




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

Ryan D Chow<sup>1–3,17</sup> , Christopher D Guzman<sup>1,2,4–6,17</sup>, Guangchuan Wang<sup>1,2,17</sup>, Florian Schmidt<sup>7,8,17</sup>, Mark W Youngblood<sup>1,3,9,18</sup>, Lupeng Ye<sup>1,2,18</sup>, Youssef Errami<sup>1,2</sup>, Matthew B Dong<sup>1–3</sup>, Michael A Martinez<sup>1,2</sup>, Sensen Zhang<sup>1,2</sup>, Paul Renauer<sup>1,2,4</sup>, Kaya Bilguvar<sup>1,10</sup>, Murat Gunel<sup>1,3,9,10</sup>, Phillip A Sharp<sup>11,12</sup>, Feng Zhang<sup>13,14</sup>, Randall J Platt<sup>7,8</sup>  & Sidi Chen<sup>1–5,15,16</sup> 

**A causative understanding of genetic factors that regulate glioblastoma pathogenesis is of central importance. Here we developed an adeno-associated virus–mediated, autochthonous genetic CRISPR screen in glioblastoma. Stereotaxic delivery of a virus library targeting genes commonly mutated in human cancers into the brains of conditional-Cas9 mice resulted in tumors that recapitulate human glioblastoma. Capture sequencing revealed diverse mutational profiles across tumors. The mutation frequencies in mice correlated with those in two independent patient cohorts. Co-mutation analysis identified co-occurring driver combinations such as *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*.**

Glioblastoma (glioblastoma multiforme, GBM) is one of the deadliest cancers<sup>1</sup>. Current standard of care fails to cure the vast majority of patients with this disease<sup>1–3</sup>, leaving them a median survival time of 12.2 to 18.2 months<sup>4–7</sup>. The first genome atlas of GBM uncovered 453 validated nonsilent somatic mutations in 223 unique genes, which were further refined to a total of 71 significantly mutated genes (SMGs)<sup>8</sup>. Subsequent integrative genomic analyses revealed comprehensive mutational landscapes in GBM, uncovering 21 to 75 SMGs across multiple different cohorts of patients<sup>8–13</sup>. Many of these newly discovered genes have never been characterized in GBM; thus, their functional roles in gliomagenesis remain largely unknown<sup>9,10</sup>. 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<sup>3,14–16</sup>. 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 of which we are aware has comprehensively and combinatorially investigated which of the mutations identified in human patients can indeed functionally drive GBM from normal cells in the brain<sup>9</sup>. Key barriers include accurate delivery; precise genome

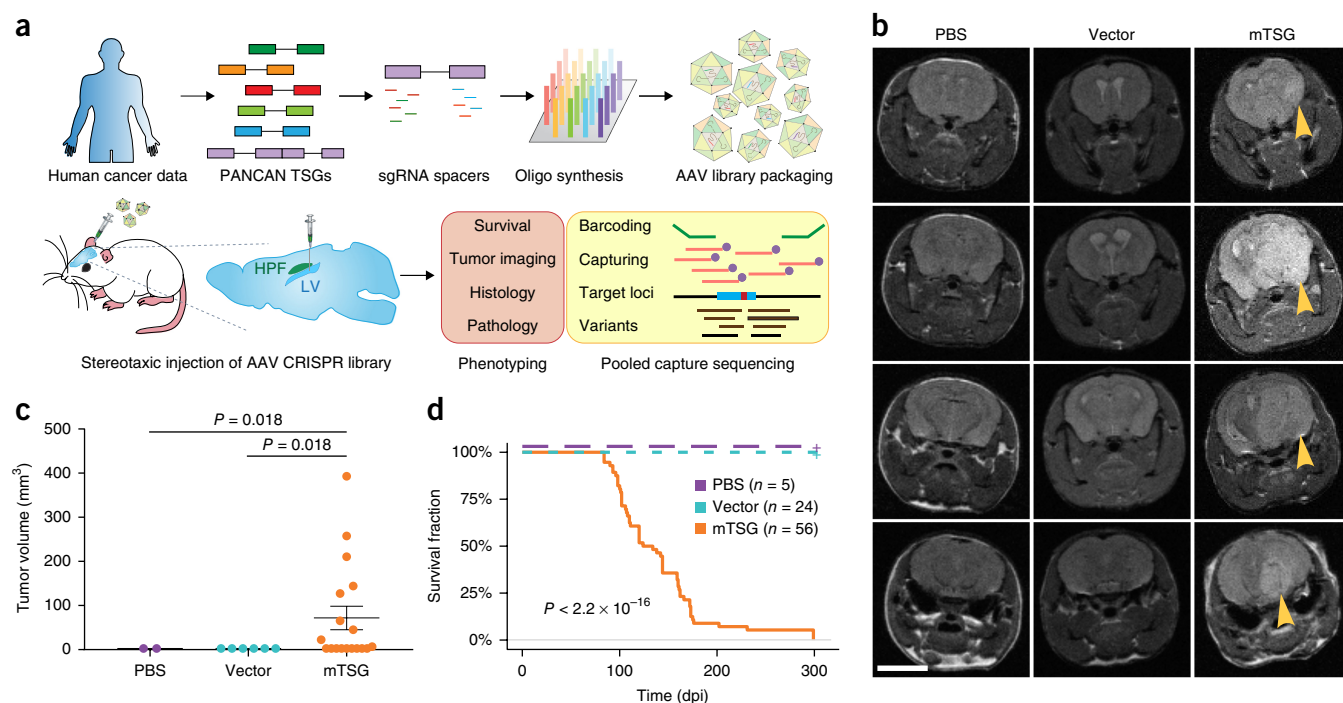
manipulation; efficient, massively parallel perturbation; and unbiased, high-sensitivity quantitative readout, all of which must be achieved simultaneously in the native brain microenvironment. We overcame these challenges through an AAV-mediated direct *in vivo* autochthonous CRISPR screen in the brain of fully immunocompetent mice, coupled with capture sequencing to achieve an ultradeep readout of all functional variants. With these data, we identified multiple drivers and co-occurring drivers, and subsequently we validated a set of such combinations. Transcriptome profiling of mouse GBMs with precisely controlled driver combinations revealed distinct expression signatures between genotypes and in response to temozolomide (TMZ) treatment. Using 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.

<sup>1</sup>Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>2</sup>Systems Biology Institute, Yale University School of Medicine, West Haven, Connecticut, USA. <sup>3</sup>Medical Scientist Training Program, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>4</sup>Biological and Biomedical Sciences Program, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>5</sup>Immunobiology Program, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>6</sup>Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>7</sup>Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. <sup>8</sup>Department of Chemistry, University of Basel, Basel, Switzerland. <sup>9</sup>Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>10</sup>Yale Center for Genome Analysis, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>11</sup>Koch Institute for Integrative Cancer Research, MIT, Cambridge, Massachusetts, USA. <sup>12</sup>Department of Biology, MIT, Cambridge, Massachusetts, USA. <sup>13</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>14</sup>Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. <sup>15</sup>Comprehensive Cancer Center, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>16</sup>Stem Cell Center, Yale University School of Medicine, New Haven, Connecticut, USA. <sup>17</sup>These authors contributed equally to this work. <sup>18</sup>These authors contributed equally to this work. Correspondence should be addressed to R.J.P. (rplatt@ethz.ch) or S.C. (sidi.chen@yale.edu).



**Figure 1** Autochthonous brain tumorigenesis induced by an AAV-mediated CRISPR library. **(a)** Schematics of direct *in vivo* AAV-CRISPR GBM screen design. Top: AAV-mTSG library design, synthesis and production. Bottom: 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 PBS- or AAV-vector-injected mice. Matching sections are shown. 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%; the curves are dashed and slightly offset for visibility; dpi, days postinjection. Log-rank (LR) test,  $P < 2.20 \times 10^{-16}$ , mTSG vs. vector or PBS.

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; Online Methods and **Supplementary Fig. 1a**). The vector also contains an sgRNA targeting *Trp53*, initially intended to generate co-mutational *Trp53* knockouts that might exhibit genome instability and thus be sensitized to tumorigenesis<sup>17–21</sup>. Local viral delivery into the brain restricts the number of transducible cells, and cancer genomes generally consist of dozens to hundreds of SMGs<sup>22–25</sup>. With these considerations in mind, we designed an sgRNA library (a mouse-homolog tumor suppressor gene (mTSG) library) targeting the mouse homologs of top-ranked pan-cancer SMGs (Online Methods), plus seven genes with essential molecular functions that we initially considered as internal controls (**Fig. 1a** and **Supplementary Table 1**). We pool-synthesized all sgRNAs, cloned them into the AAV-CRISPR vector at greater than 100 $\times$  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; **Fig. 1a** and **Supplementary Fig. 1b**). We generated high-titer AAVs ( $> 1 \times 10^{12}$  viral particles per mL) from the plasmid that contained the mTSG library (AAV-mTSG), as well as an empty vector (AAV-vector; **Fig. 1a**). We then stereotactically injected AAV-mTSG, AAV-vector or phosphate-buffered saline (PBS) into the lateral ventricle ( $n = 40$  mice) or hippocampus ( $n = 16$  mice) in the brains of LSL-Cas9 mice (Online Methods). We scanned the brains of these mice by MRI 4 months postinjection and found that half (9 of 18 = 50%)

of AAV-mTSG library-transduced animals had developed brain tumors at this time point, whereas none of the AAV-vector-injected or PBS-injected animals had tumors detectable by MRI (**Fig. 1b**, **Supplementary Fig. 1e** and **Supplementary Table 2**). Quantification of tumor volumes showed that AAV-mTSG-transduced mice had average tumor volumes of 70.2 mm<sup>3</sup> (including animals without tumors) or 140.3 mm<sup>3</sup> (excluding animals without a tumor; two-tailed Welch's *t* test,  $t_{17} = 2.62$ ,  $P = 0.018$ , all mTSG mice versus vector or PBS; **Fig. 1c** and **Supplementary Table 2**). These data suggested that the AAV-mTSG library robustly initiated tumorigenesis in the brains of LSL-Cas9 mice.

