Single-cell multi-omic integration analysis prioritizes druggable genes and reveals cell-type-specific causal effects in glioblastomagenesis

Yu-Feng Huang[1] Kun-Long Huang[1]
1 The First Clinical Medical School, Shanxi Medical University, Jinzhong, China.
*Correspondence to: Yu-Feng Huang huangyufeng@sxmu.edu.cn

Abstract

Gliomas, particularly aggressive glioblastoma (GBM), pose significant therapeutic challenges due to limited understanding of their single-cell drivers. Here, we integrate large glioma genetic data with brain multi-omics (bulk and single-cell) to identify causal genes and their cell-type-specific roles. We prioritize 11 high-confidence and 47 potential causal genes; 41 are novel associations. Analyses suggest most of these 58 genes are druggable, supported by CRISPR/RNAi screens showing essentiality/dependency for 53.7% novel candidates in glioma cell lines. Single-cell data identifies astrocytes and oligodendrocyte precursor cells (OPCs) as likely GBM cells-of-origin and reveals increased tumor microenvironment (TME) communication involving neurons. We uncover 14 cell-type-specific causal gene effects, including *EGFR* in astrocytes, *CDKN2A* in OPCs, and *JAK1* in excitatory neurons. Notably, 85.7% effects occur in non-risk populations (glial and neural), highlighting complex interplays. This study provides critical cell-resolved insights into glioma susceptibility mechanisms and identifies potential therapeutic targets within complex intratumoral interactions, advancing targeted precision therapies.

Keywords: genome-wide association studies, single-cell multi-omics, glioblastomagenesis, cell-type-specific causal genes

Introduction

Gliomas constitute nearly 30% of primary brain tumors and 80% of malignant cases, with glioblastoma (GBM) being the most prevalent and aggressive subtype, generally associated with a poor prognosis[1]. Despite extensive research, numerous experimental drugs have failed in clinical trials, highlighting the need for more precise and effective therapies[2,3]. However, a major barrier is the intratumoral heterogeneity of GBM[2,4], resulting from different cells of origin, diverse genetic and epigenetic mutations, and varied tumor microenvironments (TMEs) shaped by cell communication[5-7]. Hence, dissecting cell-type-specific mutations and molecular expression in both cancerous and non-cancerous cells is critical for deciphering gliomagenesis and developing precise interventions[5,8].

Human genetic evidence markedly improves the success rate of targeted therapies[9], making

large-scale glioma genome-wide association studies (GWAS) and subsequent analyses a powerful approach for understanding gliomagenesis mechanisms and discovering druggable molecules[10-12]. However, single-data approaches often show limited power and explain only a fraction of heritability[13,14]; lack of replication across genetic studies has led to high false-positive rates[15]; and bulk-tissue analyses fail to capture the cell-type-resolved biological processes and cell-type-specific genes[12], which are linked to higher clinical success[8].

Because most GWAS-identified risk variants lie in non-coding regions, integrating multi-omics data provides deeper insights for post-GWAS analyses. Protein quantitative trait loci (pQTL) and expression quantitative trait loci (eQTL) datasets offer complementary effects in target investigation[16]. Open chromatin regions and enhancer loops from brain-specific epigenomics data—including ATAC-seq, promoter capture Hi-C (pcHi-C), and the activity-by-contact (ABC) model—improve long-range interactional causal gene mapping[17-19]. Advances in single-cell multi-omics further extend our perspective to single-cell resolution. Large-scale single-cell and single-nucleus RNA sequencing (sc/snRNA-seq) capture global transcriptional profiles at the individual-cell level, thereby improving our understanding of cell-type differentiation and interactions[17,20]. Moreover, cell-type-specific ATAC-seq, ChIP-seq, proximity ligation-assisted ChIP-seq (PLAC-seq), and single-cell eQTL (sc-eQTL) explore cell-type-resolved regulatory consequences and help explain the missing heritability observed in bulk tissue analyses[17,21].

In this study, we integrate glioma GWAS with brain-specific multi-omics data—including transcriptomics, epigenomics, and proteomics—to identify causal genes for gliomagenesis. By evaluating druggability and target pleiotropy via sc/snRNA-seq, bulk RNA-seq, CRISPR/RNAi, and cross-phenotype GWAS, we prioritized many druggable genes. Single-cell multi-omics were further employed to elucidate cells of origin, TME interactions, and cell-type-specific effects in glioblastomagenesis. **Fig. 1** provides an overview of the main analytical steps conducted in the study.

Result

Integrating Brain Tissue-Specific Genetic Multi-Omics Data to Identify Causal Genes for Glioma

We systematically explored glioma potential causal genes by integrating large-scale GWAS with brain-specific multi-omics. Two main strategies were employed: a similarity-based method (combining gene expression, pathways, and predicted PPIs) and a locus-based method (incorporating LD data, open chromatin, enhancer loops, and QTL data).

Overall, we detected 24, 33, and 26 potential causal genes for pan-glioma, non-GBM, and GBM, respectively **(Supplementary Table 5 and 18)**. There were 41 novel glioma-associated molecules that we prioritized **(Supplementary Table 6)**. Cross-referencing two broad methods[14] yielded high-confidence causal genes: 6 in pan-glioma (*EGFR*, *TERT*, *MDM4*, *SOX8*, *CDKN2A*, *STMN3*), 5 in GBM (*EGFR*, *CDKN2A*, *JAK1*, *TERT*, *STMN3*), and 4 in non-GBM (*CDKN2B*, *IDH1*, *NPAS3*, *ZGPAT*). No high-confidence genes were shared by subtypes. Identified potential causal genes showed a significant enrichment in the cerebral cortex (FC = 2.74, P adj = 3.96E-2), indicating higher drug-target potential due to tissue-specific effects[22,23]. Pathway analyses revealed significant enrichment in cancer-related biological processes (**Fig.2a)**, while PPI analysis indicated significant associations among the identified protein-coding gene sets (pan-glioma, $P_{enrichment}$ = 1.54E-04; GBM, $P_{enrichment}$ = 7.86E-04; non-GBM, $P_{enrichment}$ = 1.18E-03) **(Fig.2b)**.

Verification of Biological Enrichment and Targetability

We assessed the targetability of the identified genes through glioma sc/snRNA-seq, bulk RNA-seq, and CRISPR/RNAi data. First, to evaluate biological enrichment in tumor, we analyzed three large-scale sc/snRNA-seq datasets, scoring GO term enrichment for the identified causal genes. We observed that the target-set-mediated pathways were significantly enriched in both malignant and non-cancerous cells (**Fig.2c-e)**, suggesting broad involvement of these genetic target-mediated pathways in glioma[24].

