GENETICS

The genomic landscape of familial glioma

Dong-Joo Choi[1], Georgina Armstrong[2], Brittney Lozzi[1], Prashanth Vijayaraghavan[3], Sharon E. Plon[4], Terence C. Wong[3], Eric Boerwinkle[5], Donna M. Muzny[6], Hsiao-Chi Chen[1], Richard A. Gibbs[6], Quinn T. Ostrom[7], Beatrice Melin[8], Benjamin Deneen[1]*, Melissa L. Bondy[2]*, The Gliogene Consortium, Genomics England Research Consortium, Matthew N. Bainbridge[3]*

Glioma is a rare brain tumor with a poor prognosis. Familial glioma is a subset of glioma with a strong genetic predisposition that accounts for approximately 5% of glioma cases. We performed whole-genome sequencing on an exploratory cohort of 203 individuals from 189 families with a history of familial glioma and an additional validation cohort of 122 individuals from 115 families. We found significant enrichment of rare deleterious variants of seven genes in both cohorts, and the most significantly enriched gene was *HERC2* ($P = 0.0006$). Furthermore, we identified rare noncoding variants in both cohorts that were predicted to affect transcription factor binding sites or cause cryptic splicing. Last, we selected a subset of discovered genes for validation by CRISPR knockdown screening and found that *DMBT1*, *HP1BP3*, *and ZCH7B3* have profound impacts on proliferation. This study performs comprehensive surveillance of the genomic landscape of familial glioma.

INTRODUCTION

Glioma is the most common type of malignant brain tumor, although it is still considered rare, with an incidence of 6.0 per 100,000 individuals. Despite being rare, glioma causes high morbidity and mortality, with a median survival of 14 months (*1*–*4*). Approximately 5% of glioma cases are thought to be familial (*5*, *6*), which is typically defined by multiple affected individuals within a family who usually have early disease onset (<50 years) (*7*–*9*). The proportion of tumors that result from germline predisposition has been reported to vary from 1% (e.g., stomach tumors) to more than 50% (e.g., thyroid tumors) (*10*–*12*). Predisposition to cancer is typically inherited or caused by loss-of-function (LoF) variants in tumor suppressor genes with modest to high penetrance (25 to 90%; e.g., *BRCA1* and *TP53*) (*13*, *14*). Recently, genome-wide sequencing has helped identify previously undescribed cancer predisposition genes (CPGs) (*15*, *16*).

Case-control studies of glioma show that having a first-degree relative with glioma increases the risk of glioma by twofold (*17*). Several large genome-wide association studies (GWAS) on the risk of inherited glioma have identified approximately 25 loci that are associated with a modest increase in glioma risk (odds ratio, 1.1 to 3.5) (*18*). Functional, high-penetrance variants in several genes have also been associated with glioma, including *NF1/2*, *TSC1/2*, *TP53*, Lynch syndrome genes, *CDKN2A*, and *CHD1* (*16*–*21*). Previously, we performed exome sequencing of 90 individuals from 55 families and identified *POT1* as a previously unidentified glioma predisposition gene (*15*, *22*); however, many cases of familial glioma remain unexplained. Whole-genome sequencing (WGS) is an increasingly inexpensive method to screen for rare variants across the genome (*23*) and provides robust coverage of coding regions. It also allows for discovery of structural and noncoding variants, the latter of which are gaining appreciation for their functional roles in inherited disease. In this study, we used WGS to examine the genomes of two cohorts with a family history of glioma. To identify putative novel glioma predisposition genes, we compared the results with those from a control cohort consisting of unselected individuals whose genomes were sequenced on the same platform. Coding and noncoding variants were examined for associations with both known and potentially novel CPGs.

RESULTS

We performed WGS on 325 individuals with glioma from 304 families with a history of familial glioma (table S1). The vast majority of participants self-reported their race as non-Hispanic white. The participants were divided into exploratory and validation cohorts consisting of 203 subjects (189 families) and 122 subjects (115 families), respectively. Analysis of single-nucleotide variants (SNVs), indels, and larger copy number variants (CNVs) identified approximately 1 billion germline variants across both cohorts. Variants were filtered and prioritized as outlined in Methods and in Fig. 1. Concomitantly, we performed WGS of 1013 non-Hispanic white participants from the United States who participated in the Atherosclerosis Risk in Communities (ARIC) study (*24*). These individuals were 45 to 65 years of age at the time of recruitment and were randomly selected from the U.S. Census data (see Methods).

Coding variants in known and novel CPGs

Glioma susceptibility has been associated with several known CPGs. Therefore, we examined whether our familial cohort had LoF or known deleterious variants in any of 32 commonly tested CPGs (table S2). We identified significant ($P = 0.0071$, Fisher's exact test) enrichment of LoF and likely pathogenic/pathogenic variants in known CPGs in our cohort (12 of 189) compared with the control group (23 of 1013) (Fig. 2 and table S3) (*25*). Heterozygous *BRIP1* and *PMS2* mutations were the most common mutations in our

[1]Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX, USA. [2]Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA. [3]Rady Children's Institute for Genomic Medicine, San Diego, CA, USA. [4]Department of Pediatrics/Hematology-Oncology, Baylor College of Medicine, Houston, TX, USA. [5]The University of Texas Health Science Center School of Public Health, Houston, TX, USA. [6]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [7]Department of Neurosurgery, Duke University School of Medicine, Durham, NC, USA. [8]Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden.
*Corresponding author. Email: mbainbridge@rchsd.org (M.N.B.); deneen@bcm.edu (B.D.); mbondy@stanford.edu (M.L.B.)

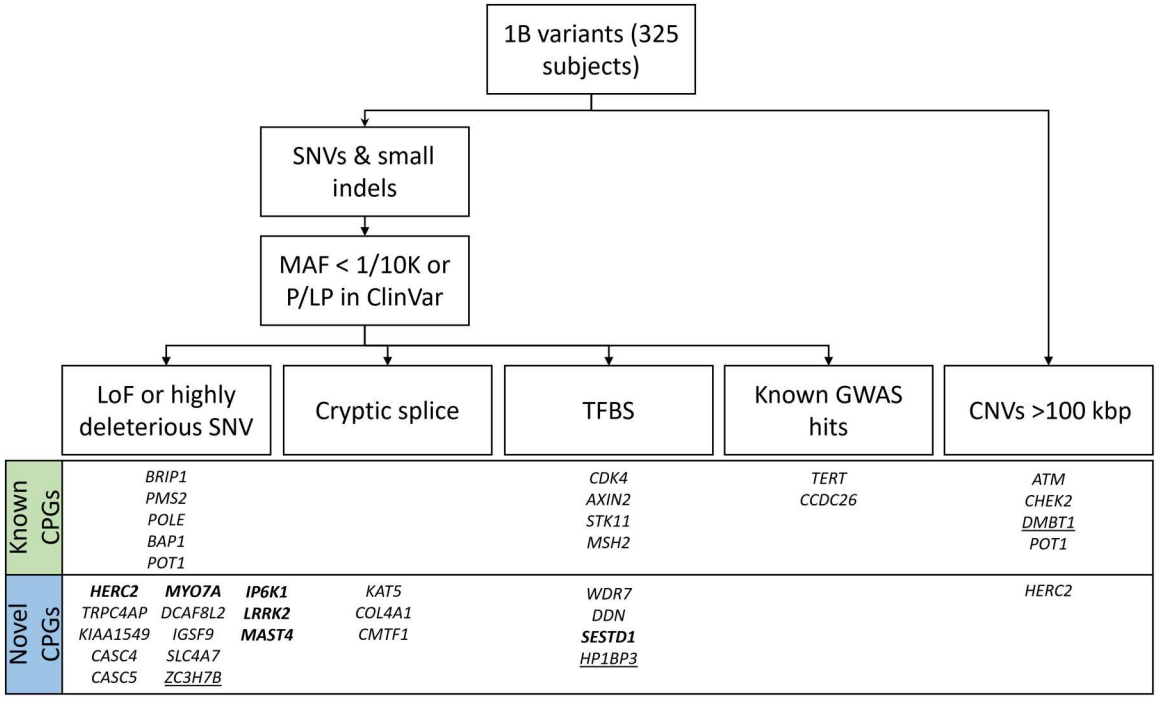

Fig. 1. Study analysis pipeline. Variants were classified as CNVs and smaller events. SNVs and indels were selected according to rarity or presence using the ClinVar database. The resulting variants were then divided into four groups: highly deleterious variants, intronic cryptic splicing variants, variants in TFBSs, and known glioma-associated GWAS SNPs (see Methods). Affected genes were stratified on the basis of whether they were known or novel CPGs. Genes with identified variants are shown below each category. Discovered genes were then validated in a second cohort (bold), and a subset of genes was examined via in vivo screening (underlined). B, billion; SNV, single-nucleotide variant; MAF, minor allele frequency in gnomAD overall; K, thousand; P/LP, pathogenic or likely pathogenic; LoF, loss of function; TFBS, transcription factor binding site; GWAS, genome-wide association study; CNV, copy number variant; CPG, cancer predisposition gene.

cohort, accounting for three samples each, and *POLE* was mutated in two samples. In contrast, none of the control cases had *BRIP1* or *POLE* variants, and only one sample had a pathogenic *PMS2* variant ($P = 0.004$, $P = 0.03$, and $P = 0.06$, respectively, Fisher's exact test). In the validation cohort, we identified additional pathogenic mutations in *ATM*, *BRCA1*, and *CDKN2A*, the latter of which was shared by two affected family members (table S3).

Next, we identified genes with a novel association with familial glioma. We filtered variants in the case and control groups according to rarity [minor allele frequency (MAF) < 0.0001], LoF, and deleteriousness [Combined Annotation Dependent Depletion phred-like (CADD PH) score ≥ 30] and identified genes that were significantly enriched for these variants. We only considered genes with two or more variants in our cohort. After correction for multiple tests, we found 19 genes that were significantly enriched, with a false discovery rate of <0.1 (Table 1). Of these, *HERC2* was the most significantly enriched, with four variants in our exploratory cohort and none in the control group ($P = 0.0006$, Fisher's exact test). In addition, six of these genes contained at least one mutation in our validation cohort (bold).

To further assess the role of HERC2 in gliomagenesis, we examined the Genomics England dataset and filtered all cases of glioma for rare (1 of 10,000 MAF) and deleterious (CADD score ≥ 30) variants in *HERC2*. We identified two additional *HERC2* variants in the Genomics England dataset (p.R4587Q and p.S4607C) in two individuals, one male who was in his 60s at the time of glioma onset and one female with an unknown age of onset. At the time of sampling (6 June 2020), there were 611 cases of glioma in the Genomics England Clinical Interpretation Partnership (GeCIP) database. The glioma cohort in the GeCIP database is not focused on familial glioma but instead agnostically includes both familial and sporadic cases. On the basis of the frequency of familial glioma (approximately 5% of all gliomas), we would expect approximately 30 of 611 samples to originate from a familial predisposition to glioma. Under this assumption, we found significant enrichment of *HERC2* mutations in this cohort ($P = 0.0009$, Fisher's exact test), thereby confirming an association between *HERC2* mutation and familial glioma in a second independent dataset. As a caveat, when considering the incidence of HERC2 variants across all glioma cases in the GeCIP database, enrichment of *HERC2* mutation was not significant ($P = 0.14$), further emphasizing the role of *HERC2* mutation in the incidence of familial, rather than sporadic, glioma.

