies, analysis, and data interpretation. G.H. Barnett, M.A. Vogelbaum, A.E. Sloan, and S. Bao assisted with the collection of patient samples, study design, and data interpretation. S.C. Mack, P.C. Scacheri, C.Y. Lin, and J.N. Rich contributed to study design, data interpretation, and manuscript preparation.

Submitted: 31 January 2019

Revised: 13 March 2019

Accepted: 15 March 2019

### References

- Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Auton, A., L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, J.O. Korbel, J. L. Marchini, S. McCarthy, G.A. McVean, and G.R. Abecasis. 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature.* 526:68–74. <https://doi.org/10.1038/nature15393>
- Bao, S., Q. Wu, R.E. McLendon, Y. Hao, Q. Shi, A.B. Hjelmeland, M.W. Dewhirst, D.D. Bigner, and J.N. Rich. 2006. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature.* 444:756–760. <https://doi.org/10.1038/nature05236>
- Barnes, J.M., S. Kaushik, R.O. Bainer, J.K. Sa, E.C. Woods, F. Kai, L. Przybyla, M. Lee, H.W. Lee, J.C. Tung, et al. 2018. A tension-mediated glycoalyx-integrin feedback loop promotes mesenchymal-like glioblastoma. *Nat. Cell Biol.* 20:1203–1214. <https://doi.org/10.1038/s41556-018-0183-3>
- Bhat, K.P.L., V. Balasubramaniyan, B. Vaillant, R. Ezhilarasan, K. Hummelink, F. Hollingsworth, K. Wani, L. Heathcock, J.D. James, L.D. Goodman, et al. 2013. Mesenchymal differentiation mediated by NF- $\kappa$ B promotes

- radiation resistance in glioblastoma. *Cancer Cell*. 24:331–346. <https://doi.org/10.1016/j.ccr.2013.08.001>
- Boyer, L.A., T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 122:947–956. <https://doi.org/10.1016/j.cell.2005.08.020>
- Brennan, C.W., R.G. Verhaak, A. McKenna, B. Campos, H. Nounshmehr, S.R. Salama, S. Zheng, D. Chakravarty, J.Z. Sanborn, S.H. Berman, et al.; TCGA Research Network. 2013. The somatic genomic landscape of glioblastoma. *Cell*. 155:462–477. <https://doi.org/10.1016/j.cell.2013.09.034>
- Cancer Genome Atlas Research Network. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 455:1061–1068. <https://doi.org/10.1038/nature07385>
- Capper, D., D.T.W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D.E. Reuss, et al. 2018. DNA methylation-based classification of central nervous system tumours. *Nature*. 555:469–474. <https://doi.org/10.1038/nature26000>
- Carro, M.S., W.K. Lim, M.J. Alvarez, R.J. Bollo, X. Zhao, E.Y. Snyder, E.P. Sulman, S.L. Anne, F. Doetsch, H. Colman, et al. 2010. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*. 463:318–325. <https://doi.org/10.1038/nature08712>
- Chen, J., Y. Li, T.S. Yu, R.M. McKay, D.K. Burns, S.G. Kernie, and L.F. Parada. 2012a. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*. 488:522–526. <https://doi.org/10.1038/nature11287>
- Chen, J., R.M. McKay, and L.F. Parada. 2012b. Malignant glioma: lessons from genomics, mouse models, and stem cells. *Cell*. 149:36–47. <https://doi.org/10.1016/j.cell.2012.03.009>
- Chen, R.Z., U. Pettersson, C. Beard, L. Jackson-Grusby, and R. Jaenisch. 1998. DNA hypomethylation leads to elevated mutation rates. *Nature*. 395:89–93. <https://doi.org/10.1038/25779>
- Cingolani, P., A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6:80–92. <https://doi.org/10.4161/fly.19695>
- Cotto, K.C., A.H. Wagner, Y.Y. Feng, S. Kiwala, A.C. Coffman, G. Spies, A. Wollam, N.C. Spies, O.L. Griffith, and M. Griffith. 2018. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 46(D1):D1068–D1073. <https://doi.org/10.1093/nar/gkx1143>
- DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–498. <https://doi.org/10.1038/ng.806>
- Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Durbin, A.D., M.W. Zimmerman, N.V. Dharria, B.J. Abraham, A.B. Iniguez, N. Weichert-Leahey, S. He, J.M. Krill-Burger, D.E. Root, F. Vazquez, et al. 2018. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.* 50:1240–1246. <https://doi.org/10.1038/s41588-018-0191-z>
- Eyler, C.E., Q. Wu, K. Yan, J.M. MacSwords, D. Chandler-Militello, K.L. Misuraca, J.D. Lathia, M.T. Forrester, J. Lee, J.S. Stamler, et al. 2011. Glioma stem cell proliferation and tumor growth are promoted by nitric oxide synthase-2. *Cell*. 146:53–66. <https://doi.org/10.1016/j.cell.2011.06.006>
- Flavahan, W.A., Q. Wu, M. Hitomi, N. Rahim, Y. Kim, A.E. Sloan, R.J. Weil, I. Nakano, J.N. Sarkaria, B.W. Stringer, et al. 2013. Brain tumor initiating cells adapt to restricted nutrition through preferential glucose uptake. *Nat. Neurosci.* 16:1373–1382. <https://doi.org/10.1038/nn.3510>
- Frattoni, V., V. Trifonov, J.M. Chan, A. Castano, M. Lia, F. Abate, S.T. Keir, A. X. Ji, P. Zoppoli, F. Niola, et al. 2013. The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* 45:1141–1149. <https://doi.org/10.1038/ng.2734>
- Garancher, A., C.Y. Lin, M. Morabito, W. Richer, N. Rocques, M. Larcher, L. Bihannic, K. Smith, C. Miquel, S. Leboucher, et al. 2018. NRL and CRX Define Photoreceptor Identity and Reveal Subgroup-Specific Dependencies in Medulloblastoma. *Cancer Cell*. 33:435–449.e6. <https://doi.org/10.1016/j.ccell.2018.02.006>
- Gilbert, M.R., J.J. Dignam, T.S. Armstrong, J.S. Wefel, D.T. Blumenthal, M.A. Vogelbaum, H. Colman, A. Chakravarti, S. Pugh, M. Won, et al. 2014. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N. Engl. J. Med.* 370:699–708. <https://doi.org/10.1056/NEJMoa1308573>
- Graf, T., and T. Enver. 2009. Forcing cells to change lineages. *Nature*. 462:587–594. <https://doi.org/10.1038/nature08533>
- Grant, C.E., T.L. Bailey, and W.S. Noble. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 27:1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>
- Griveau, A., G. Seano, S.J. Shelton, R. Kupp, A. Jahangiri, K. Obernier, S. Krishnan, O.R. Lindberg, T.J. Yuen, A.C. Tien, et al. 2018. A Glial Signaling Regulate Glioma-Vascular Interactions and Tumor Microenvironment. *Cancer Cell*. 33:874–889.e7. <https://doi.org/10.1016/j.ccell.2018.03.020>
- Hegi, M.E., A.C. Diserens, T. Gorlia, M.F. Hamou, N. de Tribolet, M. Weller, J. M. Kros, J.A. Hainfellner, W. Mason, L. Mariani, et al. 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* 352:997–1003. <https://doi.org/10.1056/NEJMoa043331>
- Hu, B., Q. Wang, Y.A. Wang, S. Hua, C.G. Sauv , D. Ong, Z.D. Lan, Q. Chang, Y. W. Ho, M.M. Monasterio, et al. 2016. Epigenetic Activation of WNT5A Drives Glioblastoma Stem Cell Differentiation and Invasive Growth. *Cell*. 167:1281–1295.e18. <https://doi.org/10.1016/j.cell.2016.10.039>
- Jin, X., L.J.Y. Kim, Q. Wu, L.C. Wallace, B.C. Prager, T. Sanvoranart, R.C. Gimple, X. Wang, S.C. Mack, T.E. Miller, et al. 2017. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat. Med.* 23:1352–1361. <https://doi.org/10.1038/nm.4415>
- John Lin, C.C., K. Yu, A. Hatcher, T.W. Huang, H.K. Lee, J. Carlson, M.C. Weston, F. Chen, Y. Zhang, W. Zhu, et al. 2017. Identification of diverse astrocyte populations and their malignant analogs. *Nat. Neurosci.* 20:396–405. <https://doi.org/10.1038/nn.4493>
- Kim, S.H., R. Ezhilarsan, E. Phillips, D. Gallego-Perez, A. Sparks, D. Taylor, K. Ladner, T. Furuta, H. Sabit, R. Chhipa, et al. 2016. Serine/Threonine Kinase MLK4 Determines Mesenchymal Identity in Glioma Stem Cells in an NF- B-dependent Manner. *Cancer Cell*. 29:201–213. <https://doi.org/10.1016/j.ccell.2016.01.005>
- Kuilman, T., A. Velds, K. Kemper, M. Ranzani, L. Bombardelli, M. Hoogstraat, E. Nevedomskaya, G. Xu, J. de Ruiter, M.P. Lolkema, et al. 2015. CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biol.* 16:49. <https://doi.org/10.1186/s13059-015-0617-1>