Next, we matched targets to DEGs (healthy vs. glioma) in five bulk RNA datasets (N total = 1,890). In pan-glioma, GBM, and non-GBM, 62.5–92.3% of identified genes were DEGs in ≥1 dataset, with 18.8–46.2% labeled as DEGs in ≥3 datasets with consistent effect directions. These findings suggest that the identified genes showed significant expression changes between tumor and normal tissues.

Since DEGs may mirror disease-driven expression changes rather than causality[25], we verified target selectivity and essentiality via genome-wide CRISPR loss-of-function and RNAi viability screens across over 1,300 tumor cell lines. In pan-glioma and its subtypes, the majority of identified genes (88.5–93.8%) were enriched in multiple tumor cell types. Among 94 glioma cell lines, 51.72% (30/58) of targets were selective or essential. Finally, we annotated the identified genes with ligandability assessments, 3D structural data, and compound information. (**Supplementary Tables 7, 11, and 15)**

Uncovering the Risk Cells in GBM Origins

To determine which cell types contribute to glioblastomagenesis, we combined GWAS data with large-scale GBM snRNA-seq[20]. Malignant cells (P adj = 1.03E-09), astrocytes (P adj = 1.14E-11), and oligodendrocyte precursor cells (OPCs, P adj = 1.51E-10) showed significant associations with GBM risk **(Fig.3b)**. Malignant cells in a proliferative state demonstrated strong differentiation capacities and elevated genetically associated pathway activity scores (gPAS), reflecting high enrichment of cell-specific pathway-based genetic variances[26]**(Fig.3c and 3d)**. Using adult-brain scATAC-seq, we assessed the cell-type-specific heritability enrichment of GBM in risk cells. OPCs displayed significant GBM heritability enrichment (44.10-fold; $P_{Enrichment} = 0.038$, $P_{Coefficient} = 0.042$), while astrocytes did not reach significance (31.97-fold; $P_{Enrichment} = 0.075$, $P_{Coefficient} = 0.082$). Pseudotime analyses, derived from GBM snRNA-seq, were consistent with the genetic findings **(Fig 3d-f)**.

Cell Interaction in the GBM TME

To investigate cell interactions in the GBM TME, we compared cell-cell communication among non-cancerous cells from patients and healthy controls. In total, 1,129,369 cells and 846 signaling genes were included. We observed significantly elevated cell-cell communication in GBM **(Supplementary Fig. 4)**, especially between neurons and risk cells (astrocytes, OPCs). **(Fig 3g)**.

Investigating Cell-Type-Specific Effects

We performed cell-type-specific causal gene analyses based on sc-eQTL data from eight brain cell types. 14 cell-type-gene pairs involving 12 genes were identified, with 58.3% being novel in brain tissue analyses. Three high-confidence causal genes—*JAK1*, *CDKN2A*, *EGFR*—showed cell-type-specific effects **(Fig 4a-e and Supplementary Fig. 18)**. In GBM risk cells, we detected cell-type-specific glioblastomagenesis effects, including *EGFR* in astrocytes and *CDKN2A* in OPCs. The majority of detected cell-type-specific effects (85.7%, 12/14) involved non-risk cells, including both glial and neural cells. Oligodendrocytes accounted for most of the cell-type-specific genes

(8/12). Excitatory neurons were enriched in high-neural GBM, and their neural-tumor interactions drive glioma growth[27]; their cell-type-specific causal gene *JAK1* was also a high-confidence causal gene in tissue-level analyses.

To evaluate cell-type-specific enrichment of identified genes, we performed conditional/unconditional enrichment analyses on large-scale GBM snRNA-seq. Among the significant cell-type-specific gene sets, only oligodendrocytes had ≥4 genes for further analysis. We found that oligodendrocyte-specific targets exhibited high specificity in oligodendrocytes (P adj = 3.92E-2). Even after controlling for OPCs, oligodendrocyte enrichment remained significant, suggesting independence from OPC enrichment. To further probe enrichment in other cell types, we evaluated cell-type-specific gene groups (significant in association studies and replicated). Each of the four cell-type-specific gene groups (neuron, astrocyte, OPC, oligodendrocyte) displayed maximal specific expression in their respective cell types; two (neuron, oligodendrocyte) reached significance (<0.05) following Bonferroni correction **(Supplementary Fig. 6)**.

We integrated GBM GWAS fine-mapping results with over 900 cell-type-specific cCREs to uncover cell-type-specific causal variants in glioblastomagenesis. We identified rs6964933, located within the *EGFR* gene body, as an astrocyte-specific causal variant (PIP = 0.10, Z-score = 1.99). This variant also appeared in the GWAS fine-mapping credible model set (P_{GWAS} = 2.19E-18, PIP = 0.18) **(Supplementary Table 20)**. Within the rs6964933 high LD SNP set identified by fine-mapping (N_{SNP}= 9, R^2 > 0.95, PIP_{sum} = 1), we detected high FORGEdb scores (range 6 to 10) , and many SNPs overlapped astrocyte-specific epigenomic peaks, suggesting a potential effect on cell-type-specific transcriptional regulation[28,29] **(Fig. 4g-h)**.

High *EGFR* Expression Related to Lower Glioblastomagenesis Risk

Epidermal growth factor receptor (*EGFR*) is a well-established clinical target for GBM. Its signaling has both tumor-promoting[30] and tumor-suppressing effects[31]. Our QTL analyses show that higher EGFR expression (mRNA/protein) correlates with lower GBM risk, particularly through astrocyte-specific effects **(Fig. 5a) (Supplementary Tables 11 and 19)**. Furthermore, robust colocalization of QTL and GWAS data was mediated by variants exhibiting opposing influences on *EGFR* levels and GBM risk **(Fig. 5b-f)**. In supervised pseudotime analyses, with astrocytes as root cells, *EGFR* expression varied across stages of astrocyte differentiation into malignant cells, culminating in AC-like cells **(Fig. 5g-h)**. Comparing normal brain cells to GBM samples, we observed significant activation of the EGF pathway (including *EGFR* and ligands) in the TME, notably in astrocytes **(Fig. 5i and Supplementary Fig. 7)**. This implies that both constitutive and ligand-driven *EGFR* signaling may function at different stages of glioblastomagenesis[32].