Next, we sought to understand the role of CNVs in glioma susceptibility. We identified CNVs in known CPGs and large [>100 kilo–base pairs (kbp)] copy number losses across all genes (Fig. 3 and table S4). We identified deletions in three known cancer genes: a heterozygous deletion of four exons in *CHEK2*, a homozygous complex deletion in *DMBT1*, and a deletion of the terminal coding exon in *ATM*. Each deletion was specific to a family. We also identified internal tandem duplications in *POT1* (exons 4 and 5) and *HERC2* (exons 3 to 65), which are predicted to cause a frameshift insertion, leading to nonsense-mediated decay, or may otherwise be pathogenic and cause haploinsufficiency, bringing the total number of deleterious variants found in *HERC2* to six

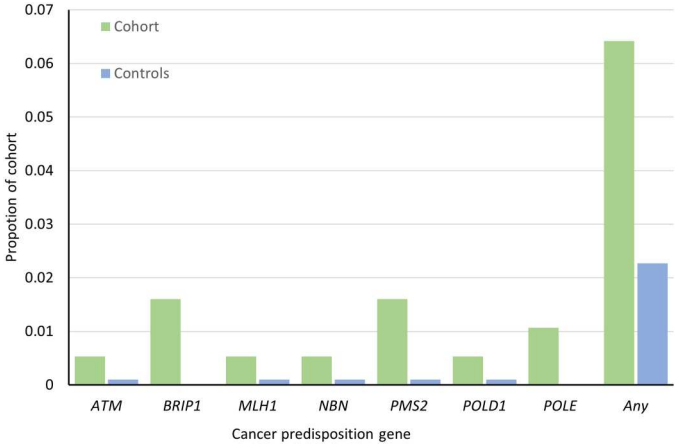

Fig. 2. Proportion of the glioma exploratory cohort (green) or control cohort (blue) with variants in seven cancer genes and overall. *P* values were calculated with Fisher's exact test and Bonferroni correction.

(Fig. 4) (*26*, *27*). We also identified four large deletions that did not involve known cancer susceptibility genes (table S4). No genes with deletions were shared across families, but for each deletion, at least one gene was both LoF intolerant and involved in cell growth or differentiation (*KAT2B*, *DIP2A*, *PRKCA*, and *FOXJ2*). Of these, *FOXJ2* has been suggested to act as a tumor suppressor in glioma cells and other cancers (*28*) and lowered expression is associated with worse outcomes in glioma ($P = 0.011$) (*29*). Low expression of *DIP2A* is associated with decreased temozolomide sensitivity, but the role of *DIP2A* and *PRKCA* in tumor initiation is not clear and may be incidental. Last, *KAT2B* is a histone acetyl transferase putatively involved in many cellular processes, and lower expression is also correlated with worse outcomes ($P = 0.03$) (*30*). No large structural variants were identified in the validation cohort.

In our exploratory cohort, we identified 17 rare variants that likely affect gene expression by altering transcription factor binding sites (TFBSs) (see Table 2). In five cases, these variants affected four known CPGs (*CDK4*, *AXIN2*, *STK11*, and *MSH2*), with *MSH2* variants occurring in two families. We identified a TFBS variant upstream of *HP1BP3* in one family with a notable history. However, TFBS variants were not considered significantly enriched compared with those in our control cohort. We also identified three genes (*WDR7*, *DDN*, and *PC*) that had TFBS variants in three or more families in our cohort and no variants in the control cohort ($P = 0.011$, Fisher's exact test, Bonferroni-corrected). One family, FG110, had two TFBS variants, and FG128 had a CNV in *POT1*. We also identified upstream variants in the previously identified *TRPC4AP* (Table 1) gene and *SESTD1*, the primary docking gene for the TRPC4 complex. The *SESTD1* variant also occurred in two of the families in the validation cohort. When examining the types of TFBSs with identified variants, we found that the most commonly mutated TFBS was the Myc-associated zinc finger protein, *MAZ*, followed by *EGR1* and *SP1*, with two variants each.

Last, we examined 42 known glioma predisposition GWAS signals that were identified in our previous study (table S6) (*18*) and found that, after correction for multiple tests, five sites were significantly enriched in our exploratory cohort compared with their MAFs in the general population. The two strongest associations were found for telomerase reverse transcriptase (*TERT*) (40% versus 50%, rs2736100, $P = 0.00014$) and an intergenic single-nucleotide polymorphism (SNP) often associated with *CCDC26* (12% versus 6%, rs55705857, $P = 0.00011$). The former variant occurred in an MITF TFBS. In our validation cohort, the *CCDC26*-associated SNP was observed at a similar rate (approximately 12%), but the *TERT*-associated SNP was even more depleted than it was in the discovery cohort, with only 18% of the cohort carrying this SNP.

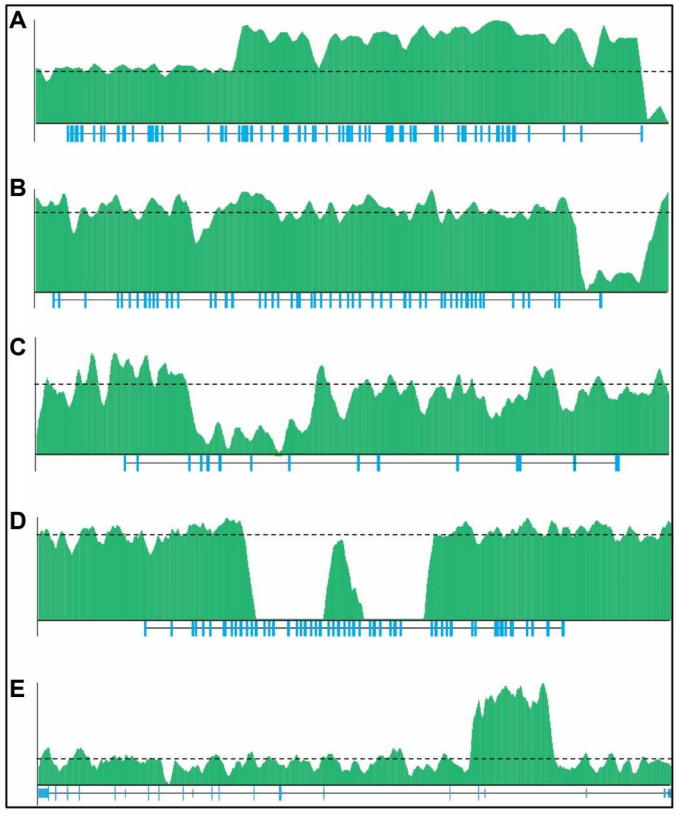

Fig. 3. CNVs identified in the glioma exploratory cohort. The relative coverage of *HERC2* (**A**), *ATM* (**B**), *CHEK2* (**C**), *DMBT1* (**D**), and *POT1* (**E**) is shown in green; the average coverage is shown by a dashed black line. Exons (blue lines) and introns (thin lines) are shown below the coverage plots.

Functional analyses

We prioritized 72 genes containing prospective LoF mutations for validation studies in an established mouse model of glioma. To examine the contribution of this gene cohort to gliomagenesis, we performed CRISPR-Cas–based LoF screening in an in utero electroporation (IUE) model of glioma (table S7). The potential for gliomagenesis was assayed using barcoded CRISPR-Cas9 screening in the IUE model (*31*). We generated barcoded guide RNAs (gRNAs) for each of the candidate genes and combined this library with our established IUE model, which contains gRNAs for *TP53*, *PTEN*, and *NF1* (i.e., 3xCr), along with Cas9 and green fluorescent protein (GFP) expression constructs (Fig. 5A, see Methods). This DNA cocktail was introduced via IUE into the cortex, and after glioma formation, we used targeted sequencing of tumors to detect barcodes as a surrogate for enrichment of a given gRNA within the tumor. This analysis revealed that gRNAs for *DMBT*, *ZC3H7B*, and *HP1BP3* were enriched in tumors (Fig. 6B),

Fig. 4. Schematic of variants identified in *HERC2*. Copy gain (blue bar) and variants (arrows: red, orange, and gray indicate CADD PH scores of 35, ≥30 but <35, and frameshift mutations, respectively) are shown relative to the regulator of chromosome condensation protein domains (green bar). Approximately 99% of all variants identified by WGS were noncoding; however, the functional consequences and MAF distributions of these variants are not as well understood as those of coding variants. To avoid numerous spurious associations, we restricted our analysis to deep-intronic (i.e., >5 bp from the exon) variants that were predicted to cause cryptic splicing and variants in the immediate upstream region of genes that overlapped a TFBS. We identified 16 deep-intronic variants in LoF-intolerant and novel CPGs that were predicted to cause cryptic splicing in our exploratory cohort (see table S5). Identical variants in two genes, *NCAM1* and *SMG6*, were identified in the validation cohort.

suggesting that loss of these genes promotes tumorigenesis. These gRNAs were subsequently validated by Sanger sequencing, which showed that they cut at the intended regions (fig. S5).

To confirm that loss of these genes promotes tumorigenesis, we individually introduced the gRNAs in our established IUE glioma model and found that individual loss of each gene decreased overall survival (Fig. 5C). In parallel, we performed overexpression and gain-of-function (GoF) studies by introducing piggyBac (PB) constructs containing each open reading frame (ORF) in our established IUE glioma model (*32*). Complementary to our LoF results, we found that overexpression of each gene extended overall survival (fig. S1), further supporting the notion that these genes suppress glioma tumorigenesis. Next, we harvested tumors and performed immunostaining to confirm loss of each gene while also validating the histopathological features of glioma (fig. S2). The decrease in overall survival suggested that loss of these genes accelerates malignant progression of glioma and associated proliferation. Bromodeoxyuridine (BrdU) labeling of mice bearing end-stage tumors showed that the number of BrdU-labeled cells was significantly increased in tumors lacking *ZCH7B*, *DMBT1*, and *HP1BP3* (Fig. 5D). Notably, loss of *ZCH7B* and *DMBT1* had a more profound effect on proliferation than loss of *HP1BP3* (Fig. 5D), which is consistent with the observed relative reductions in overall survival (Fig. 5C). Together, these data indicate that familial glioma-associated genes contribute to tumorigenesis and that *DMBT1*, *ZC3H7B*, and *HP1BP3* play key roles in this process.