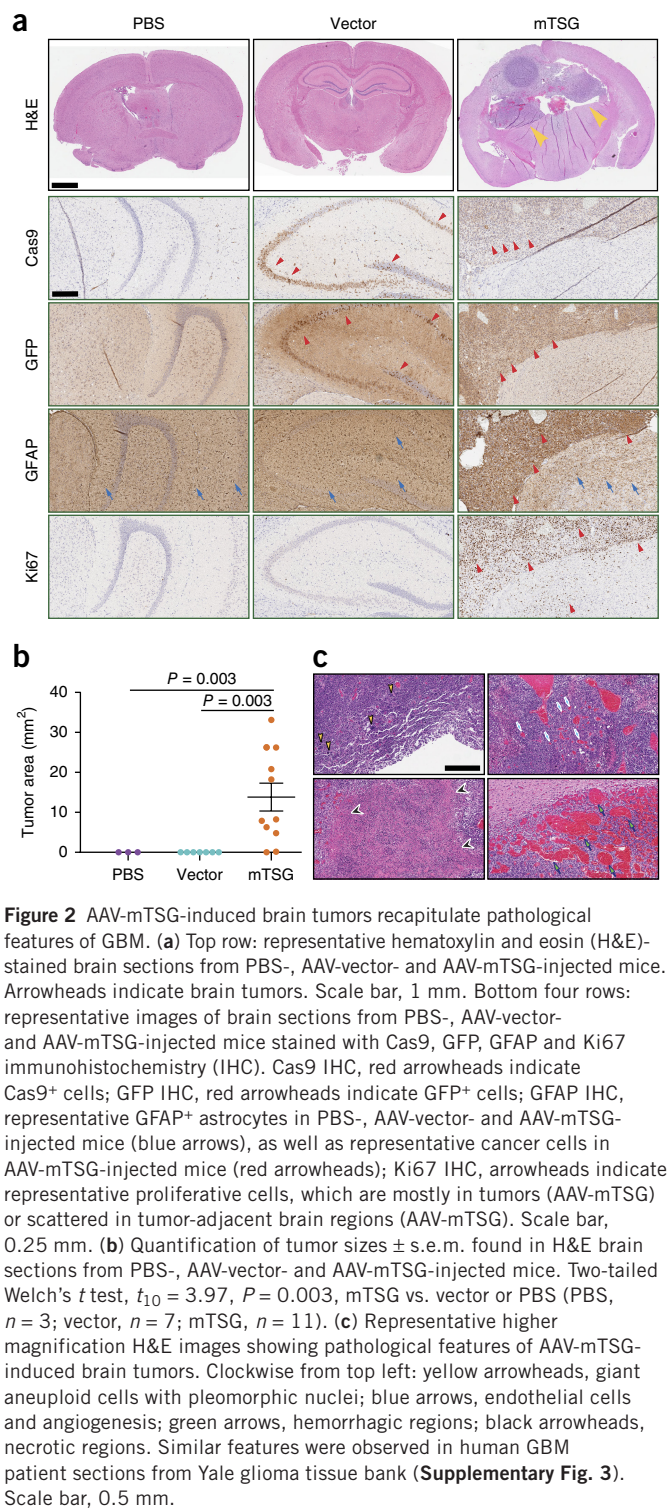
We analyzed the overall survival of a cohort of LSL-Cas9 mice injected with AAV-mTSG, AAV-vector or PBS (**Supplementary Table 3**). In this screen, injection location did not affect the rate of tumor development as reflected by overall survival (two-sided Mann-Whitney *U* test of hippocampus versus lateral ventricle,  $P = 0.054$ ; **Supplementary Table 3**) and thus were considered as one group (AAV-mTSG). For the AAV-mTSG-transduced group, the first three animals died 84 days postinjection, 90% of animals did not survive 176 d and all 56 AAV-mTSG-transduced animals reached their survival endpoints within 299 d (i.e., died or had a poor body condition score (BCS)  $< 2$  and thus were euthanized; **Fig. 1d** and **Supplementary Table 3**). The median survival time of the AAV-mTSG group was 129 d (95% confidence interval = 111–159 d; **Fig. 1d**), consistent with the presence of tumors in half of the mice at 4 months by MRI. In contrast, all 24 AAV-vector-injected and all 5 PBS-injected animals survived the duration of the study and maintained good body condition (BCS = 5; log-rank test,  $P < 2.2 \times 10^{-16}$ , mTSG versus vector or PBS;

log-rank test,  $P = 1$ , vector versus PBS; **Fig. 1d**), suggesting Trp53 knockout alone does not lead to gliomagenesis. For the vast majority (96.4% or 54 of 56) of AAV-mTSG-injected mice, macrocephaly was observed at the survival endpoint (**Supplementary Fig. 1c**), suggesting that they had developed brain tumors. On the contrary, macrocephaly was observed in none of the AAV-vector-injected (0 of 24) or PBS-injected (0 of 5) mice during the study (two-tailed Fisher's exact test,  $P < 1 \times 10^5$ , mTSG versus vector or PBS;  $P = 1$ , vector versus PBS). These data indicate that the brain tumors induced by the AAV-mTSG viral library were typically lethal.

We observed that AAV-mTSG mice had GFP<sup>+</sup> masses that deformed their brains (100% or 6 of 6; **Fig. 2a** and **Supplementary Fig. 1d**). AAV-vector mice had diffuse GFP<sup>+</sup> regions in the brain with fully normal morphology, suggesting that these were AAV-transduced cells expressing Cas9-GFP induced by Cre expression and that had not become tumors ( $n = 2$ ; **Fig. 2a** and **Supplementary Fig. 1d**). PBS-injected or uninjected mice had no detectable GFP expression even at long exposure ( $n = 3$ ; **Supplementary Fig. 1d**). Immunohistochemistry 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 (**Fig. 2a**). These tumors were also positive for GFAP, an astrocytic marker (**Fig. 2a**), and for Ki67, a proliferation marker (**Fig. 2a**). AAV-vector-transduced brains stained positive for Cas9 and GFP in a subset of cells at the injection site (**Fig. 2a**), but these cells were not proliferative (Ki67<sup>-</sup>) and did not have tumor-like pathological features (**Fig. 2a**). PBS-injected mice stained negative for Cas9, GFP and Ki67 (**Fig. 2a**). Endpoint histopathology showed that the vast majority of AAV-mTSG mice developed brain tumors (10 of 11 = 91%), whereas none of the AAV-vector (0 of 7 = 0%) or PBS (0 of 3 = 0%) mice had detectable tumors (two-tailed Fisher's exact test:  $P = 0.0003$ , mTSG versus vector;  $P = 0.011$ , mTSG versus PBS; **Fig. 2a,b** and **Supplementary Table 4**). The mean endpoint tumor size, as measured by area in brain sections, for the AAV-mTSG group was 13.9 mm<sup>2</sup>, compared to 0 mm<sup>2</sup> in the two control groups (two-tailed Welch's  $t$  test,  $t_{10} = 3.97$ ,  $P = 0.003$ , mTSG versus vector or PBS; **Fig. 2b**). Brain tumors in AAV-mTSG mice showed pathological features of dense cellular structure with proliferative spindles, nuclear aneuploidy and pleomorphism, giant cells, regions of necrosis, angiogenesis and hemorrhage (**Fig. 2c**), all of which are hallmark features of human GBM<sup>2</sup>. 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, Wright Giemsa, Masson and Alcian blue periodic acid–Schiff (**Supplementary Fig. 2**). We investigated a panel of human GBM clinical samples from the Yale Brain Tumor Program ([http://medicine.yale.edu/brain\\_tumor\\_research/](http://medicine.yale.edu/brain_tumor_research/)) and confirmed the observation of these pathological features (**Supplementary Fig. 3a,b**). These data suggest that the AAV-mTSG library-induced autochthonous brain tumors recapitulated features of human GBM.

### Targeted-capture sequencing reveals diverse mutational profiles across tumors

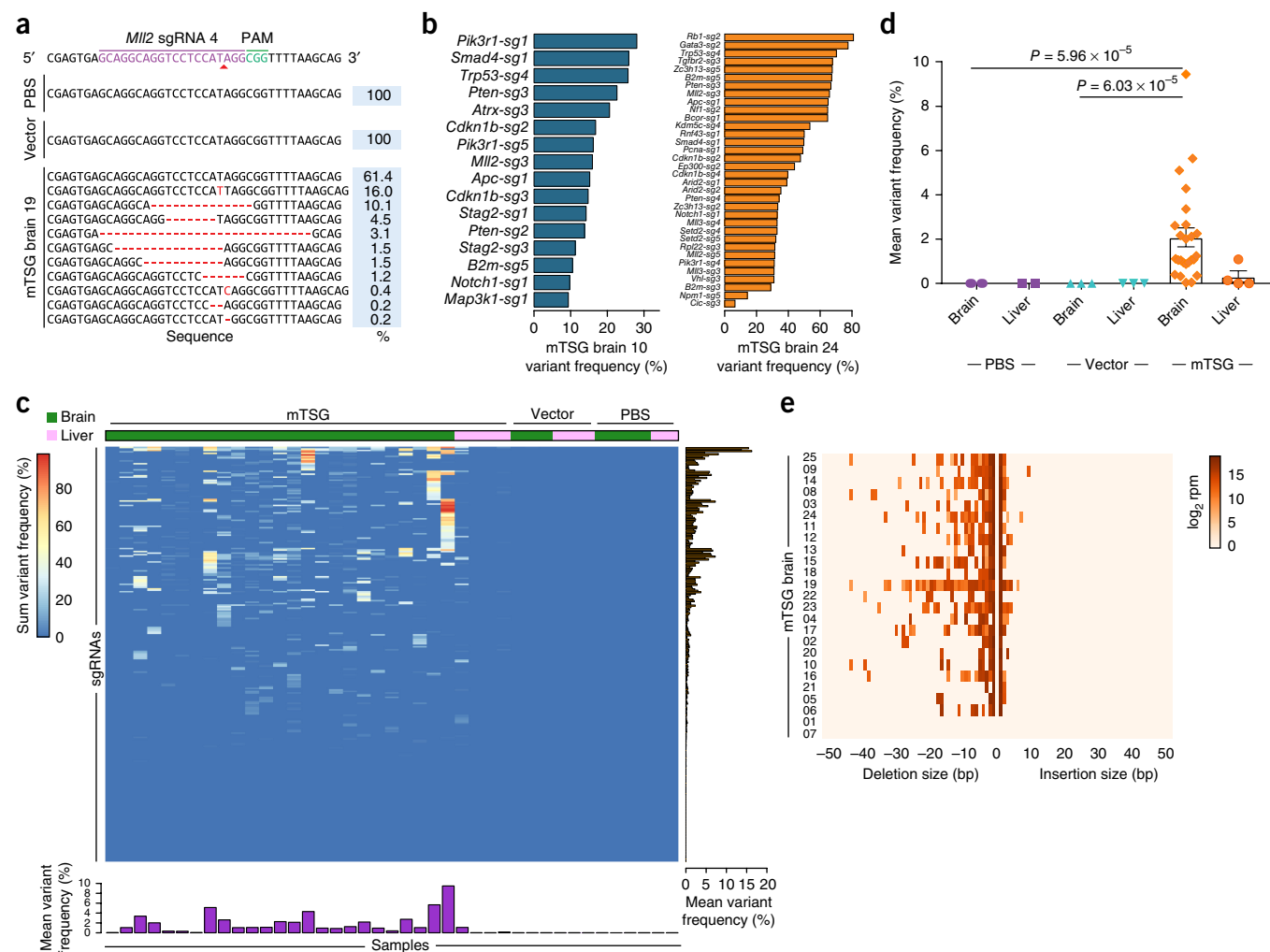
Because AAVs usually do not integrate into the genome, direct sequencing of the targeted regions was needed to determine which mutations were in each tumor. To map the molecular landscape of these brain tumors, we designed a probe set (mTSG-Amplicon probes) covering the target regions of all library sgRNAs (Online Methods and **Supplementary Table 5**). 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



**Figure 2** AAV-mTSG-induced brain tumors recapitulate pathological features of GBM. **(a)** Top row: representative hematoxylin and eosin (H&E)-stained brain sections from PBS-, AAV-vector- and AAV-mTSG-injected mice. Arrowheads indicate brain tumors. Scale bar, 1 mm. Bottom four rows: representative images of brain sections from PBS-, AAV-vector- and AAV-mTSG-injected mice stained with Cas9, GFP, GFAP and Ki67 immunohistochemistry (IHC). Cas9 IHC, red arrowheads indicate Cas9<sup>+</sup> cells; GFP IHC, red arrowheads indicate GFP<sup>+</sup> cells; GFAP IHC, representative GFAP<sup>+</sup> 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  $\pm$  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 (**Supplementary Fig. 3**). Scale bar, 0.5 mm.