Exploring Causal Phenotypes and Target Gene Pleiotropy

Safety concerns and side effects frequently terminate oncology clinical trials[23], yet glioma risk factors remain unclear. To explore glioma causal phenotypes and genetic pleiotropy of target genes, we undertook several analyses. First, we examined the genetic architecture of 3,000 phenotypes, identifying 137 (pan-glioma), 67 (GBM), and 130 (non-GBM) significantly associated phenotypes (P adj < 0.05). GWAS summary data were collected from 260 phenotypes (126 pan-glioma, 105 GBM, 152 non-GBM). Using a two-sample MR framework, we performed a Mendelian randomization phenome-wide association study (MR-PheWAS) **(Supplementary Fig. 8)** After heterogeneity, pleiotropy, and MR sensitivity assessments, we found 7, 7, and 9 potential causal traits for pan-glioma, GBM, and non-GBM, respectively. In light of widespread genetic heritability, we applied the LCV analysis on potential genetically correlated (P<0.05) glioma-phenotype pairs[33].

There were 5, 6, and 5 suggestive causal traits identified for pan-glioma, GBM, and non-GBM, respectively **(Supplementary Tables 22)**. Only telomere length showed robust causal evidence for pan-glioma after Bonferroni correction (OR = 1.888, 95% CI (1.521 to 2.344), P_{IVW}= 8.37E-9, P_{HDL} = 3.31E-1). Notable trends emerged among other potential traits across various data. For instance, traits reflecting cognitive proficiency (e.g., good cognitive function, high educational attainment, high intelligence) correlated with reduced non-GBM risk. Conversely, schizophrenia, a disease characterized by cognitive impairment, aligned with a higher non-GBM risk. Finally, we conducted a gene-based PheWAS to characterize target pleiotropy[34]. Reported pleiotropy mostly concerned potential causal phenotypes. **(Supplementary Tables 23)**.

Drug Repurposing of Target Genes

To facilitate drug repurposing for identified targets, we queried databases to detect drugs with high gene-specific interaction scores (>0.9). We then annotated drug characteristics along with the latest clinical details. We collected 87 drugs targeting 28 genes in more than one clinical trial: 57 (65.51%) for tumor therapy, 21 (24.13%) for brain diseases, and 18 (20.69%) for glioma therapy. **(Supplementary Tables 24)**

Discussion

In this study, we integrated large-scale GWAS with brain-specific multi-omics (transcriptomic, proteomic, and epigenomic) data and employed two broad categories of gene prioritization strategies to investigate the causal genes for gliomagenesis. 47 potential and 11 high-confidence causal genes were identified. Evidence from large-scale sc/snRNA, bulk RNA, and CRISPR/miRNA data showed that these targets were significantly enriched in glioma pathways, differentially expressed (patients vs. controls), and selective or essential in cancer cell lines. To elucidate cell-type-specific effects, we integrated single-cell multi-omics data with GBM GWAS, enabling us to pinpoint cells of origin, examine TME cell interactions, and detect multiple cell-type-specific causal genes. We identified 12 causal genes mediated by cell-type-specific effects in risk and other non-malignant cells (both glial and neural cells), potentially linked to elevated cell interactions in the GBM TME. This underscores the importance of cell-type-level precision therapeutics. To refine target pleiotropy assessment and advance drug repositioning, we conducted PheWAS and drug repurposing analyses.

Compared to previous tissue-level studies[10-12,35-39], we prioritized 41 novel glioma-associated molecules, Notably, two of these genes were in genome-wide significant loci that had not been linked to any genes[10]. All previous glioma gene prioritization studies using locus-based methods have especially focused on the QTL-based method, which may relate to lower precision (14–46%) due to a fraction of heritability explained by QTLs and static power limitations[13]. We combined two different methods and highlighted 11 high-confidence causal targets for gliomagenesis, which provide a precision of at least 67% and up to 79% in previous benchmarked analyses. Notably, no high-confidence targets overlapped between GBM and non-GBM, reflecting glioma subtype heterogeneity[7].

GBM heterogeneity correlates with multiple cells of origin[1,7,40]. Here, astrocytes and OPCs emerged as GBM origin cells, and related cell-type-specific causal genes expand our understanding of glioblastomagenesis. *CDKN2A* is deleted in approximately 60% of GBM, sensitizing malignant cells to lipid peroxidation and ferroptosis[41]. Recent work demonstrates that methylation-silenced *CDKN2A*, along with OPC insulator loss, further drives gliomagenesis[42]. Here,

we determined its OPC-specific causal effect on glioblastomagenesis. *EGFR* is one of the most promising GBM therapy targets[3] , amplified in ~50% of GBM patients[43]. Recent study indicates EGFR-amplified malignancies tend to differentiate into astrocytic lineages, and in some cases, *EGFR* amplification is not involved in early glioblastomagenesis[44]. Here, we pinpointed an astrocyte-specific causal effect of *EGFR* and its variant rs6964933 in glioblastomagenesis. Interestingly, increased astrocyte-specific *EGFR* expression was associated with lower GBM risk, possibly associated with *EGFR* ligand activity across malignant transformation and TME development stages[31,32].

The majority of detected cell-type-specific effects (85.7%, 12/14) involved non-risk cells, including both glial and neural cells. Oligodendrocytes, which are abundant and drive neuronal activity in the GBM TME[27], accounted for 66.7% cell-type-specific genes. Neurons constitute the largest brain cell population, and intratumoral neurons often exhibit high excitability, contributing to glioma development and progression[27]. By comparing non-malignant TME cells with healthy controls, we found significantly elevated communication—particularly between neurons, astrocytes, and OPCs—likely linked to neuronal hijacking in tumorigenesis[45]. Intratumoral neuronal expression also affects GBM immunotherapy efficacy[46]. Here, we identified *JAK1*, a component central to many pro-inflammatory cytokine signaling and its neuronal expression regulates neuropeptide expression and neuroinflammation across various tissues[47], as an excitatory-neuron-specific causal gene for glioblastomagenesis.

Many genetic molecules associated with gliomagenesis have been used in glioma therapy or clinical experiments such as *IDH1*, *EGFR*, and *TERT*. To improve the clinical translation, we assessed the targets based on tumor data. The identified target was significantly enriched in glioma biological pathways and differential expression between glioma patients and healthy individuals, most importantly, CRISPR and RNAi screens showed that 53.6% of the novel genes exhibited either selective dependency or essentiality in glioma cell lines. Besides, the robust genetic evidence and brain tissue/cell-type enrichment in our identified molecules are all associated with higher clinical success[8,9,22]. The targets were annotated with chemical structure information and genetic pleiotropy to identify potential side effects[23]. The up-to-date clinical experiments were collected to promote drug repurposing; several promising drugs for glioma were uncovered; for example, Tertomotide hydrochloride, an approved TERT inhibitor for pancreatic cancer[48] , crosses the blood-brain barrier (BBB) effectively and demonstrated promising results for intracranial lesions in recent trials[49].