Expression profiling of *ZC3H7B*, *DMBT1*, and *HP1BP3* LoF glioma tumors

To understand the molecular changes that occur in glioma tumors lacking *ZC3H7B*, *DMBT1*, and *HP1BP3*, we performed RNA sequencing (RNA-seq) of tumors expressing *ZC3H7B*-LOF (gRNA vector for LoF), *DMBT1*-LOF, *HP1BP3*-LOF, and 3xCr constructs. Bioinformatics analysis of the differentially expressed genes (DEGs) revealed that 277 genes were down-regulated and 250 genes were up-regulated by *ZC3H7B*-LOF, 178 genes were down-regulated and 252 genes were up-regulated by DMBT1-LOF, and 736 genes were down-regulated and 147 genes were up-regulated by

Table 1. Genes enriched for variants in both cohorts. Gene, number of variants in the exploratory and validation cohorts (comma separated) and the control cohort set, *P* value of enrichment (Fisher's exact test) of variants, and families bearing the variant. Genes with variants in the validation cohort are shown in bold; variants validated by in vivo testing are underlined.

Gene	Families in cohort	Control counts	*P* value	Variants	Family
HERC2	4, 1	0	0.0006	p.R4293Q, p.F3704L, p.F3642L, p.D523N, p.L2384fs	FG101, FG118, FG136, FG140, VC23
TRPC4AP	3	0	0.0037	p.Q479X, p.R637Q, p.R307Q	FG104, FG121, FG139
KIAA1549	3	1	0.013	c.5248-2A>G, p.R1275X, p.V1469M	FG102, FG119, FG137
MYO7A	3, 2	1	0.013	p.R740Q, p.R830H, p.R63Q, p.E1579K, p.E602K	FG103, FG120, FG113, VC24, VC34
PIK3R4	2	0	0.024	p.V478I,p.P460L	FG105, FG122
ZC3H7B	2, 0	0	0.024	p.K854R, p.R923H	**FG110**, **FG129**
PCNXL4	2	0	0.024	p.P363L, p.A906V	FG106, FG123
THSD7A	2	0	0.024	p.E284X, p.R1253Q	FG107, FG124
PCYT1A	2	0	0.024	p.K186E, p.A99V	FG108, FG125
IP6K1	2, 1	0	0.024	p.R139Q, p.P71L, p.K53R (cryptic splice)	FG109, FG126, VC97
ADAMTS8	2, 1	0	0.024	p.R551H, p.S406C, p.M1I	FG110, FG127, VC54
TRPM1	2	0	0.024	p.L1021F, p.D82N	FG111, FG128
LRRK2	2, 1	0	0.024	p.R1334Q, p.F2059L, p.E2508X	FG112, FG129, VC84
FLOT2	2	0	0.024	p.A332V, p.A325V	FG113, FG130
PPP1R16B	2	0	0.024	p.R24Q, p.E175K	FG114, FG112
CTNND1	2	0	0.024	p.E73K, p.R653H	FG115, FG132
IFIH1	2	0	0.024	p.V453L, p.P330S	FG109, FG133
MAST4	2, 1	0	0.024	p.A1190S, p.P1439L, p.G1519S	FG116, FG134, VC17
TFAP2E	2	0	0.024	p.S229T, p.G367R	FG117, FG135

Table 2. Variants identified in upstream Factorbook TFBSs. The affected gene, transcription factor type, and family with the variant are shown. Genes shown in bold were validated in the second cohort. Genes underlined were validated by in vivo screening.

Gene	Variant	TF	Family
AXIN2	17:63557882T>A	EGR1	FG141
CDK4	12:58146267G>A	SP1	FG129
DDN	12:49393353G>A	EGR1, MAZ	FG143
DDN	12:49393365G>A	MAZ	FG110
DDN	12:49393561G>A	UA2	FG144
MSH2	2:47630140G>C	E2F4	FG128
MSH2	2:47629890GT>G	ZNF263	FG146
PC	11:66726049C>G	E2F1	FG147
PC	11:66726027G>C	E2F4	FG148
PC	11:66726101C>T	SP1	FG135
HP1BP3	1:21113282A>G	E2F4	FG198
SESTD1	2:180129654C>G	MAZ	FG150, VC46, VC65
STK11	19:1205343G>C	UA3	FG151
TRPC4AP	20:33680727G>C	ZNF143	FG152
WDR7	18:54318543T>C	MAZ	FG153
WDR7	18:54318485C>T	NRF1	FG110
WDR7	18:54318419G>C	YY1	FG154

HP1BP3-LOF compared with tumors expressing 3xCr (fig. S3 and table S8). Next, we performed Gene Ontology (GO) analysis of the DEGs from the RNA-seq data of the ZC3H7B, DMBT1, and HP1BP3 tumor models. GO analyses revealed that chemokine-binding, antigen receptor–mediated signaling pathways, regulation of cytosolic calcium ion concentration, regulation of cytokine production, and cytokine signaling pathways are regulated by ZC3H7B; extracellular matrix organization, cell proliferation, and cell-matrix adhesion are regulated by DMBT1; and neutrophil activation, cytokine production, and regulation of interleukin-6 are regulated by HP1BP3 (fig. S3 and table S9).

Because the loss of each of these genes accelerates glioma tumorigenesis and is associated with increased glioma risk, we hypothesized that these genes regulate a common set of pathways. Therefore, we evaluated the overlapping genes of each DEG set and found that 167 genes are common to each of the LOF tumors, and the GO terms were related to calcium ion homeostasis, inflammatory response, and regulation of angiogenesis (Fig. 6, A to C, and table S10).

To determine whether these conserved DEGs are altered in glioma by loss of *ZC3H7B*, *DMBT1*, and *HP1BP3*, we examined the vascular endothelial growth factor (VEGF) and fibroblast growth factor 2 (FGF2) genes, which are associated with angiogenesis and calcium ion homeostasis. By immunostaining our glioma tumors, we found that *VEGF* and *FGF2* expression was significantly increased in glioma tumors lacking *ZC3H7B*, *DMBT1*, or *HP1BP3* (Fig. 6, D to F). Together, these data indicate that the familial glioma–associated genes *DMBT1*, *ZC3H7B*, and *HP1BP3* contribute to tumorigenesis by influencing the glioma tumor microenvironment through regulation of multiple processes, including angiogenesis.

DISCUSSION

Familial glioma occurs rarely and has few associated genes. Our two cohorts of 189 and 115 glioma families represent the largest cohorts to undergo WGS. Previously, we identified *POT1* variants in three families, including a large deletion in *MSH2* in one family and variants in *TP53* in two families, in a smaller cohort of subjects who underwent exome sequencing (*15*, *33*). Here, we identified 54 genetic variants in 28 genes, or unique genetic loci in the case of CNVs, that contained rare, deleterious variants, noncoding functional variants, or CNVs that occurred in both our exploratory cohort and our validation cohort or in known CPGs. In total, 37 families (20%) in our exploratory cohort and 13 families (10%) in our validation cohort had at least one such alteration.

HERC2 and other enriched genes

We identified potential novel drivers of gliomagenesis that were statistically enriched in our familial glioma cohorts compared with the control cohort (Table 1). *HERC2* showed the highest enrichment with deleterious variants compared with our control cohort. In total, six mutations of this gene were found: five rare coding variants that were predicted to be highly deleterious SNVs and one internal copy gain variant (see Fig. 4). Furthermore, we identified two additional, rare, deleterious variants in the Genome England dataset (p.R4587Q and p.S4607C). In accordance with our findings, *HERC2* mutation was enriched in the putative familial glioma subset of this dataset ($P < 0.001$). All coding variants occurred in highly conserved regions of the gene; two occurred in a regulation of chromosome condensation protein domain, and two occurred in the homologous to the E6-AP C terminus (HECT) domain. *HERC2* is highly conserved throughout the human population, and we found marked selection against LoF [predicted LoF intolerance (pLoFI) = 1.0] and missense variants ($Z = 4.42$). However, HERC2 is an extremely large protein (4834 amino acids) that frequently harbors many variants in normal individuals. Identification of *HERC2* as a possible gliomagenesis gene in this study was only possible through aggressive filtering to determine the rarest and most deleterious variants. However, in future studies, less stringent filtering could be used to identify more causal variants, especially if these variants can be tested functionally. HERC2 conducts several important cellular functions. It regulates ubiquitin-dependent retention of repair proteins on damaged chromosomes, affecting both the double-stranded break and excision repair pathways, and modulates p53 activity (*34*, *35*). It also indirectly regulates the insulin-like growth factor receptor pathway and cell cycle and may be involved in immune function (*36*). *HERC2* is also commonly mutated and focally deleted in glioma (*37*). *HERC2* was not detected in our CRISPR screening, possibly because cells with HERC2 knockdown become nonviable and are suppressed because of cell competition. This hypothesis is supported by literature demonstrating that HERC2 is essential for embryonic growth (*38*). Furthermore, HERC2 may only act as a tumor suppressor in the context of wild-type p53; thus, using a model system that has normal p53 levels may be necessary to observe the tumorigenic potential of HERC2 mutation (*39*). The second most enriched gene, *MYO7A*, was mutated three times in