AAV-mTSG-, AAV-vector- and PBS-injected mice ( $n = 25$ , 3 and 4 brain samples, respectively; **Supplementary Table 6**). We captured 277 of 278 (99.6%) unique sgRNA target regions for all samples from this experiment, the exception being *Arid1a*-sg5 (due to the unavailability of qualified regions in capture probe design). Across all 41 brain and liver samples, the average of mean coverage across all probes was  $19,405 \pm 180$  s.e.m. (**Supplementary Table 7**). We analyzed the mutant variants at the predicted cutting sites of the 277 successfully



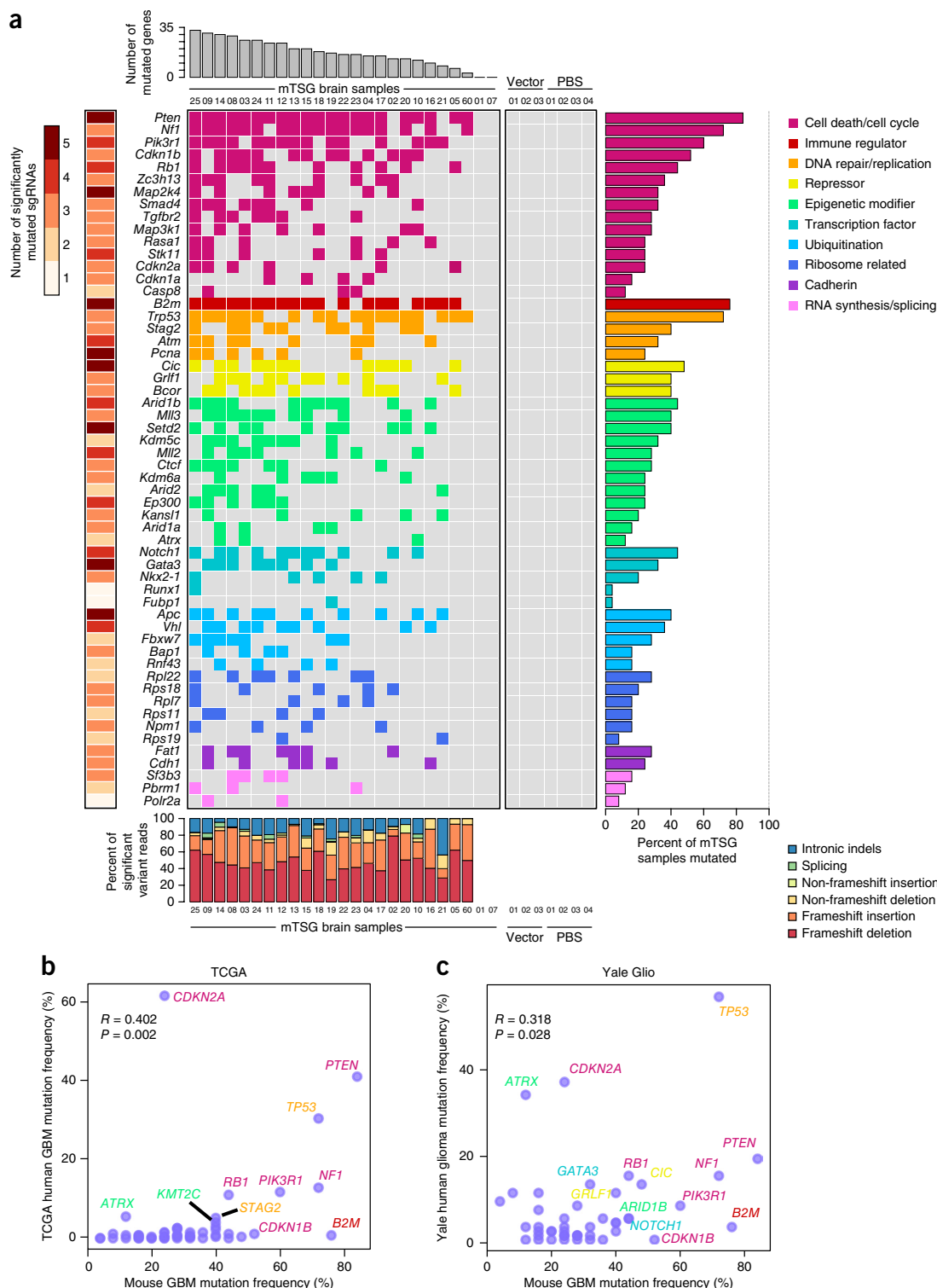


**Figure 3** Targeted-capture sequencing of sgRNA sites in AAV-mTSG-induced mouse GBM. (a) Indel variants observed at the genomic region targeted by *Mli2* sgRNA-4 in representative PBS-, AAV-vector- and AAV-mTSG-injected mouse brain samples. (b) Bar plots of variant frequencies in significantly mutated sgRNA target regions from two representative AAV-mTSG-injected mouse brain samples. (c) Heat map of variant frequency across all targeted capture samples ( $n = 41$ ). Rows denote individual sgRNAs, while columns correspond to samples from mice stereotactically injected with PBS, AAV-vector or AAV-mTSG. The liver was considered an off-target organ and thus was used as a background control. Of note, one liver sample contains modest levels of indels. Bar plots of the mean variant frequencies for each sgRNA (right, orange bars) and each sample (bottom, purple bars) are shown. (d) Dot plot of mean variant frequency  $\pm$  s.e.m., grouped by treatment condition and tissue type. AAV-mTSG-injected brains had significantly higher mean variant frequencies ( $2.087 \pm 0.429$ ,  $n = 25$ ) compared to vector ( $0.005 \pm 0.001$ ,  $n = 3$ ) or PBS ( $0.003 \pm 0.001$ ,  $n = 4$ ) injected brains (two-tailed Welch's  $t$  test,  $t_{24} = 4.85$  and  $t_{24} = 4.86$ ,  $P = 6.03 \times 10^{-5}$  and  $P = 5.96 \times 10^{-5}$  for mTSG vs. vector and mTSG vs. PBS, respectively). In AAV-mTSG-injected mice, mean variant frequencies of brains ( $2.087 \pm 0.429$ ) were significantly higher than those of livers ( $0.309 \pm 0.261$ ,  $n = 4$ ;  $t_{21.48} = 3.54$ ,  $P = 0.002$ ). (e) Indel size distribution for all filtered variants in each mTSG brain sample ( $n = 25$ ).

captured unique sgRNAs across all samples (Online Methods and **Supplementary Table 8**). At the single-sgRNA level, for example, at the predicted cutting site of sgRNA-4 (sg4) in the *Mli2* locus (also known as *Kmt2d*), various insertions and deletions (indels) were detected in AAV-mTSG but not AAV-vector or PBS mice (**Fig. 3a**). As gliomagenesis takes multiple months in mice (**Fig. 1d**), we surveyed three mice at 3.5 weeks postinjection and performed capture sequencing to reveal early mutation profiles, as an approximation for *in vivo* sgRNA cutting efficiency (**Supplementary Fig. 4a,b** and **Supplementary Table 9**). We found that even lower-efficiency sgRNAs could end up being highly enriched in the process of tumorigenesis if the mutations they generated were strongly oncogenic (**Fig. 3b,c** and **Supplementary Fig. 4c**). After removing germline variants, we determined whether the regions flanking each sgRNA target site would be classified as significantly mutated sgRNA sites (SMSs;

Online Methods and **Supplementary Table 10**). We implemented a false-discovery rate (FDR) approach based on the 12 PBS/vector samples (FDR  $< 1/12$ , or 8.33%) as well as a flat 5% variant-frequency cutoff, and we confirmed that the choice of alternative cutoffs did not alter the final SMS calls (**Supplementary Fig. 6a**). With these criteria, we observed a diverse mutational landscape across most mice that were capture-sequenced (**Fig. 3b,c**, **Supplementary Fig. 5** and **Supplementary Table 11**). As an example, one AAV-mTSG mouse (mTSG brain 10) had significant mutations at the predicted cutting sites of 16 of 277 captured gene-targeting sgRNAs in the mTSG library, covering 12 significantly mutated genes (mouse SMGs, or mSMGs; **Fig. 3b**). A second example (mTSG brain 24) showed a more diverse mutational profile (34 SMSs for 26 mSMGs; **Fig. 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





**Figure 4** Integrative analysis of functional mutations in driving tumorigenesis. **(a)** Gene-level mutational landscape of AAV-mTSG induced primary mouse GBM. Top: bar plots of the total number of significantly mutated genes identified in each AAV-mTSG sample. 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. Right: bar plots of the percentage of GBM samples that were called as significantly mutated for each gene. Left: heat map of the numbers of unique SMSs for each gene. Bottom: stacked bar plots describing the types 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 Glioma; Pearson correlation  $R = 0.318$ ,  $P = 0.028$ ).

of variant frequencies generated by this sgRNA pool (Fig. 3c and Supplementary Table 10). Comparing brain samples between treatment groups, AAV-mTSG-injected brains had significantly higher mean variant frequencies ( $2.087 \pm 0.429$  s.e.m.,  $n = 25$ ) compared to vector-injected ( $0.005 \pm 0.001$ ,  $n = 3$ ) or PBS-injected ( $0.003 \pm 0.001$ ,  $n = 4$ ) brains (two-tailed Welch's  $t$  test,  $t_{24} = 4.85$  and  $t_{24} = 4.86$ ,  $P = 6.03 \times 10^{-5}$  and  $P = 5.96 \times 10^{-5}$  for mTSG versus vector and mTSG versus PBS, respectively; Fig. 3c,d). Comparing targeted versus nontargeted organs in AAV-mTSG-injected mice, the mean variant frequencies of brains ( $2.087 \pm 0.429$ ,  $n = 25$ ) were significantly higher than those of livers ( $0.309 \pm 0.261$ ,  $n = 4$ ; two-tailed Welch's  $t$  test,  $t_{21.48} = 3.54$ ,  $P = 0.002$ ; Fig. 3c,d). The predominant indels were deletions for virtually all samples, and most insertions at SMS sites were 1 bp in size (Fig. 3e and Supplementary Table 8). We identified distinct variant-frequency clusters of sgRNA-induced indels that may serve as an approximation to the clonality of these tumors (Online Methods). From this analysis, we found that only 2 of 22 of the brains had single-cluster tumors, with the majority (20 of 22) comprising multiple clusters (Supplementary Fig. 6b). 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 (Supplementary Table 12) and created an oncomap of all mTSG brain samples (Fig. 4a). Across all mice with mSMGs (23 of 25), the detected variants were predominantly frameshift indels (frameshift reads per total variant fraction > 60% in 22 of 23 mice) compared to non-frameshift indels, splicing indels and intronic indels (Fig. 4a, bottom panel). Unexpectedly, all 56 genes had at least one associated SMS, and eight of them (*Pten*, *Map2k4*, *B2m*, *Pcna*, *Cic*, *Setd2*, *Gata3* and *Apc*) had all 5 SMSs (Fig. 4a). The mSMGs 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 (Fig. 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*, *Mill3*, *Setd2*, *Mill2*, *Kdm5c*, *Kdm6a*, *Arid2* and *Ctcf* (Fig. 4a), highlighting the role of epigenetic regulators in brain tumorigenesis. Notably, *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 of 25 = 76% of mice; Fig. 4a). 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 nonsilent 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 patients with GBM in The Cancer Genome Atlas (TCGA) (Pearson correlation  $R = 0.402$ ,  $P = 2.1 \times 10^{-3}$ ; Fig. 4b and Supplementary Table 13). To further investigate this correlation, we used the clinical cancer genomics data of a cohort of patients (Yale Glio) from the Yale Brain Tumor Program, a source independent of TCGA (Supplementary Tables 14 and 15). Collectively, the mouse mutation frequencies again significantly correlated with those in human patients ( $R = 0.318$ ,  $P = 0.0277$ ; Fig. 4c and Supplementary Table 16). 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