Despite notable findings, several limitations exist. First, our glioma GWAS subtypes are based on classifications established before the WHO 2021 guidelines, which integrate molecular data, thereby reclassifying IDH-mutant high-grade astrocytoma under the non-GBM group[50,51]. However, earlier studies suggests that potential misclassification did not bias PRS estimates from this GWAS[52], and we found that the characteristic molecular marker *IDH1* was only identified in the non-GBM data. This may suggest that most GBM and non-GBM cases were likely IDH-wildtype and IDH-mutant, respectively. Given non-GBM subtype heterogeneity[50], we primarily focused on GBM cell-type-specific genes. Future efforts should refine GWAS using updated glioma subtype definitions. Second, our analyses relied on brain-specific data[53], and the detected genes show significant enrichment in brain tissues and cells, potentially boosting clinical trial success[22] while lowering systemic side-effect risks[23]. However, the BBB remains a major hurdle in glioma therapy, impeding drug penetration[2]. In contrast, plasma targets may not require BBB crossing, facilitating

delivery. To address this, we assembled current clinical trial data on the identified targets and highlighted drugs tested in CNS diseases. Third, considering the high false positive rate in QTL association study[14], we conducted replicate studies based on different QTL and GWAS. Although we set a relaxed replication threshold (P=0.05), the disparity in sample sizes between discovery and replication could inflate false negatives. Lastly, our analysis mainly focused on European ancestry and many of the glioma germline variants are European specific, which constrains generalizability to other populations. Finally, our study was based on multi-omics evidence from the largest datasets, which is highly reproducible and easy to replicate, but future experimental validation remains essential for clinical translation.

In summary, we applied tissue- and cell-type-specific multi-omics data to systematically dissect causal genes in gliomagenesis, shedding light on cell-type-level mechanisms and advancing future precision medicine efforts.

Method

GWAS data

We used the GLIOGENE consortium glioma meta GWAS of 30,686 European (EUR) individuals (6,183 GBM, 5,820 non-GBM, and 18,169 controls) for the main analyses[10]. We included two additional EUR glioma GWAS from FinnGen R9 (243 GBM, 84 non-GBM, 287,137 controls)[54] and the Australian Genomics and Clinical Outcomes of Glioma (AGOG) (560 glioma, 2,237 controls)[36] for replicated QTL-based analyses (SMR, PWAS, TWAS). GWAS data for causal inference included 209 phenotypes **(Supplementary Table 1 and 21)**.

Gene prioritization using similarity-based methods

(1) PoPS

Polygenic Priority Score (PoPS, v0.2) integrates GWAS, gene expression, biological pathways, and predicted protein-protein interactions (PPIs) to prioritize genes[14]. Analyses were performed independently for all-tissue and brain-specific features. We partitioned the genotype into 1,703 LD blocks using LDetect[55], then we ranked the top 500 (<3%) genes and reported the highest-scoring gene per genome-wide significant LD block (where the number of genome-wide significant SNPs > 1).

Gene prioritization using locus-based methods

(1) Mapgen

We fine-mapped genome-wide significant blocks with L=10 via mapgen (v0.5.8)[56]. We integrated brain-specific epigenomics to capture long-range regulation. The gene-level posterior inclusion probability (PIP) is the weighted sum of PIPs for all linked SNPs. We reported the gene with the highest PIP in genome-wide significant LD block when $PIP_{gene} > 0.8$.

(2) TWAS

We conducted a transcriptome-wide association study (TWAS) using FUSION[57], which integrates GWAS with precomputed gene expression models[58] to identify phenotype-related genes. The largest brain-specific cis-eQTL datasets from whole brain (N = 3,154)[59] and five brain regions (N =108 to 2,683)[60] were included **(Supplementary Table 2)**. A robust GCTA-GREML (v1.94.2)[58,61] was used to generate heritability estimates per feature and the likelihood ratio test P value. Only features with P<0.05 and H2>0.1 were retained for analysis. An ACAT-O omnibus test (v0.91)[62] combined results from all predictive models, generating meta-P values and effect sizes[63].

(3) PWAS

We conducted a proteome-wide association study (PWAS) using four brain-specific pQTL datasets from the EUR population: Wingo et al.[64], with 10,198 proteins and N = 716; The Religious Orders Study and Rush Memory and Aging Project (ROS/MAP) cohort[65], with 8,356 proteins and N = 376; The Banner Sun Health Research Institute (Banner) cohort[65], with 11,518 proteins and N = 152; and Yang et al.[66], with 1,079 proteins and N = 380 **(Supplementary Table 3)**. This method connects genes and phenotypes through functional protein variation.

(4) SMR

We performed summary-level Mendelian randomization (SMR, v1.3.1) analyses[67] based on previous QTL data. HEIDI tests (P_{HEIDI} >0.05) were applied to distinguish pleiotropy from linkage.

(5) Replication studies

We conducted two separate replication studies using different QTLs and GWAS datasets in the QTL-based association analyses (SMR, PWAS, and TWAS): (1) various QTL sources and GLIOGENE glioma GWAS, (2) GWAS datasets from FinnGen and AGOG, along with prior QTL data. The AGOG GWAS was used in the GBM replication study because the majority of its cases were GBM[36]. Gene-disease pairs that satisfied (P_{SMR} <0.05 &P_{HEIDI} >0.05 or $P_{TWAS/PWAS}$ <0.05) in two duplicate studies were considered successfully replicated.

(6) Colocalization

Colocalization analyses enhance QTL association study results by excluding genetic confounding[68]. We applied colocalization focusing on the cis gene region using coloc (v5.2.3). Colocalization analysis evaluates five hypotheses regarding the association of the variants with the traits. Posterior probability H4 represents the hypothesis that the variants are associated with both traits and share the same causal variant. Genes displaying significant association (P < 2.12e-6, Bonferroni correction for 23,517 genes) in the QTL-based association analyses, successfully replicated, and demonstrating colocalization evidence (SNP.PP.H4 > 0.70) were reported.

Enrichment Analyses and PPI

We first assessed tissue enrichment of the identified potential causal gene set via TissueEnrich (v1.24.1)[69], adjusting P values by Bonferroni correction. PPI analyses utilized the STRING database (v12.0)[70]. The biological functions of the gene set were annotated using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases through the clusterProfiler[71]. To evaluate pathway activation in tumors, we performed single-cell Gene Set Enrichment Analysis (scGSEA) using the tumor-adapted method JASMINE[72] on glioma subtype sc/snRNA-seq data. Three sc/snRNA-seq datasets were included: (1) Pan-glioma scRNA-seq containing 55,284 cells from 11 glioma patients[73] ; (2) GBM snRNA-seq from GBmap (core), containing 338,564 cells from 110 IDH-wild-type GBM patients[20]; and (3) Non-GBM scRNA-seq containing 76,639 cells from 16 IDH-mutant glioma patients[74]. **(Supplementary Table 5)**. Gene sets with significant GO terms ($P_{Bonferroni}$ <0.05) were chosen for scoring.