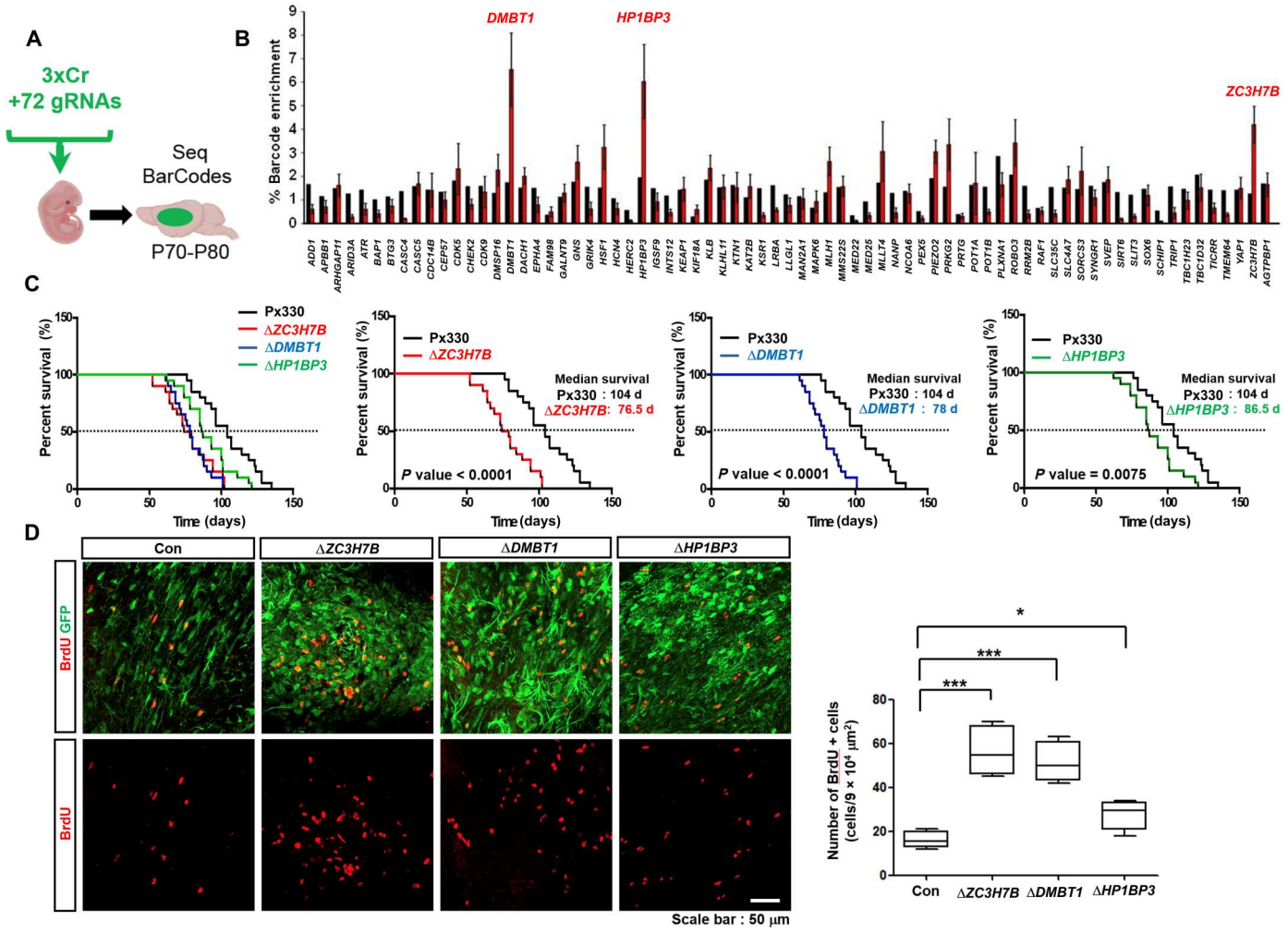

Fig. 5. In vivo functional screening identified regulators of glioma tumorigenesis. (**A**) Schematic of barcoded screening, where the 3xCr glioma system was combined with 72 barcoded gRNAs and co-electroporated into the embryonic cortex. (**B**) Next-generation sequencing to determine barcode amplification. The barcode for each gRNA (red) and the input signal (black) are shown ($n = 3$ tumors). Data are indicated as the mean and SEM. (**C**) Kaplan-Meier curves from individual validation studies ($n = 20$, px330 control, median survival 104 days; $n = 20$, ΔZC3H7B, median survival 76.5 days; $n = 20$, ΔDMBT1, median survival 78 days; $n = 20$, ΔHP1BP3, median survival 86.5 days). (**D**) BrdU staining of end-stage tumors; quantification for each group is derived from five different tumors. $^*P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.005$. Scale bar, 50 μm.

our exploratory cohort and twice in the validation cohort. This gene is an unconventional myosin with a very short tail (*40*). Mutations in this gene are associated with recessive and dominant deafness and Usher syndrome (MIM:276900). The role of MYO7A outside of hearing loss is not well understood, although it is ubiquitously expressed (*40*). LoF mutations in this gene lead to the "Shaker" mouse phenotype, and MYO7A appears to be essential in other mammals and fruit flies (*41*, *42*).

Known CPGs

Known CPGs were found in 18 families. Our exploratory cohort showed significant enrichment of LoF or pathogenic mutations of *BRIP1* and *POLE*, although this was not observed in our validation cohort. Heterozygous mutations of *BRIP1* are commonly associated with a small increase in the risk of ovarian cancer susceptibility and may be associated with colon cancer (*43*). Gliomas have been identified in colorectal cancer families with *POLE* missense variants (*44*, *45*). LoF variants of *POLE* are not generally considered pathogenic; however, both *POLE* variants occurred in the background of an additional variant (*ATM* and *MLH1* in families FG189 and FG155, respectively), which may indicate an alternate mechanism of pathogenicity wherein haploinsufficiency of *POLE* exacerbates variants in other CPGs in a multigenic manner (*46*). Further CNVs were identified in *CHEK2* and *ATM*. These results indicate that known cancer genes, not previously associated with glioma, play a role in familial glioma. However, it remains unclear why these families presented with familial glioma and not with the canonical cancers associated with the respective genes. This is true for both the personal history of the proband and the family history of all members in the family.

Glioma linkage and telomeres

Previous attempts to perform linkage analysis to identify glioma predisposition genes identified several linkage loci (17q12-21.32,

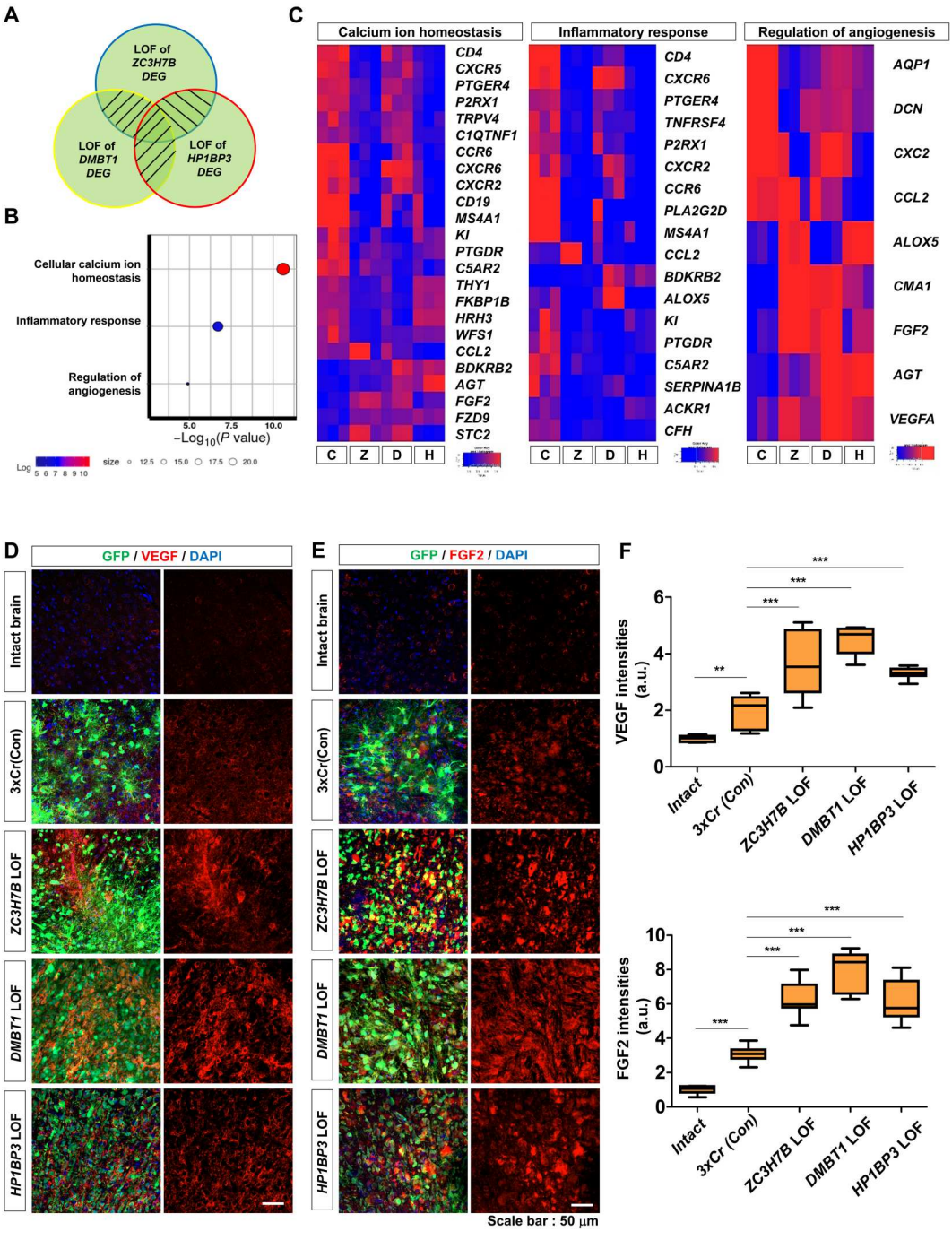

Fig. 6. ZC3H7B, DMBT1, and HP1BP3 are key regulators of glioma tumorigenesis–related gene populations. (**A**) Venn diagram of DEGs affected by *ZC3H7B*-LOF, *DMBT1*-LOF, and *HP1BP3*-LOF. (**B**) GO analysis of the common DEGs. (**C**) Heatmap of GO-related genes from the RNA-seq results of 3xCr [control (C)], *ZC3H7B*-LOF (Z), *DMBT1*-LOF (D), and *HP1BP3*-LOF (H). (**D** and **E**) VEGF and FGF2 staining from a postnatal day 60 tumor. DAPI, 4′,6-diamidino-2-phenylindole. (**F**) Intensities of VEGF and FGF2 staining determined by ImageJ software. The quantification of each group is derived from three different tumors. **P* < 0.05, ***P* < 0.01, and ****P* < 0.005. Scale bars, 50 μm. a.u., arbitrary units.

15q23-q26.3, and 1q23) but were unable to identify an underlying causative gene (*47*–*49*). Several genes (*BRIP1*, *PMS2*, *PRKCA*, *ACACA*, *SMG6*, *MAP2K6*, *HCN4*, *PRTG*, and *IGSF9*) identified in this study are in or proximal to these linkage sites.

This study reaffirms the importance of telomere maintenance in gliomagenesis. We identified two novel variants (a frameshift and an internal duplication variant) in the telomere gene *POT1*, a gene that we previously associated with glioma susceptibility (*15*). We further identified variants in *SMG6*, a component of the telomerase ribonucleoprotein complex responsible for the replication and maintenance of chromosome ends and identified enrichment of a known gene identified by a GWAS upstream of *TERT*. In addition to the variant upstream of *TERT*, we showed significant enrichment of another GWAS signal, rs55705857, an intergenic SNP associated with oligodendroglioma. This latter variant occurs in the *MITF* TFBS. Hi-C data from astrocytes (ENCSR011GNI) indicated that regions around this variant can associate with an upstream enhancer of MYC (*50*, *51*). MYC dysregulation in tumors has been associated with individuals who carry this SNP (*52*).

Functional validation

We identified multiple novel genes as candidate CPGs. Of these, three were selected in a high-throughput in vivo screening assay and exhibited tumor suppressor–like function toward glioma tumorigenesis when individually manipulated in our model. These observations suggest that hypomorphic alleles or complete LoF of these candidates can promote tumorigenesis and may explain why variants in these genes predispose individuals carrying these alleles to glioma. Future studies will be geared toward understanding how these genes influence neural progenitor expansion during early tumor-initiating stages that better reflect cancer predisposition states in humans. In addition, we performed functional analysis of glioma tumors lacking *DMBT1*, *HP1BP3*, and *ZC3H7B* and identified large-scale dysregulation of multiple genes compared with those of the triple CRISPR knockout model alone. By focusing on common pathways dysregulated by each of these genes, we found that the angiogenesis pathway is conserved across these genes and validated the up-regulation of VEGF and FGF2 in each of these tumor models. These observations, coupled with the dysregulation of immune and cytokine signaling signatures from individual *ZC3H7B* and *HP1BP3* knockout tumors, suggest an interaction between these genes and the tumor microenvironment. It will be important to further determine how these genes and associated variants influence the immune and vascular environment during tumorigenesis and during early tumor-initiating events.

