MIT Open Access Articles

AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma

AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma

Ryan D. Chow[*,1,2,3], **Christopher D. Guzman**[*,1,2,4,5,6], **Guangchuan Wang**[*,1,2], **Florian Schmidt**[*,7,8], **Mark W. Youngblood**[**,1,3,9], **Lupeng Ye**[**,1,2], **Youssef Errami**[1,2], **Matthew B. Dong**[1,2,3], **Michael A. Martinez**[1,2], **Sensen Zhang**[1,2], **Paul Renauer**, **Kaya Bilguvar**[1,10], **Murat Gunel**[1,3,9,10], **Phillip A. Sharp**[11,12], **Feng Zhang**[13,14], **Randall J. Platt**[7,8,#], and **Sidi Chen**[1,2,3,4,5,6,15,16,#]

[1]Department of Genetics, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [2]System Biology Institute, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [3]Medical Scientist Training Program, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [4]Biological and Biomedical Sciences Program, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [5]Immunobiology Program, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [6]Department of Immunobiology, Yale University School of Medicine, 333 Cedar Street, SHM I-308, New Haven, CT 06520, USA [7]Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland [8]Department of Chemistry, University of Basel, Petersplatz 1, 4003 Basel, Switzerland [9]Department of Neurosurgery, Yale University School of Medicine, 300 Cedar Street, New Haven, CT 06520-8043 [10]Yale Center for Genome Analysis, Yale University School of Medicine, 300 Cedar Street, New Haven, CT 06520-8043 [11]Koch Institute for Integrative Cancer Research, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA [12]Department of Biology, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA [13]Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, MA 02142, USA [14]Department of Biological Engineering, MIT, 75 Ames Street, Cambridge, MA 02142, USA [15]Comprehensive Cancer Center, Yale University School of Medicine, Integrated Science & Technology Center, 850 West Campus Drive, Room 361, West Haven, CT 06516, USA [16]Stem

[#]Correspondence: RJP (rplatt@ethz.ch), SC (sidi.chen@yale.edu), +1-203-737-3825 (office), +1-203-737-4952 (lab).
[*]Co-first authors
[**]Co-second authors

Cell Center, Yale University School of Medicine, Integrated Science & Technology Center, 850 West Campus Drive, Room 361, West Haven, CT 06516, USA

Abstract

A causative understanding of genetic factors that regulate glioblastoma (GBM) pathogenesis is of central importance. Here, we developed an adeno-associated virus (AAV)-mediated autochthonous CRISPR screen in GBM. Stereotaxic delivery of an AAV library targeting genes commonly mutated in human cancers into the brains of conditional Cas9 mice resulted in tumors that recapitulate human GBM. Capture sequencing revealed diverse mutational profiles across tumors. The mutation frequencies in mice correlate with those in two independent patient cohorts. Co-mutation analysis identified co-occurring driver combinations such as *Mll2, B2m-Nf1, Mll3-Nf1* and *Zc3h13-Rb1*, which were subsequently validated using AAV minipools. Distinct from *Nf1*-mutant tumors, *Rb1*-mutant tumors are undifferentiated and aberrantly express *Homeobox* gene clusters. The addition of *Zc3h13* or *Pten* mutations altered the gene expression profiles of *Rb1* mutants, rendering them more resistant to temozolomide. Our study provides a functional landscape of gliomagenesis suppressors *in vivo*.

Introduction

Glioblastoma (Glioblastoma multiforme, GBM) is one of the deadliest cancers[1]. Current standard of care fails to cure the vast majority of patients with this disease[1–3], leaving them a median survival of 12.2 to 18.2 months[4–7]. The first genome atlas of GBM uncovered 453 validated non-silent somatic mutations in 223 unique genes, which were further refined to a total of 71 significantly mutated genes (SMGs)[8]. Subsequent integrative genomic analyses revealed comprehensive mutational landscapes in GBM, uncovering 21 to 75 SMGs across multiple different cohorts of patients[8–13]. Many of the newly discovered genes have never been characterized in GBM; thus, their functional roles in gliomagenesis remain largely unknown[9, 10]. Further complicating the interpretation of causality, mutations can occur in novel combinations across individual patients, leading to drastically different pathological features, prognoses, and therapeutic responses[3, 14–16]. Thus, a deeper functional understanding of gliomagenesis and a quantitative measurement of phenotypic effects across various combinations of drivers are both of central importance.

To date, no study has comprehensively and combinatorially investigated which of the mutations identified in human patients can indeed functionally drive GBM from normal cells in the brain[9]. Key barriers include accurate delivery, precise genome manipulation, efficient massively parallel perturbation, and unbiased, high-sensitivity quantitative readout, all of which have to be achieved simultaneously in the native brain microenvironment. We overcame these challenges through an AAV-mediated direct *in vivo* autochthonous CRISPR screen in the brain of fully immunocompetent mice, coupled with capture sequencing to achieve an ultra-deep readout of all functional variants. With these data, we identified multiple new drivers and co-occurring drivers, and subsequently validated a set of such combinations. Transcriptome profiling of these driver combinations revealed distinct

expression signatures between genotypes and in response to temozolomide (TMZ) treatment. Utilizing this approach, we mapped the functional landscape of GBM suppressors in the native microenvironment of the mouse brain.

Results

Stereotaxic injection of an AAV-CRISPR library drives robust gliomagenesis

To directly test the function of putative SMGs in the mouse brain, we set out to develop a direct *in vivo* autochthonous screening strategy, which necessitates pooled mutagenesis of normal cells directly in the native organ and subsequent deconvolution of mutant phenotypes. Because GBM is a disease originating from astrocytes, we generated an AAV-CRISPR vector that encodes Cre recombinase under a *Glial fibrillary acidic protein* (*GFAP*) promoter, resulting in conditional expression of Cas9 and GFP in astrocytes when injected into a conditional Rosa26-LSL-Cas9-GFP mouse (LSL-Cas9 mouse) (Methods) (Figure S1a). The vector also contains an sgRNA targeting *Trp53*, with the initial intent to generate co-mutational *Trp53* knockouts that might exhibit genome instability and thus be sensitized to tumorigenesis[17–21]. Local viral delivery into the brain restricts the number of transducible cells, and cancer genomes generally consist of dozens to hundreds of SMGs[22–25]. With these considerations in mind, we designed an sgRNA library (mTSG library) targeting the mouse homologs of top-ranked pan-cancer SMGs (Methods), plus 7 genes with essential molecular functions that we initially considered as internal controls (Figure 1a) (Table S1). We pool-synthesized all sgRNAs, cloned them into the AAV-CRISPR vector at greater than 100× coverage, and deep-sequenced the library to ensure all sgRNAs were fully covered and represented with a tight lognormal distribution (99% within two orders of magnitude) (Figure 1a, Figure S1b). We generated high-titer AAVs ($> 1 * 10^{12}$ viral particles per mL) from the plasmid that contained the mTSG library (AAV-mTSG), as well as the empty vector (AAV-vector) (Figure 1a). We then stereotaxically injected AAV-mTSG, AAV-vector or PBS into the lateral ventricle (LV, n = 40 mice) or hippocampus (HPF, n = 16 mice) in the brains of LSL-Cas9 mice (Methods). We performed magnetic resonance imaging (MRI) to scan the brains of these mice at four-months post-injection, and found that half (9/18 = 50%) of AAV-mTSG library transduced animals developed brain tumors at this time point, whereas none of the AAV-vector or PBS injected animals had detectable tumors by MRI (Figure 1b) (Figure S1e) (Table S2). Quantification of tumor volumes showed that AAV-mTSG transduced mice had average tumor volumes of 70.2 mm³ (including animals without tumors), or 140.3 mm³ (excluding animals without a tumor) (two-tailed Welch's *t*-test, $t_{17} = 2.62$, $p = 0.018$, mTSG vs. vector or PBS) (Figure 1c) (Table S2). These data suggested that the AAV-mTSG library robustly initiated tumorigenesis in the brains of LSL-Cas9 mice.

We analyzed the overall survival of a cohort of LSL-Cas9 mice injected with AAV-mTSG, AAV-vector or PBS (Table S3). In this screen, injection location did not affect the rate of tumor development as reflected by overall survival (two-sided Mann-Whitney *U* test of HPF vs LV, $p = 0.054$) (Table S3), and thus were considered as one group (AAV-mTSG). For the AAV-mTSG transduced group, the first three animals died 84 days post injection (dpi), 90% of animals did not survive 176 dpi, and all 56 AAV-mTSG transduced animals reached their survival endpoints within 299 days (*i.e.* died, or had a poor body condition score BCS < 2,

and thus were euthanized) (Table S3) (Figure 1d). The median survival time of the AAV-mTSG group was 129 days (95% confidence interval (CI) = 111–159 days) (Figure 1d), consistent with the presence of tumors in half of the mice at 4 months by MRI. In contrast, all 24 AAV-vector and all 5 PBS injected animals survived the duration of the study and maintained good body condition (BCS = 5) (log-rank (LR) test, $p < 2.2 * 10^{-16}$, mTSG vs. vector or PBS; LR test, $p = 1$, vector vs PBS) (Figure 1d). For the vast majority (96.4%, or 54/56) of AAV-mTSG injected mice, macrocephaly was observed at the survival endpoint (Figure S1c), suggesting that they had developed brain tumors. On the contrary, macrocephaly was observed in none of the AAV-vector (0/24) or PBS (0/5) injected mice during the study (two-tailed Fischer's exact test, $p < 1 * 10^5$, mTSG vs. vector or PBS; $p = 1$, vector vs. PBS). These data indicated that the brain tumors induced by the AAV-mTSG viral library were typically lethal.

We observed that AAV-mTSG mice had GFP-positive masses that deformed their brains (100%, or 6/6) (Figure S1d, Figure 2a). AAV-vector mice had diffuse GFP-positive regions in the brain with fully normal morphology, suggesting these were AAV-transduced cells expressing Cas9-GFP induced by Cre expression that had not become tumors (n = 2) (Figure S1d, Figure 2a). PBS injected or uninjected mice had no detectable GFP expression even at long exposure (n = 3) (Figure S1d). Immunohistochemistry (IHC) analysis showed that AAV-mTSG induced tumors stained positive for Cas9 and GFP, consistent with them having arisen from cells with activation of Cas9-GFP expression (Figure 2a). These tumors were also positive for GFAP, an astrocytic marker (Figure 2a), and for Ki67, a proliferation marker (Figure 2a). AAV-vector transduced brains stained positive for Cas9 and GFP in a subset of cells at the injection site (Figure 2a), but these cells were not proliferative (Ki67 negative) and did not have tumor-like pathological features (Figure 2a). PBS-injected mice stained negative for Cas9, GFP and Ki67 (Figure 2a). Endpoint histopathology showed that the vast majority of AAV-mTSG mice developed brain tumors (10/11 = 91%), whereas none of AAV-vector (0/7 = 0%) or PBS (0/3 = 0%) mice had detectable tumors (two-tailed Fischer's exact test: $p = 0.0003$, mTSG vs. vector; $p = 0.011$, mTSG vs. PBS) (Figure 2a–b) (Table S4). The mean endpoint tumor size as measured by area in brain sections for the AAV-mTSG group was 13.9 mm^2, as compared to 0 mm^2 in the two control groups (two-tailed Welch's t-test, $t_{10} = 3.97$, $p = 0.003$, mTSG vs. vector or PBS) (Figure 2b). Brain tumors in AAV-mTSG mice showed pathological features of dense cellular structure with proliferative spindles, nuclear aneuploidy and pleiomorphism, giant cells, regions of necrosis, angiogenesis and hemorrhage (Figure 2c), all of which are hallmark features of human GBM[2]. Clinical features such as deformation of the brain, invasion, loss of neuronal bundles, necrosis and hemorrhage were further corroborated by special staining methods such as Luxol fast blue Cresyl violet (LFB/CV), Wright Giemsa, Masson and Alcian Blue Periodic acid – Schiff (AB/PAS) (Figure S2). We investigated a panel of human GBM clinical samples from Yale Glioma tissue bank, and confirmed the observation of these pathological features (Figure S3a–b). These data suggested that the AAV-mTSG library-induced autochthonous brain tumors recapitulate features of human GBM.

Targeted capture sequencing reveals diverse mutational profiles across tumors

Because AAVs usually do not integrate into the genome, direct sequencing of the targeted regions was needed to determine which mutations were in each tumor. To map the molecular landscape of these brain tumors, we designed a probe set (mTSG-Amplicon probes) covering the target regions of all library sgRNAs (Methods) (Table S5). We used these probes to perform targeted capture sequencing for whole brain and liver samples (as a control organ not being directly transduced) in a cohort of AAV-mTSG, AAV-vector and PBS injected mice (n = 25, 3, and 4 brain samples, respectively) (Table S6). 277/278 (99.6%) of unique sgRNA target regions were captured for all samples from this experiment, the exception being Arid1a-sg5 (due to unavailability of qualified regions in capture probe design). Across all 41 brain and liver samples, the average of mean coverage across all probes was 19,405 ± 180 s.e.m. (Table S7). We analyzed the mutant variants at the predicted cutting sites of the 277 successfully captured unique sgRNAs across all samples (Methods) (Table S8). At the single sgRNA level, for example, at the predicted cutting site of sgRNA-4 in the *Mll2* locus (also known as *Kmt2d*), various insertions and deletions (indels) were detected in AAV-mTSG but not AAV-vector or PBS mice (Figure 3a). As gliomagenesis takes multiple months in mice (Figure 1d), we surveyed 3 mice at 3.5 weeks post-injection and performed capture sequencing to reveal early mutation profiles, as an approximation for *in vivo* sgRNA cutting efficiency (Figure S4a–b, Table S9). We found that even lower-efficiency sgRNAs can end up being highly enriched in the process of tumorigenesis if the mutations they generated are strongly oncogenic (Figure S4c, Figure 3b–c). After removing germline variants, we determined whether the regions flanking each sgRNA target site would be classified as significantly mutated sgRNA sites (SMSs) (Methods) (Table S10). We implemented a false-discovery-rate (FDR) approach with (FDR < 1/12 PBS/vector samples, or 8%) as well as a flat 5% variant frequency cutoff, and confirmed that the choice of alternative cutoffs did not alter the final SMS calls (Figure S6a). With these criteria, we observed a diverse mutational landscape across most mice that were capture-sequenced (Figure 3b–c, Figure S5, Table S11). As an example, one AAV-mTSG mouse (mTSG brain 10) had significant mutations at the predicted cutting sites of 16 out of 277 captured gene-targeting sgRNAs in the mTSG library, covering 12 significantly mutated genes (mSMGs) (Figure 3b). A second example (mTSG brain 24) showed a more diverse mutational profile (34 SMSs for 26 mSMGs) (Figure 3b). The raw indel frequencies were also summed across all detected variants for each sgRNA target site in each sample, revealing a highly diverse pattern of variant frequencies generated by this sgRNA pool (Figure 3c) (Table S10). Comparing brain samples between treatment groups, AAV-mTSG injected brains had significantly higher mean variant frequencies (2.087 ± 0.429 s.e.m., n = 25) compared to vector (0.005 ± 0.001, n = 3) or PBS (0.003 ± 0.001, n = 4) injected brains (two-tailed Welch's *t*-test, $t_{24} = 4.85$ and $t_{24} = 4.86$, $p = 6.03 * 10^{-5}$ and $p = 5.96 * 10^{-5}$ for mTSG vs. vector and mTSG vs. PBS, respectively) (Figure 3c–d). Comparing targeted vs. non-targeted organs in AAV-mTSG injected mice, the mean variant frequencies of brains (2.087 ± 0.429, n = 25) were significantly higher than livers (0.309 ± 0.261, n = 4) (two-tailed Welch's *t*-test, $t_{21.48} = 3.54$, $p = 0.002$) (Figure 3c–d). The predominant indels were deletions for virtually all samples, and most insertions at SMS sites were 1 bp in size (Figure 3e) (Table S8). We identified distinct variant frequency clusters of sgRNA-induced indels that may serve as an approximation to the clonality of these tumors (Methods). From this analysis, we

found that only 2/22 of the brains had single-cluster tumors, with the majority (20/22) being comprised of multiple clusters (Figure S6b). These data demonstrate on-target, pooled genome editing in the brain at a library scale, stochastically generating loss-of-function mutations in native glial cells and priming them for selection during gliomagenesis.