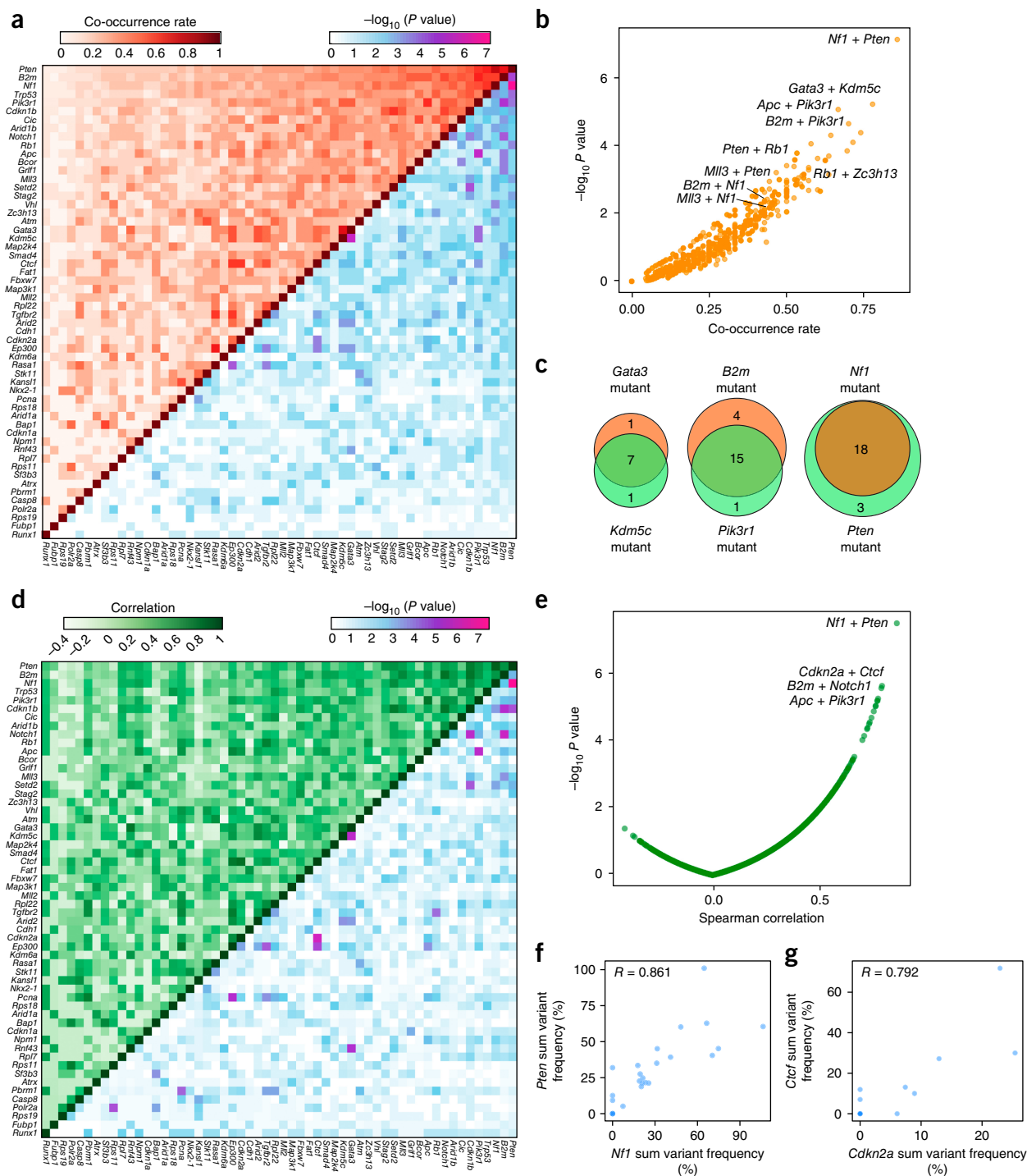
To generate an unbiased map of co-drivers, we calculated the co-occurrence rate of double mutations for each gene pair (Fig. 5a,b, Online Methods and Supplementary Table 17). This analysis showed that 76 gene pairs out of a total of 1,540 possible pairs were statistically significant in terms of co-occurrence (hypergeometric test, FDR adjusted  $q < 0.05$ ). The *Nf1*+*Pten* pair emerged as the top pair (co-occurrence rate = 18 of 21 = 85.7%, hypergeometric test,  $P = 7.53 \times 10^{-8}$ ; Fig. 5a–c). Notably, several previously undocumented combinations emerged, such as *Kdm5c*+*Gata3* (co-occurrence rate = 77.8%, hypergeometric test,  $P = 6.04 \times 10^{-6}$ ) and *B2m*+*Pik3r1* (70.0%,  $P = 2.28 \times 10^{-5}$ ; Fig. 5a–c). In addition, we performed correlation analysis of summed mutant frequencies for each pair of genes across all mice (Fig. 5d,e, Online Methods and Supplementary Table 18). We found that 22.9% (352 of 1,540) of the gene pairs were positively correlated (Spearman correlation > 0, FDR adjusted  $q < 0.05$ ; Fig. 5d,e). The most significantly correlated gene pair was again *Nf1*+*Pten* (Spearman correlation  $r = 0.861$ ,  $P = 3.34 \times 10^{-8}$ ; Fig. 5d–f and Supplementary Table 18), along with other representative pairs such as *Cdkn2a*+*Ctcf* (correlation  $r = 0.792$ ,  $P = 2.41 \times 10^{-6}$ ; Fig. 5g), *B2m*+*Notch1* (correlation  $r = 0.789$ ,  $P = 2.82 \times 10^{-6}$ ) and *Apc*+*Pik3r1* (correlation  $r = 0.774$ ,  $P = 5.77 \times 10^{-6}$ ; Fig. 5d,e). Exclusion of *Trp53* revealed largely identical results for the remaining genes (Supplementary Fig. 7a,b). Of note, a subset of the significantly co-occurring pairs were also found to be co-mutated in human GBM, including *RB1*+*TP53*, *PTEN*+*RB1*, *RASA1*+*STK11*, *B2M*+*MAP2K4*, *PTEN*+*STAG2*, *CDKN1B*+*TP53* and *CDKN1B*+*NF1* (Supplementary Fig. 7c,d). These data revealed the 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 (Fig. 6a and Online Methods). All of the uninjected ( $n = 2$ ), EYFP ( $n = 4$ ) and empty-vector ( $n = 3$ ) mice survived and maintained good body condition for the whole duration of the study. Control mice examined 4–11 months after treatment were devoid of any observable tumors by histology analysis (Fig. 6b, Supplementary Fig. 8 and Supplementary Table 19), again indicating that without mutagenesis, or with *Trp53* disruption alone, LSL-Cas9 mice did not develop brain tumors. In contrast, within 11 months postinjection, 50% (4 of 8) of mice receiving AAVs containing *Nf1* sgRNA minipool developed macrocephaly, poor BCS and tumors (compared to all 9 control mice; two-tailed Fisher's exact test,  $P = 0.029$ ). All mice receiving *Nf1*+*Pten* (9 of 9, 100%,  $P = 4.11 \times 10^{-5}$ ) and *Nf1*+*B2m* minipools (4 of 4, 100%,  $P = 0.0014$ ) developed macrocephaly, poor BCS and large tumors (Fig. 6c and Supplementary Table 19). Notably, mice receiving *Nf1*+*B2m* minipools had significantly worse survival times than mice receiving *Nf1* minipools alone ( $P = 0.0067$ ; Supplementary Fig. 9a), 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 of 3, 100%,  $P = 0.0045$  for all three groups) developed macrocephaly, poor BCS and large tumors (Fig. 6d and Supplementary Table 19). For the same duration of study (maximum 11 months), smaller fractions of mice receiving the AAV sgRNA minipools targeting *Arid1b*+*Nf1* (4 of 9), *Mill3*+*Nf1* (2 of 5), *Mill2* (2 of 10), *Cic* (1 of 5), *Cic*+*Pten*

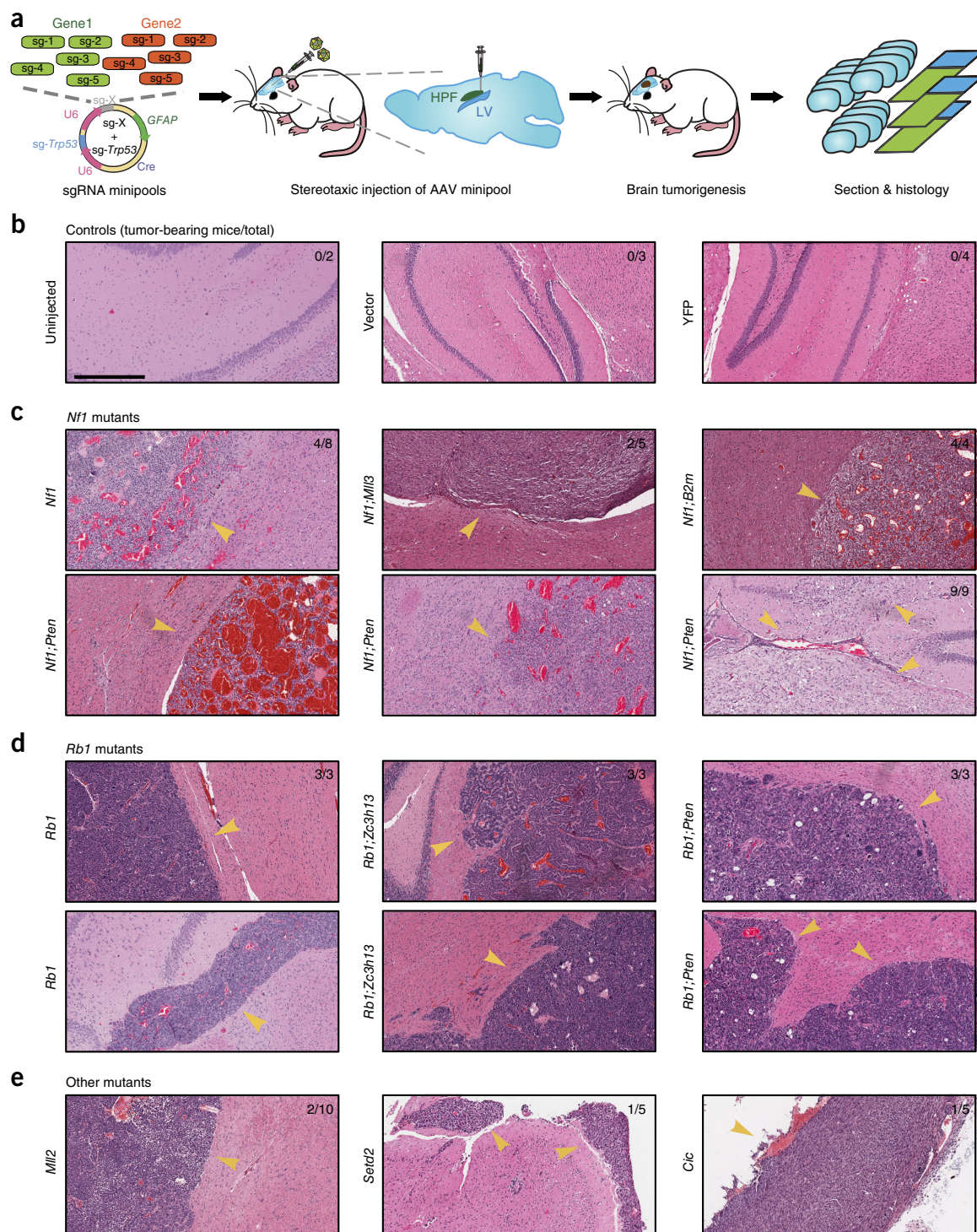
(1 of 4), *Setd2* (1 of 5) and *Gata3;Mll3* (1 of 5) developed tumors (Supplementary Table 19). Collectively, half (40 of 80, or 50%) of the mice receiving AAV sgRNA minipools targeting any of the single

genes or gene pairs developed brain tumors within 11 months (collective validation versus all controls, two-tailed Fisher's exact test,  $P = 0.004$ ). These data indicated that mutating these individual genes