Noncoding variants

Few studies have used WGS to examine the causes of familial cancer, and this study is one of the first to examine rare, functional, noncoding variants in depth. The noncoding portion of the genome is less well understood than the coding portion; therefore, we necessarily took a conservative approach to analyze these variants, only considering very high-quality (score > 0.8) cryptic splicing events and well-annotated, high-quality TFBSs in the immediate upstream region of a gene. This led to the identification of noncoding, potentially deleterious variants in several known CPGs. In both cohorts, we identified cryptic splicing variants of *NCAM1* and *SMG6*, both of which are highly LoF intolerant (*53*). SMG6 is associated with telomere maintenance (*54*), and NCAM1 regulates neurogenesis, neurite outgrowth, cell migration, and immune surveillance (*55*, *56*). We identified identical mutations in a cis-regulatory element upstream of *SESTD1* in three unrelated families. This variant is absent in the Genome Aggregation Database (gnomAD), likely indicating a population frequency of much less than 1 in 10,000. We also identified a mutation in a cis-regulatory element in *HP1BP3*, a gene that maintains heterochromatin and influences the G_1 phase duration during the cell cycle (*57*). In general, noncoding variants may have a more subtle effect on gene function. Genes that are extremely important to cellular functions may contain few or no coding variants across the population, as these may be lethal. Noncoding variants that modulate expression may, therefore, be an important and unappreciated mechanism of disease for these genes.

Limitations

Because of the rarity of mutation of any particular gene in familial glioma, validation of findings through a second cohort can be challenging, highlighting the important role of functional validation. Here, we selected a subset of identified genes for functional validation through CRISPR knockdown. This approach identified three genes that increased tumor proliferation in vivo—*DMBT1*, *HP1BP3*, and *ZC3H7B*—with the latter two having no previous association with cancer. Other genes that were not detected during screening may also be involved in gliomagenesis but may have a mechanism that is more complex than LoF or require wild-type versions of genes that are knocked out in our model, or complete knockout of the gene may be incompatible with life. Furthermore, we chose to only use a single guide for each gene; thus, it may be possible that some guides are not as effective as others, leading to an increase in our false-negative rate. Thus, genes that pass our screen should be considered good candidates for gliomagenesis, and those that fail the screen cannot be discounted entirely.

Another limitation of this study is the lack of multiple family members available for study because of the rapid and aggressive nature of glioma and because glioma may not be classified as familial until many relatives have already died from the disease. The availability of first-degree relatives can eliminate approximately half of all rare variants considered (100 to 400 coding variants depending on filtering criteria). This can greatly reduce the number of variants for consideration and has been used to great effect in studying other rare diseases (*58*).

Although we potentially identified causal variants in many families, most of the families had no identified causal variant. This may be due to several factors, including stringent filtering of variants and the small cohort size. Furthermore, some families may not have had familial glioma but only coincidentally had multiple family members with glioma. This aspect may be further exacerbated in our validation cohort because most of those families only had self-reported second cases that were not confirmed by independent examination of the relevant medical records. We are also extremely limited in our understanding of noncoding variants, and any such variants require functional testing to determine their effects.

In conclusion, few genes are currently associated with familial glioma. We showed that genes associated with other heritable cancers may also increase glioma susceptibility. Furthermore, we identified novel roles for some genes, including *HERC2* and *TRPC4AP*, as possible causes of familial glioma. Last, we showed significant enrichment of noncoding variants in several families with glioma. Noncoding variants, while difficult to annotate

functionally, are a largely unstudied source of variation that may account for numerous cases of glioma.

METHODS

Case cohorts

The Gliogene Consortium was established in 2004 to identify susceptibility genes in high-risk familial brain tumor pedigrees (*59*). The consortium assembled information on 376 glioma families in 14 centers in the United States, Israel, Sweden, and Denmark between 2007 and 2011. The current report presents data from 189 families who are not part of our previous analysis (*15*). The eligibility criteria for inclusion were that the family had at least two members with pathologically confirmed primary glioma according to the 2007 World Health Organization Classification of Tumors of the Central Nervous System (*60*). Families with a known inherited genetic syndrome such as *NF1* were excluded. We collected extensive pedigree information from each family, including cancer diagnoses, age, age at death, degree of relationship to the proband, size of the family, and number of generations, as well as blood or saliva samples from first- and selected second-degree relatives of the proband. The validation cohort included 122 familial glioma cases selected from the Glioma International Case Control Study. All patients completed an interviewer-administered questionnaire that included a detailed family cancer history report. Staff were trained to probe and to exclude any participant who was likely to have a non-primary brain tumor or brain metastasis. Included participants reported a family history of glioma in first-, second-, or third-degree relatives. All participating institutions received institutional review board approval.

Control cohort

A set of 1013 non-Hispanic white control individuals from the ARIC cohort, aged 45 to 64 years, were randomly selected from four communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban Minneapolis, Minnesota; and Washington County, Maryland. These samples were not specifically chosen to be age- or sex-matched with the case cohort but were chosen because of identical sequencing and processing to the case cohort to serve as a filter for artifactual findings. Samples from these participants underwent WGS concurrently with those from case participants and were analyzed in an identical manner to act as a control cohort (*61*).

Sequencing and alignment

WGS of DNA extracted from the peripheral blood was performed using the HiSeqX (Illumina Inc.) platform (*62*). Genomes were sequenced to an average redundant depth of 30×. Reads were aligned to HG19 using Burrows-Wheeler alignment, with duplicate reads removed before variant calling using platypus (*63*). Reads were annotated with CASSANDRA (*64*).

Coding and noncoding small and structural variant annotation and filtering

Variants were filtered out if they were not shared by all affected, sequenced family members, where applicable. All variants were filtered on the basis of MAF < 1 of 10,000 as measured by gnomAD "overall" (*65*). In addition, known pathogenic or likely pathogenic variants (any level of evidence) in known CPGs in ClinVar (*25*) were selected regardless of the MAF. Variants were then selected on the basis of the following criteria:

1) Coding variants were selected for either LoF (1-methionine, frameshift, canonical splice, stop-gain) or highly deleterious nonsynonymous variants. Deleterious variants were defined as having a CADD PH score of 30 or greater, which is equivalent to the top 0.1% most deleterious variants (*66*).

2) Intronic variants were filtered for deep-intronic variants (>5 bp from the exon, i.e., excluding the canonical splice region) and based on predicted effects on splicing by generating a cryptic splice acceptor or donor (score > 0.8 by spliceAI) (*67*).

3) TFBS variants were selected if they were 1 kbp upstream of a transcription start site and occurred in a Factorbook high-quality site (score > 2.0) (*68*).

4) Noncoding variants that were previously identified by GWAS as associated with glioma predisposition were selected (*18*).

5) Structural variants were identified using both the DELLY suite of tools and an in-house developed tool that uses a normalized set of genomes as a background for identifying CNVs (*69*).

Variant gene selection

All variants were further filtered if they affected a gene with the following properties: (i) known CPGs, (ii) novel CPGs as determined by an increase in pLoFI when considering The Cancer Genome Atlas (TCGA) germline samples versus the entire gnomAD cohort (see below), and (iii) genes that were LoF intolerant (pLoFI > 0.5) and missense intolerant (Z > 3.0).

Novel CPGs

gnomAD (r03 March 2016 release) data were downloaded for the entire dataset as well as a dataset without TCGA data. Genes with increased pLoFI values in the TCGA-cleaned dataset compared with those in the gnomAD dataset in its entirety were depleted of LoF variants, i.e., an increased number of variants existed among TCGA germline samples. We defined these cancer-specific, LoF-intolerant genes as any genes with pLoFI > 0.8 in the TCGA-cleaned data, a difference of >0.2 between the pLoFI in the TCGA-cleaned data, or a full ExAc OR difference of >0.4.

Validation

All insertion or deletion events were validated by Sanger sequencing. SNVs were validated by Sanger sequencing if they had fewer than five reads (or <25% of the total reads) supporting the variant.

Plasmid constructs

All gRNAs used for the in vivo barcode enrichment competition assay LoF screening were designed using CRISPOR (http://crispor.tefor.net/crispor.py). The gRNAs were generated with barcode sequences through a previously described HiTMMoB approach in a PB (PBCAG) vector by gateway cloning (*1*). For individual LoF studies with *ZC3H7B*, *DMBT1*, and *HP1BP3* gRNAs, each gRNA was generated in a Px330-Crispr-mCherry vector by gateway cloning (*ZC3H7B*, 5′-GGTCGTTGCCCTCGGCAAAC-3′; *DMBT1*, 5′-CTCCACAGGCCACGTCATCC-3′; and *HP1BP3*, 5′-GAATTCACGGCTCGACGTAT-3′). For individual GoF studies, the *ZC3H7B* ORF was obtained from Origin (catalog no. RC219566). The *DMBT1* and *HP1BP3* ORFs were obtained from Thermo Fisher Scientific (Ultimate ORF clones). The efficiency of individual

vectors was confirmed by staining the brain (LoF vectors) and NIH3T3 cells (GoF vectors) (fig. S4).

In vivo functional screening

All mouse gliomas were generated on the CD-1 IGS mouse background as previously described (*31*). IUE was performed on embryonic day 15. Previously generated CRISPR constructs were used to knock out *NF1*, *PTEN*, and *p53* (1.5 μg/μl each). gRNAs were genomically integrated and overexpressed using the PB transposase (PBase) system. The pGlast-PBase plasmid (2.0 μg/μl) (*34*) was co-electroporated with a PBCAG-gRNA construct (1.0 μg/μl).

In vivo barcode enrichment competition assay

A PB transposable vector (PBCAG-EGFP-T2A-GWR1R4) was engineered from the PBCAG-EGFP construct (*32*). The EGFP STOP codon was removed, and an in-frame T2A sequence followed by attR1/attR4 Gateway cloning sites was inserted. These attR sites flanked chloramphenicol and ccdB selection cassettes. A V5 tag sequence was also inserted downstream of the attR4 site. gRNAs and barcode sequences were cloned using the HiTMMoB approach (*70*). The gRNAs were designed using standard approaches and cloned into a PB transposase system containing a unique barcode identifier. Pooled injection cocktails were assembled such that the pool of tested gRNA totaled 1.0 μg/μl. Briefly, an equal number of moles of each plasmid was mixed together, ethanol-precipitated, and redissolved/concentrated in double distilled H_2O. This pooled plasmid mixture was diluted to 1.0 μg/μl in the final IUE injection cocktail. After IUE of the pooled cocktail (with 3xCr, PBase, and GFPt2a-Luc), mice were born and observed for symptoms suggestive of brain tumors (see below). Upon demonstration of symptoms, the animal was euthanized, and tumors were examined using a fluorescence reporter. All procedures were approved by the Institutional Animal Care and Use Committee at the Baylor College of Medicine and conformed to the U.S. Public Health Service Policy on Human Care and Use of Laboratory Animals.