- Kundaje, A., W. Meuleman, J. Ernst, M. Bilienky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, et al.; Roadmap Epigenomics Consortium. 2015. Integrative analysis of 111 reference human epigenomes. *Nature*. 518:317–330. <https://doi.org/10.1038/nature14248>
- Lathia, J.D., S.C. Mack, E.E. Mulkearns-Hubert, C.L. Valentim, and J.N. Rich. 2015. Cancer stem cells in glioblastoma. *Genes Dev.* 29:1203–1217. <https://doi.org/10.1101/gad.261982.115>
- Lek, M., K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, et al.; Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285–291. <https://doi.org/10.1038/nature19057>
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Z., S. Bao, Q. Wu, H. Wang, C. Eyler, S. Sathornsumetee, Q. Shi, Y. Cao, J. Lathia, R.E. McLendon, et al. 2009b. Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. *Cancer Cell*. 15:501–513. <https://doi.org/10.1016/j.ccr.2009.03.018>
- Lin, C.Y., S. Erkek, Y. Tong, L. Yin, A.J. Federation, M. Zapatka, P. Haldipur, D. Kawachi, T. Risch, H.J. Warnatz, et al. 2016. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*. 530:57–62. <https://doi.org/10.1038/nature16546>
- Liu, F., G.C. Hon, G.R. Villa, K.M. Turner, S. Ikegami, H. Yang, Z. Ye, B. Li, S. Kuan, A.Y. Lee, et al. 2015. EGFR Mutation Promotes Glioblastoma through Epigenome and Transcription Factor Network Remodeling. *Mol. Cell*. 60:307–318. <https://doi.org/10.1016/j.molcel.2015.09.002>
- Lov n, J., H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, and R.A. Young. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 153:320–334. <https://doi.org/10.1016/j.cell.2013.03.036>
- Lu, Q.R., T. Sun, Z. Zhu, N. Ma, M. Garcia, C.D. Stiles, and D.H. Rowitch. 2002. Common developmental requirement for Olig function indicates

- a motor neuron/oligodendrocyte connection. *Cell*. 109:75–86. [https://doi.org/10.1016/S0092-8674\(02\)00678-5](https://doi.org/10.1016/S0092-8674(02)00678-5)
- Mack, S.C., K.W. Pajtler, L. Chavez, K. Okonechnikov, K.C. Bertrand, X. Wang, S. Erkek, A. Federation, A. Song, C. Lee, et al. 2018. Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature*. 553:101–105. <https://doi.org/10.1038/nature25169>
- Makova, K.D., and R.C. Hardison. 2015. The effects of chromatin organization on variation in mutation rates in the genome. *Nat. Rev. Genet.* 16: 213–223. <https://doi.org/10.1038/nrg3890>
- Mansour, M.R., B.J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A.D. Durbin, J. Etchin, L. Lawton, S.E. Sallan, L.B. Silverman, et al. 2014. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*. 346: 1373–1377. <https://doi.org/10.1126/science.1259037>
- Mao, P., K. Joshi, J. Li, S.H. Kim, P. Li, L. Santana-Santos, S. Luthra, U.R. Chandran, P.V. Benos, L. Smith, 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. <https://doi.org/10.1073/pnas.1221478110>
- Matys, V., O.V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, et al. 2006. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34:D108–D110. <https://doi.org/10.1093/nar/gkj143>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Miller, T.E., B.B. Liau, L.C. Wallace, A.R. Morton, Q. Xie, D. Dixit, D.C. Factor, L.J.Y. Kim, J.J. Morrow, Q. Wu, et al. 2017. Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. *Nature*. 547:355–359. <https://doi.org/10.1038/nature23000>
- Noushmehr, H., D.J. Weisenberger, K. Diefes, H.S. Phillips, K. Pujara, B.P. Berman, F. Pan, C.E. Pelloski, E.P. Sulman, K.P. Bhat, et al.; Cancer Genome Atlas Research Network. 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 17:510–522. <https://doi.org/10.1016/j.ccr.2010.03.017>