**Figure 5** Co-mutation analysis uncovers synergistic gene pairs in GBM. **(a)** Top left: heat map of pairwise mutational co-occurrence rates. Bottom-right: heat map 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,  $P = 6.04 \times 10^{-6}$ ), *B2m* and *Pik3r1* (70.0%,  $P = 2.28 \times 10^{-5}$ ), and *Nf1* and *Pten* (85.7%,  $P = 7.53 \times 10^{-8}$ ). **(d)** Top-left: heat map of the pairwise Spearman correlation of variant frequency for each gene, summed across sgRNAs. Bottom right: heat map 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 stereotactically injected into the lateral ventricle of LSL-Cas9 mice. **(b–e)** Endpoint histology (H&E) of representative brain sections from mice treated with AAV sgRNA minipools or relevant controls. In this endpoint analysis, mice were killed for analysis when either macrocephaly or poor body condition (BCS < 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 of 3), EYFP (0 of 4) or uninjected (0 of 2) groups. **(c)** Representative histology of brain sections from mice treated with various *Nf1* minipools, such as *Nf1* alone (4 of 8 mice developed tumors within 11 months), *Nf1;Pten* (9 of 9 mice developed tumors within 6 months), *Nf1;Mli3* (2 of 5 mice developed tumors within 6 months) and *Nf1;B2m* (4 of 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 of 3 mice developed tumors within 6 months), *Rb1;Zc3h13* (3 of 3 mice developed tumors within 6 months) and *Rb1;Pten* (3 of 3 mice developed tumors within 6 months). **(e)** Representative histology of brain sections from mice treated with other minipools, such as *Mli2* alone (2 of 10 mice developed tumors within 11 months), *Setd2* (1 of 5 mice developed tumors within 6 months) and *Cic* (1 of 5 mice developed tumors within 6 months).

or combinations in combination with *Trp53* causes GBM in fully immunocompetent animals.

Notably, brain tumors with *Nf1* mutations displayed highly polymorphic pathological features, with diverse fibroblastic cell morphologies, regions of necrosis and large hemorrhages (Fig. 6c) yet were almost always GFAP<sup>+</sup> (Supplementary Fig. 8). 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 pleomorphism but rarely with regions of necrosis or large-area hemorrhage (Fig. 6d), and they often contained mixtures of GFAP<sup>+</sup> and GFAP<sup>-</sup> cells (Supplementary Fig. 8).

### Transcriptomic characterization of tumors with differing mutational backgrounds

We then investigated the molecular underpinnings of gliomagenesis driven by different combinations of drivers (Fig. 7a and Supplementary Fig. 9b). 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) (Supplementary Tables 20 and 21). 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 (Fig. 7b and Supplementary Table 22). 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, biological adhesion, neuron differentiation, hormone metabolic process, cell motion and cell-cell signaling (Fig. 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 unexpectedly included regionalization, anterior-posterior pattern formation, transcription factor activity, embryonic morphogenesis, cell adhesion, extracellular matrix, neuron differentiation and GTPase regulator activity (Fig. 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 (Fig. 7e and Supplementary Table 23). 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 (Fig. 7f). Comparing *Rb1;Zc3h13* to *Rb1* cells revealed 703 upregulated and 166 downregulated genes (Fig. 7g and Supplementary Table 24). *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 (Fig. 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 important. 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

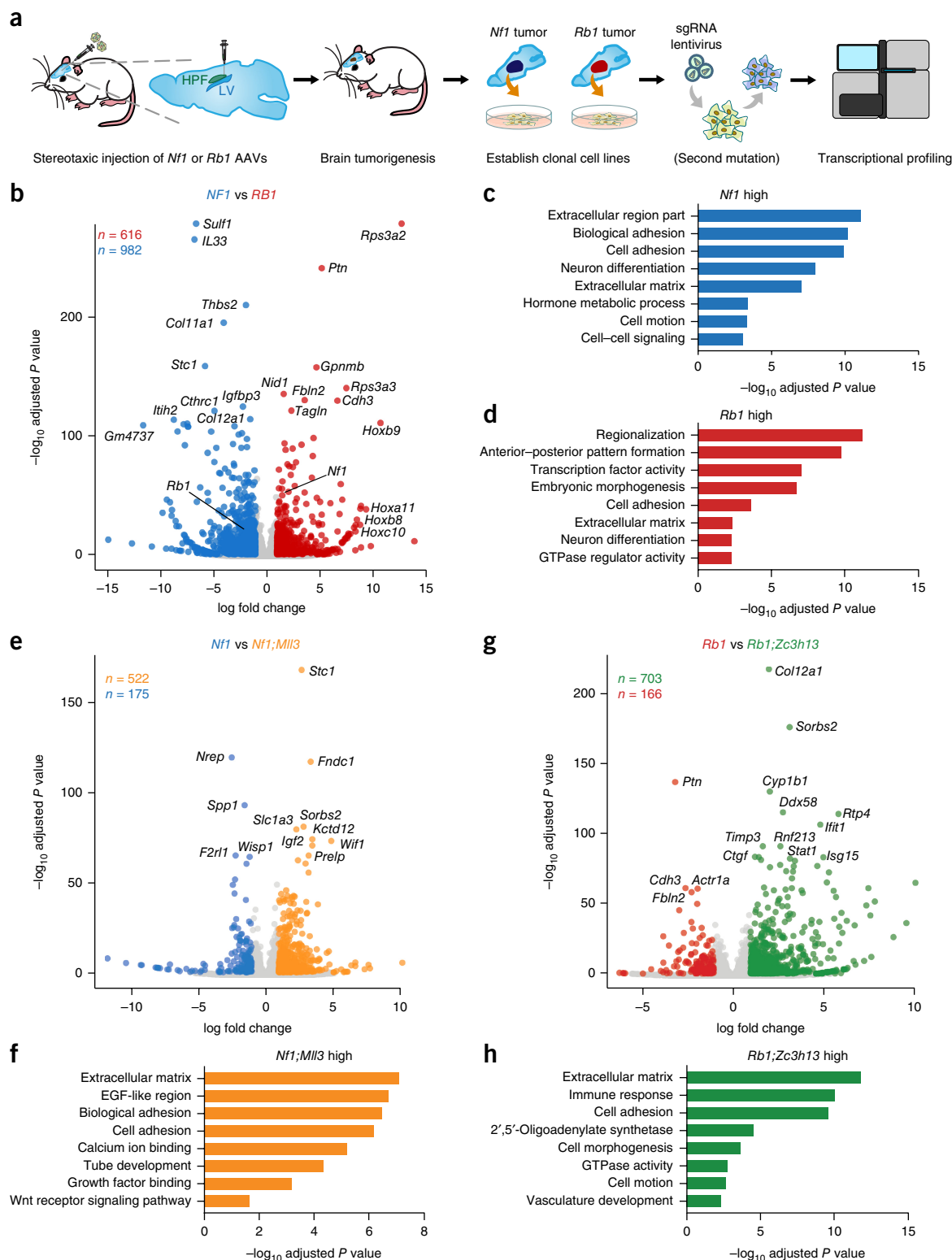
only four currently approved drugs for this disease (Online Methods and Fig. 8a). Drug-response phenotyping showed that *Zc3h13* loss-of-function rendered *Rb1* cells significantly more resistant to 1 mM TMZ, similarly to *Pten* loss-of-function (two-tailed *t* test, *t*<sub>4</sub> = 31.32 and *t*<sub>4</sub> = 23.51, *P* = 6.20 × 10<sup>-6</sup> and *P* = 1.94 × 10<sup>-5</sup>, for *Rb1;Pten* versus *Rb1* and *Rb1;Zc3h13* versus *Rb1*, respectively; Fig. 8b). These differences were also observed with 2 mM TMZ (*t*<sub>4</sub> = 50.69 and *t*<sub>4</sub> = 38.10, *P* = 9.06 × 10<sup>-7</sup> and *P* = 2.84 × 10<sup>-6</sup>, for *Rb1;Pten* versus *Rb1* and *Rb1;Zc3h13* versus *Rb1*, respectively; Fig. 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 dimethyl sulfoxide (DMSO)-treated controls (Supplementary Tables 20 and 21). Differential expression analyses of TMZ- and DMSO-treated cells from each of the three genotypes revealed systematic changes in gene expression (Fig. 8d-f and Supplementary Tables 25-27). Collectively, the differentially expressed genes in the TMZ versus DMSO comparisons uncovered the transcriptomic differences between genotypes in response to TMZ treatment (Fig. 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 (Fig. 8h and Supplementary Table 28), 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 (Fig. 8h and Supplementary Table 28). As for the genes that were significantly induced by TMZ, a total of 42 genes were common in all three genotypes (Fig. 8i and Supplementary Table 29), representing a shared TMZ-induced gene signature. Notably, 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<sup>26</sup>, and *Cd274* (also known as *PD-L1*), which encodes the ligand for an inhibitory receptor, PD-1, that is currently a major focus of investigation in cancer immunotherapy<sup>27</sup>. Taken together, the transcriptomic analyses provide unbiased molecular signatures underlying the increased TMZ-resistance upon *Zc3h13* or *Pten* mutations in *Rb1*-mutant glioma cells, suggesting that the precise combinations of mutational drivers present in individual GBMs directly influence therapeutic responses.