We next summarized the mutational data from the SMS level to the mSMG level (Table S12) and created an oncomap of all mTSG brain samples (Figure 4a). Across all mice with SMGs (23/25), the detected variants were predominantly frameshift indels (frameshift reads / total variant fraction >60% in 22/23 mice) compared to non-frameshift indels, splicing indels and intronic indels (Figure 4A, bottom panel). Surprisingly, all 56 genes have at least one associated SMSs, with eight of them (*Pten, Map2k4, B2m, Pcna, Cic, Setd2, Gata3* and *Apc*) having 5/5 SMSs (Figure 4a left and middle panel). The SMGs encode functionally diverse categories of proteins, including cell death or cell cycle regulators, immunological regulators, DNA repair and replication regulators, transcriptional repressors, epigenetic regulators, transcription factors, cadherin type proteins and ubiquitin ligases (Figure 4a). Many of the genes were significantly mutated in 20% to 50% of mice, with most of the epigenetic regulators in this range, such as *Arid1b, Mll3, Setd2, Mll2, Kdm5c, Kdm6a, Arid2* and *Ctcf* (Figure 4a), highlighting the role of epigenetic regulators in brain tumorigenesis. Interestingly, *B2m*, a core component of major histocompatibility complex (MHC) class I that is essential for antigen presentation, appeared as the second most frequently mutated gene (19/25 = 76% of mice) (Figure 4a, right panel). This analysis revealed, in a quantitative manner, the relative phenotypic strength of specific loss-of-function mutations in driving gliomagenesis *in vivo*.

We compared the mutational frequencies in mice to the variant frequencies of their homologous genes in human GBM with their frequencies of non-silent mutation and deletion. For these 56 genes, the mutation frequencies in mouse GBMs (an end-product of pooled mutagenesis and *in vivo* gliomagenesis) significantly correlated with the mutation frequencies in TCGA GBM patients (Pearson correlation R = 0.402, $p = 2.1 * 10^{-3}$) (Figure 4b) (Table S13). To further investigate this correlation, we utilized the clinical cancer genomics data from the Yale Glioma tissue bank, a source independent from TCGA (Tables S14, S15). Collectively, the mouse mutation frequencies again significantly correlated with those in human patients (R = 0.318, $p = 0.0277$) (Figure 4c) (Table S16). These data suggest that the AAV-CRISPR autochthonous GBM mouse model revealed a quantitative phenotypic profile of tumor suppressors reflecting the genomic landscape of human GBM patients.

Co-mutation analysis identifies frequently co-occurring driver combinations

To generate an unbiased map of co-drivers, we calculated the co-occurrence rate of double-mutations for each gene pair (Methods) (Figure 5a–b, Table S17). This analysis showed that 76 gene pairs out of a total of 1540 possible pairs were statistically significant in terms of co-occurrence (hypergeometric test, FDR adjusted $q < 0.05$). The *Nf1+Pten* pair emerged as the top pair (co-occurrence rate = 18/21 = 85.7%, hypergeometric test, $p = 7.53 * 10^{-8}$) (Figure 5a–c). Interestingly, several previously undocumented combinations emerged, such as *Kdm5c+Gata3* (co-occurrence rate = 77.8%, hypergeometric test, $p = 6.04 * 10^{-6}$), and *B2m+Pik3r1* (70.0%, $p = 2.28 * 10^{-5}$) (Figure 5a–c). In addition, we performed correlation

analysis of summed mutant frequencies for each pair of genes across all mice (Methods) (Figure 5d–e, Table S18). 22.9% (352/1540) of the gene pairs were positively correlated (Spearman correlation > 0, FDR adjusted $q < 0.05$) (Figure 5d–e). The most significantly correlated gene pair was again *Nf1+Pten* (Spearman correlation = 0.861, $p = 3.34 * 10^{-8}$) (Figure 5d–f, Table S18), along with other representative pairs such as *Cdkn2a+Ctcf* (correlation = 0.792, $p = 2.41 * 10^{-6}$) (Figure 5g), *B2m+Notch1* (correlation = 0.789, $p = 2.82 * 10^{-6}$), and *Apc+Pik3r1* (correlation = 0.774, $p = 5.77 * 10^{-6}$) (Figure 5d–e). Exclusion of *Trp53* revealed largely identical results for the remaining genes (Figure S7a-b). Of note, a subset of the significantly co-occurring pairs were also found to be co-mutated in human GBM, including the pairs of *RB1+TP53, PTEN+RB1, RASA1+STK11, B2M +MAP2K4, PTEN+STAG2, CDKN1B+TP53* and *CDKN1B+NF1* (Figure S7C–D). These data revealed co-occurrence and correlation relationships of specific mutations during glioblastoma progression *in vivo*.

Minipool validation of individual drivers and combinations

We went on to test several of the highly represented individual drivers or combinations using an sgRNA minipool validation approach (Methods) (Figure 6a). All of the un-injected (n = 2), EYFP (n = 4) and empty vector (n = 3) injected mice survived and maintained good body condition for the whole duration of the study. Control mice sacrificed from 4 to 11 months after treatment were devoid of any observable tumors by histology analysis (Figure 6b, Figure S8, Table S19), again indicating that without mutagenesis, or with *Trp53* disruption alone, LSL-Cas9 mice did not develop brain tumors. In contrast, within 11 months post injection, 50% (4/8) of mice receiving AAVs containing *Nf1* sgRNA minipool developed macrocephaly, poor body condition score and tumors (compared to all 9 control mice, two-tailed Fisher's exact test, $p = 0.029$). All mice receiving *Nf1;Pten* (9/9, 100%, $p = 4.11 * 10^{-5}$) and *Nf1;B2m* minipools (4/4, 100%, $p = 0.0014$) developed macrocephaly, poor body condition score and large tumors (Figure 6c, Table S19). Notably, mice receiving *Nf1;B2m* minipools had significantly worse survival than mice receiving *Nf1* minipools alone ($p = 0.0067$) (Figure S9a), implying that loss of antigen presentation in cancer cells likely accelerates GBM progression in immunocompetent mice. All mice receiving *Rb1, Rb1;Pten*, or *Rb1;Zc3h13* minipools (3/3, 100%, $p = 0.0045$ for all three groups) developed macrocephaly, poor body condition score and large tumors (Figure 6D, Table S19). For the same duration of study (maximum 11 months), smaller fractions of mice receiving the AAV sgRNA minipools targeting *Arid1b;Nf1* (4/9), *Mll3;Nf1* (2/5), *Mll2* (2/10), *Cic* (1/5), *Cic;Pten* (1/4), *Setd2* (1/5), and *Gata3;Mll3* (1/5) developed tumors (Table S19). Collectively, half (40/80, or 50%) of the mice receiving AAV sgRNA minipools targeting any of the single genes or gene pairs developed brain tumors within 11 months (collective validation vs. all controls, two-tailed Fisher's exact test, $p = 0.004$). These data indicated that mutating these individual genes or combinations in combination with *Trp53* causes GBM in fully immunocompetent animals.

Interestingly, brain tumors with *Nf1* mutations displayed highly polymorphic pathological features, with diverse fibroblastic cell morphologies, regions of necrosis and large hemorrhages (Figure 6c), yet were almost always GFAP-positive (Figure S8). In sharp contrast, tumors with *Rb1* mutations were composed of round cells with dense nuclei,

frequently with proliferative spindles and giant cells with massive nuclear aneuploidy and pleiomorphism, but rarely with regions of necrosis or large-area hemorrhage (Figure 6d), and often contained mixtures of GFAP-positive and GFAP-negative cells (Figure S8).

Transcriptomic characterization of tumors with differing mutational backgrounds

We then investigated the molecular underpinnings of gliomagenesis driven by different combinations of drivers (Figure 7a, Figure S9b). We performed mRNA-seq to profile the transcriptome of these mutant glioma cells (*Nf1, Nf1;Mll3, Rb1*, and *Rb1;Zc3h13*, n = 3 cell replicates each) (Tables S20, S21). Comparing *Nf1*-mutant and *Rb1*-mutant cells, we found that 616 genes were more highly expressed in *Rb1* cells (Benjamini-Hochberg adjusted $p <$ 0.05 and log fold change 1), while 982 genes were more highly expressed in *Nf1* cells (Figure 7b, Table S22). Gene ontology analysis of the genes associated with higher expression in *Nf1*-mutant cells revealed multiple enriched categories (Benjamini-Hochberg adjusted $p < 0.05$), including extracellular region part, biological adhesion, neuron differentiation, hormone metabolic process, cell motion, and cell-cell signaling (Figure 7c). Gene ontology analysis of the genes associated with higher expression in *Rb1*-mutant cells revealed a distinct set of enriched categories (adjusted $p < 0.05$), which surprisingly included regionalization, anterior/posterior pattern formation, transcription factor activity, embryonic morphogenesis, cell adhesion, extracellular matrix, neuron differentiation, and GTPase regulator activity (Figure 7d). Strikingly, a total of 13 *Homeobox* genes were among the top-40 upregulated genes in *Rb1* mutants.

To understand the direct effect of additional mutations on the transcriptome of these cells, we next compared *Nf1;Mll3* to *Nf1* cells, and *Rb1;Zc3h13* to *Rb1* cells. We found that 522 genes were upregulated in *Nf1;Mll3* compared to *Nf1* cells, while 175 were downregulated (Figure 7e, Table S23). Gene ontology analysis of the upregulated genes in *Nf1;Mll3* cells revealed enrichment of extracellular matrix, EGF-like region, biological adhesion, calcium ion binding, tube development, and growth factor binding (Figure 7f). Comparing *Rb1;Zc3h13* to *Rb1* cells revealed 703 upregulated and 166 downregulated genes (Figure 7g, Table S24). *Rb1;Zc3h13*-high genes were enriched in categories such as extracellular matrix, immune response, cell adhesion, 2–5-oligoadenylate synthetase, cell morphogenesis, GTPase activity, cell motion, and vasculature development (Figure 7h). Collectively, these findings indicate that the addition of an *Mll3* mutation significantly alters the transcriptome of *Nf1* mutant cells, as does the addition of a *Zc3h13* mutation on *Rb1* mutant cells.

Secondary mutations influence the transcriptome and engender chemotherapeutic resistance

As GBM remains a very challenging cancer type to treat, understanding the molecular changes underlying drug response is of profound importance. We thus performed drug-treatment-RNA-seq experiments to investigate the transcriptome responses of AAV-CRISPR-induced GBM cells (*Rb1, Rb1;Pten* and *Rb1;Zc3h13*) to TMZ, a chemotherapeutic with significant albeit small survival benefit for GBM patients, among the only 4 currently approved drugs for this disease (Methods) (Figure 8a). Drug response phenotyping showed that *Zc3h13* loss-of-function rendered *Rb1* cells significantly more resistant to 1mM TMZ, similar to *Pten* loss-of-function (two-tailed *t*-test, $t_4 = 31.32$ and $t_4 = 23.51$, $p = 6.20 * 10^{-6}$

and $p = 1.94 * 10^{-5}$, for *Rb1;Pten* vs. *Rb1* and *Rb1;Zc3h13* vs. *Rb1*, respectively) (Figure 8b). These differences were also observed with 2mM TMZ ($t_4 = 50.69$ and $t_4 = 38.10$, $p = 9.06 * 10^{-7}$ and $p = 2.84 * 10^{-6}$, for *Rb1;Pten* vs. *Rb1* and *Rb1;Zc3h13* vs. *Rb1*, respectively) (Figure 8c). Given the differential responses among these three genotypes, we performed mRNA-seq to profile the transcriptome of these mutant cells treated with TMZ as compared to DMSO controls (Tables S20, S21). Differential expression analyses of TMZ and DMSO treated cells from each of the three genotypes revealed systematic changes in gene expression (Figure 8d–f, Table S25–27). Collectively, the differentially expressed genes in the TMZ vs. DMSO comparisons uncovered a molecular map of the transcriptomic differences between genotypes in response to TMZ treatment (Figure 8g).

Of the genes that were significantly reduced upon TMZ treatment in each group, a total of 69 genes were shared among all three genotypes (Figure 8h, Table S28), indicating that these genes are a common transcriptional response to TMZ. As *Rb1;Zc3h13* and *Rb1;Pten* cells exhibited greater survival fractions with TMZ treatment when compared to *Rb1* cells, we identified 37 genes that were significantly reduced upon TMZ treatment in *Rb1;Zc3h13* and *Rb1;Pten* cells, but not in *Rb1* cells (Figure 8h, Table S28). As for the genes that were significantly induced by TMZ, a total of 42 genes were common in all three genotypes (Figure 8i, Table S29), representing a shared TMZ-induced gene signature Interestingly, we identified 60 genes that were upregulated upon TMZ treatment in *Rb1;Zc3h13* and *Rb1;Pten* cells, but not in *Rb1* cells. These included *Arl6ip1*, which encodes a protein that has been shown to suppress cisplatin-induced apoptosis in cancer cells[26], and *Cd274* (also known as *PD-L1*), which encodes the ligand for the inhibitory receptor PD-1 that is currently a major focus of investigation in cancer immunotherapy[27]. Taken together, the transcriptomic analyses provide unbiased molecular signatures underlying the increased TMZ-resistance upon *Zc3h13* or *Pten* mutations in *Rb1*-oncotype glioma cells, suggesting that the precise combinations of mutational drivers present in individual GBMs directly influence therapeutic responses.