DEG

We conducted two differential gene expression (DEG) analyses across five bulk RNA-seq datasets **(Table S4)**. RNA base DEG analysis was carried out, to compare healthy ($N_{control}$ = 1,167) most of from GTEX[75] with glioma (N_{case} = 1,728) from the CGGA[76] and TCGA[77], with provided classifications. Briefly, datasets were normalized with edgeR (v4.0.3) using TMM (trimmed mean of M-values)[78]. Then log2CPM values were computed by voom. Batch effects across datasets were estimated using surrogate variable analysis (SVA) and included as covariates in the limma (v3.58.1)

model[79]. Two gene expression microarray datasets were utilized for within-dataset contrasts (normal vs. tumor) in DEG analyses via the online tool GEO2R[80]. Limma precision weights (voom) estimated the mean-variance relationship and produced observational-level weights. A gene was considered differentially expressed when the logFC > 1 and P adj < 0.05.

DepMap

To evaluate gene selectivity and essentiality for tumor cells, we examined the identified genes in the DepMap (24Q2) dataset (https://depmap.org/portal), which offers genome-wide CRISPR and miRNA loss-of-function screen data from characterized cell lines[81]. Target selectivity and essentiality (gene effect < -0.5) from CRISPR and RNAi screens on 72 glioma cell lines.

CellChat

To compare cell-cell communication between the GBM and healthy brain, we computed the communication probabilities among non-cancerous cell types in brain regions (frontal, temporal, and parietal lobes) with high GBM incidence[82] using CellChat (v2.1.2)[83] and compared them between GBM patients and healthy controls. Two large-scale snRNA-seq datasets were included in: (1) Case: GBmap (expanded), including 289,395 cells from 240 IDH-wild-type GBM patients[20]; and (2) Control: the Human Brain Cell Atlas, containing 839,975 cells from 3 healthy adults[84]**(Supplementary Table 5)**. A custom human ligand-receptor (L-R) database was used, which was generated by merging CellChatDB and CellPhoneDB (v5.0.1)[85], increasing curated L-R interactions from 3,234 to 6,007 and complexes from 338 to 700.

Genetic Risk Cell Analyses

To distinguish critical cell types for glioblastomagenesis, we applied scPagwas (v1.3.0), a pathway-based polygenic regression method integrating GWAS signals with sc/snRNA-seq transcriptional features[26]. The GBM meta GWAS and GBmap (core) snRNA-seq data[20] were used as inputs; cell types with $P_{Bonferroni}$ <0.05 were deemed GBM risk cells. Then, we assessed the GBM risk cell-type-specific (astrocytes, OPCs) heritability enrichments (proportion of heritability explained divided by annotation size) and estimated the proportion of the cell-type annotations by applying S-LDSC (v1.0.1)[86] in adult brain scATAC-seq[87]. The latest version of the baseline model (v2.3)[88] was used.

Pseudotime Analyses in snRNA-seq

We implemented a deep learning-based method, CytoTRACE2 (v1.0.0)[89], to predict cellular potency and developmental potential in GBM. The results were projected onto a UMAP. Unsupervised inference methods VECTOR (v0.0.4)[90] and PAGA[91] identified starting cells and inferred developmental trajectories in glioblastomagenesis. We defined astrocytes and OPCs as root cells and inferred their differentiation in GBM using semi-supervised method Monocle 3 (v1.4.18)[92].

Cell-Type-Specific Causal Target Genes Analyses

We incorporated two sc-eQTL datasets spanning eight major brain cell types[21,93] to pinpoint cell-type-specific causal genes in glioblastomagenesis, one by Fujita et al.[93], involving 1,509,626 cells from 424 EUR individuals, and another by Bryois et al.[21], involving 577,115 cells from 192 EUR individuals **(Supplementary Table 2)**. Cell-type-specific TWAS and SMR were performed, the genes displaying significant association considered significant (P < 2.78e-6, Bonferroni correction for 17,964 genes). Two replicate studies used independent QTLs and alternative GWAS data. FinnGen and AGOG GWAS replicated our QTL-based association analyses (TWAS, SMR). Genes that replicated (P_{TWAS} <0.05 or P_{SMR} <0.05 and P_{HEIDI} >0.05) were labeled as potential

cell-type-specific causal genes. Colocalization analysis (PP.H4 >0.7) was conducted to improve identification of cell-specific effects[56]. cis-MRcML (v0.0.09), which an MR method based on conditional/joint SNP effects and robust to IV violations, was applied to confirm gene-trait causal relationships when $P_{Bonferroni}$ <0.05[94]. A merged reference panel of 10K UKBB and 1KG EUR (N = 4285)[95,96] was used in GCTA[97] to derive conditional/joint SNP groups.

Cell Type Enrichment Tests

To evaluate whether identified cell-type-specific causal genes are overexpressed in corresponding GBM cell types beyond chance, we applied expression-weighted cell type enrichment (EWCE) v1.10.2[98]. Specificity values, indicating the proportion of a gene's total expression in one cell type versus others, were calculated using snRNA for GBmap (core)[20]. We selected cell-type-specific gene groups with more than four genes, including significant and potential causal genes (passing association and replication studies), along with specificity matrices. After controlling for transcript length and GC-content, we executed EWCE with 10,000 bootstrap replicates, and applied Bonferroni correction to P-values.

CT-FM-SNP

To pinpoint cell-type-specific causal variants in glioblastomagenesis, we implemented Cell Type Fine-Mapping SNP (CT-FM-SNP), which integrates GWAS fine-mapping data with cell-type-specific cCRE annotations (N = 927)[99]. The significant SNPs in the credible set underwent fine-mapping with S-LDSC results and were overlapped with candidate cis-regulatory elements (cCREs).

Peak overlap

We overlapped the fine-mapping identified credible set in *EGFR* with astrocyte epigenetic data peaks. Bulk ATAC-seq[18], snATAC-seq[17], regulatory element data from ChIP-seq and PLAC-seq[17] provided a comprehensive view of astrocyte-specific epigenetic effects. FORGEdb scores were retrieved via FORGEdb[29].

iCPAGdb

We used the interactive Cross-Phenotype Analysis of GWAS database (iCPAGdb) v1.1[100] to identify glioma genetically shared traits across 3,793 phenotypes.