Subsequently, genomic DNA (gDNA) was prepared with the EZNA Tissue DNA Kit (Omega Bio-tek, D3396), according to the manufacturer′s instructions. Samples were prepared in biological and technical replicates ($3 \leq N_{biological} \leq 5$; $N_{technical} = 3$). In addition, the IUE injection cocktail was used for input and prepared in technical duplicates. Following gDNA isolation, barcode libraries were prepared as previously reported (*31*, *71*). Polymerase chain reactions (PCRs) were used to amplify the barcode pools from 50 ng of gDNA (experimental samples) or 2 ng of the plasmid pool (input control) using Platinum Super Mix (Thermo Fisher Scientific, 12532016) with primers targeting the T3 promoter site (directly upstream of the barcode) (5′-CAATTAACCCTCACTAAAGG-3′) and the V5 tag (5′-ACCGAGGAGAGGGTTAGGGAT-3′). The amplification parameters were as follows: 1× (94°C, 4 min); 35× (94°C, 1 min; 54°C, 1 min; 68°C, 1 min); 1× (68°C, 10 min); 10°C, hold. The PCR products were purified with the PureLink PCR Purification Kit (Thermo Fisher Scientific, K310001), processed using the Ion Plus Library Kit (Thermo Fisher Scientific, 4471252), subsequently purified, and ligated to unique Ion Xpress Barcode Adaptors (Thermo Fisher Scientific, 4474517). The resulting Ion Xpress barcoded libraries were amplified, purified, and pooled for Personal Genome Machine sequencing (318 V2 Chip) following the manufacturer′s recommendations. Raw data were concatenated into one "reference" file and indexed using the Burrows-Wheeler alignment tool40 for alignment of barcode sequences (with parameters "-l7 -t12 -N -n3") to count the occurrence of each barcode. Barcode enrichment was assessed by quantitating the number of occurrences of each barcode sequence as a ratio to the total number of barcode reads in each sample. The standard error was calculated across replicates and plotted as error bars on the barcode enrichment graphs.

To assay in vivo cell proliferation, 4 hours before harvesting, 100 μg of BrdU [in phosphate-buffered saline (PBS)] per gram body mass was intraperitoneally injected. Mouse brains were collected, frozen, and sectioned as described above. Before blocking, the sections were incubated in 2 N HCl at 37°C for 30 min and neutralized with 3.8% sodium borate for 10 min at room temperature.

Validation of gRNA vectors

For full methods, see Yu *et al.* (*31*). Briefly, for genomic editing validation of individual gRNA vectors, tumor tissues were collected for analysis at post-natal day 60, and tumors were resected and subjected to gDNA isolation using the DNeasy Blood and Tissue Kit (Qiagen) for sequencing. Genomic specific primers were used to amplify amplicons surrounding the putative CRISPR-Cas9–modified sites from gDNA-derived ΔZC3H7B, ΔDMBT1, and ΔHP1BP3 tumors. The targeted genomic areas flanking single guide RNA (sgRNA) binding sites were PCR-amplified using Phusion High-Fidelity DNA Polymerase (NEB M0530S). PCR products were purified using Ampure XP magnetic beads at 1.8× volume to separate amplicons from primers and were inspected using the dsDNA 900 Reagent Kit (DNF-900-K0500, AATI) on the Advanced Analytical 12-Capillary Fragment Analyzer. The verified product was then tagmented, amplified, and cleaned up using the Nextera XT DNA Library Prep Kit (FC-131-1024, Illumina). The regulating library was validated using the Standard Sensitivity NGS Fragment Analysis Kit on the 12-Capillary Fragment Analyzer, and samples were normalized to 2 nM using Ampure XP bead-based normalization. Samples were sequenced on Illumina NextSeq550 following the manufacturer′s instructions. Subsequently, reads were aligned to reference sequences of the PCR amplicons and evaluated for the presence of genomic alterations using custom script (fig. S5).

Tissue preparation

For immunostaining, mice were anesthetized and transcardially perfused first with saline solution and then with 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.2). Brains were stored at 4°C in 4% paraformaldehyde for 2 days and then stored in a 20% sucrose solution until they sank. Six separate series of 40-μm coronal brain sections were obtained using a Model CM3050S cryostat (Leica, Wetzlar, Germany) and stored in an anti-freeze stock solution (phosphate buffer containing 30% glycerol and 30% ethylene glycol, pH 7.2) at 4°C before use.

Immunohistochemistry and immunofluorescence staining

For immunofluorescence staining, every sixth serial section in each set (approximately eight sections) was collected, rinsed twice with PBS containing 0.2% Triton X-100 (PBST), and rinsed once with PBST. After nonspecific binding was blocked by incubating the sections with 1% bovine serum albumin in PBST, the sections were incubated overnight at room temperature with primary antibodies [GFP (Abcam, ab13970), ZC3H7B (Invitrogen, Carlsbad, CA, USA, PA5-106537), DMBT1 (Invitrogen, PA5-115313), HP1BP3

(Oroteintech, 24556-I-AP), FGF2 (Invitrogen, PA5-116495), and VEGF (Invitrogen, MA5-13182)]. Immunoreactive proteins were visualized using Alexa Fluor 488– or Alexa Fluor 555–conjugated secondary antibodies (1:600 dilution; Invitrogen). Images were captured using a confocal microscope (Carl Zeiss, Oberkochen, Germany).

For hematoxylin and eosin staining, 10-μm paraffin-embedded sections were processed using the following solutions: xylene (10 min ×2), 100% EtOH (10 min), 95% EtOH (10 min), 80% EtOH (10 min), 70% EtOH (10 min), ddH_2O (5 min), Harris hematoxylin (2 min; Poly Scientific R&D Corp., S212A), running tap water wash, 95% EtOH (30 s), eosin (2 min, Poly Scientific R&D Corp., S176), 95% EtOH (2 min x2), 100% EtOH (2 min ×2), and xylene (10 min ×2).

A BrdU assay was performed to confirm in vivo cell proliferation. Four hours before harvesting, 100 μg of BrdU (in PBS) per gram body mass was intraperitoneally injected.

RNA extraction, library preparation, sequencing, and bioinformatic analysis

Mouse brains were harvested from each experimental group at postnatal day 60. Tissues were rinsed with PBS and dissociated in TRIzol (Thermo Fisher Scientific, 15596018). Total RNA was extracted using the RNeasy Mini Kit (Qiagen, 74106) according to the manufacturer's instructions. Illumina sequencing libraries with 6-bp single indices were constructed from 300 ng of total RNA using a TruSeq Stranded mRNA LT kit (Illumina, RS-122-2101). Equal concentrations (2 nM) of libraries were pooled and subjected to paired-end sequencing of approximately 20 million to 30 million reads per sample using a Mid Output v2 kit (Illumina, 20024904) on Illumina NextSeq550 according to the manufacturer's instructions.

Sequencing files from each flow cell lane were downloaded in fastq files and merged. Quality control was performed using fastQC (v0.11.7) and MultiQC (v1.11). Reads were mapped to the mouse genome mm10 assembly using STAR (v2.5.0a) (*72*). Mapped reads were used to build count matrices using the Bioconductor packages GenomicAlignments (v1.26.0) and GenomicFeatures (v1.42.2) in R (v4.1.2) (*73*). UCSC transcripts were downloaded from Illumina iGenomes in GTF file format. DESeq2 (v1.30.1) (*74*) was used for normalization and DEG analysis. GOs were determined using Enrichr and the GO resource (http://geneontology.org/), and significant GO terms with P values of <0.05 were selected for visualization using ggplot2 (v3.3.5) and GOplot (v1.0.2). Gene expression heatmaps were generated using ComplexHeatmap (v2.6.2).

Supplementary Materials

This PDF file includes:
Figs. S1 to S5
Legends for tables S1, S8 to S10
Tables S2 to S7, S11 and S12

Other Supplementary Material for this manuscript includes the following:
Tables S1, S8 to S10

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. M. N. Bainbridge, Determining the incidence of rare diseases. *Hum. Genet.* **139**, 569–574 (2020).
2. E. R. Laws, I. F. Parney, W. Huang, F. Anderson, A. M. Morris, A. Asher, K. O. Lillehei, M. Bernstein, H. Brem, A. Sloan, M. S. Berger, S. Chang, Survival following surgery and prognostic factors for recently diagnosed malignant glioma: Data from the Glioma Outcomes Project. *J. Neurosurg.* **99**, 467–473 (2003).
3. H. Ohgaki, P. Kleihues, Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J. Neuropathol. Exp. Neurol.* **64**, 479–489 (2005).
4. Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-Sloan, CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2013-2017. *Neuro-Oncol.* **22**, iv1–iv96 (2020).
5. B. Malmer, H. Grönberg, A. T. Bergenheim, P. Lenner, R. Henriksson, Familial aggregation of astrocytoma in northern Sweden: An epidemiological cohort study. *Int. J. Cancer* **81**, 366–370 (1999).
6. M. Wrensch, M. Lee, R. Miike, B. Newman, G. Barger, R. Davis, J. Wiencke, J. Neuhaus, Familial and personal medical history of cancer and nervous system conditions among adults with glioma and controls. *Am. J. Epidemiol.* **145**, 581–593 (1997).
7. C. Jones, L. Perryman, D. Hargrave, Paediatric and adult malignant glioma: Close relatives or distant cousins? *Nat. Rev. Clin. Oncol.* **9**, 400–413 (2012).
8. S. Sadetzki, R. Bruchim, B. Oberman, G. N. Armstrong, C. C. Lau, E. B. Claus, J. S. Barnholtz-Sloan, D. Il'yasova, J. Schildkraut, C. Johansen, R. S. Houlston, S. Shete, C. I. Amos, J. L. Bernstein, S. H. Olson, R. B. Jenkins, D. Lachance, N. A. Vick, R. Merrell, M. Wrensch, F. G. Davis, B. J. McCarthy, R. Lai, B. S. Melin, M. L. Bondy; Gliogene Consortium, Description of selected characteristics of familial glioma patients—Results from the Gliogene Consortium. *Eur. J. Cancer* **49**, 1335–1345 (2013).
9. X. Sun, J. Vengoechea, R. Elston, Y. Chen, C. I. Amos, G. Armstrong, J. L. Bernstein, E. Claus, F. Davis, R. S. Houlston, D. Il'yasova, R. B. Jenkins, C. Johansen, R. Lai, C. C. Lau, Y. Liu, B. J. McCarthy, S. H. Olson, S. Sadetzki, J. Schildkraut, S. Shete, R. Yu, N. A. Vick, R. Merrell, M. Wrensch, P. Yang, B. Melin, M. L. Bondy, J. S. Barnholtz-Sloan; Gliogene Consortium, A variable age of onset segregation model for linkage analysis, with correction for ascertainment, applied to glioma. *Cancer Epidemiol. Biomark. Prev.* **21**, 2242–2251 (2012).
10. K. Czene, P. Lichtenstein, K. Hemminki, Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. *Int. J. Cancer* **99**, 260–266 (2002).
11. P. Lichtenstein, N. V. Holm, P. K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe, K. Hemminki, Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
12. D. Malkin, J. E. Garber, L. C. Strong, S. H. Friend, The cancer predisposition revolution. *Science* **352**, 1052–1053 (2016).
13. L. S. Friedman, E. A. Ostermeyer, C. I. Szabo, P. Dowd, E. D. Lynch, S. E. Rowell, M.-C. King, Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat. Genet.* **8**, 399–404 (1994).
14. S. Srivastava, Z. Q. Zou, K. Pirollo, W. Blattner, E. H. Chang, Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome. *Nature* **348**, 747–749 (1990).
15. M. N. Bainbridge, G. N. Armstrong, M. M. Gramatges, A. A. Bertuch, S. N. Jhangiani, H. Doddapaneni, L. Lewis, J. Tombrello, S. Tsavachidis, Y. Liu, A. Jalali, S. E. Plon, C. C. Lau, D. W. Parsons, E. B. Claus, J. Barnholtz-Sloan, D. Il'yasova, J. Schildkraut, F. Ali-Osman, S. Sadetzki, C. Johansen, R. S. Houlston, R. B. Jenkins, D. Lachance, S. H. Olson, J. L. Bernstein, R. T. Merrell, M. R. Wrensch, K. M. Walsh, F. G. Davis, R. Lai, S. Shete, K. Aldape, C. I. Amos, P. A. Thompson, D. M. Muzny, R. A. Gibbs, B. S. Melin, M. L. Bondy; Gliogene Consortium, Germline mutations in shelterin complex genes are associated with familial glioma. *J. Natl. Cancer Inst.* **107**, 384 (2014).
16. N. J. Roberts, A. L. Norris, G. M. Petersen, M. L. Bondy, R. Brand, S. Gallinger, R. C. Kurtz, S. H. Olson, A. K. Rustgi, A. G. Schwartz, E. Stoffel, S. Syngal, G. Zogopoulos, S. Z. Ali, J. Axilbund, K. G. Chaffee, Y.-C. Chen, M. L. Cote, E. J. Childs, C. Douville, F. S. Goes, J. M. Herman, C. Iacobuzio-Donahue, M. Kramer, A. Makohon-Moore, R. W. McCombie, K. W. McMahon, N. Niknafs, J. Parla, M. Pirooznia, J. B. Potash, A. D. Rhim, A. L. Smith, Y. Wang, C. L. Wolfgang, L. D. Wood, P. P. Zandi, M. Goggins, R. Karchin, J. R. Eshleman, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, R. H. Hruban, A. P. Klein, Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov.* **6**, 166–175 (2016).
17. M. E. Scheurer, C. J. Etzel, M. Liu, R. El-Zein, G. E. Airewele, B. Malmer, K. D. Aldape, J. S. Weinberg, W. K. A. Yung, M. L. Bondy, Aggregation of cancer in first-degree relatives of patients with glioma. *Cancer Epidemiol. Biomark. Prev.* **16**, 2491–2495 (2007).