Discussion

To date, there has been no unbiased phenotypic picture of which genetic factors and combinations are necessary or sufficient to drive tumorigenesis in the brain. In order to answer these questions, it is critical to directly test the hypotheses gleaned from cancer genomics in a controlled experimental setting to find causative genes and to quantitatively measure their phenotypes *in vivo*. Determining whether these alterations are *bona fide* driver mutations requires direct *in vivo* testing, and has historically been performed using cell lines or genetically engineered mouse models[28]. Development and applications of pioneering mouse models of GBM have led to profound progress in our understanding of tumor initiation, stem cell populations, progression and therapeutic responses of GBM[29–40] driven by a variety of important mutations. The CRISPR system[41–43] has been utilized to directly mutate oncogenes and tumor suppressor genes in somatic cells for modeling genetic events in various cancer types (reviewed in[44]). We now demonstrate high-throughput gene editing for multiplexed autochthonous GBM models in fully immunocompetent mice. Our study provides a massively parallel view of tumor suppressors *in vivo*, revealing the relative selective strength of mutations in these genes when competing in the brain, as well as the

driver variations between individuals. As CRISPR targeting might lead to off-target effects at other loci, we performed exome sequencing for a subset of the mTSG brain samples. This dataset revealed unbiased, exome-wide measurement of other mutations (Figure S7e, Tables S30–31). These mutations might be caused by off-target effects of CRISPR, rare AAV integrations[45], or by the unstable genomes of GBM cells, especially those with *p53* loss[20], throughout tumorigenesis.

Across all genes tested, the mutation frequencies in this highly complex mouse model of GBM significantly correlate with the mutation frequencies in human patients from two large independent cohorts (TCGA and Yale Glioma bank), suggesting the clinical relevance of the findings. Several of the novel SMGs highly enriched in this mouse study have also been associated with GBM in the clinical setting, such as *B2M, CIC, MLL2, MLL3, SETD2, ZC3H13* and *ARID1B*[11, 46–49]. Because differences in driver mutations can dramatically affect treatment efficacies in pre-clinical animal models and in patients, a functional understanding of cancer drivers is therefore essential for precision medicine[50].

We also performed a lentiCRISPR direct *in vivo* screen in GBM, using the same mTSG library (Methods, Table S32, Figure S9c–d). We found that the AAV-mTSG CRISPR library resulted in more robust gliomagenesis *in vivo* compared to lentiCRISPR mTSG, in terms of latency (death of first animal, 84 vs. 200 days), survival (median survival, 4 vs. 10 months), and penetrance (100% vs. 67%). However, AAVs usually do not integrate into the genome, except under certain circumstances (low rate of integration at *AAVS1* locus)[45]. Thus, AAV-encoded transgenes such as exogenously supplied sgRNAs do not replicate as cells divide during tumor progression, limiting the readout of the mTSG library by PCR amplification of the sgRNA cassette itself. Instead, we achieved successful readout of driver mutations by sequencing the predicted sgRNA cutting sites using ultra-deep targeted captured sequencing. A key advantage of this approach is the ability to perform high-throughput mutagenesis in an autochthonous model of GBM, in which tumors directly evolve from normal cells at the organ site *in situ* in immunocompetent mice, without cellular transplantation. This platform can be readily extended to study other types of cancer for tumor progression, as well as therapeutic responses *in vivo*. Taken together, our study provides a systematic and unbiased molecular landscape of functional tumor suppressors in an autochthonous mouse model of GBM, opening new paths for high-throughput analysis of cancer genetics directly *in vivo*.

Online Methods

Please see the Life Sciences Reporting Summary for additional information about the methods used in this study.

Design, synthesis and cloning of the mTSG library

Briefly, pan-cancer mutation data from 15 cancer types were retrieved from The Cancer Genome Atlas (TCGA portal) via cBioPortal[51] and Synapse (https://www.synapse.org). Significantly mutated genes were calculated similar to previously described methods[22–25]. Known oncogenes were excluded and only known or predicted tumor suppressor genes (TSGs) were included. The top 50 TSGs were chosen, and their mouse homologs (mTSG) were retrieved from mouse genome informatics (MGI) (http://www.informatics.jax.org). A

total of 49 mTSGs were found. A total of 7 known housekeeping genes were chosen as internal controls. We designed sgRNAs against these 56 genes using a previously described method[52, 53] with our custom scripts. Five sgRNAs were chosen for each gene, plus 8 non-targeting controls (NTCs), making a total 288 sgRNAs in the mTSG library. There were two sets of duplicate sgRNAs, Cdkn2a-sg2 / Cdkn2a-sg5, and Rpl22-sg4 / Rpl22-sg5, leaving a total of 286 unique sgRNAs.

Design, cloning of an AAV-CRISPR GBM vector and mTSG sgRNA library cloning

An AAV-CRISPR vector was designed for astrocyte-specific genome editing. This vector contains a cassette specifically expresses Cre recombinase under the control of a *GFAP* promoter for conditional induction of Cas9 expression in astrocytes in the brain when delivered to LSL-Cas9 mice[54]. Two sgRNA cassettes were built in this vector, one encoding an sgRNA targeting *Trp53*, the most frequently mutated gene in cancer[22, 23, 25], with the other being an empty sgRNA cassette (double SapI sites for sgRNA cloning) enabling flexible targeting of genes of interest in either individual or pooled manner. The vector was generated by gBlock gene fragment synthesis (IDT) followed by Gibson assembly (NEB). The mTSG libraries were generated by oligo synthesis, pooled, and cloned into the double SapI sites of the AAV-CRISPR GBM vector. The library cloning was done at over 100× coverage to ensure proper representation. Plasmid library representation was readout by barcoded Illumina sequencing as described previously[55] with primers customized to this vector.

AAV-mTSG viral library production

The AAV-CRISPR GBM plasmid vector (AAV-vector) and library (AAV-mTSG) were subjected to AAV9 production and chemical purification. Briefly, HEK 293FT cells (ThermoFisher) were transiently transfected with transfer (AAV-vector or AAV-mTSG), serotype (AAV9) and packaging (pDF6) plasmids using polyethyleneimine (PEI). Each replicate consist of five of 80% confluent HEK 293FT cells in 15-cm tissue culture dishes or T-175 flasks (Corning). Multiple replicates were pooled to enhance production yield. Approximately 72 hours post transfection, cells were dislodged and transferred to a conical tube in sterile PBS. 1/10 volume of pure chloroform was added and the mixture was incubated at 37°C and vigorously shaken for 1 hour. NaCl was added to a final concentration of 1 M and the mixture was shaken until dissolved and then pelleted at 20k g at 4°C for 15 minutes. The chloroform layer was discarded while the aqueous layer was transferred to another tube. PEG8000 was added to 10% (w/v) and shaken until dissolved. The mixture was incubated at 4°C for 1 hour and then spun at 20k g at 4° C for 15 minutes. The supernatant was discarded and the pellet was resuspended in DPBS plus MgCl$_2$ and treated with Benzonase (Sigma) and incubated at 37°C for 30 minutes. Chloroform (1:1 volume) was then added, shaken, and spun down at 12k g at 4C for 15 min. The aqueous layer was isolated and passed through a 100 kDa MWCO (Millipore). The concentrated solution was washed with PBS and the filtration process was repeated. Virus was titered by qPCR using custom Taqman assays (ThermoFisher) targeted to Cre.

Design, cloning of lentiCRISPR GBM vectors and mTSG sgRNA library, and lentivirus production

Two lentiCRISPR vectors were designed, one for constitutive, and the other for astrocyte-specific genome editing. These vectors contain a cassette specifically expresses Cre recombinase under the control of an *EFS* promoter or a *GFAP* promoter for conditional induction of Cas9 expression in the brain when deliver to LSL-Cas9 mice. Two sgRNA cassettes were built in this vector, one encoding an sgRNA targeting *Trp53*, with the other being an empty sgRNA cassette (double BsmbI sites for sgRNA cloning) enabling flexible targeting of genes of interest in either individual or pooled manner. These vectors were generated by gBlock gene fragment synthesis (IDT) followed by Gibson assembly (NEB). The mTSG libraries were generated by oligo synthesis, pooled, and cloned into the double BsmbI sites of the lentiCRISPR GBM vectors. The library cloning was done at over 100× coverage to ensure proper representation. Plasmid library representation was readout by barcoded Illumina sequencing as described above, with primers customized to the vectors. The LentiCRISPR GBM plasmid vector (Lenti-vector) and library (Lenti-mTSG) were subjected to high-titre lentivirus production and purification. Briefly, HEK 293FT cells (ThermoFisher) were transiently transfected with transfer (Lenti-vector or Lenti-mTSG), and packaging (psPAX and pMD2.G) plasmids using PEI or Lipofectamine. Each replicate consist of five of 80% confluent HEK 293FT cells in 15-cm tissue culture dishes or T-175 flasks (Corning). Multiple replicates were pooled to enhance production yield. Approximately 48 hours post transfection, virus-containing media was collected, and purified via sucrose gradient ultracentrifugation at >=30,000 rpm for 2–3h. The supernatant was discarded and the pellet was dried and resuspended with 100 µl sterile PBS in 4 °C overnight. Virus was titered by viral protein p24 ELISA (RnD).

Animal work statements

All animal work was performed under the guidelines of Yale University Institutional Animal Care and Use Committee (IACUC) and Massachusetts Institute of Technology Committee for Animal Care (CAC), with approved protocols (Chen-2015-20068, Zhang-0414-024-17, and Sharp-0914-091-17), and were consistent with the Guide for Care and Use of Laboratory Animals, National Research Council, 1996 (institutional animal welfare assurance no. A-3125-01).

Stereotaxic surgery and virus transduction in the mouse brain

Conditional LSL-Cas9 knock-in mice were bred in a mixed 129/C57BL/6 background. Mixed gender (randomized males and females) 6–14 week old mice were used in the experiment. Animals were maintained and breed in standard individualized cages with maximum of 5 mice per cage, with regular room temperature (65–75°F, or 18–23°C), 40–60% humidity, and a 12h:12h light cycle. Mice were anesthetized by intraperitoneally injection of ketamine (100 mg/kg) and xylazine (10 mg/kg), or by inhalation of isoflurane at approximately 2% for 20–30 minutes. We also administered buprenorphine HCl (0.1 mg/kg), or carprofen (5.0 mg/kg) intraperitoneally as a pre-emptive analgesic. Reflexes were tested before surgical procedures. Once subject mice were in deep anesthesia, they were immobilized in a stereotaxic apparatus (Kopf, or Stoelting) using intra-aural positioning

studs and a tooth bar to immobilize the skull. Heat is provided for warmth by a standard heating pad, or a heatlamp. According to the mouse brain stereotaxic coordinates[56], we drilled a 1–2 mm hole on the surface of the skull, and used a 33 G Nanofil syringe needle (World Precision Instrument) to inject into the lateral ventricle (LV) at 0.6–1.0 mm caudal/ posterior to Bregma, 0.8–1.5 mm right-side lateral to Bregma, and 2.0–3.0 mm deep from the pial surface for injection (coordinates: A/P −0.6 to −1.0, M/L 0.8 to 1.5, D/V −2.0 to −3.0). For a small fraction of animals, injections were made into hippocampus (HPF) at the following coordinates (A/P −1.3, M/L 0.6, D/V −1.7). PBS, 8 uL AAV (Between 1×10^{10} – 1×10^{11} viral genome copies, or Cre copy number equivalent), or 8 uL lentivirus (Between 8×10^{9} – 8×10^{10} viral particles, or p24 equivalent) was injected into the right hemisphere of the brain for each mouse. Injection rates were monitored by an UltraMicroPump3 (World Precision Instruments). After injection, the incision site was closed with 6–0 Ethilon sutures (Ethicon by Johnson & Johnson), or a VetBond tissue glue (3M). Animal were postoperatively hydrated with 1 mL lactated Ringer's solution (subcutaneous) and housed on warmed cages or in a temperature controlled (37°C) environment until achieving ambulatory recovery. Meloxicam (1–2 mg/kg) was also administered subcutaneously directly after surgery.

MRI

MRI imaging was performed using standard imaging protocol with MRI machines (Varian 7T/310/ASR-whole mouse MRI system, or Bruker 9.4T horizontal small animal systems). Briefly, animals were anesthetized using isoflurane, and setup in the imaging bed with a nosecone providing constant isoflurane. A total of 20 – 30 views were acquired for each mouse brain using a custom setting: echo time (TE) = 20, repetition time (TR) = 2000, slicing = 0.5 mm. Raw image stacks were processed using Osirix or Slicer tools[57]. Rendering and quantification were performed using Slicer (www.slicer.org). For all mice with brain tumors, only 1 tumor was observed per mouse. Tumors were approximate as spheres and their sizes were calculated with the following formula:

$$\text{Volume (mm}^3) = 0.5 * \text{length (mm)} * \text{height (mm)} * \text{depth (mm)}$$

Survival analysis

We observed that mice that developed brain tumors rapidly deteriorate in their body condition scores. Mice with observed macrocephaly and body condition score 1 were euthanized and the euthanasia date was recorded as the last survival date. Occasionally mice bearing brain tumors died unexpectedly early, and the date of death was recorded as the last survival date. Cohorts of mice stereotaxically injected with PBS, AAV-vector or AAV-mTSG virus were monitored for their survival. Survival analysis was analyzed using standard Kaplan-Meier analysis. Of note, several AAV-vector or PBS injected mice were sacrificed at time points earlier than 299 days (at times when a certain AAV-mTSG mice were found dead or euthanized due to poor body conditions), to provide time-matched histology, but those mice were healthy without brain tumors or other signs of detectable symptoms. Mice euthanized early in healthy state were excluded from calculation of survival percentage.