Bidirectional MR-PheWAS

To determine genetic causal relationships, we performed bidirectional MR-PheWAS, including the largest available GWAS (before May 20, 2024) on genetic shared traits ($P_{Bonferroni}$ <0.05). GWAS summaries from 208 traits were collected. Genome-wide significant (P <5×10^-8), independent (LD r^2 < 0.001, 10,000 kb) SNPs were chosen as instrument variables (IVs) via TwoSampleMR (v0.5.8)[101]. SNPs palindromic with intermediate allele frequencies or missing in the outcome GWAS were excluded. Radial MR (v1.1) detected and removed outlier SNPs[102]. We retained strong IVs (F>10) and phenotypes with >3 SNPs post-filtering to maintain statistical power. A Q-test P > 0.05 indicated no significant IV heterogeneity. MR-PRESSO (v1.0)[103] evaluated horizontal pleiotropy; P>0.05 in global tests denoted no pleiotropic effects. The random-IVW method served as the main analysis given its higher power[104] , with P values Bonferroni-corrected. Weighted median (tolerating up to half pleiotropic SNPs)[104] and 200-time simulation MR-cML-MA v0.0.0.9 (robust to correlated/uncorrelated pleiotropic effects)[105] were used as sensitivity analyses. Reverse MR swapped exposure/outcome roles but applied the same filtering criteria; traits not significant (P_{IVW} > 0.05) were deemed non-bidirectional. Potential causal traits were required to (1) pass heterogeneity/pleiotropy checks, (2) lack bidirectional effects, and (3) be nominally significant

(P<0.05) with consistent effect direction across all MR sensitivities.

HDL

We estimated cross-phenotype genetic correlation between potential causal traits and glioma via High-Definition Likelihood (HDL) v1.4.0[106] , utilizing precomputed EUR LD matrices from UK Biobank. We focused on potential causal traits, P_{HDL} >0.05 indicated suggestive causal phenotypes, and genetically correlated (P_{HDL} <0.05) traits underwent further analysis.

LCV

We employed the Latent Causal Variable (LCV) model[33] on genetically correlated traits to filter false positives arising from genetic correlations. Traits with P < 0.05 and the same effect direction as MR analyses were considered suggestive causal.

Gene-based PheWAS

We performed gene-based PheWAS analyses using GWAS ATLAS (3,302 unique traits from 4,155 GWASs)[107] and UK Biobank TOPMed-imputed PheWeb (1,419 unique PheWAS codes from 400,000 EUR individuals)[108] to annotate genetic pleiotropy. Genome-wide significant traits (P <5×10^-8) were reported.

Drug Repurposing Analysis

We queried predicted drug-gene interactions from the Drug-Gene Interaction database (DGIdb v5.0)[109], retaining only those with high interaction scores (> 0.8). DrugBank v6.0[110], canSAR.ai v1.6.7[111] , Chinese Clinical Trial Registry (ChiCTR, https://www.chictr.org.cn/), ClinicalTrials.gov (https://www.clinicaltrials.gov/), and the WHO (https://www.who.int/) supplied information on chemical structure, late-stage clinical trials (pre–2024 July 20), and disease indications for drug repositioning.

Ethics

Ethical approval for this study was not required as our analyses were based on summary statistics from published GWAS or the data were publicly accessible.

Acknowledgments

We thank the consortium studies such as GLIOGENE, AGOG, UKBB, FinnGen, and others. Their collaborative efforts and dedication were instrumental in enabling this study. We also appreciate Dr. Beatrice S Melin and Dr. Karen Alpen for their assistance in accessing the GWAS datasets.

Author contributions

H.Y.F. and H.K.L. designed the analysis. H.Y.F. analyzed the data. H.Y.F. wrote the paper. H.Y.F., and H.K.L. provided feedback and approved the final version of the paper.

Data available

The GLIOGENE consortium glioma GWAS summary statistics are available from the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number

EGAS00001003372. AGOG glioma GWAS summary statistics are applied from AGOG. The FinnGen R9 glioma GWAS summary statistics are available from https://www.finngen.fi/en/access_results. Summary QTL database reference panel weights generated for this study used in TWAS and PWAS are available in https://figshare.com/articles/dataset/_b_TWAS_PWAS_gene_expression_prediction_models_b_/28080035. The reference panel, which is the merged genotype data from UK10K and 1000 Genomes Project Phase 3, is available from the EGA under accession number EGAD00001000776. The PLAC-seq, snATAC-seq and ChIP-seq data were downloaded from https://github.com/nottalexi/brain-cell-type-peak-files. pcHi-C was downloaded from GSE113481, bulk ATAC was downloaded from GSE113480. The remaining data are available within the Article, Supplementary Information or Source Data files.

Code availability

We used publicly available software for the analyses. The software used is listed in the Methods section and supplement table. Custom analysis scripts are available at github (https://github.com/1667857557/Glioma-Cell-Type-Specific-Causal-Genes).

1. Weller, M. *et al.* Glioma. *Nat Rev Dis Primers* **1**, 15017 (2015).
2. Dewdney, B. *et al.* From signalling pathways to targeted therapies: unravelling glioblastoma's secrets and harnessing two decades of progress. *Signal Transduct Target Ther* **8**, 400 (2023).
3. Cruz Da Silva, E., Mercier, M.C., Etienne-Selloum, N., Dontenwill, M. & Choulier, L. A Systematic Review of Glioblastoma-Targeted Therapies in Phases II, III, IV Clinical Trials. *Cancers (Basel)* **13**(2021).
4. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
5. Venkatesh, H.S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539-545 (2019).
6. Kondo, T. Glioblastoma-initiating cell heterogeneity generated by the cell-of-origin, genetic/epigenetic mutation and microenvironment. *Semin Cancer Biol* **82**, 176-183 (2022).
7. Laug, D., Glasgow, S.M. & Deneen, B. A glial blueprint for gliomagenesis. *Nat Rev Neurosci* **19**, 393-403 (2018).
8. Dann, E. *et al.* Estimating the impact of single-cell RNA sequencing of human tissues on drug target validation. *medRxiv*, 2024.04.04.24305313 (2024).
9. Minikel, E.V., Painter, J.L., Dong, C.C. & Nelson, M.R. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624-629 (2024).
10. Melin, B.S. *et al.* 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).
11. Atkins, I. *et al.* Transcriptome-Wide Association Study Identifies New Candidate Susceptibility Genes for Glioma. *Cancer Res* **79**, 2065-2071 (2019).
12. Thornton, Z.A. *et al.* Multi-omics Mendelian randomisation using expression, splicing and protein quantitative trait loci: identification of novel drug targets for gliomagenesis. *medRxiv*, 2022.05.11.22274959 (2024).