### DISCUSSION

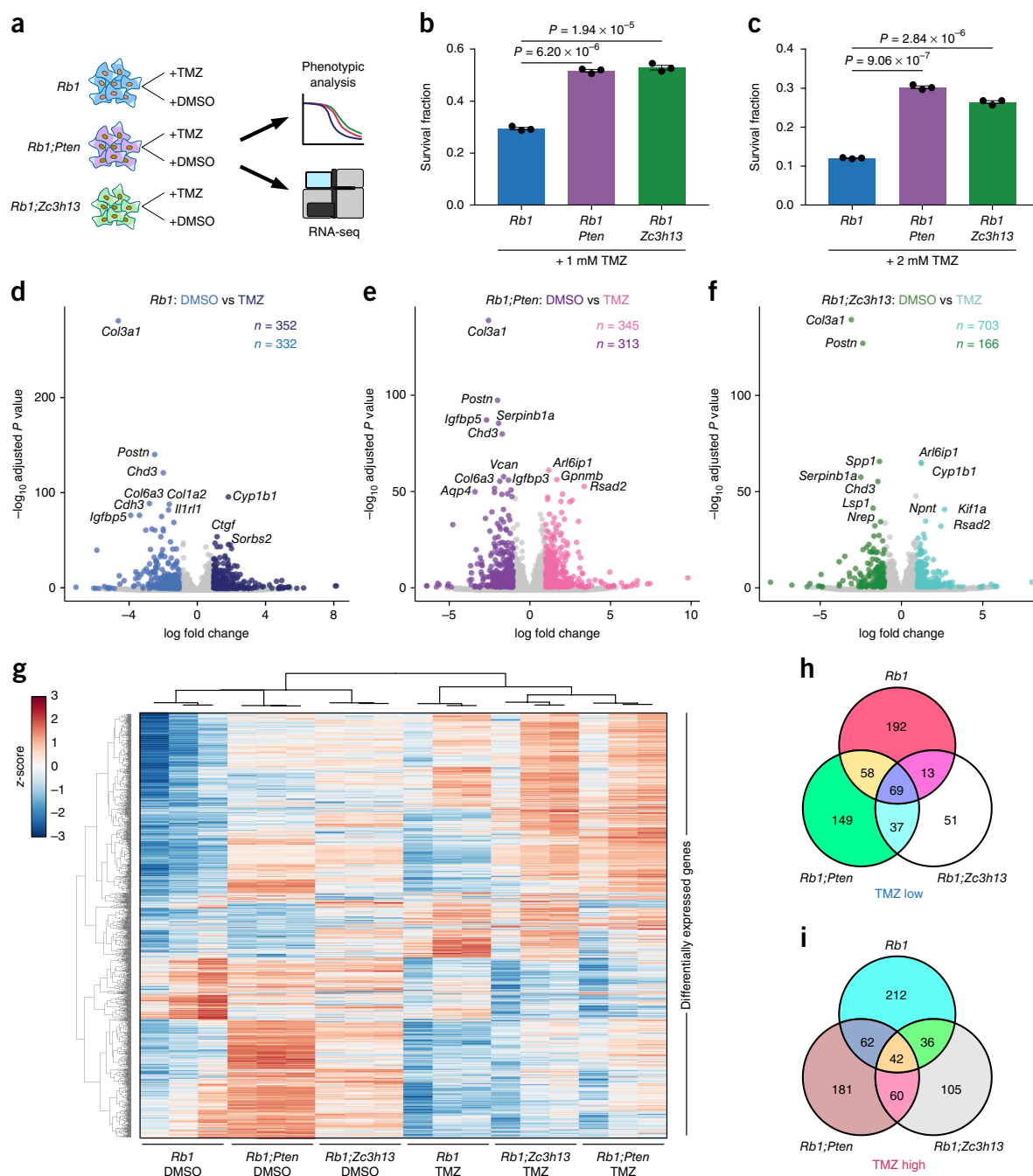
A systematic phenotypic picture of which genetic factors and combinations are necessary or sufficient to drive tumorigenesis in the brain is critical to better understanding GBM. To answer these questions, it is critical to directly test the hypotheses gleaned from cancer genomics in a controlled experimental 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<sup>28</sup>. 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 driven by a variety of important mutations<sup>29-40</sup>. Recently, the CRISPR system<sup>41-43</sup> has been used to directly mutate oncogenes and tumor suppressor genes in somatic cells for modeling genetic events in various cancer types (reviewed in ref. 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





**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;MII3* mutant (orange) to *Nf1*-mutant (blue) GBM cells. 522 genes were significantly higher in *Nf1;MII3* cells, and 175 genes were significantly higher in *Nf1* cells. **(f)** Enriched gene ontology categories among *Nf1;MII3*-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  $\geq 1$  or  $\leq -1$ .





**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 TMZ or DMSO. Cells were then subjected to phenotypic analysis and RNA-seq. **(b,c)** Survival fraction  $\pm$  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,  $P = 6.20 \times 10^{-6}$  and  $P = 1.94 \times 10^{-5}$ , respectively). **(c)** *Rb1;Pten* and *Rb1;Zc3h13* cells had significantly higher survival fractions than *Rb1* cells with 2 mM TMZ (two-sided  $t$  test,  $P = 9.06 \times 10^{-7}$  and  $P = 2.84 \times 10^{-6}$ , respectively). **(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)** Heat map 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 three 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 three groups upon TMZ treatment, there exist differential responses to TMZ treatment. Differentially expressed genes were defined as Benjamini-Hochberg adjusted  $P < 0.05$  and log fold change  $\geq 1$  or  $\leq -1$ .

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 an unbiased measurement of other coding mutations (**Supplementary Fig. 7e** and **Supplementary Tables 30** and **31**). These mutations might be caused by off-target effects of CRISPR/Cas9, rare AAV integrations<sup>45</sup> or by the unstable genomes of GBM cells, especially those with *p53* loss<sup>20</sup>, throughout tumorigenesis.

Across all genes tested, the mutation frequencies in this highly complex mouse model of GBM significantly correlated with the mutation frequencies in human patients from two large independent cohorts (TCGA and Yale Glio), suggesting the clinical relevance of the findings. Several of the novel mSMGs 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*<sup>11,46–49</sup>. Because differences in driver mutations can dramatically affect treatment efficacies in preclinical animal models and in human patients<sup>50</sup>, a functional understanding of cancer drivers is therefore essential for precision medicine.

We also performed a lentiCRISPR direct *in vivo* screen in GBM, using the same mTSG library (Online Methods, **Supplementary Table 32** and **Supplementary Fig. 9c,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 versus 200 d), survival (median survival, 4 versus 10 months) and penetrance (100% versus 67%). However, AAVs usually do not integrate into the genome, except under certain circumstances (low rate of integration at *AAVS1* locus)<sup>45</sup>. 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 ultradeep 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*.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank all members of the Chen, Sharp, Zhang and Platt laboratories, as well as our colleagues in the Yale Department of Genetics, Systems Biology Institute, Yale Cancer Center and Stem Cell Center, Koch Institute and Broad Institute at MIT for assistance and/or discussions. We thank the Center for Genome Analysis, Center for Molecular Discovery, High Performance Computing Center, West Campus Analytical Chemistry Core and West Campus Imaging Core and Keck Biotechnology Resource Laboratory at Yale, as well as Swanson Biotechnology Center at MIT, for technical support. S.C. is supported by Yale SBI/Genetics Startup Fund, Damon Runyon (DRG-2117-12; DFS-13-15), Melanoma Research

Alliance (412806, 16-003524), St. Baldrick's Foundation (426685), American Cancer Society (IRG 58-012-54), Breast Cancer Alliance, Cancer Research Institute (CLIP), AACR (499395), DoD (W81XWH-17-1-0235) and NIH/NCI (1U54CA209992, 5P50CA196530-A10805, 4P50CA121974-A08306). R.J.P. is supported by NCCRMSE and ETH Zurich, the McGovern Institute and NSF (1122374). P.A.S. is supported by NIH (R01-CA133404, R01-GM034277, CCNE), Skoltech Center and the Casimir-Lambert Fund. F.Z. is supported by the NIH/NIMH (5DP1-MH100706 and 1R01-MH110049), NSF, NY Stem Cell Foundation, HHMI, Poitras, Simons, Paul G. Allen Family, Vallee Foundations, D.R. Cheng and B. Metcalfe. C.D.G. and P.R. are supported by an NIH Graduate Training Grant (T32GM007499). R.D.C., M.B.D. and M.W.Y. are supported by an NIH MSTP training grant (T32GM007205). F.S. is supported by NCCRMSE and ETH Zurich. G.W. is supported by RJ Anderson and CRI Irvington Postdoctoral Fellowships.

## AUTHOR CONTRIBUTIONS

S.C. and R.J.P. conceived the study and performed the initial set of experiments. R.D.C. developed the algorithms and performed integrative analyses of all the data. C.D.G. performed validation, performed histology and established primary cell lines. G.W. performed exome-capture, mutant cell line generation, drug treatment and RNA-seq. F.S. performed AAV production. S.C. performed MRI. M.W.Y. contributed to data analysis. L.Y., Y.E., M.B.D., M.A.M., S.Z. and P.R. contributed to experiments including mouse breeding, genotyping, cloning, cell culture, virus prep, injection, necropsy and sample prep. K.B. assisted in captured and exome sequencing. M.G. provided clinical insights. P.A.S., F.Z., R.J.P. and S.C. jointly supervised the work. R.D.C. and S.C. wrote the manuscript with inputs from all authors.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Sturm, D. *et al.* Paediatric and adult glioblastoma: multifactorial (epi)genomic culprits emerge. *Nat. Rev. Cancer* **14**, 92–107 (2014).
2. Louis, D.N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007).
3. Krex, D. *et al.* Long-term survival with glioblastoma multiforme. *Brain* **130**, 2596–2606 (2007).
4. American Brain Tumor Association. *Glioblastoma and Malignant Astrocytoma* <http://www.abta.org/secure/glioblastoma-brochure.pdf> (2016).
5. Claus, E.B. & Black, P.M. Survival rates and patterns of care for patients diagnosed with supratentorial low-grade gliomas: data from the SEER program, 1973–2001. *Cancer* **106**, 1358–1363 (2006).
6. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
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.* **10**, 459–466 (2009).
8. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
9. Brennan, C.W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
10. Verhaak, R.G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
11. Bai, H. *et al.* Integrated genomic characterization of *IDH1*-mutant glioma malignant progression. *Nat. Genet.* **48**, 59–66 (2016).
12. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
13. Parsons, D.W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
14. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
15. Hainaut, P. & Plymoth, A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Curr. Opin. Oncol.* **25**, 50–51 (2013).
16. Cairncross, J.G. *et al.* Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. *J. Natl. Cancer Inst.* **90**, 1473–1479 (1998).
17. Napoli, M. & Flores, E.R. The p53 family orchestrates the regulation of metabolism: physiological regulation and implications for cancer therapy. *Br. J. Cancer* **116**, 149–155 (2017).
18. Muller, P.A. & Vousden, K.H. p53 mutations in cancer. *Nat. Cell Biol.* **15**, 2–8 (2013).