Mouse brain dissection, fluorescent imaging, and histology

Mice were sacrificed by carbon dioxide asphyxiation or deep anesthesia with isoflurane followed by cervical dislocation. Mouse brains were manually dissected under a fluorescent stereoscope (Zeiss, Olympus or Leica). Brightfield and/or GFP fluorescent images were taken for the dissected brain, and overlaid using ImageJ[58]. Brains were then fixed in 4% formaldehyde or 10% formalin for 48 to 96 hours, embedded in paraffin, sectioned at 6 μm and stained with hematoxylin and eosin (H&E) for pathology. For tumor size quantification, H&E slides were scanned using an Aperio digital slidescanner (Leica). Tumors were manually outlined as region-of-interest (ROI), and subsequently quantified using ImageScope (Leica). Sections were de-waxed, rehydrated and stained using standard immunohistochemistry (IHC) protocols as previously[55, 59]. The following commonly used antibodies were used for IHC: rabbit anti-Ki67 (abcam ab16667, 1:500), rabbit anti-GFP (ThermoFisher Scientific A11122, 1:300) rabbit anti-GFAP (Dako, 1:500), and mouse anti-Cas9 (Diagenode, 1:300).

Mouse tissue collection for molecular biology

Mouse brain (targeting organ) and liver (non-targeting organ) were dissected and collected manually. For molecular biology, tissues were flash frozen with liquid nitrogen, ground in 24 Well Polyethylene Vials with metal beads in a GenoGrinder machine (OPS diagnostics). Homogenized tissues were used for DNA/RNA/protein extractions using standard molecular biology protocols.

Genomic DNA extraction from cells and mouse tissues

For genomic DNA (gDNA) extraction, 50–200 mg of frozen ground tissue was resuspended in 6 ml of Lysis Buffer (50 mM Tris, 50 mM EDTA, 1% SDS, pH 8) in a 15 ml conical tube, and 30 μl of 20 mg/ml Proteinase K (Qiagen) was added to the tissue/cell sample and incubated at 55 °C overnight. The next day, 30 μl of 10 mg/ml RNAse A (Qiagen) was added to the lysed sample, which was then inverted 25 times and incubated at 37 °C for 30 minutes. Samples were cooled on ice before addition of 2 ml of pre-chilled 7.5M ammonium acetate (Sigma) to precipitate proteins. The samples were vortexed at high speed for 20 seconds and then centrifuged at $4,000 \times g$ for 10 minutes. Then, a tight pellet was visible in each tube and the supernatant was carefully decanted into a new 15 ml conical tube. Then 6 ml 100% isopropanol was added to the tube, inverted 50 times and centrifuged at $4,000 \times g$ for 10 minutes. Genomic DNA was visible as a small white pellet in each tube. The supernatant was discarded, 6 ml of freshly prepared 70% ethanol was added, the tube was inverted 10 times, and then centrifuged at $4,000 \times g$ for 1 minute. The supernatant was discarded by pouring; the tube was briefly spun, and remaining ethanol was removed using a P200 pipette. After air-drying for 10–30 minutes, the DNA changed appearance from a milky white pellet to slightly translucent. Then, 500 μl of ddH$_2$O was added, the tube was incubated at 65 °C for 1 hour and at room temperature overnight to fully resuspend the DNA. The next day, the gDNA samples were vortexed briefly. The gDNA concentration was measured using a Nanodrop (Thermo Scientific).

Targeted capture sequencing probe design

Targeted capture sequencing probes were designed as following: the predicted cutting sites (3bp 5′ of PAM) of the 280 gene-targeting sgRNAs in the mTSG library plus the *Trp53*-targeting sgRNA in the vector were retrieved from mouse genome (mm10). The 140bp sequences of the flanking regions of the cutting sites (5′- 70bp and 3′- 70bp) were retrieved using Bedtools[60]. The regions were consolidated using NimbleDesign (Roche / NimbleGen), and probe matches were set with these parameters (Preferred Close Matches = 3, where initial selection of probes for a given region will only include probes with 3 close matches or less; and Maximum Close Matches = 20, where if there are insufficient probes available for a given region at the Preferred Close Match number, the threshold will be incrementally increased to 20 until adequate coverage is achieved. After consolidation, a number of 178 regions covering 277 sgRNAs, with a total of 33638bp were covered in the probeset, with Target Bases Covered = 32239 (95.8%) and one target sgRNA without coverage due to a lack of qualified candidate probes in the region.

Targeted capture sequencing

The mTSG-Amplicon targeted capture sequencing probes were synthesized using the SeqCap EZ Probe Pool synthesis procedure (Roche). The capture sequencing was done following standard Illumina-Roche-Illumina protocols. Genomic DNA samples from mouse organs were subjected to fragmentation, followed by a library preparation step using KAPA Library Preparation Kit (Illumina). The libraries were then amplified using LM-PCR, hybridized to the mTSG-Amplicon probe pool, washed and recovered, and amplified with multiplexing barcodes using LM-PCR. The multiplexed library was then QC'ed using qPCR, and subjected to high-throughput sequencing using the Hiseq-2500 or Hiseq-4000 platforms (Illumina) at Yale Center for Genome Analysis. 277/278 (99.6%) of unique targeted sgRNAs were captured for all samples from this experiment, with the missing one being Arid1a-sg5.

mTSG sgRNA cutting efficiency measurement

The mouse mTSG sgRNA cutting efficiency measurement was performed similar to the screen with the exception of early sampling. Briefly, AAV-mTSG library virus was injected in to LV of LSL-Cas9 mice, but instead of end-point tumor, mice were sacrificed at early time point (3.5 weeks post injection) and examined under fluorescent stereoscope, and GFP + regions from the brain were dissected. Genomic DNA was extracted, and subjected to capture sequencing.

Mouse whole-exome capture sequencing

The mouse whole-exome capture was performed using SeqCap EZ exome kit (Roche). Briefly, capture sequencing was done following standard Illumina-Roche-Illumina protocols. Genomic DNA samples from mouse organs were subjected to fragmentation, followed by a library preparation step using KAPA Library Preparation Kit (Illumina). The libraries were then amplified using LM-PCR, hybridized to the exome probe pool, washed and recovered, and amplified with multiplexing barcodes using LM-PCR. The multiplexed library was then

QC'ed using qPCR, and subjected to high-throughput sequencing using the Hiseq-2500 or Hiseq-4000 platforms (Illumina) at Yale Center for Genome Analysis.

Illumina sequencing data processing and variant calling

FASTQ reads were mapped to the mm10 genome using the bwa mem function in BWA v0.7.13[61]. Bam files were merged, sorted, and indexed using bamtools v2.4.0[62] and samtools v1.3[63]. For each sample, indel variants were called using samtools and VarScan v2.3.9[64]. Specifically, we used samtools mpileup (–d 1000000000 –B –q 10), and piped the output to VarScan pileup2indel (–min-coverage 1 –min-reads2 1 –min-var-freq 0.001 –p-value 0.05). To link each indel to the sgRNA that most likely caused the mutation, we took the center position of each indel and mapped it to the closest sgRNA cut site.

Calling significantly mutated sgRNAs and significantly mutated genes

We further filtered all detected indels by requiring that each indel must overlap the ± 3 base pairs flanking the closest sgRNA cut site, as Cas9-induced double-strand breaks are expected to occur within a narrow window of the predicted cut site. We then utilized a series of criteria to identify high-confidence mutations: 1) As an initial pass to exclude possible germline mutations, we removed any sgRNAs with indels present in more than half of the control samples with greater than 5% variant frequency. In our data, this filter specifically removed Rps19_sg5 from further consideration. 2) To determine significantly cutting sgRNAs in each sample, we used a false-discovery approach based on the PBS and vector control samples. For each sgRNA, we first took the highest % variant read frequency across all control samples; in order for a mutation to be called in an mTSG sample, the % variant read frequency had to exceed the control sample cutoff. However, since the base vector contained a *Trp53* sgRNA (*Trp53* sg8) whose cut site was only 1 bp away from the target site of *Trp53* sg4 (from mTSG library), we only considered PBS samples when calculating the false-discovery cutoff for *Trp53* sg4. Nevertheless, in the current study this exception was unnecessary because of our third filter: 3) As we were most interested in identifying the dominant clones in each sample, we further set a 5% variant frequency cutoff on top of the false-discovery cutoff. These criteria gave us a binary table (i.e. not significantly mutated vs. significantly mutated) detailing each sgRNA and whether its target site was significantly mutated in each sample. None of the AAV-vector samples passed the 5% cutoff at the p53 sg4/8 target site, which is consistent with our observation that no tumors were found in vector-treated animals. To convert significantly cutting sgRNAs into significantly mutated genes, we simply collapsed the binary sgRNA scores by gene, such that if any of the sgRNAs for a gene were found to be significantly cutting, the gene would be called as significantly mutated.

Exome sequencing data analysis

For exome sequencing analysis, we imposed a modified set of criteria on each detected variant: 1)   10 supporting reads for the reference allele; 2)   10 supporting reads for the variant allele; 3) the variant is within ± 6 bp of a Cas9 PAM, NGG (or CCN on the reverse strand); 4) a variant allele frequency < 75%, as this was the maximum detected variant frequency out of the mTSG brain samples; 5) the variant was not detected in any sequenced control samples, which were considered as germline variants.

Clustering of variant frequencies to infer clonality of tumors

For each mTSG brain sample, we extracted the individual variants that comprised the SMS calls in that sample, with a cutoff of 5% variant frequency to eliminate low-abundance variants. Because of these cutoffs, 3 sequenced mTSG brain samples were not eligible for variant frequency clustering analysis. To identify clusters of variant frequencies in an unbiased manner, we modeled the variant frequency distribution with a Gaussian kernel density estimate, using the Sheather-Jones method to select the smoothing bandwidth. From the kernel density estimate, we then identified the number of local maxima (i.e. "peaks") within the density function. The number of peaks thus represents the number of variant frequency clusters for an individual sample, which is an approximation for the clonality of the tumors.

Coding frame analysis

For coding frame and exonic/intronic analysis, we only considered the indels that were associated with a sgRNA which had been considered significantly mutated in that particular sample. This final set of significant indels was converted to .avinput format and subsequently annotated using ANNOVAR v. 2016Feb01, using default settings[65].

Co-occurrence and correlation analysis

Co-occurrence analysis was performed by first generating a double-mutant count table for each pairwise combination of genes in the mTSG library. Statistical significance of the co-occurrence was assessed by hypergeometric test. For correlation analysis, we first collapsed the % variant frequency tables on the gene level (in other words, summing the % variant frequencies for all 5 of the targeting sgRNAs for each gene). Using these summed % variant frequency values, we calculated the Spearman correlation between all gene pairs, across each mTSG sample. Statistical significance of the correlation was determined by converting the correlation coefficient to a t-statistic, and then using the t-distribution to find the associated probability. A similar approach was used to analyze co-occurring mutations in human TCGA GBM data.

Testing driver combinations with sgRNA minipool

Mixtures of five sgRNAs targeting each gene were cloned as sgRNA minipool into the same astrocyte-specific AAV-CRISPR vector. For gene pair targeting, the five-sgRNA single gene minipools from both genes were mixed 1:1. Plasmid mixes were then packaged into AAV1/2. Briefly, HEK293FT cells were transfected with the minipools plasmids, pAAV1 plasmid, pAAV2 plasmid, helper plasmid pDF6, and PEI Max (Polysciences, Inc. 24765-2) in DMEM (ThermoFisher, 10569-010). 72 hours post transfection, cell culture media was discarded and cells were rinsed and pelleted via low speed centrifugation. Cells were then lysed and the supernatant containing viruses was applied to HiTrap heparin columns (GE Biosciences 17-0406-01) and washed with a series of salt solutions with increasing molarities. During the final stages, the eluates from the heparin columns were concentrated using Amicon ultra-15 centrifugal filter units (Millipore UFC910024). Titering of viral particles was executed by quantitative PCR using custom Cre-targeted Taqman probes (ThermoFisher). After packaging, AAV minipools were stereotaxically injected into the

ventricle of LSL-Cas9 mice. Survival and histology analysis followed injection as described above. Several control (uninjected, EYFP and vector) mice were sacrifice as surrogate histology although they were in good body condition and were subsequently found devoid of tumor.

Generation of Nf1 and Rb1 mutant cell lines from primary GBMs induced by AAV-CRISPR minipools

Autochthonous mouse GBMs were induced by stereotaxic injection with the *Nf1* or *Rb1* AAV minipool (in the AAV9-sg*Trp53*-sgX-GFAP-Cre vector described above). Tumor-containing brains were visually inspected under a fluorescent dissecting scope, made into single-cell suspension through physical dissociation plus Collagenase/DNase digestion, and cultured in DMEM supplemented with 10% FBS and Pen/Strep. Growing clones were further established as autochthonous mouse GBM cell lines.

Single sgRNA knockout lentiviral production

Lenti-pHKO-U6-sgBsmBI-EF1a-Puro-P2A-FLuc was generated by subcloning P2A-Fluc expression cassette into lentiviral CRISPR knockout vector by Gibson assembly. For the cloning of sgRNA targeting individual genes such as *Pten*, *Arid1b*, *Mll3 (Kmt2c)*, *B2m*, and *Zc3h13* (Table S1), the corresponding oligos were synthesized, annealed and cloned into BsmBI linearized lentiviral knockout vector. Lentiviruses were produced by transfecting lentiviral knockout plasmids, together with pMD2.G and psPAX2 into 80–90% confluent HEK293FT cells, with viral supernatants collected 48 and 72 h post-transfection, aliquoted and stored in −80°C.

Generation of NF1+geneX and RB1+geneX knockout cell lines

The *Nf1* and *Rb1* knockout tumor cells were infected by single sgRNA knockout lentiviruses at M.O.I <= 0.3 to further knockout desired geneX. 24 h post-infection, lentiviral transduced cells were selected by the addition of 4–8 μg/ml puromycin and were split 2–3 days.

Temozolomide (TMZ) treatment, Cell viability assay, and RNAseq

After 7–9 days' culture under puromycin selection, lentiviruses-infected *Nf1* and *Rb1* knockout tumor cells were plated in triplicates into 96-well plate at a density of 2.5×10^3 cells per well, and ~5×10^6 cells were collected at the same time and used for cutting efficiency analysis. One day after plating, either TMZ or DMSO was added at a concentration of 10 μM, 100 μM, 500 μM, 1 mM, and 2 mM. After 3 days of drug/vehicle treatment, cell viability was measured using CellTiter Glo (Promega) according to the manufactures' protocol. Briefly, we first equilibrated the CellTiter Glo at room temperature for 1 h before use. The media of 96-well plates was aspirated, and then 50 μL fresh DMEM +10%FBS and 50 μL CellTiter Glo was added. The luminescent signals were readout using EnVision plate reader (PerkinElmer). For RNAseq samples preparation, cell lines harboring specific gene knockouts were cultured for 7–9 days under the selection pressure of puromycin, and then plated into 6-well plates at a density of 2×10^5 cells per well in triplicates. 24 h after plating, 1 mM TMZ or DMSO in fresh DEME + 10% FBS was added

and cultured for another 48 h. Then, cellular RNA of control or treated cells were extracted by adding 350 μL TRIzol Reagent (Invitrogen) directly into 6-well plates to lyse the cells, followed by gently shaking the plates and incubation for 5–10 min to complete and lysis and homogenization. Then, 70 μL chloroform was added, vigorously mixed, and centrifuged at 16,000 g for 15 min. Transfer the RNA containing aqueous phase to a new tube and further purified using RNeasy mini kit (Qiagen). After eluting RNA from column using Nuclease free water, the concentrations of sample RNA were normalized into 150–300 ng/uL for RNAseq.