13. Yao, D.W., O'Connor, L.J., Price, A.L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet* **52**, 626-633 (2020).
14. Weeks, E.M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *medRxiv*, 2020.09.08.20190561 (2020).
15. Marigorta, U.M., Rodríguez, J.A., Gibson, G. & Navarro, A. Replicability and Prediction: Lessons and Challenges from GWAS. *Trends Genet* **34**, 504-517 (2018).
16. Cope, A.L., Schraiber, J.G. & Pennell, M. Phylogenetic analysis reveals how selection and mutation shape the coevolution of mRNA and protein abundances. *bioRxiv*, 2024.07.08.602411 (2024).
17. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134-1139 (2019).
18. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262 (2019).
19. Fulco, C.P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669 (2019).
20. Ruiz-Moreno, C. *et al.* Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma. *bioRxiv*, 2022.08.27.505439 (2022).
21. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104-1112 (2022).
22. Nelson, M.R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-60 (2015).
23. Razuvayevskaya, O., Lopez, I., Dunham, I. & Ochoa, D. Genetic factors associated with reasons for clinical trial stoppage. *Nat Genet* **56**, 1862-1867 (2024).
24. Jacobs, D.I. *et al.* Elucidating the molecular pathogenesis of glioma: integrated germline and somatic profiling of a familial glioma case series. *Neuro Oncol* **20**, 1625-1633 (2018).
25. Porcu, E. *et al.* Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat Commun* **12**, 5647 (2021).
26. Ma, Y. *et al.* Polygenic regression uncovers trait-relevant cellular contexts through pathway activation transformation of single-cell RNA sequencing data. *Cell Genom* **3**, 100383 (2023).
27. Drexler, R. *et al.* A prognostic neural epigenetic signature in high-grade glioma. *Nat Med* **30**, 1622-1635 (2024).
28. Gaffney, D.J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol* **13**, R7 (2012).
29. Breeze, C.E. *et al.* FORGEdb: a tool for identifying candidate functional variants and uncovering target genes and mechanisms for complex diseases. *Genome Biol* **25**, 3 (2024).
30. An, Z., Aksoy, O., Zheng, T., Fan, Q.W. & Weiss, W.A. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. *Oncogene* **37**, 1561-1575 (2018).
31. Guo, G. *et al.* EGFR ligand shifts the role of EGFR from oncogene to tumour suppressor in EGFR-amplified glioblastoma by suppressing invasion through BIN3 upregulation. *Nat Cell Biol* **24**, 1291-1305 (2022).
32. Hu, C. *et al.* Glioblastoma mutations alter EGFR dimer structure to prevent ligand bias. *Nature* **602**, 518-522 (2022).
33. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* **50**, 1728-1734 (2018).
34. Minikel, E.V. & Nelson, M.R. Human genetic evidence enriched for side effects of approved drugs.

medRxiv, 2023.12.12.23299869 (2023).
35. Eckel-Passow, J.E. *et al.* Adult diffuse glioma GWAS by molecular subtype identifies variants in D2HGDH and FAM20C. *Neuro Oncol* **22**, 1602-1613 (2020).
36. Alpen, K. *et al.* Australian genome-wide association study confirms higher female risk for adult glioma associated with variants in the region of CCDC26. *Neuro Oncol* **25**, 1355-1365 (2023).
37. Kinnersley, B., Houlston, R.S. & Bondy, M.L. Genome-Wide Association Studies in Glioma. *Cancer Epidemiol Biomarkers Prev* **27**, 418-428 (2018).
38. Thornton, Z.A. *et al.* Brain multi-omic Mendelian randomisation to identify novel drug targets for gliomagenesis. *Hum Mol Genet* **34**, 178-192 (2025).
39. Nakase, T. *et al.* EPCO-58. INTEGRATIVE PROTEOGENOMIC ANALYSES REVEAL INSIGHTS INTO SUBTYPE-SPECIFIC GLIOMA RISK. *Neuro-Oncology* **26**, viii15-viii15 (2024).
40. Verdugo, E., Puerto, I. & Medina, M. An update on the molecular biology of glioblastoma, with clinical implications and progress in its treatment. *Cancer Commun (Lond)* **42**, 1083-1111 (2022).
41. Minami, J.K. *et al.* CDKN2A deletion remodels lipid metabolism to prime glioblastoma for ferroptosis. *Cancer Cell* **41**, 1048-1060.e9 (2023).
42. Rahme, G.J. *et al.* Modeling epigenetic lesions that cause gliomas. *Cell* **186**, 3674-3685.e14 (2023).
43. Snuderl, M. *et al.* Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810-7 (2011).
44. Mathur, R. *et al.* Glioblastoma evolution and heterogeneity from a 3D whole-tumor perspective. *Cell* **187**, 446-463.e16 (2024).
45. Venkataramani, V. *et al.* Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* **185**, 2899-2917.e31 (2022).
46. Tomaszewski, W.H. *et al.* Neuronal CaMKK2 promotes immunosuppression and checkpoint blockade resistance in glioblastoma. *Nat Commun* **13**, 6483 (2022).
47. Tamari, M. *et al.* Sensory neurons promote immune homeostasis in the lung. *Cell* **187**, 44-61.e17 (2024).
48. Jo, J.H. *et al.* KG 4/2015: A randomized, controlled, multicenter, open-label phase III clinical trial of GV1001 with gemcitabine/capecitabine in previous untreated, eotaxin-high patients with advanced pancreatic ductal adenocarcinoma. *Journal of Clinical Oncology* **39**, 4020-4020 (2021).
49. Koh, S.H. *et al.* Efficacy and safety of GV1001 in patients with moderate-to-severe Alzheimer's disease already receiving donepezil: a phase 2 randomized, double-blind, placebo-controlled, multicenter clinical trial. *Alzheimers Res Ther* **13**, 66 (2021).
50. Louis, D.N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* **23**, 1231-1251 (2021).
51. Louis, D.N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* **114**, 97-109 (2007).
52. Nakase, T. *et al.* Genome-wide Polygenic Risk Scores Predict Risk of Glioma and Molecular Subtypes. *medRxiv*, 2024.01.10.24301112 (2024).
53. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
54. Kurki, M.I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508-518 (2023).
55. Berisa, T. & Pickrell, J.K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283-5 (2016).