19. Feng, Z., Hu, W., Rajagopal, G. & Levine, A.J. The tumor suppressor p53: cancer and aging. *Cell Cycle* **7**, 842–847 (2008).
20. Levine, A.J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
21. Berns, A. Cancer: the blind spot of p53. *Nature* **468**, 519–520 (2010).
22. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
23. Lawrence, M.S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
24. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
25. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).
26. Guo, F., Li, Y., Liu, Y., Wang, J. & Li, G. ARL6IP1 mediates cisplatin-induced apoptosis in CaSki cervical cancer cells. *Oncol. Rep.* **23**, 1449–1455 (2010).
27. Ribas, A. Tumor immunotherapy directed at PD-1. *N. Engl. J. Med.* **366**, 2517–2519 (2012).
28. Frese, K.K. & Tuveson, D.A. Maximizing mouse cancer models. *Nat. Rev. Cancer* **7**, 645–658 (2007).
29. Holland, E.C. Gliomagenesis: genetic alterations and mouse models. *Nat. Rev. Genet.* **2**, 120–129 (2001).
30. Huse, J.T. & Holland, E.C. Genetically engineered mouse models of brain cancer and the promise of preclinical testing. *Brain Pathol.* **19**, 132–143 (2009).
31. Alcantara Llaguno, S. *et al.* Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer Cell* **15**, 45–56 (2009).
32. Friedmann-Morvinski, D. *et al.* Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. *Science* **338**, 1080–1084 (2012).
33. Marumoto, T. *et al.* Development of a novel mouse glioma model using lentiviral vectors. *Nat. Med.* **15**, 110–116 (2009).
34. Perry, A. *et al.* Malignant gliomas with primitive neuroectodermal tumor-like components: a clinicopathologic and genetic study of 53 cases. *Brain Pathol.* **19**, 81–90 (2009).
35. Reilly, K.M., Loisel, D.A., Bronson, R.T., McLaughlin, M.E. & Jacks, T. Nf1;Trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. *Nat. Genet.* **26**, 109–113 (2000).
36. Schmid, R.S., Vitucci, M. & Miller, C.R. Genetically engineered mouse models of diffuse gliomas. *Brain Res. Bull.* **88**, 72–79 (2012).
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.* **1210**, 37–49 (2014).
38. Agemy, L. *et al.* Targeted nanoparticle enhanced proapoptotic peptide as potential therapy for glioblastoma. *Proc. Natl. Acad. Sci. USA* **108**, 17450–17455 (2011).
39. Chow, L.M. *et al.* Cooperativity within and among Pten, p53, and Rb pathways induces high-grade astrocytoma in adult brain. *Cancer Cell* **19**, 305–316 (2011).
40. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522–526 (2012).
41. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
42. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
43. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
44. Sánchez-Rivera, F.J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat. Rev. Cancer* **15**, 387–395 (2015).
45. Weitzman, M.D., Kyöstiö, S.R.M., Kotin, R.M. & Owens, R.A. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. USA* **91**, 5808–5812 (1994).
46. Yeung, J.T. *et al.* LOH in the HLA class I region at 6p21 is associated with shorter survival in newly diagnosed adult glioblastoma. *Clin. Cancer Res.* **19**, 1816–1826 (2013).
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* **6**, 43452–43471 (2015).
48. Bale, T.A. *et al.* Genomic characterization of recurrent high-grade astroblastoma. *Cancer Genet.* **209**, 321–330 (2016).
49. Aithal, M.G. & Rajeswari, N. Validation of housekeeping genes for gene expression analysis in glioblastoma using quantitative real-time polymerase chain reaction. *Brain Tumor Res. Treat.* **3**, 24–29 (2015).
50. Friedman, A.A., Letai, A., Fisher, D.E. & Flaherty, K.T. Precision medicine for cancer with next-generation functional diagnostics. *Nat. Rev. Cancer* **15**, 747–756 (2015).



## ONLINE METHODS

**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<sup>51</sup> and Synapse (<https://www.synapse.org>). SMGs were calculated similar to previously described methods<sup>22–25</sup>. 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 seven known housekeeping genes were initially chosen as internal controls. We designed sgRNAs against these 56 genes using a previously described method<sup>52,53</sup> with our custom scripts. Five sgRNAs were chosen for each gene, plus eight nontargeting controls (NTCs), for a total of 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 expressing Cre recombinase under the control of a *GFAP* promoter for conditional induction of Cas9 expression in brain astrocytes when delivered to LSL-Cas9 mice<sup>54</sup>. Two sgRNA cassettes were built in this vector, one encoding an sgRNA targeting *Trp53*, the most frequently mutated gene in cancer<sup>22,23,25</sup>, and the other an empty sgRNA cassette (double SapI sites for sgRNA cloning) enabling flexible targeting of genes of interest in either an individual or a pooled manner. The vector was generated by gBlock gene fragment synthesis (IDT) followed by Gibson assembly (NEB). The mTSG libraries were generated by oligonucleotide 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 read out by barcoded Illumina sequencing as described previously<sup>55</sup> 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, HEK293FT cells (ThermoFisher) were transiently transfected with transfer (AAV-vector or AAV-mTSG), serotype (AAV9) and packaging (pDF6) plasmids using polyethylenimine (PEI). Each replicate consist of five 80% confluent HEK293FT cells in 15-cm tissue culture dishes or T-175 flasks (Corning). Multiple replicates were pooled to enhance production yield. Approximately 72 h post-transfection, cells were dislodged and transferred to a conical tube in sterile PBS. We added 1/10 volume of pure chloroform and the mixture was incubated at 37 °C and vigorously shaken for 1 h. NaCl was added to a final concentration of 1 M, and the mixture was shaken until dissolved and then pelleted at 20,000g at 4 °C for 15 min. 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 h and then spun at 20,000g at 4 °C for 15 min. The supernatant was discarded and the pellet was resuspended in DPBS + MgCl<sub>2</sub>, treated with benzonase (Sigma) and incubated at 37 °C for 30 min. Chloroform (1:1 volume) was then added, shaken and spun down at 12,000g at 4 °C 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 expressing Cre recombinase under the control of an *EFS* promoter or a *GFAP* promoter for conditional induction of Cas9 expression in the brain when delivered to LSL-Cas9 mice. Two sgRNA cassettes were built in this vector, one encoding an sgRNA targeting *Trp53* and the other 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 oligonucleotide 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 read out by barcoded Illumina sequencing as described above, with primers customized to the vectors. The lentiCRISPR GBM plasmid vector (lentiCRISPR) and library (lenti-mTSG) were subjected to high-titer lentivirus production and purification. Briefly, HEK293FT cells (ThermoFisher) were transiently transfected with transfer (lentiCRISPR or lenti-mTSG) and packaging (psPAX and pMD2.G) plasmids using PEI or Lipofectamine. Each replicate consist of five 80% confluent HEK293FT cells in 15-cm tissue culture dishes or T-175 flasks (Corning). Multiple replicates were pooled to enhance production yield. Approximately 48 h post-transfection, virus-containing media was collected and purified via sucrose gradient ultracentrifugation at ≥ 30,000 rpm for 2–3 h. 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). The experiments in **Supplementary Figure 9** were done using the EFS version of the vector.

**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 was 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- to 14-week-old mice were used in experiments. Animals were maintained and bred 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 12-h:12-h light cycle. Mice were anesthetized by intraperitoneal injection of ketamine (100 mg/kg) and xylazine (10 mg/kg) or by inhalation of isoflurane at approximately 2% for 20–30 min. 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 was provided by a standard heating pad or a heat lamp. According to the mouse brain stereotaxic coordinates<sup>56</sup>, we drilled a 1–2 mm hole on the surface of the skull and used a 33-gauge 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). We injected PBS, AAV (1 × 10<sup>10</sup> to 1 × 10<sup>11</sup> viral genome copies or Cre copy-number equivalent) or lentivirus (8 × 10<sup>9</sup> to 8 × 10<sup>10</sup> viral particles or p24 equivalent) into the right hemisphere of the brain for each mouse (8 μL for LV, 2 μL for HPF). 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 VetBond tissue glue (3M). Animal were postoperatively hydrated with 1 mL lactated Ringer's solution (subcutaneous) and housed in 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) = 2,000, slicing = 0.5 mm. Raw image stacks were processed using Osirix or Slicer tools<sup>57</sup>. Rendering and quantification were performed using Slicer (<https://www.slicer.org/>). For all mice with brain tumors, only one tumor was observed per mouse. Tumors were approximated as spheres and their sizes were calculated with the following formula:

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

**Survival analysis.** We observed that mice that developed brain tumors rapidly deteriorated in their body condition scores. Mice with observed macrocephaly and body condition score  $\leq 1$  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 stereotactically injected with PBS, AAV-vector or AAV-mTSG virus were monitored for their survival. Survival analysis was analyzed using a standard Kaplan-Meier analysis. Of note, several AAV-vector or PBS injected mice were killed for analysis at time points earlier than 299 d (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 killed for analysis early in healthy states were excluded from the calculation of survival percentage.

**Mouse brain dissection, fluorescence imaging and histology.** Mice were killed for analysis by carbon dioxide asphyxiation or deep anesthesia with isoflurane followed by cervical dislocation. Mouse brains were manually dissected under a fluorescence stereoscope (Zeiss, Olympus or Leica). Brightfield and/or GFP fluorescence images were taken for the dissected brain and overlaid using ImageJ<sup>58</sup>. Brains were then fixed in 4% formaldehyde or 10% formalin for 48–96 h, embedded in paraffin, sectioned at 6  $\mu\text{m}$  and stained with hematoxylin and eosin (H&E) for pathology. For tumor size quantification, H&E slides were scanned using an Aperio digital slide scanner (Leica). Tumors were manually outlined as regions of interest (ROI) and subsequently quantified using ImageScope (Leica). Sections were dewaxed, rehydrated and stained using standard immunohistochemistry (IHC) protocols as previously<sup>55,59</sup>. 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 (the targeted organ) and liver (a nontargeted organ) were dissected and collected manually. For molecular biology, tissues were flash-frozen with liquid nitrogen and 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  $\mu\text{L}$  of 20-mg/mL Proteinase K (Qiagen) was added to the tissue/cell sample and incubated at 55 °C overnight. The next day, 30  $\mu\text{L}$  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 min. Samples were cooled on ice before addition of 2 mL of prechilled 7.5-M ammonium acetate (Sigma) to precipitate proteins. The samples were vortexed at high speed for 20 s and then centrifuged at  $\geq 4,000g$  for 10 min. At this point a tight pellet was visible in each tube, and the supernatant was carefully decanted into a new 15-mL conical tube. We added 6 mL 100% isopropanol to the tube, inverted it 50 times and centrifuged it at  $\geq 4,000g$  for 10 min. 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, and the tube was inverted 10 times and then centrifuged at  $\geq 4,000g$  for 1 min. The supernatant was poured off and discarded; the tube was briefly spun, and remaining ethanol was removed using a P200 pipette. After air-drying for 10–30 min, the DNA changed appearance from a milky white pellet to slightly translucent. We then added 500  $\mu\text{L}$  of ddH<sub>2</sub>O, and the tube was incubated at 65 °C for 1 h 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 follows: the predicted cutting sites (3 bp 5' of PAM) of the 280 gene-targeting sgRNAs in the mTSG library, plus the *Trp53*-targeting sgRNA in the vector, were retrieved from the mouse reference genome (mm10). The 140-bp sequences of the flanking regions of the cutting sites (5', 70 bp; and 3', 70 bp) were retrieved using Bedtools<sup>60</sup>. 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 three or fewer close matches; 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, 178 regions covering 277 sgRNAs, with a total of 33,638 bp, were covered in the probe set, with target bases covered = 32,239 (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 fragmented, followed by library preparation using a 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 quality-controlled using qPCR, and subjected to high-throughput sequencing using the HiSeq-2500 or HiSeq-4000 platforms (Illumina) at the Yale Center for Genome Analysis. We captured 277 of 278 (99.6%) of unique targeted sgRNAs for all samples from this experiment, missing only *Arid1a*-sg5 due to unavailability of qualified probes in this locus.

**mTSG sgRNA cutting efficiency measurement.** The mouse mTSG sgRNA cutting efficiency measurement was performed similarly to the screen, with the exception of early sampling. Briefly, AAV-mTSG library virus was injected in to the LV of LSL-Cas9 mice, but instead of waiting until mice reached tumor endpoints, they were killed for analysis at an early time point (3.5 weeks postinjection) and examined under a fluorescence stereoscope to dissect GFP<sup>+</sup> regions from the brain. Genomic DNA was extracted and subjected to capture sequencing.