18. B. S. Melin, J. S. Barnholtz-Sloan, M. R. Wrensch, C. Johansen, D. Il'yasova, B. Kinnersley, Q. T. Ostrom, K. Labreche, Y. Chen, G. Armstrong, Y. Liu, J. E. Eckel-Passow, P. A. Decker, M. Labussière, A. Idbaih, K. Hoang-Xuan, A.-L. Di Stefano, K. Mokhtari, J.-Y. Delattre, P. Broderick, P. Galan, K. Gousias, J. Schramm, M. J. Schoemaker, S. J. Fleming, S. Herms, S. Heilmann, M. M. Nöthen, H.-E. Wichmann, S. Schreiber, A. Swerdlow, M. Lathrop, M. Simon, M. Sanson, U. Andersson, P. Rajaraman, S. Chanock, M. Linet, Z. Wang, M. Yeager; GliomaScan Consortium, J. K. Wiencke, H. Hansen, L. McCoy, T. Rice, M. L. Kosel, H. Sicotte, C. I. Amos, J. L. Bernstein, F. Davis, D. Lachance, C. Lau, R. T. Merrell, J. Shildkraut, F. Ali-Osman, S. Sadetzki, M. Scheurer, S. Shete, R. K. Lai, E. B. Claus, S. H. Olson, R. B. Jenkins, R. S. Houlston, M. L. Bondy, Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* **49**, 789–794 (2017).
19. F. S. Collins, P. O'Connell, B. A. Ponder, B. R. Seizinger, Progress towards identifying the neurofibromatosis (NF1) gene. *Trends Genet.* **5**, 217–221 (1989).
20. T. Kobayashi, Y. Hirayama, E. Kobayashi, Y. Kubo, O. Hino, A germline insertion in the tuberous sclerosis (Tsc2) gene gives rise to the Eker rat model of dominantly inherited cancer. *Nat. Genet.* **9**, 70–74 (1995).
21. W. D. Foulkes, T. Y. Flanders, P. M. Pollock, N. K. Hayward, The CDKN2A (p16) gene and human cancer. *Mol. Med.* **3**, 5–20 (1997).
22. A. Jalali, K. Yu, V. Beechar, N. A. B. Huerta, A. Grichuk, D. Mehra, B. Lozzi, K. Kong, K. L. Scott, G. Rao, M. N. Bainbridge, M. L. Bondy, B. Deneen, POT1 regulates proliferation and confers sexual dimorphism in glioma. *Cancer Res.* **81**, 2703–2713 (2021).
23. M. N. Bainbridge, W. Wiszniewski, D. R. Murdock, J. Friedman, C. Gonzaga-Jauregui, I. Newsham, J. G. Reid, J. K. Fink, M. B. Morgan, M.-C. Gingras, D. M. Muzny, L. D. Hoang, S. Yousaf, J. R. Lupski, R. A. Gibbs, Whole-genome sequencing for optimized patient management. *Sci. Transl. Med.* **3**, 87re3 (2011).
24. J. Bressler, M. Fornage, E. W. Demerath, D. S. Knopman, K. L. Monda, K. E. North, A. Penman, T. H. Mosley, E. Boerwinkle, Fat mass and obesity gene and cognitive decline: The atherosclerosis risk in communities study. *Neurology* **80**, 92–99 (2013).
25. M. J. Landrum, J. M. Lee, M. Benson, G. R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J. B. Holmes, B. L. Kattman, D. R. Maglott, ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
26. R. G. H. Lindeboom, F. Supek, B. Lehner, The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118 (2016).
27. G. Reznick Levi, G. Larom, V. Ofen Glassner, N. Ekhilevitch, N. Sharon Swartzman, T. Paperna, H. Baris-Feldman, K. Weiss, A recurrent pathogenic BRCA2 exon 5–11 duplication in the Christian Arab population in Israel. *Fam. Cancer* **21**, 289–294 (2022).
28. G.-Z. Jin, Y. Zhang, W.-M. Cong, X. Wu, X. Wang, S. Wu, S. Wang, W. Zhou, S. Yuan, H. Gao, G. Yu, W. Yang, Phosphoglucomutase 1 inhibits hepatocellular carcinoma progression by regulating glucose trafficking. *PLOS Biol.* **16**, e2006483 (2018).
29. Expression of FOXJ2 in glioma—The Human Protein Atlas; www.proteinatlas.org/ENSG00000065970-FOXJ2/pathology/glioma.
30. Expression of KAT2B in glioma—The Human Protein Atlas; www.proteinatlas.org/ENSG00000114166-KAT2B/pathology/glioma.
31. K. Yu, C.-C. J. Lin, A. Hatcher, B. Lozzi, K. Kong, E. Huang-Hobbs, Y.-T. Cheng, V. B. Beechar, W. Zhu, Y. Zhang, F. Chen, G. B. Mills, C. A. Mohila, C. J. Creighton, J. L. Noebels, K. L. Scott, B. Deneen, PIK3CA variants selectively initiate brain hyperactivity during gliomagenesis. *Nature* **578**, 166–171 (2020).
32. F. Chen, J. LoTurco, A method for stable transgenesis of radial glia lineage in rat neocortex by piggyBac mediated transposition. *J. Neurosci. Methods* **207**, 172–180 (2012).
33. U. Andersson, C. Wibom, K. Cederquist, S. Aradottir, A. Borg, G. N. Armstrong, S. Shete, C. C. Lau, M. N. Bainbridge, E. B. Claus, J. Barnholtz-Sloan, R. Lai, D. Il'yasova, R. S. Houlston, J. Schildkraut, J. L. Bernstein, S. H. Olson, R. B. Jenkins, D. H. Lachance, M. Wrensch, F. G. Davis, R. Merrell, C. Johansen, S. Sadetzki; Gliogene Consortium, M. L. Bondy, B. S. Melin, Germline rearrangements in families with strong family history of glioma and malignant melanoma, colon, and breast cancer. *Neuro-Oncol.* **16**, 1333–1340 (2014).
34. S. Bekker-Jensen, N. Mailand, Assembly and function of DNA double-strand break repair foci in mammalian cells. *DNA Repair* **9**, 1219–1228 (2010).
35. M. Cubillos-Rojas, F. Amair-Pinedo, R. Peiró-Jordán, R. Bartrons, F. Ventura, J. L. Rosa, The E3 ubiquitin protein ligase HERC2 modulates the activity of tumor protein p53 by regulating its oligomerization. *J. Biol. Chem.* **289**, 14782–14795 (2014).
36. M. Zhu, H. Zhao, J. Liao, X. Xu, HERC2/USP20 coordinates CHK1 activation by modulating CLASPIN stability. *Nucleic Acids Res.* **42**, 13074–13081 (2014).
37. V. Frattini, V. Trifonov, J. M. Chan, A. Castano, M. Lia, F. Abate, S. T. Keir, A. X. Ji, P. Zoppoli, F. Niola, C. Danussi, I. Dolgalev, P. Porrati, S. Pellegatta, A. Heguy, G. Gupta, D. J. Pisapia, P. Canoll, J. N. Bruce, R. E. McLendon, H. Yan, K. Aldape, G. Finocchiaro, T. Mikkelsen, G. G. Privé, D. D. Bigner, A. Lasorella, R. Rabadan, A. Iavarone, The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* **45**, 1141–1149 (2013).
38. M. Cubillos-Rojas, T. Schneider, O. Hadjebi, L. Pedrazza, J. R. de Oliveira, F. Langa, J.-L. Guénet, J. Duran, J. M. de Anta, S. Alcántara, R. Ruiz, E. M. Pérez-Villegas, F. J. Aguilar, Á. M. Carrión, J. A. Armengol, E. Baple, A. H. Crosby, R. Bartrons, F. Ventura, J. L. Rosa, The HERC2 ubiquitin ligase is essential for embryonic development and regulates motor coordination. *Oncotarget* **7**, 56083–56106 (2016).
39. J. García-Cano, S. Sánchez-Tena, J. Sala-Gaston, A. Figueras, F. Viñals, R. Bartrons, F. Ventura, J. L. Rosa, Regulation of the MDM2-p53 pathway by the ubiquitin ligase HERC2. *Mol. Oncol.* **14**, 69–86 (2020).
40. M. H. Orme, G. Liccardi, N. Moderau, R. Feltham, S. Wicky-John, T. Tenev, L. Aram, R. Wilson, K. Bianchi, O. Morris, C. Monteiro Domingues, D. Robertson, M. Tare, A. Wepf, D. Williams, A. Bergmann, M. Gstaiger, E. Arama, P. S. Ribeiro, P. Meier, The unconventional myosin CRINKLED and its mammalian orthologue MYO7A regulate caspases in their signalling roles. *Nat. Commun.* **7**, 10972 (2016).
41. J. L. Sallee, J. M. Crawford, V. Singh, D. P. Kiehart, Mutations in *Drosophila* crinkled/myosin VIIA disrupt denticle morphogenesis. *Dev. Biol.* **470**, 121–135 (2021).
42. M. F. L. Derks, H.-J. Megens, W. L. Giacomini, M. A. M. Groenen, M. S. Lopes, A natural knockout of the MYO7A gene leads to pre-weaning mortality in pigs. *Anim. Genet.* **52**, 514–517 (2021).
43. M. Ali, C. D. Delozier, U. Chaudhary, BRIP-1 germline mutation and its role in colon cancer: Presentation of two case reports and review of literature. *BMC Med. Genet.* **20**, 75 (2019).
44. F. Bellido, M. Pineda, G. Aiza, R. Valdés-Mas, M. Navarro, D. A. Puente, T. Pons, S. González, S. Iglesias, E. Darder, V. Piñol, J. L. Soto, A. Valencia, I. Blanco, M. Urioste, J. Brunet, C. Lázaro, G. Capellá, X. S. Puente, L. Valle, POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: Review of reported cases and recommendations for genetic testing and surveillance. *Genet. Med.* **18**, 325–332 (2016).
45. A. Rohlin, T. Zagoras, S. Nilsson, U. Lundstam, J. Wahlström, L. Hultén, T. Martinsson, G. B. Karlsson, M. Nordling, A mutation in POLE predisposing to a multi-tumour phenotype. *Int. J. Oncol.* **45**, 77–81 (2014).
46. C. Palles, J.-B. Cazier, K. M. Howarth, E. Domingo, A. M. Jones, P. Broderick, Z. Kemp, S. L. Spain, E. Guarino, I. Salguero, A. Sherborne, D. Chubb, L. G. Carvajal-Carmona, Y. Ma, K. Kaur, S. Dobbins, E. Barclay, M. Gorman, L. Martin, M. B. Kovac, S. Humphray; CORGI Consortium; WGS500 Consortium, A. Lucassen, C. C. Holmes, D. Bentley, P. Donnelly, J. Taylor, C. Petridis, R. Roylance, E. J. Sawyer, D. J. Kerr, S. Clark, J. Grimes, S. E. Kearsey, H. J. W. Thomas, G. McVean, R. S. Houlston, I. Tomlinson, Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45**, 136–144 (2013).
47. B. Malmer, S. Haraldsson, E. Einarsdottir, P. Lindgren, D. Holmberg, Homozygosity mapping of familial glioma in Northern Sweden. *Acta Oncol.* **44**, 114–119 (2005).
48. S. Shete, C. C. Lau, R. S. Houlston, E. B. Claus, J. Barnholtz-Sloan, R. Lai, D. Il'yasova, J. Schildkraut, S. Sadetzki, C. Johansen, J. L. Bernstein, S. H. Olson, R. B. Jenkins, P. Yang, N. A. Vick, M. Wrensch, F. G. Davis, B. J. McCarthy, E. H.-C. Leung, C. Davis, R. Cheng, F. J. Hosking, G. N. Armstrong, Y. Liu, R. K. Yu, R. Henriksson; Gliogene Consortium, B. S. Melin, M. L. Bondy, Genome-wide high-density SNP linkage search for glioma susceptibility loci: Results from the Gliogene Consortium. *Cancer Res.* **71**, 7568–7575 (2011).
49. N. Paunu, P. Lahermo, P. Onkamo, V. Ollikainen, I. Rantala, P. Helén, K. O. J. Simola, J. Kere, H. Haapasalo, A novel low-penetrance locus for familial glioma at 15q23-q26.3. *Cancer Res.* **62**, 3798–3802 (2002).
50. B. R. Lajoie, J. Dekker, N. Kaplan, The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods* **72**, 65–75 (2015).
51. Y. Wang, F. Song, B. Zhang, L. Zhang, J. Xu, D. Kuang, D. Li, M. N. K. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, F. Yue, The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).
52. Y. Oktay, E. Ülgen, Ö. Can, C. B. Akyerli, Ş. Yüksel, Y. Erdemgil, İ. M. Durası, O. I. Henegariu, E. P. Nanni, N. Selevsek, J. Grossmann, E. Z. Erson-Omay, H. Bai, M. Gupta, W. Lee, Ş. Turcan, A. Özpınar, J. T. Huse, M. A. Sav, A. Flanagan, M. Günel, O. U. Sezerman, M. C. Yakıcıer, M. N. Pamir, K. Özduman, IDH-mutant glioma specific association of rs55705857 located at 8q24.21 involves MYC deregulation. *Sci. Rep.* **6**, 27569 (2016).
53. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski; Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, D. G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