T7 Endonuclease I (T7E1) Assays

The genomic DNA of these cells that were collected after 9 days of puromycin selection was extracted by using QuickExtract™ DNA Extraction Solution (Epicentre), mixed well and incubated at 65°C for 30–60 min. Then, 1–2 μL of genomic DNA from parental or Lenti-sgRNA transduced cells was used as the template to amplify gene of interest using surveyor primers (Table S33) with thermocycling conditions as 98°C for 1 min, 35 cycles of (98°C for 1 s, 60°C for 5 s, 72°C for 10 s), and 72°C for 1 min. The PCR products were gel-purified using QIAquick Gel Extraction Kit from 2% E-gel EX and quantified, followed by PCR products denaturing at 95°C for 5 min, and annealing by using following conditions: ramp from 95 to 85 °C at a rate of −2°C per seconds; from 85°C to 25°C at a rate of −0.1°C per second, and 4°C hold. 1 μL of T7 endonuclease I was added into annealed oligo, and incubated at 37°C for 60 min to digest the mismatched sites. The digested PCR products were loaded into 2% E-gel EX, and the amount of DNA fragments were quantified. The cutting efficiency was calculated to estimate gene editing using the following formula: Indels (%) = 100 × (1 – (1- fraction cleaved)1/2).

Transcriptome profiling of different driver combinations in the presence and absence of chemotherapy

Mixtures of five sgRNAs targeting each gene were cloned as sgRNA minipool into the same astrocyte-specific AAV-CRISPR vector. After packaging, AAV minipools were stereotaxically injected into the lateral ventricle of LSL-Cas9 mice. Cell lines were derived from mouse GBMs by single-cell isolation, plating and culture in DMEM media. Additional driver mutations were introduced by lentiCRISPR where applicable. GBM cells with different drivers were treated with DMSO or TMZ for 48h, and harvested for mRNA-seq for transcriptome profiling. Briefly, total RNA was extracted from cancer cells derived from AAV-CRISPR minipools induced GBM treated with DMSO or TMZ, using commercially available kits (Qiagen / Thermofisher). PolyA-mRNA library was constructed using Illumina TruSeq mRNA library prep kit, and sequenced on Illumina Hiseq 2500 and/or Hiseq 4000 platform.

RNA-seq differential expression analysis

Strand-specific single-end RNA-seq read files were analyzed to obtain transcript level counts using Kallisto[66], with the settings –rf-stranded -b 100. The counts were subsequently passed to the *tximport* R package to collapse to gene level counts. Pairwise differential expression analysis between groups was then performed using edgeR with default settings[67].

Pathway enrichment analysis of differentially expressed transcripts

Using an adjusted p-value cutoff of 0.05, and a log fold change threshold of ±1, we determined the set of genes that were significantly upregulated or downregulated. We then used the resultant gene sets for DAVID functional annotation analysis[68]. We considered a GO category as statistically significant if the Benjamini-Hochberg adjusted p-value was less than 0.05.

GBM comparative cancer genomics analysis using TCGA datasets

Somatic mutation calls, copy number variation calls, RNA-seq expression z-scores, and clinical data containing patient survival information were obtained through cBioPortal for GBM on November 15, 2016. Pearson correlation coefficients were calculated comparing mouse and human mutation frequencies; statistical significance was calculated by converting the correlation coefficient to a t-statistic, and then using the t-distribution to calculate significance.

GBM comparative cancer genomics analysis using Yale Glioma datasets

Somatic mutation calls and copy number variation calls and partial clinical data containing diagnostic information were obtained from Yale Glioma tissue bank and data bank. All patient samples were de-identified. The general description, demographics and tumor characteristics are noted in the supplemental tables. Total event for each patient was calculated as the sum of mutation events and copy number variant events. Pearson correlation coefficients were calculated comparing mouse and human mutation frequencies; statistical significance was calculated by converting the correlation coefficient to a t-statistic, and then using the t-distribution to calculate significance.

Histology analysis of clinical GBM samples from Yale Glioma tissue bank

Histology sections were obtained from Yale Glioma tissue bank. All patient samples were de-identified. The mutations associated with specific samples were obtained from Yale Glioma data bank. Slides stained with H&E or anti-GFAP were subsequently scanned using a slidescanner (Leica) and subjected to pathological analysis.

Statistical tests

In addition to the statistical tests detailed above, we used a two-tailed Welch's *t*-test for comparisons in which group variances were unequal. If the variances were found to be sufficiently equal and the data was normally distributed, we used a standard two-tailed *t*-test. For evaluating differences in the incidence of tumors in different groups, we used Fischer's exact test.

Blinding statement

Investigators were blinded for histology scoring, capture sequencing and RNA-seq collection, but not blinded for dissection, MRI, survival analysis, or sequencing analysis.

Determination of sample sizes

No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications cited here.

Code availability

Key custom scripts used to process and analyze the data are included in Supplementary Software. Other accessory scripts are available on request.

Data accession

Genomic sequencing (targeted capture, exome) and RNA-seq data have all been deposited in NCBI SRA (project PRJNA393202): targeted capture sequencing (SUB2842710), RNA-seq (SUB2843781), and exome sequencing (SUB2843864). CRISPR reagents (AAV-CRISPR and lentiCRISPR backbone plasmids and mTSG libraries) are available to the academic community and have been submitted through Addgene.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Sturm D, et al. Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. Nat Rev Cancer. 2014; 14:92–107. [PubMed: 24457416]

2. Louis DN, et al. The 2007 WHO ClassiWcation of Tumours of the Central Nervous System. Acta Neuropathol. 2007; 114:97–109. [PubMed: 17618441]

3. Krex D, et al. Long-term survival with glioblastoma multiforme. Brain. 2007; 130:2596–2606. [PubMed: 17785346]

4. ABTA. Glioblastoma and Astrocytoma Malignant. 2016.

5. Claus EB, Black PM. Survival rates patterns of care for patients diagnosed with supratentorial low-grade gliomas data from the SEER program, 1973-2001. Cancer. 2006; 106:1358–1363. [PubMed: 16470608]

6. Stupp R, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med. 2005; 352:987–996. [PubMed: 15758009]

7. Stupp R, et al. 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. 2009; 10:459–466. [PubMed: 19269895]

8. TCGA-Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]

9. Brennan CW, et al. The Somatic Genomic Landscape of Glioblastoma. Cell. 2013; 155:462–477. [PubMed: 24120142]

10. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010; 17:98–110. [PubMed: 20129251]

11. Bai H, et al. Integrated genomic characterization of IDH1-mutant glioma malignant progression. Nat Genet. 2016; 48:59–66. [PubMed: 26618343]

12. Ceccarelli M, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell. 2016; 164:550–563. [PubMed: 26824661]

13. Parsons DW, et al. An integrated genomic analysis of human glioblastoma multiforme. Science. 2008; 321:1807–1812. [PubMed: 18772396]

14. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011; 144:646–674. [PubMed: 21376230]

15. Hainaut P, Plymoth A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. Curr Opin Oncol. 2013; 25:50–51. [PubMed: 23150341]

16. Cairncross JG, et al. Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. J Natl Cancer Inst. 1998; 90:1473–1479. [PubMed: 9776413]

17. Napoli M, Flores ER. The p53 family orchestrates the regulation of metabolism: physiological regulation and implications for cancer therapy. Br J Cancer. 2016

18. Muller PA, Vousden KH. p53 mutations in cancer. Nat Cell Biol. 2013; 15:2–8. [PubMed: 23263379]

19. Feng Z, Hu W, Rajagopal G, Levine AJ. The tumor suppressor p53: cancer and aging. Cell Cycle. 2008; 7:842–847. [PubMed: 18414039]

20. Levine AJ. p53 the cellular gatekeeper for growth and division. Cell. 1997; 88:323–331. [PubMed: 9039259]

21. Berns A. Cancer: The blind spot of p53. Nature. 2010; 468:519–520. [PubMed: 21107421]

22. Kandoth C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502:333-+. [PubMed: 24132290]

23. Lawrence MS, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505:495-+. [PubMed: 24390350]

24. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–218. [PubMed: 23770567]

25. Davoli T, et al. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Cell. 2013; 155:948–962. [PubMed: 24183448]

26. Guo F, Li Y, Liu Y, Wang J, Li G. ARL6IP1 mediates cisplatin-induced apoptosis in CaSki cervical cancer cells. Oncol Rep. 2010; 23:1449–1455. [PubMed: 20372863]

27. Ribas A. Tumor immunotherapy directed at PD-1. N Engl J Med. 2012; 366:2517–2519. [PubMed: 22658126]

28. Frese KK, Tuveson DA. Maximizing mouse cancer models. Nature Reviews Cancer. 2007; 7:645–658. [PubMed: 17687385]

29. Holland EC. Gliomagenesis: genetic alterations and mouse models. Nat Rev Genet. 2001; 2:120–129. [PubMed: 11253051]

30. Huse JT, Holland EC. Genetically engineered mouse models of brain cancer and the promise of preclinical testing. Brain Pathol. 2009; 19:132–143. [PubMed: 19076778]

31. Alcantara Llaguno S, et al. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. Cancer Cell. 2009; 15:45–56. [PubMed: 19111880]

32. Friedmann-Morvinski D, et al. Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. Science. 2012; 338:1080–1084. [PubMed: 23087000]

33. Marumoto T, et al. Development of a novel mouse glioma model using lentiviral vectors. Nat Med. 2009; 15:110–116. [PubMed: 19122659]

34. Perry A. Malignant gliomas with primitive neuroectodermal tumor-like components: a clinicopathologic and genetic study of 53 cases. Brain Pathol. 2009; 19:81–90. [PubMed: 18452568]

35. Reilly KM, Loisel DA, Bronson RT, McLaughlin ME, Jacks T. Nf1;Trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. Nat Genet. 2000; 26:109–113. [PubMed: 10973261]

36. Schmid RS, Vitucci M, Miller CR. Genetically engineered mouse models of diffuse gliomas. Brain Res Bull. 2012; 88:72–79. [PubMed: 21684324]

37. Qazi M, et al. Generation of murine xenograft models of brain tumors from primary human tissue for in vivo analysis of the brain tumor-initiating cell. Methods Mol Biol. 2014; 1210:37–49. [PubMed: 25173159]

38. Agemy L, et al. Targeted nanoparticle enhanced proapoptotic peptide as potential therapy for glioblastoma. P Natl Acad Sci USA. 2011; 108:17450–17455.

39. Chow LM, et al. Cooperativity within and among Pten, p53 and Rb pathways induces high-grade astrocytoma in adult brain. Cancer Cell. 2011; 19:305–316. [PubMed: 21397855]

40. Chen J, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature. 2012; 488:522–526. [PubMed: 22854781]

41. Jinek M, et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012; 337:816–821. [PubMed: 22745249]

42. Cong L, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013; 339:819–823. [PubMed: 23287718]

43. Mali P, et al. RNA-Guided Human Genome Engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]

44. Sanchez-Rivera FJ, Jacks T. Applications of the CRISPR-Cas9 system in cancer biology. Nature Reviews Cancer. 2015; 15:387–395. [PubMed: 26040603]

45. Weitzman MD, Kyostio SRM, Kotin RM, Owens RA. Adenoassociated Virus (Aav) Rep Proteins Mediate Complex-Formation between Aav DNA and Its Integration Site in Human DNA. P Natl Acad Sci USA. 1994; 91:5808–5812.

46. Yeung JT, et al. LOH in the HLA class I region at 6p21 is associated with shorter survival in newly diagnosed adult glioblastoma. Clin Cancer Res. 2013; 19:1816–1826. [PubMed: 23401227]

47. Patil V, Pal J, Somasundaram K. Elucidating the cancer-specific genetic alteration spectrum of glioblastoma derived cell lines from whole exome and RNA sequencing. Oncotarget. 2015; 6:43452–43471. [PubMed: 26496030]

48. Bale TA, et al. Genomic characterization of recurrent high-grade astroblastoma. Cancer Genet. 2016; 209:321–330. [PubMed: 27425854]

49. Aithal MG, Rajeswari N. Validation of housekeeping genes for gene expression analysis in glioblastoma using quantitative real-time polymerase chain reaction. Brain Tumor Res Treat. 2015; 3:24–29. [PubMed: 25977903]

50. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. Nature Reviews Cancer. 2015; 15:747–756. [PubMed: 26536825]

51. Gao JJ, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal. 2013; 6

52. Shalem O, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014; 343:84–87. [PubMed: 24336571]

53. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 2014; 343:80–84. [PubMed: 24336569]

54. Platt RJ, et al. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell. 2014; 159:440–455. [PubMed: 25263330]

55. Chen S, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell. 2015; 160:1246–1260. [PubMed: 25748654]

56. Franklin, KBJ., Paxinos, G. Paxinos and Franklin's The mouse brain in stereotaxic coordinates. Academic Press, an imprint of Elsevier; Amsterdam: 2013.