- Ostrom, Q.T., H. Gittleman, P.M. de Blank, J.L. Finlay, J.G. Gurney, R. McKean-Cowdin, D.S. Stearns, J.E. Wolff, M. Liu, Y. Wolinsky, et al. 2016. American Brain Tumor Association Adolescent and Young Adult Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008–2012. *Neuro-oncol.* 18(Suppl 1):i1–i50. <https://doi.org/10.1093/neuonc/nov297>
- Ott, C.J., A.J. Federation, L.S. Schwartz, S. Kasar, J.L. Klitgaard, R. Lenci, Q. Li, M. Lawlor, S.M. Fernandes, A. Souza, et al. 2018. Enhancer Architecture and Essential Core Regulatory Circuitry of Chronic Lymphocytic Leukemia. *Cancer Cell*. 34:982–995.e7. <https://doi.org/10.1016/j.ccell.2018.11.001>
- Parsons, D.W., S. Jones, X. Zhang, J.C. Lin, R.J. Leary, P. Angenendt, P. Mankoo, H. Carter, I.M. Siu, G.L. Gallia, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 321: 1807–1812. <https://doi.org/10.1126/science.1164382>
- Patel, A.P., I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, D.P. Cahill, B.V. Nahed, W.T. Curry, R.L. Martuza, et al. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 344:1396–1401. <https://doi.org/10.1126/science.1254257>
- Phillips, H.S., S. Kharbanda, R. Chen, W.F. Forrester, R.H. Soriano, T.D. Wu, A. Misra, J.M. Nigro, H. Colman, L. Soroceanu, 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. <https://doi.org/10.1016/j.ccr.2006.02.019>
- Quail, D.F., and J.A. Joyce. 2017. The Microenvironmental Landscape of Brain Tumors. *Cancer Cell*. 31:326–341. <https://doi.org/10.1016/j.ccell.2017.02.009>
- Quail, D.F., R.L. Bowman, L. Akkari, M.L. Quick, A.J. Schuhmacher, J.T. Huse, E.C. Holland, J.C. Sutton, and J.A. Joyce. 2016. The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. *Science*. 352:aad3018. <https://doi.org/10.1126/science.aad3018>
- Segerman, A., M. Niklasson, C. Haglund, T. Bergström, M. Jarvius, Y. Xie, A. Westermark, D. Sönmez, A. Hermansson, M. Kastemar, et al. 2016. Clonal Variation in Drug and Radiation Response among Glioma-Initiating Cells Is Linked to Proneural-Mesenchymal Transition. *Cell Reports*. 17:2994–3009. <https://doi.org/10.1016/j.celrep.2016.11.056>
- Singh, D.K., R.K. Kollipara, V. Vemireddy, X.L. Yang, Y. Sun, N. Regmi, S. Klingler, K.J. Hatanpaa, J. Raisanen, S.K. Cho, et al. 2017. Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Reports*. 18:961–976. <https://doi.org/10.1016/j.celrep.2016.12.064>
- Singh, S.K., C. Hawkins, I.D. Clarke, J.A. Squire, J. Bayani, T. Hide, R.M. Henkelman, M.D. Cusimano, and P.B. Dirks. 2004. Identification of human brain tumour initiating cells. *Nature*. 432:396–401. <https://doi.org/10.1038/nature03128>
- Stupp, R., M.E. Hegi, W.P. Mason, M.J. van den Bent, M.J. Taphoorn, R.C. Janzer, S.K. Ludwin, A. Allgeier, B. Fisher, K. Belanger, et al.; National Cancer Institute of Canada Clinical Trials Group. 2009. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* 10: 459–466. [https://doi.org/10.1016/S1470-2045\(09\)70025-7](https://doi.org/10.1016/S1470-2045(09)70025-7)
- Stupp, R., S. Taillibert, A. Kanner, W. Read, D. Steinberg, B. Lhermitte, S. Toms, A. Idbaih, M.S. Ahluwalia, K. Fink, et al. 2017. Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. *JAMA*. 318:2306–2316. <https://doi.org/10.1001/jama.2017.18718>
- Suvà, M.L., E. Rheinbay, S.M. Gillespie, A.P. Patel, H. Wakimoto, S.D. Rabkin, N. Riggi, A.S. Chi, D.P. Cahill, B.V. Nahed, et al. 2014. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell*. 157:580–594. <https://doi.org/10.1016/j.cell.2014.02.030>
- Tachon, G., K. Masliantsev, P. Rivet, C. Petropoulos, J. Godet, S. Milin, M. Wager, P.O. Guichet, and L. Karayan-Tapon. 2019. Prognostic significance of MEOX2 in gliomas. *Mod. Pathol.* <https://doi.org/10.1038/s41379-018-0192-6>
- Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 126:663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Van der Auwera, G.A., M.O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. 43:1–33.