54. T. Li, Y. Shi, P. Wang, L. M. Guachalla, B. Sun, T. Joerss, Y.-S. Chen, M. Groth, A. Krueger, M. Platzer, Y.-G. Yang, K. L. Rudolph, Z.-Q. Wang, Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* **34**, 1630–1647 (2015).
55. R. Kleene, M. Mzoughi, G. Joshi, I. Kalus, U. Bormann, C. Schulze, M.-F. Xiao, A. Dityatev, M. Schachner, NCAM-induced neurite outgrowth depends on binding of calmodulin to NCAM and on nuclear import of NCAM and fak fragments. *J. Neurosci.* **30**, 10784–10798 (2010).
56. T. N. Caza, S. I. Hassen, M. Kuperman, S. G. Sharma, Z. Dvanajscak, J. Arthur, R. Edmondson, A. Storey, C. Herzog, D. J. Kenan, C. P. Larsen, Neural cell adhesion molecule 1 is a novel autoantigen in membranous lupus nephritis. *Kidney Int.* **100**, 171–181 (2021).
57. B. Dutta, Y. Ren, P. Hao, K. H. Sim, E. Cheow, S. Adav, J. P. Tam, S. K. Sze, Profiling of the chromatin-associated proteome identifies HP1BP3 as a novel regulator of cell cycle progression*. *Mol. Cell. Proteomics* **13**, 2183–2197 (2014).
58. I. Ionita-Laza, R. Ottman, Study designs for identification of rare disease variants in complex diseases: The utility of family-based designs. *Genetics* **189**, 1061–1068 (2011).
59. B. Malmer, P. Adatto, G. Armstrong, J. Barnholtz-Sloan, J. L. Bernstein, E. Claus, F. Davis, R. Houlston, D. Il'yasova, R. Jenkins, C. Johansen, R. Lai, C. Lau, B. M. Carthy, H. Nielsen, S. H. Olson, S. Sadetzki, S. Shete, F. Wiklund, M. Wrensch, P. Yang, M. Bondy, GLIOGENE an International Consortium to Understand Familial Glioma. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1730–1734 (2007).
60. D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, P. Kleihues, The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007).
61. L. E. Chambless, G. Heiss, A. R. Folsom, W. Rosamond, M. Szklo, A. R. Sharrett, L. X. Clegg, Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: The Atherosclerosis Risk in Communities (ARIC) Study, 1987–1993. *Am. J. Epidemiol.* **146**, 483–494 (1997).
62. S. Bennett, Solexa Ltd. *Pharmacogenomics* **5**, 433–438 (2004).
63. A. Rimmer, H. Phan, I. Mathieson, Z. Iqbal, S. R. F. Twigg; WGS500 Consortium, A. O. M. Wilkie, G. McVean, G. Lunter, Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
64. M. N. Bainbridge, H. Hu, D. M. Muzny, L. Musante, J. R. Lupski, B. H. Graham, W. Chen, K. W. Gripp, K. Jenny, T. F. Wienker, Y. Yang, V. R. Sutton, R. A. Gibbs, H. H. Ropers, De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med.* **5**, 11 (2013).
65. T. Soussi, B. Leroy, M. Devir, S. Rosenberg, High prevalence of cancer-associated TP53 variants in the gnomAD database: A word of caution concerning the use of variant filtering. *Hum. Mutat.* **40**, 516–524 (2019).
66. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
67. K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J. F. McRae, S. F. Darbandi, D. Knowles, Y. I. Li, J. A. Kosmicki, J. Arbelaez, W. Cui, G. B. Schwartz, E. D. Chow, E. Kanterakis, H. Gao, A. Kia, S. Batzoglou, S. J. Sanders, K. K.-H. Farh, Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e24 (2019).
68. Factorbook; https://factorbook.org/.
69. T. Rausch, T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, J. O. Korbel, DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
70. Y. H. Tsang, T. Dogruluk, P. M. Tedeschi, J. Wardwell-Ozgo, H. Lu, M. Espitia, N. Nair, R. Minelli, Z. Chong, F. Chen, Q. E. Chang, J. B. Dennison, A. Dogruluk, M. Li, H. Ying, J. R. Bertino, M.-C. Gingras, M. Ittmann, J. Kerrigan, K. Chen, C. J. Creighton, K. Eterovic, G. B. Mills, K. L. Scott, Functional annotation of rare gene aberration drivers of pancreatic cancer. *Nat. Commun.* **7**, 10500 (2016).
71. T. Dogruluk, Y. H. Tsang, M. Espitia, F. Chen, T. Chen, Z. Chong, V. Appadurai, A. Dogruluk, A. K. Eterovic, P. E. Bonnen, C. J. Creighton, K. Chen, G. B. Mills, K. L. Scott, Identification of variant-specific functions of PIK3CA by rapid phenotyping of rare mutations. *Cancer Res.* **75**, 5341–5354 (2015).
72. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
73. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
74. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgments: We would like to thank the Gliogene, ARIC, and Genomics England study participants. **Funding:** This work was funded in part by an NIH grant (R01CA217105) awarded to M.N.B., M.L.B., and B.D. and an NHGRI grant (X01HG009883) to M.L.B. Samples were originally collected by the Gliogene consortium (table S11) (R01CA119215) and Glioma International Case Control Study (R01CA139020). Sequencing was funded in party by NHGRI grants (U54 HG003273 and UM1 HG008898) awarded to R.A.G. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care) (table S12). The 100,000 Genomes Project is funded by the National Institute for Health Research and National Health Service England. The Wellcome Trust, Cancer Research UK, and the Medical Research Council have also funded the research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. **Author contributions:** Conceived the study: M.L.B., G.A., B.D., B.M., and M.N.B. Wrote the manuscript: M.N.B. Data analysis: M.N.B., P.V., and T.C.W. Produced sequencing data: E.B., R.A.G., and D.M.M. Edited the manuscript: S.E.P., M.N.B., M.L.B., Q.T.O., B.M., T.C.W., and B.D. In vivo studies: B.D., H.-C.C., B.L., and D.-J.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Sequence data are available from dbGaP phs002250.v1.p1 (https://ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002250.v1.p1).

Submitted 4 August 2022
Accepted 22 March 2023
Published 28 April 2023
10.1126/sciadv.ade2675