57. Fedorov A, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012; 30:1323–1341. [PubMed: 22770690]

58. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: An open platform for biomedical image analysis. Mol Reprod Dev. 2015; 82:518–529. [PubMed: 26153368]

59. Chen S, et al. Global microRNA depletion suppresses tumor angiogenesis. Genes Dev. 2014; 28:1054–1067. [PubMed: 24788094]

60. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]

61. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

62. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011; 27:1691–1692. [PubMed: 21493652]

63. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

64. Koboldt DC, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–576. [PubMed: 22300766]

65. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38

66. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34:525–527. [PubMed: 27043002]

67. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]

68. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
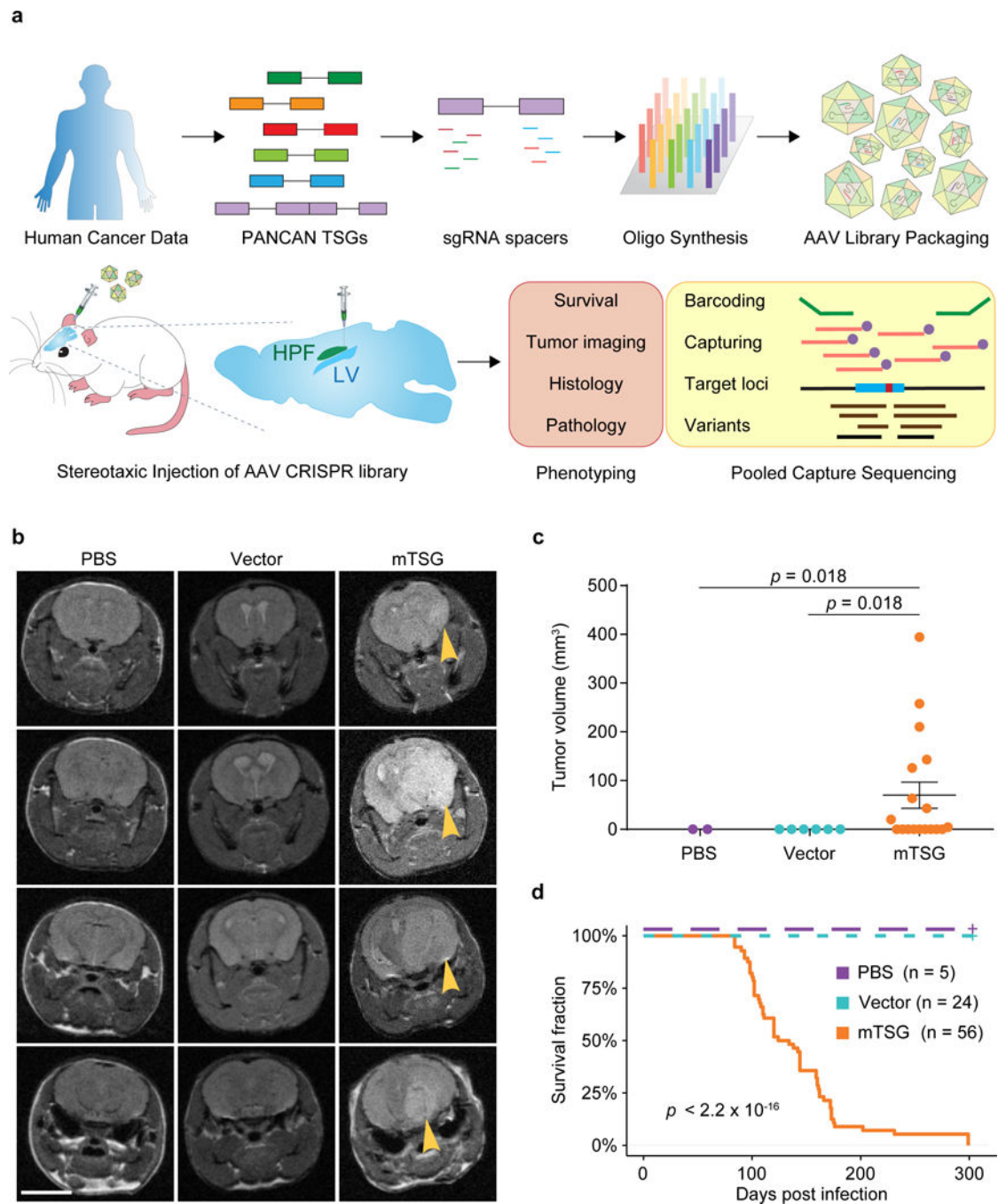
Figure 1. Autochthonous brain tumorigenesis induced by an AAV-mediated CRISPR library

(**a**) Schematics of direct *in vivo* AAV-CRISPR GBM screen design. Top panel, AAV-mTSG library design, synthesis and production. Bottom panel, stereotaxic injection of AAV library and subsequent analysis. HPF, hippocampus; LV, lateral ventricle.

(**b**) MRI sections show brain tumors in AAV-mTSG injected mice, but not in matching sections from PBS or AAV-vector injected mice. Arrowheads indicate brain tumors. Scale bar, 5 mm.

(**c**) MRI-based volumetric quantification of time-matched tumor size $\pm$ s.e.m. Two-tailed Welch's t-test, $t_{17} = 2.62$, $p = 0.018$, mTSG vs. vector or PBS (PBS, n = 2 mice; Vector, n = 6; mTSG, n = 18).

(**d**) Kaplan-Meier curves for overall survival (OS) of mice injected with PBS (n = 5), AAV-vector (n = 24) or AAV-mTSG library (n = 56). OS for PBS and vector groups are both 100%, where the curves are dashed and slightly offset for visibility. Log-rank (LR) test, $p < 2.20 * 10^{-16}$, mTSG vs. vector or PBS.
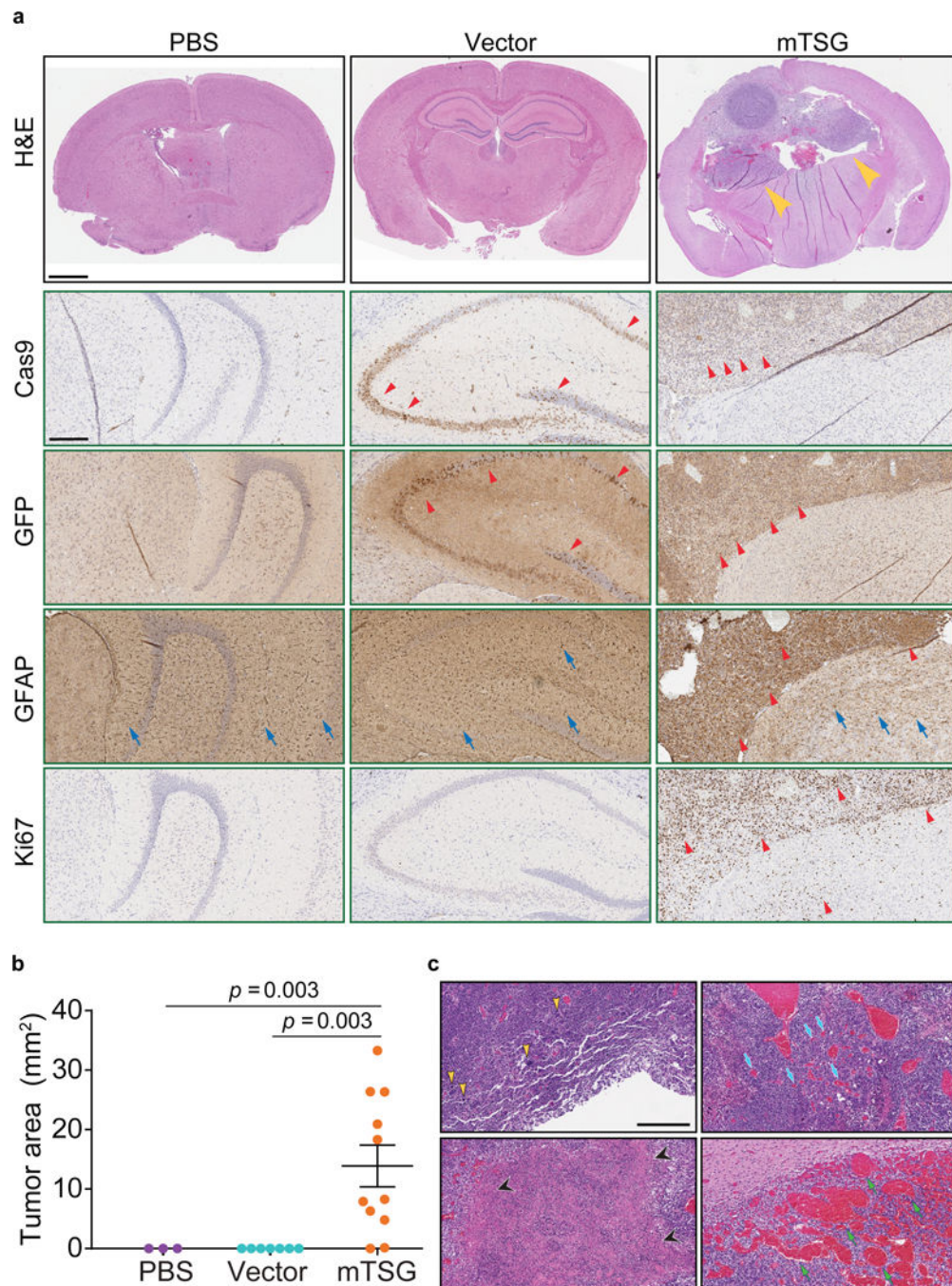
Figure 2. AAV-mTSG induced brain tumors recapitulate pathological features of GBM
(**a**) Top panel, representative H&E brain sections from PBS, AAV-vector and AAV-mTSG injected mice. Arrowheads indicate brain tumors. Scale bar = 1 mm. Lower panels, representative images of brain sections from PBS, AAV-vector and AAV-mTSG injected mice stained by Cas9, GFP, GFAP, and Ki67 immunohistochemistry. Cas9 IHC, red arrowheads indicate Cas9-positive cells; GFP IHC, red arrowheads indicate GFP-positive cells; GFAP IHC, representative GFAP-positive astrocytes in PBS, AAV-vector and AAV-mTSG injected mice (blue arrows), as well as representative cancer cells in AAV-mTSG

injected mice (red arrowheads); Ki67 IHC, arrowheads indicate representative proliferative cells, which are mostly in tumors (AAV-mTSG) or scattered in tumor-adjacent brain regions (AAV-mTSG). Scale bar = 0.25 mm.

(**b**) Quantification of tumor sizes ± s.e.m. found in H&E brain sections from PBS, AAV-vector and AAV-mTSG injected mice. Two-tailed Welch's *t*-test, $t_{10} = 3.97$, $p = 0.003$, mTSG vs. vector or PBS (PBS, n = 3; Vector, n = 7; mTSG, n = 11).

(**c**) Representative higher magnification H&E images showing pathological features of AAV-mTSG induced brain tumors. Clockwise from top left: yellow arrowheads, giant aneuploid cells with pleomorphic nuclei; blue arrows, endothelial cells and angiogenesis; green arrows, hemorrhagic regions; black arrowheads, necrotic regions. Similar features were observed in human GBM patient sections from Yale Glioma tissue bank (Figure S3). Scale bar = 0.5 mm.
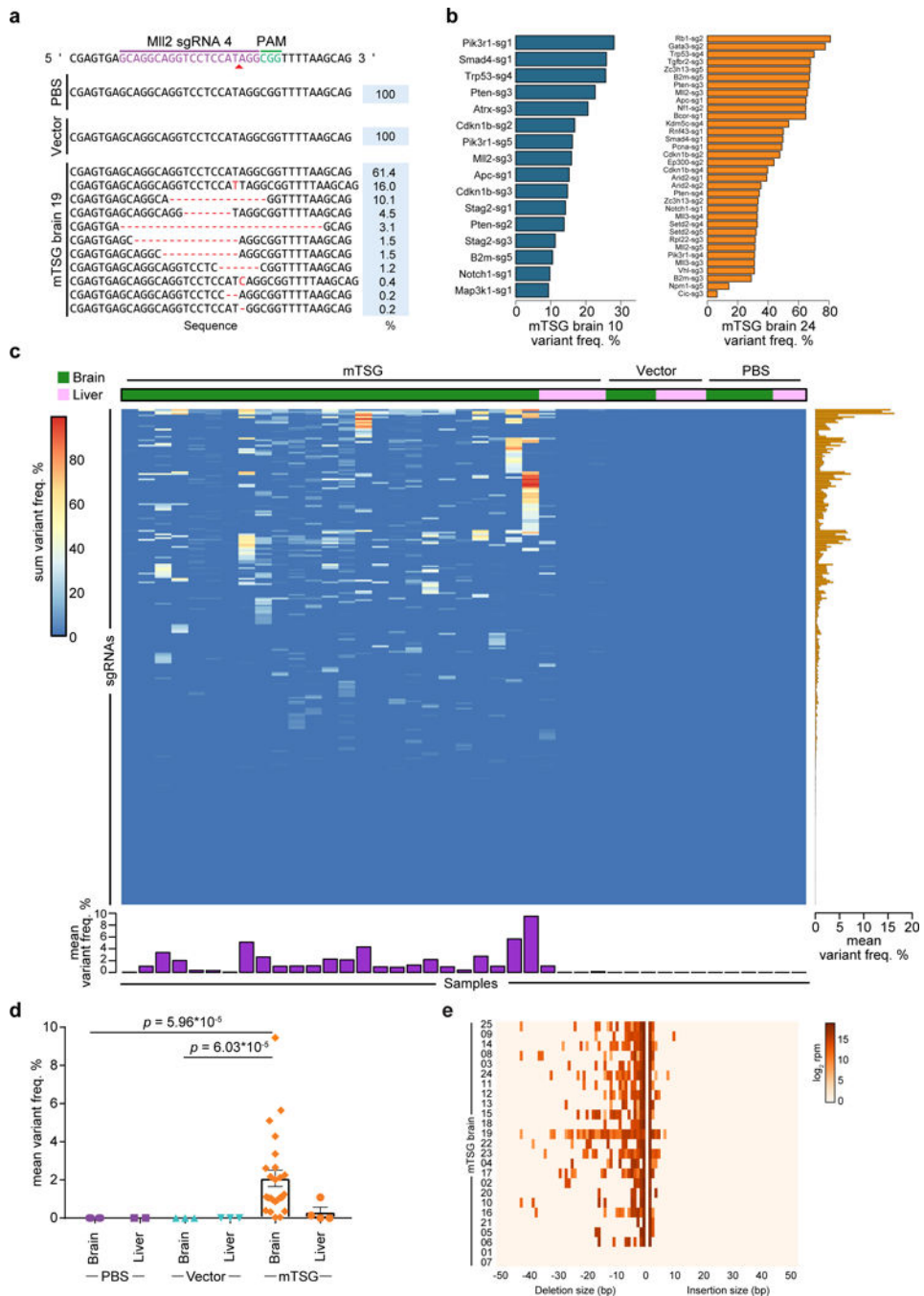
Figure 3. Targeted captured sequencing of sgRNA sites in AAV-mTSG induced mouse GBM

(**a**) Indel variants observed at the genomic region targeted by *Mll2* sgRNA 4 in representative PBS, AAV-vector, and AAV-mTSG injected mouse brain samples.

(**b**) Bar plots of variant frequencies in significantly mutated sgRNA target regions from two representative AAV-mTSG injected mouse brain samples.