**Mouse whole-exome capture sequencing.** The mouse whole-exome capture was performed using the SeqCap EZ exome kit (Roche). Briefly, capture sequencing was done following standard Illumina-Roche-Illumina protocols. Genomic DNA samples from mouse organs were fragmented, followed by library preparation using KAPA Library Preparation Kit (Illumina). The libraries were then amplified using LM-PCR, hybridized to the exome probe pool, washed and recovered, and then amplified with multiplexing barcodes using LM-PCR. The multiplexed library was then quality-controlled using qPCR and subjected to high-throughput sequencing using the HiSeq-2500 or HiSeq-4000 platforms (Illumina) at the 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 (ref. 61). Bam files were merged, sorted and indexed using bamtools<sup>62</sup> v2.4.0 and samtools<sup>63</sup> v1.3. For each sample, indel variants were called using samtools and VarScan<sup>64</sup> v2.3.9. Specifically, we used samtools mpileup (-d 1000000000 -B -q 10) and piped the output to VarScan pileup2indel (--min-coverage 1 --min-reads 2 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  $\pm 3$  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 used a series of criteria to identify high-confidence mutations. (i) 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. (ii) 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 percent of variant-read frequency across all control samples; for a mutation to be called in an mTSG sample, the percent 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 the 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. (iii) 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 versus 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: (i)  $\geq 10$  supporting reads for the reference allele; (ii)  $\geq 10$  supporting reads for the variant allele; (iii) the variant must be within  $\pm 6$  bp of a Cas9 PAM or NGG (or CCN on the reverse strand); (iv) a variant allele frequency  $< 75\%$ , as this was the maximum detected variant frequency out of the mTSG brain samples; and (v) the variant must not be detected in any sequenced control samples, which were considered 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, three 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 represented 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<sup>65</sup>.

**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 percent variant frequency tables on the gene level (in other words, summing the percent variant frequencies for all five of the targeting sgRNAs for each gene). Using these summed percent 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 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 minipools.** Mixtures of five sgRNAs targeting each gene were cloned as sgRNA minipools into the same astrocyte-specific AAV-CRISPR vector. For gene pair targeting, the five-sgRNA single gene minipools from both genes were mixed at 1:1. Plasmid mixes were then packaged into AAV1/2. Briefly, HEK293FT cells were transfected with the minipool plasmids, pAAV1 plasmid, pAAV2 plasmid, helper plasmid pDF6 and PEI Max (Polysciences, Inc. 24765-2) in DMEM (ThermoFisher, 10569-010). Seventy-two 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). Viral particles were titered by quantitative PCR using custom Cre-targeted Taqman probes

(ThermoFisher). After packaging, AAV minipools were stereotactically injected into the ventricle of LSL-Cas9 mice. Survival and histology analysis followed injection as described above. Several control mice (uninjected, EYFP and vector) were killed for analysis as surrogate histology samples although they were in good body condition and were subsequently found to be devoid of tumors.

**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-sgTrp53-sgX-GFAP-Cre vector described above). Tumor-containing brains were visually inspected under a fluorescence 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 a P2A-Fluc expression cassette into a lentiviral CRISPR knockout vector by Gibson assembly. To clone sgRNA targeting individual genes, such as *Pten*, *Arid1b*, *Mll3* (*Kmt2c*), *B2m* and *Zc3h13* (Supplementary Table 1), the corresponding oligos were synthesized, annealed and cloned into BsmBI linearized lentiviral knockout vectors. 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 at  $-80^{\circ}\text{C}$ .

**Generation of *NF1*<sup>-/-</sup> geneX and *RBI*<sup>-/-</sup> geneX knockout cell lines.** The *Nf1*- and *Rb1*-knockout tumor cells were infected by single-sgRNA knockout lentiviruses at M.O.I  $\leq 0.3$  to further knockout desired gene *X*. Twenty-four hours postinfection, lentiviral transduced cells were selected by the addition of 4–8  $\mu\text{g}/\text{mL}$  puromycin; they were then split after 2–3 d.

**TMZ treatment, cell viability assay and RNA preparation.** After 7–9 d culture under puromycin selection, lentivirus-infected *Nf1*- and *Rb1*-knockout tumor cells were plated in triplicates into 96-well plate at a density of  $2.5 \times 10^3$  cells per well, and  $\sim 5 \times 10^6$  cells were collected at the same time and for cutting efficiency analysis. One day after plating, either TMZ or DMSO was added at a concentration of 10  $\mu\text{M}$ , 100  $\mu\text{M}$ , 500  $\mu\text{M}$ , 1 mM or 2 mM. After 3 d of drug/vehicle treatment, cell viability was measured using CellTiter Glo (Promega) according to the manufacturers' protocol. Briefly, we first equilibrated the CellTiter Glo at room temperature for 1 h before use. Media from the 96-well plates was aspirated, and 50  $\mu\text{L}$  fresh DMEM + 10% FBS and 50  $\mu\text{L}$  CellTiter Glo was added. The luminescent signals were readout using EnVision plate reader (PerkinElmer). For RNA-seq sample preparation, cell lines harboring specific gene knockouts were cultured for 7–9 d under the selection pressure of puromycin, and then plated into 6-well plates at a density of  $2 \times 10^5$  cells per well in triplicates. Twenty-four hours after plating, 1 mM TMZ or DMSO in fresh DME + 10% FBS was added and cultured for another 48 h. Cellular RNA of control or treated cells was then extracted by adding 350  $\mu\text{L}$  TRIzol Reagent (Invitrogen) directly into the 6-well plates to lyse the cells, followed by gently shaking the plates and incubating them for 5–10 min to complete and lysis and homogenization. Then, 70  $\mu\text{L}$  chloroform was added, vigorously mixed and centrifuged at 16,000g for 15 min. The RNA containing the aqueous phase was transferred to a new tube and further purified using RNeasy mini kit (Qiagen). After eluting RNA from the column using nuclease-free water, the concentrations of sample RNA were normalized into 150–300 ng/ $\mu\text{L}$  for RNA-seq.

**T7 endonuclease I (T7E1) assays.** The genomic DNA of cells collected after 9 d of puromycin selection was extracted using QuickExtract DNA Extraction Solution (Epicentre), mixed well and incubated at  $65^{\circ}\text{C}$  for 30–60 min. Then, 1–2  $\mu\text{L}$  of genomic DNA from parental or lenti-sgRNA-transduced cells was used as the template to amplify the gene of interest using surveyor primers (Supplementary Table 33) with thermocycling conditions set at  $98^{\circ}\text{C}$  for 1 min;  $35^{\circ}\text{C}$  for 98  $^{\circ}\text{C}$  for 1 s,  $60^{\circ}\text{C}$  for 5 s and  $72^{\circ}\text{C}$  for 10 s; and  $72^{\circ}\text{C}$  for 1 min. The PCR products were gel-purified using QIAquick Gel Extraction Kit from 2% E-gel EX and quantified, followed by PCR product denaturing at  $95^{\circ}\text{C}$  for 5 min and annealing under the following conditions: ramp from  $95$  to  $85^{\circ}\text{C}$  at a rate of  $-2^{\circ}\text{C}$  per s; from  $85^{\circ}\text{C}$  to  $25^{\circ}\text{C}$  at a rate of  $-0.1^{\circ}\text{C}$  per s and hold at  $4^{\circ}\text{C}$ . We added 1  $\mu\text{L}$  of T7



endonuclease I into annealed oligo and incubated it 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:

$$\text{indels}(\%) = 100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$$

**Transcriptome profiling of different driver combinations in the presence and absence of chemotherapy.** Mixtures of five sgRNAs targeting each gene were cloned as sgRNA minipools into the same astrocyte-specific AAV-CRISPR vector. After packaging, AAV minipools were stereotactically injected into the lateral ventricles 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 48 h and harvested for mRNA-seq for transcriptome profiling. Briefly, total RNA was extracted from cancer cells derived from AAV-CRISPR minipool-induced GBM treated with DMSO or TMZ, using commercially available kits (Qiagen / ThermoFisher). A poly-A mRNA library was constructed using Illumina TruSeq mRNA library prep kit and sequenced on an 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<sup>66</sup>, 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<sup>67</sup>.

**Pathway enrichment analysis of differentially expressed transcripts.** Using an adjusted *P* value cutoff of 0.05 and a log fold-change threshold of  $\pm 1$ , we determined the set of genes that were significantly upregulated or downregulated. We then used the resultant gene sets for DAVID functional annotation analysis<sup>68</sup>. We considered a GO category statistically significant if the Benjamini-Hochberg adjusted *P* < 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 15 November, 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, copy-number variation calls and partial clinical data containing diagnostic information were obtained from the Yale Brain Tumor Program. All patient samples were deidentified. The general description, demographics and tumor characteristics are noted in **Supplementary Tables 14 and 15**. Total events for each patient were 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 the Yale Brain Tumor Program. All patient samples were deidentified. The mutations associated with specific samples were obtained from the Yale glioma databank. Slides stained with H&E or anti-GFAP were subsequently scanned using a slide scanner (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. To evaluate differences in the incidence of tumors in different groups, we used Fisher's exact test. See the **Life Sciences Reporting Summary** for additional information about the methods used in this study.

**Blinding statement.** Investigators were blinded with respect to histology scoring, capture sequencing and RNA-seq collection but not for dissection, MRI, survival analysis or sequencing analysis.

**Determination of sample sizes.** No statistical methods were used to predetermine 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 availability.** Genomic sequencing (targeted capture, exome) and RNA-seq data have all been deposited in NCBI SRA ([PRJNA393202](https://www.ncbi.nlm.nih.gov/sra/PRJNA393202)). CRISPR reagents (AAV-CRISPR and lentiCRISPR backbone plasmids and mTSG libraries) are available to the academic community through Addgene (<https://www.addgene.org/browse/article/28190058/>) or upon request.

51. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pii1 (2013).
52. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
53. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
54. Platt, R.J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
55. Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
56. Franklin, K.B.J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* (Academic Press, 2013).
57. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).
58. Schindelin, J., Rueden, C.T., Hiner, M.C. & Eliceiri, K.W. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
59. Chen, S. *et al.* Global microRNA depletion suppresses tumor angiogenesis. *Genes Dev.* **28**, 1054–1067 (2014).
60. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
62. Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P. & Marth, G.T. BamTools: a C API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
63. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
65. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
66. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
67. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
68. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## ► Experimental design

## 1. Sample size

Describe how sample size was determined.

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

## 2. Data exclusions

Describe any data exclusions.

Mice that were euthanized early in a healthy state for the purpose of obtaining time-matched histology were excluded from calculation of survival percentage. No other data was excluded.

## 3. Replication

Describe whether the experimental findings were reliably reproduced.

The findings reported in this study are the aggregate of multiple independent experiments, demonstrating the reproducibility of the findings.

## 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Mixed gender (randomized males and females) 6-14 week old mice were used in the experiment.

## 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

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

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ A statement indicating how many times each experiment was replicated
- ☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- ☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

A combination of custom scripts (available as Supplementary Software) and published software was used to analyze the data.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions on availability of unique materials. All plasmids have been deposited to Addgene.

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The following 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). These are all commonly used antibodies with numerous references in literature.

### 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No cell lines were used.

b. Describe the method of cell line authentication used.

No cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No cell lines were used.

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

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.

Policy information about [studies involving human research participants](#)

### 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

N/A