- van Groningen, T., J. Koster, L.J. Valentijn, D.A. Zwijnenburg, N. Akogul, N.E. Hasselt, M. Broekmans, F. Haneveld, N.E. Nowakowska, J. Bras, et al. 2017. Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.* 49:1261–1266. <https://doi.org/10.1038/ng.3899>
- Venkatesh, H.S., L.T. Tam, P.J. Woo, J. Lennon, S. Nagaraja, S.M. Gillespie, J. Ni, D.Y. Duveau, P.J. Morris, J.J. Zhao, et al. 2017. Targeting neuronal activity-regulated neuropilin-3 dependency in high-grade glioma. *Nature*. 549:533–537. <https://doi.org/10.1038/nature24014>
- Verhaak, R.G., K.A. Hoadley, E. Purdom, V. Wang, Y. Qi, M.D. Wilkerson, C.R. Miller, L. Ding, T. Golub, J.P. Mesirov, 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. <https://doi.org/10.1016/j.ccr.2009.12.020>
- Wang, Q., B. Hu, X. Hu, H. Kim, M. Squatrito, L. Scarpace, A.C. deCarvalho, S. Lyu, P. Li, Y. Li, et al. 2018a. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*. 33:152. <https://doi.org/10.1016/j.ccell.2017.12.012>
- Wang, X., B.C. Prager, Q. Wu, L.J.Y. Kim, R.C. Gimple, Y. Shi, K. Yang, A.R. Morton, W. Zhou, Z. Zhu, et al. 2018b. Reciprocal Signaling between Glioblastoma Stem Cells and Differentiated Tumor Cells Promotes Malignant Progression. *Cell Stem Cell*. 22:514–528.e5. <https://doi.org/10.1016/j.stem.2018.03.011>
- Weintraub, H., S.J. Tapscott, R.L. Davis, M.J. Thayer, M.A. Adam, A.B. Lassar, and A.D. Miller. 1989. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc. Natl. Acad. Sci. USA*. 86:5434–5438. <https://doi.org/10.1073/pnas.86.14.5434>
- Whyte, W.A., D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 153: 307–319. <https://doi.org/10.1016/j.cell.2013.03.035>
- Yan, H., D.W. Parsons, G. Jin, R. McLendon, B.A. Rasheed, W. Yuan, I. Kos, I. Batinic-Haberle, S. Jones, G.J. Riggins, et al. 2009. IDH1 and IDH2

- mutations in gliomas. *N. Engl. J. Med.* 360:765-773. <https://doi.org/10.1056/NEJMoa0808710>
- Yu, Y., Y. Chen, B. Kim, H. Wang, C. Zhao, X. He, L. Liu, W. Liu, L.M. Wu, M. Mao, et al. 2013. Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell.* 152:248-261. <https://doi.org/10.1016/j.cell.2012.12.006>
- Zeid, R., M.A. Lawlor, E. Poon, J.M. Reyes, M. Fulciniti, M.A. Lopez, T.G. Scott, B. Nabet, M.A. Erb, G.E. Winter, et al. 2018. Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat. Genet.* 50:515-523. <https://doi.org/10.1038/s41588-018-0044-9>
- Zheng, C.L., N.J. Wang, J. Chung, H. Moslehi, J.Z. Sanborn, J.S. Hur, E.A. Collisson, S.S. Vemula, A. Naujokas, K.E. Chiotti, et al. 2014. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Reports.* 9:1228-1234. <https://doi.org/10.1016/j.celrep.2014.10.031>
- Zhou, W., S.Q. Ke, Z. Huang, W. Flavahan, X. Fang, J. Paul, L. Wu, A.E. Sloan, R.E. McLendon, X. Li, et al. 2015. Periostin secreted by glioblastoma stem cells recruits M2 tumour-associated macrophages and promotes malignant growth. *Nat. Cell Biol.* 17:170-182. <https://doi.org/10.1038/ncb3090>
- Zhu, Z., M.A. Khan, M. Weiler, J. Blaes, L. Jestaedt, M. Geibert, P. Zou, J. Gronych, O. Bernhardt, A. Korshunov, et al. 2014. Targeting self-renewal in high-grade brain tumors leads to loss of brain tumor stem cells and prolonged survival. *Cell Stem Cell.* 15:185-198. <https://doi.org/10.1016/j.stem.2014.04.007>