(**c**) Heatmap of variant frequency across all targeted capture samples (n = 41). Rows denote individual sgRNAs, while columns correspond to samples from mice stereotaxically injected with PBS, AAV-vector, or AAV-mTSG. The liver was considered an off-target organ and

thus was used as a background control. Bar plots of the mean variant frequencies for each sgRNA (right panel, orange bars) and each sample (bottom panel, purple bars) are shown. (**d**) Dot plot of mean variant frequency ± s.e.m., grouped by treatment condition and tissue type. AAV-mTSG injected brains had significantly higher mean variant frequencies (2.087 ± 0.429, n = 25) compared to vector (0.005 ± 0.001, n = 3) or PBS (0.003 ± 0.001, n = 4) injected brains (two-tailed Welch's t-test, t_{24} = 4.85 and t_{24} = 4.86, p = 6.03 * 10^{-5} and p = 5.96 * 10^{-5} for mTSG vs. vector and mTSG vs. PBS). Comparing brain vs. liver in AAV-mTSG injected mice, mean variant frequencies of brains (2.087 ± 0.429) were significantly higher than livers (0.309 ± 0.261, n = 4) ($t_{21.48}$ = 3.54, p = 0.002).

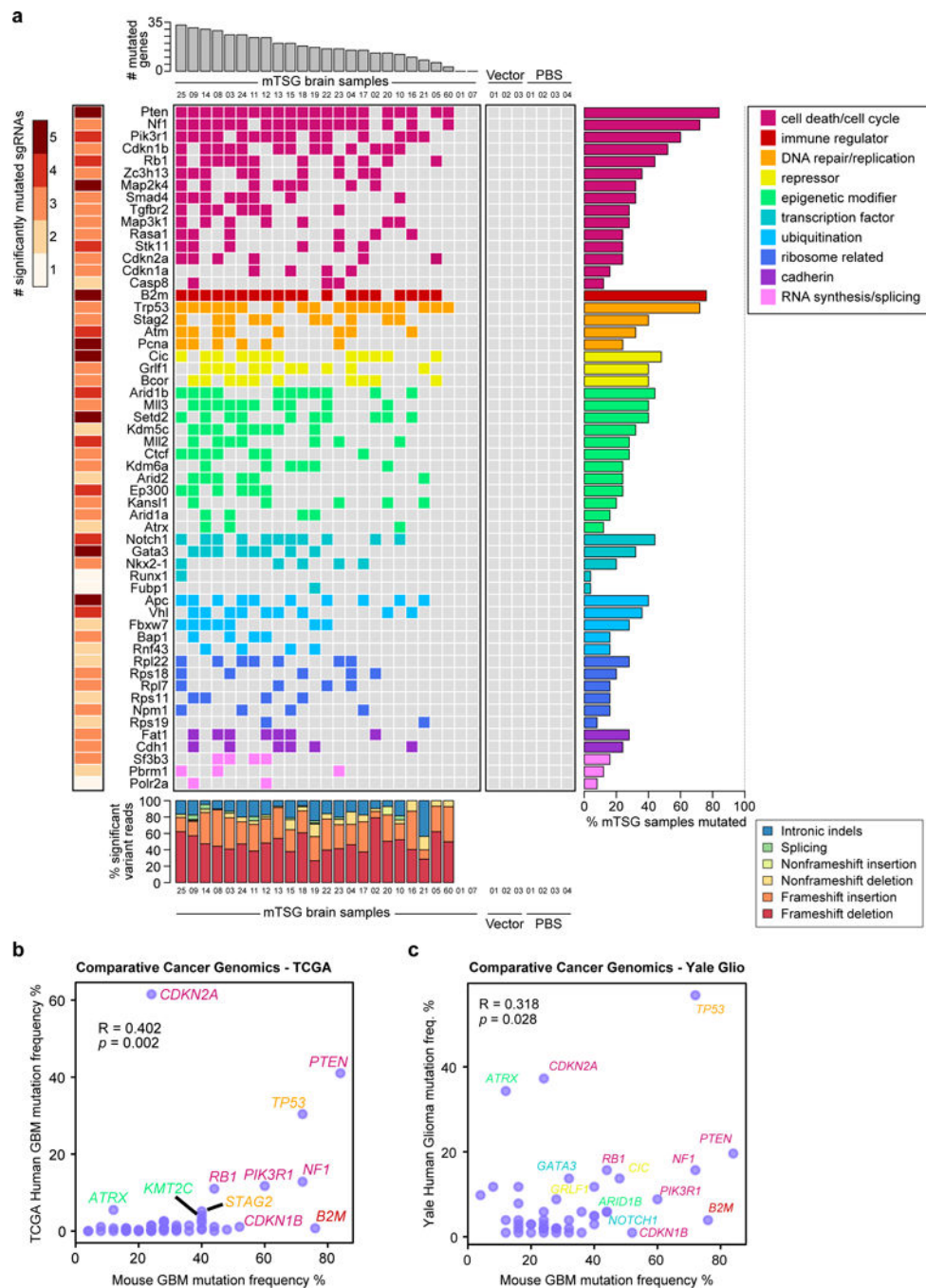(**e**) Indel size distribution for all filtered variants in each mTSG brain sample (n = 25).

Figure 4. Integrative analysis of functional mutations in driving tumorigenesis

(**a**) Gene-level mutational landscape of AAV-mTSG induced primary mouse GBM.

Center: Tile chart depicting the mutational landscape of primary brain samples from LSL-Cas9 mice injected with the AAV-mTSG library (n = 25), AAV-vector (n = 3) or PBS (n = 4). Genes are grouped and colored according to their functional classifications as noted in the top-right legend.

Top: Bar plots of the total number of significantly mutated genes identified in each AAV-mTSG sample.

Right: Bar plots of the percentage of GBM samples that were called as significantly mutated for each gene.

Left: Heatmap of the numbers of unique significantly mutated sgRNAs (SMSs) for each gene.

Bottom: Stacked bar plots describing the type of indels observed in each sample, colored according to the bottom-right legend.

(**b–c**) Comparative cancer genomics in GBM using the TCGA (b) and Yale Glioma (c) datasets. Scatterplot of population-wide mutant frequencies for the genes in the mTSG library, comparing AAV-mTSG treated mouse brain samples to human samples. Representative strong drivers in both species are labeled, with gene names color-coded based on their functional classification (as in a).

(**b**) Mutant frequencies in AAV-mTSG treated mouse brain samples correlated with patients in the TCGA GBM dataset (Pearson correlation R = 0.402, p = 0.002).

(**c**) Mutant frequencies in AAV-mTSG treated mouse brain samples correlated with patients in the Yale Glioma dataset (Yale Glio) (Pearson correlation R = 0.318, p = 0.028).
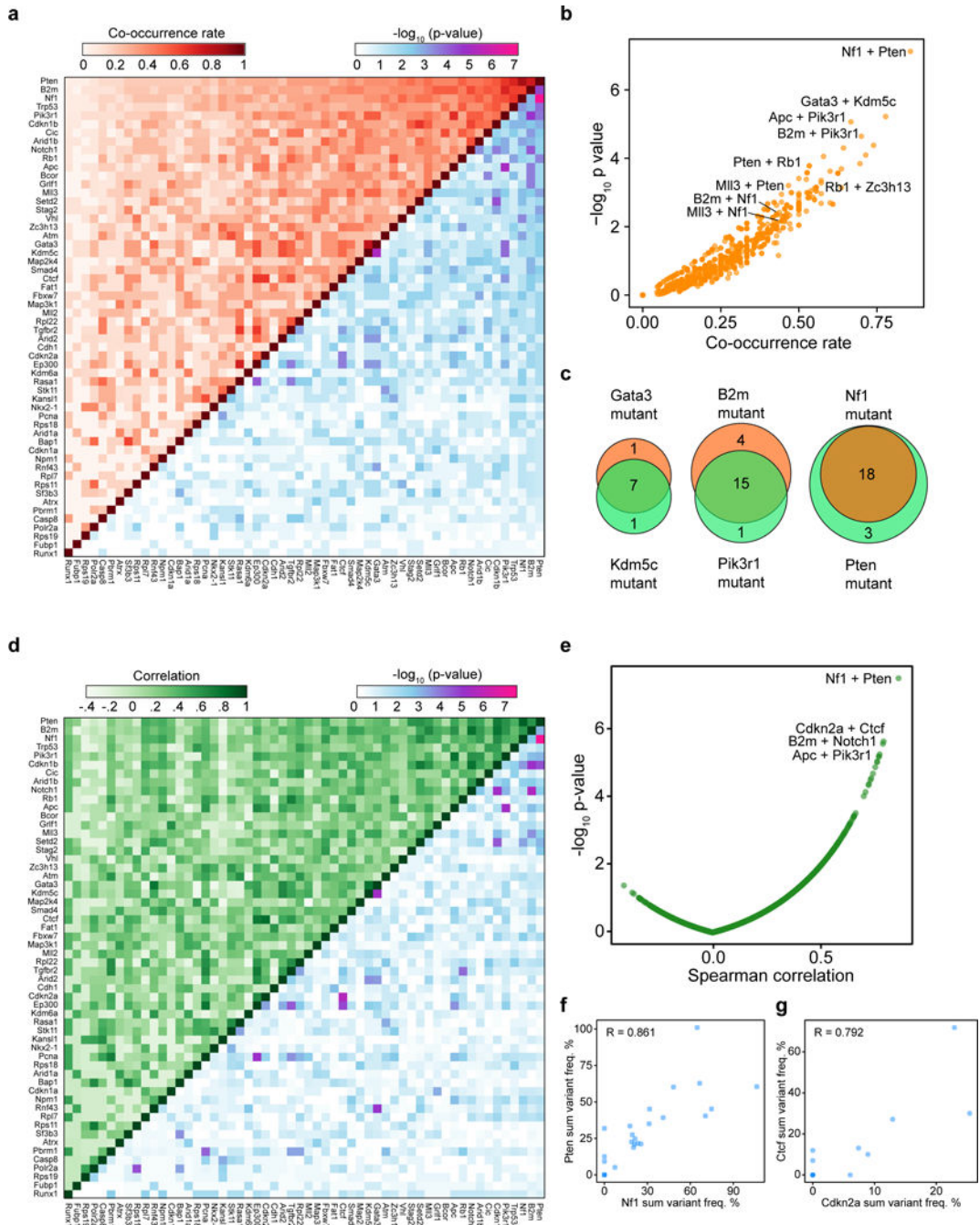
Figure 5. Co-mutation analysis uncovers synergistic gene pairs in GBM

(**a**) Upper-left half: heatmap of pairwise mutational co-occurrence rates. Lower-right half: heatmap of -log$_{10}$ p-values by hypergeometric test for statistical co-occurrence.

(**b**) Scatterplot of the co-occurrence rate of each gene pair, plotted against −log$_{10}$ p-values.

(**c**) Venn diagrams showing representative strongly co-occurring mutated gene pairs such as *Kdm5c* and *Gata3* (co-occurrence rate = 77.8%, hypergeometric test, $p = 6.04 * 10^{-6}$), *B2m* and *Pik3r1* (70.0%, $p = 2.28 * 10^{-5}$), as well as *Nf1* and *Pten* (85.7%, $p = 7.53 * 10^{-8}$).

(**d**) Upper-left half: heatmap of the pairwise Spearman correlation of variant frequency for each gene, summed across sgRNAs. Lower-right half: heatmap of $-\log_{10}$ p-values to evaluate the statistical significance of the pairwise correlations.

(**e**) Scatterplot of pairwise Spearman correlations plotted against $-\log_{10}$ p-values.

(**f–g**) Scatterplots showing representative strongly correlated gene pairs when comparing variant frequencies summed across sgRNAs, such as *Nf1+Pten* (f) and *Cdkn2a+Ctcf* (g). Spearman correlation coefficients are noted on the plot.
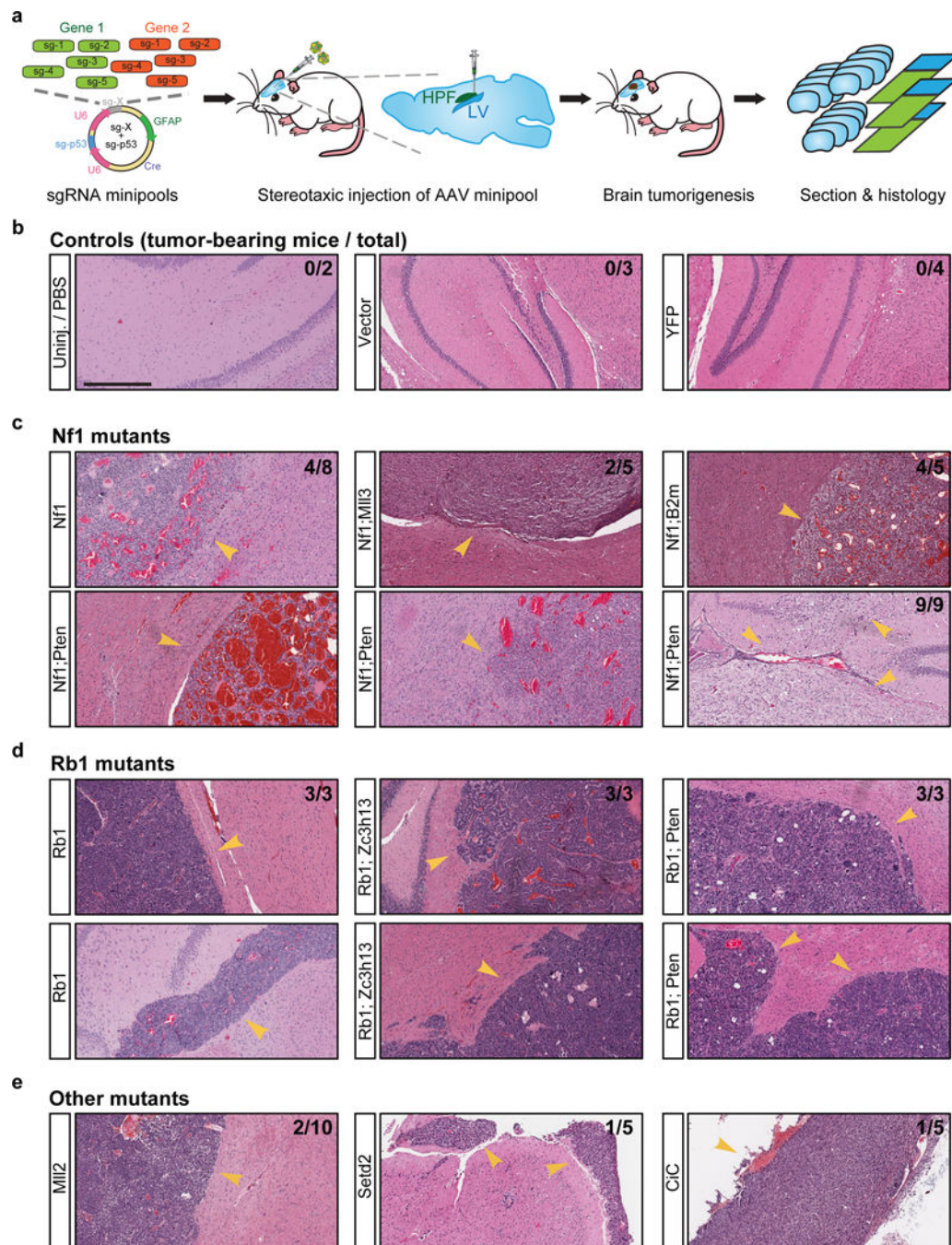
Figure 6. Validation of driver combinations

(**a**) Schematic representation of experiment design. Mixtures of five sgRNAs targeting each gene were cloned as minipools into the astrocyte-specific AAV-CRISPR vector. After packaging, AAV minipools were stereotaxically injected into the lateral ventricle of LSL-Cas9 mice.

(**b–e**) End-point histology (H&E) of representative brain sections from mice treated with AAV sgRNA minipools or relevant controls. In this end-point analysis, mice were euthanized when either macrocephaly or poor body condition score (< 2) was observed, with

survival time ranging from 3 to 11 months. Treatments are indicated to the left of each image. Arrowheads indicate the presence of brain tumors. The proportion of tumor-bearing to total mice is indicated in the top right corner of the images. Scale bar = 0.5 mm.

(**b**) Representative histology of brain sections from control mice. No tumors were observed in mice from the vector (0/3), EYFP (0/4) or uninjected (0/2) groups.

(**c**) Representative histology of brain sections from mice treated with various *Nf1* minipools, such as *Nf1* alone (4/8 mice developed tumors within 11 months), *Nf1;Pten* (9/9 mice developed tumors within 6 months), *Nf1;Mll3* (2/5 mice developed tumors within 6 months), and *Nf1;B2m* (4/4 mice developed tumors within 11 months).

(**d**) Representative histology of brain sections from mice treated with various *Rb1* minipools, such as *Rb1* alone (3/3 mice developed tumors within 6 months), *Rb1;Zc3h13* (3/3 mice developed tumors within 6 months), and *Rb1;Pten* (3/3 mice developed tumors within 6 months).

(**e**) Representative histology of brain sections from mice treated with other minipools, such as *Mll2* alone (2/10 mice developed tumors within 11 months), *Setd2* (1/5 mice developed tumors within 6 months), and *Cic* (1/5 mice developed tumors within 6 months).
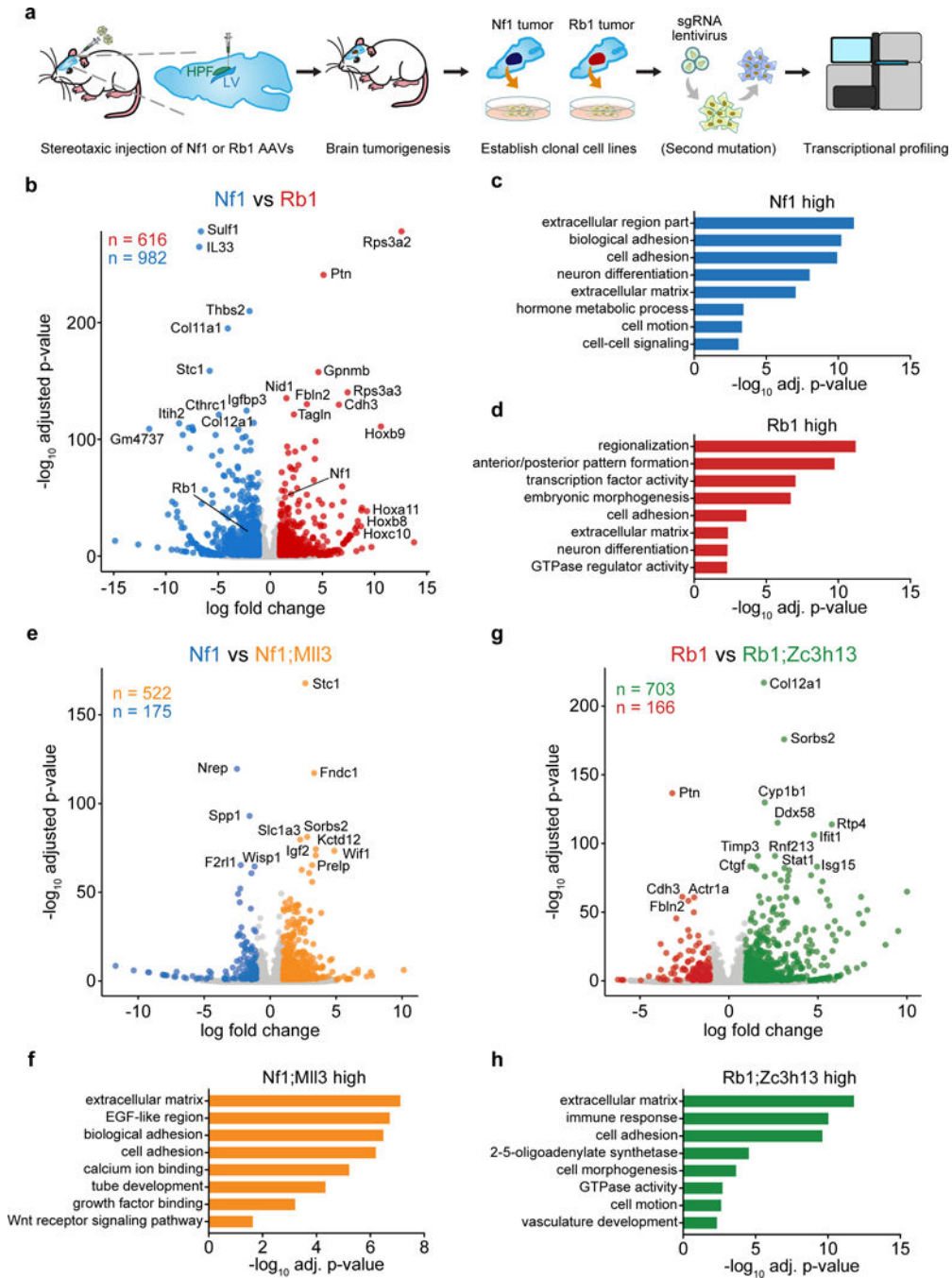
Figure 7. Transcriptional profiling of mouse GBM driver combinations
(**a**) Schematic of mouse GBM RNA-seq experimental design. *Rb1* or *Nf1* AAV minipools
were stereotaxically injected into the lateral ventricle of LSL-Cas9 mice. Cell lines were
derived from mouse GBMs by single-cell isolation. Additional driver mutations were
introduced by lentiCRISPR where applicable. Cells were then transcriptionally profiled by
RNA-seq (n = 3 samples per condition).

(**b**) Volcano plot comparing gene expression profiles in *Rb1* mutant (red) to *Nf1* mutant (blue) GBM cells. 616 genes were significantly higher in *Rb1* cells, and 982 genes were significantly higher in *Nf1* cells

(**c**) Enriched gene ontology categories among *Nf1*-high genes.

(**d**) Enriched gene ontology categories among *Rb1*-high genes.

(**e**) Volcano plot comparing *Nf1;Mll3* mutant (orange) to *Nf1* mutant (blue) GBM cells. 522 genes were significantly higher in *Nf1;Mll3* cells, and 175 genes were significantly higher in *Nf1* cells.

(**f**) Enriched gene ontology categories among *Nf1;Mll3*-high genes.

(**g**) Volcano plot comparing *Rb1;Zc3h13* mutant (green) to *Rb1* mutant (red) GBM cells. 703 genes were significantly higher in *Rb1;Zc3h13*, and 166 genes were significantly higher in *Rb1* cells.

(**h**) Enriched gene ontology categories among *Rb1;Zc3h13*-high genes.

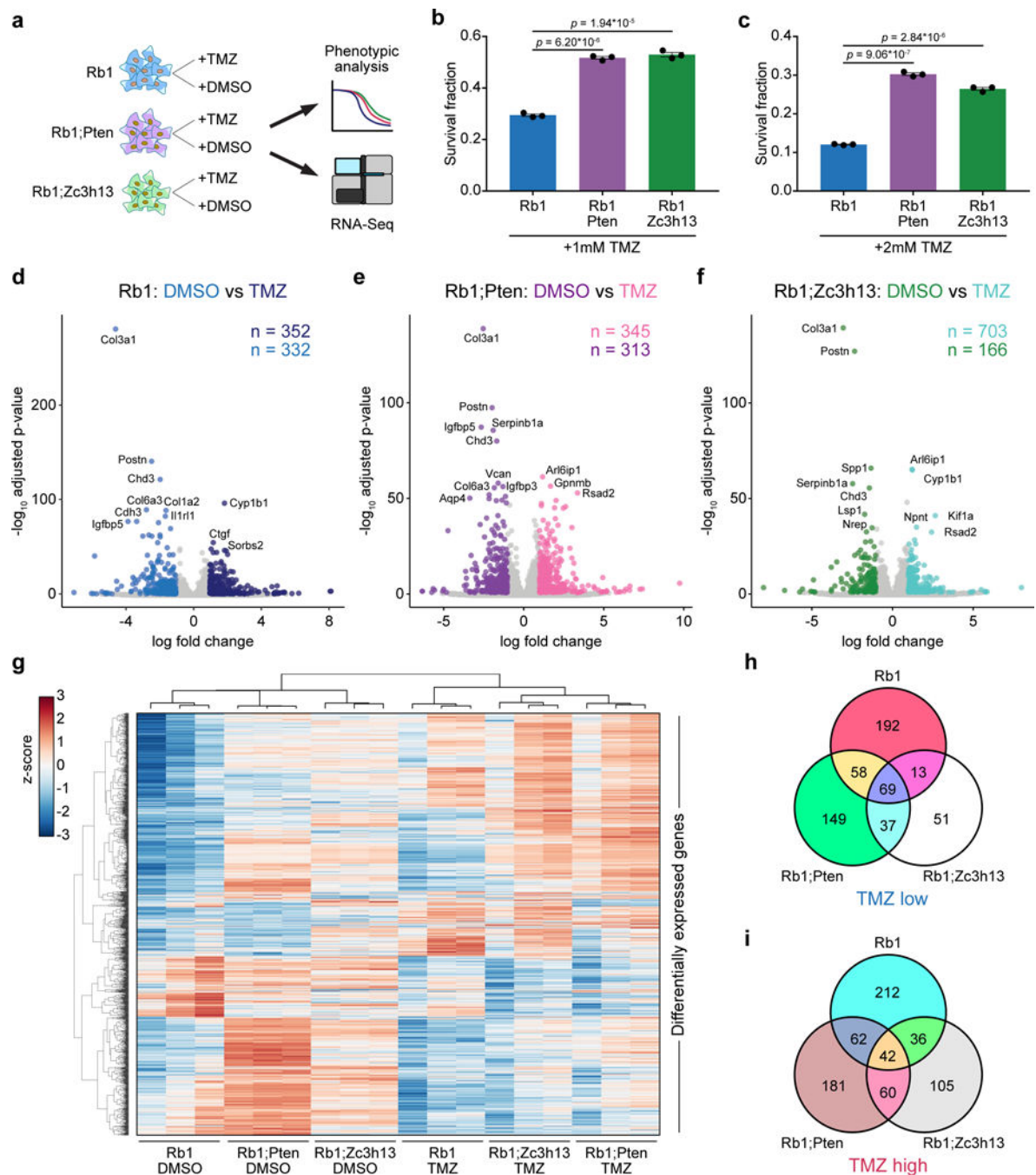Differentially expressed genes were defined as Benjamini-Hochberg adjusted $p < 0.05$ and log fold change ≥ 1 or ≤ −1.

Figure 8. Transcriptional profiling of mouse GBM driver combinations in the presence and absence of a chemotherapeutic agent

(**a**) Schematic of drug treatment RNA-Seq experimental design. *Rb1, Rb1;Pten*, and *Rb1;Zc3h13* GBM cells were treated with either temozolomide (TMZ) or DMSO. Cells were then subjected to phenotypic analysis and RNA-seq.

(**b–c**) Survival fraction ± s.e.m of *Rb1* (blue), *Rb1;Pten* (purple), and *Rb1;Zc3h13* (green) cells with 1 mM (b) or 2 mM (c) TMZ treatment. Individual cell replicates are shown (n = 3 for all conditions).

(**b**) *Rb1;Pten* and *Rb1;Zc3h13* cells had significantly higher survival fractions than *Rb1* cells with 1 mM TMZ (two-sided t-test, $p = 6.20 * 10^{-6}$ and $p = 1.94 * 10^{-5}$, respectively).

(**c**) *Rb1;Pten* and *Rb1;Zc3h13* cells had significantly higher survival fractions than *Rb1* cells with 2 mM TMZ (two-sided t-test, $p = 9.06 * 10^{-7}$ and $p = 2.84 * 10^{-6}$, respectively). Individual cell replicates are shown (n = 3 for all conditions).

(**d**) Volcano plot comparing *Rb1* cells treated with TMZ (dark blue) or DMSO (blue). 352 genes were significantly higher in TMZ-treated cells (TMZ-induced genes), and 332 genes were significantly higher in DMSO-treated cells (TMZ-reduced genes).

(**e**) Volcano plot comparing *Rb1;Pten* cells treated with TMZ (pink) or DMSO (purple). 345 genes were significantly higher in TMZ-treated cells, and 313 genes were significantly higher in DMSO-treated cells.

(**f**) Volcano plot comparing *Rb1;Zc3h13* cells treated with TMZ (turquoise) or DMSO (green). 703 genes were significantly higher in TMZ-treated cells, and 166 genes were significantly higher in DMSO-treated cells

(**g**) Heatmap of all differentially expressed genes among the TMZ vs. DMSO comparisons. Clustering was performed by average linkage using Pearson correlations. Values are shown in terms of z-scores, scaled by each gene.

(**h**) Venn diagram of TMZ-reduced genes for each tested genotype. While 69 genes were similarly downregulated among all 3 genotypes upon TMZ treatment, the differential expression signatures were nevertheless distinct, suggesting differential responses to TMZ treatment.

(**i**) Venn diagram of TMZ-induced genes for each tested genotype. Though 42 genes were consistently upregulated in all 3 groups upon TMZ treatment, numerous transcriptional differences were nevertheless apparent, suggesting differential responses to TMZ treatment. Differentially expressed genes were defined as Benjamini-Hochberg adjusted $p < 0.05$ and log fold change ≥ 1 or ≤ -1.