56. Selewa, A. *et al.* Single-cell genomics improves the discovery of risk variants and genes of atrial fibrillation. *Nat Commun* **14**, 4999 (2023).
57. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
58. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
59. Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat Genet* **54**, 161-169 (2022).
60. de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet* **55**, 377-388 (2023).
61. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet* **50**, 538-548 (2018).
62. Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* **104**, 410-421 (2019).
63. Dai, Q. *et al.* OTTERS: a powerful TWAS framework leveraging summary-level reference data. *Nat Commun* **14**, 1271 (2023).
64. Wingo, A.P. *et al.* Sex differences in brain protein expression and disease. *Nat Med* **29**, 2224-2232 (2023).
65. Wingo, A.P. *et al.* Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet* **53**, 143-146 (2021).
66. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci* **24**, 1302-1312 (2021).
67. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
68. Hukku, A. *et al.* Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am J Hum Genet* **108**, 25-35 (2021).
69. Jain, A. & Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966-1967 (2019).
70. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638-d646 (2023).
71. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284-7 (2012).
72. Noureen, N., Ye, Z., Chen, Y., Wang, X. & Zheng, S. Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *Elife* **11**(2022).
73. Johnson, K.C. *et al.* Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nat Genet* **53**, 1456-1468 (2021).
74. Wei, Y. *et al.* Stalled oligodendrocyte differentiation in IDH-mutant gliomas. *Genome Med* **15**, 24 (2023).
75. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
76. Zhao, Z. *et al.* Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. *Genomics Proteomics Bioinformatics* **19**, 1-12 (2021).
77. Weinstein, J.N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-20 (2013).

78. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
79. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).
80. Davis, S. & Meltzer, P.S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846-7 (2007).
81. DepMap, B. DepMap 24Q2 Public. (Figshare+, 2024).
82. Chakrabarti, I., Cockburn, M., Cozen, W., Wang, Y.P. & Preston-Martin, S. A population-based description of glioblastoma multiforme in Los Angeles County, 1974-1999. *Cancer* **104**, 2798-806 (2005).
83. Jin, S., Plikus, M.V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. *bioRxiv*, 2023.11.05.565674 (2023).
84. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
85. Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506 (2020).
86. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
87. Kim, S.S. *et al.* Leveraging single-cell ATAC-seq and RNA-seq to identify disease-critical fetal and adult brain cell types. *Nat Commun* **15**, 563 (2024).
88. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* **49**, 1421-1427 (2017).
89. Kang, M. *et al.* Mapping single-cell developmental potential in health and disease with interpretable deep learning. *bioRxiv*, 2024.03.19.585637 (2024).
90. Zhang, F., Li, X. & Tian, W. Unsupervised Inference of Developmental Directions for Single Cells Using VECTOR. *Cell Rep* **32**, 108069 (2020).
91. Wolf, F.A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019).
92. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
93. Fujita, M. *et al.* Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet* **56**, 605-614 (2024).
94. Lin, Z. & Pan, W. A robust cis-Mendelian randomization method with application to drug target discovery. *Nat Commun* **15**, 6072 (2024).
95. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
96. Abecasis, G.R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
97. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
98. Skene, N.G. & Grant, S.G. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci* **10**, 16 (2016).

99. Kim, A. *et al.* Inferring causal cell types of human diseases and risk variants from candidate regulatory elements. *medRxiv*, 2024.05.17.24307556 (2024).
100. Wang, L. *et al.* An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Med* **13**, 83 (2021).
101. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).
102. Bowden, J. *et al.* Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int J Epidemiol* **47**, 1264-1278 (2018).
103. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693-698 (2018).
104. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res* **4**, 186 (2019).
105. Xue, H., Shen, X. & Pan, W. Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am J Hum Genet* **108**, 1251-1269 (2021).
106. Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat Genet* **52**, 859-864 (2020).
107. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
108. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
109. Cannon, M. *et al.* DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res* **52**, D1227-d1235 (2024).
110. Knox, C. *et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* **52**, D1265-d1275 (2024).
111. Gingrich, P.W. *et al.* canSAR 2024-an update to the public drug discovery knowledgebase. *Nucleic Acids Res* (2024).

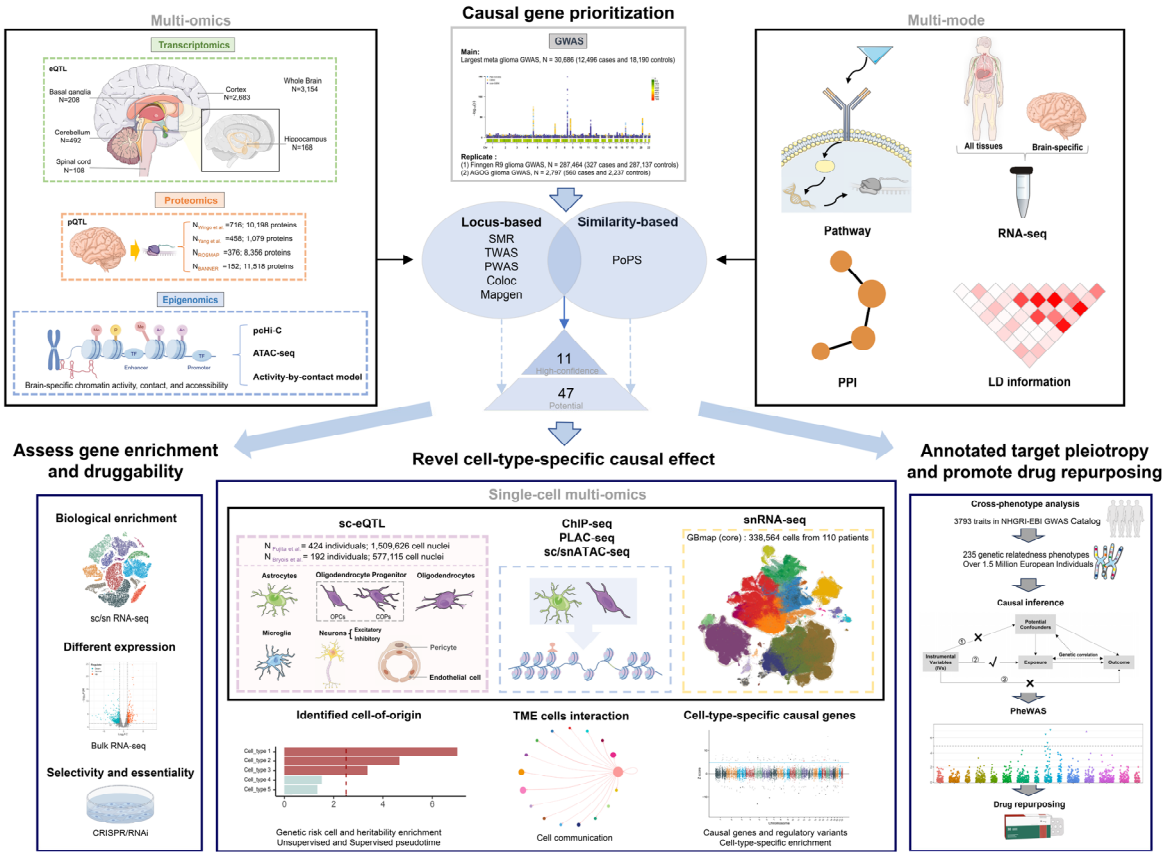

Fig 1. General study workflow. Overview of the main analytical steps conducted in the study.

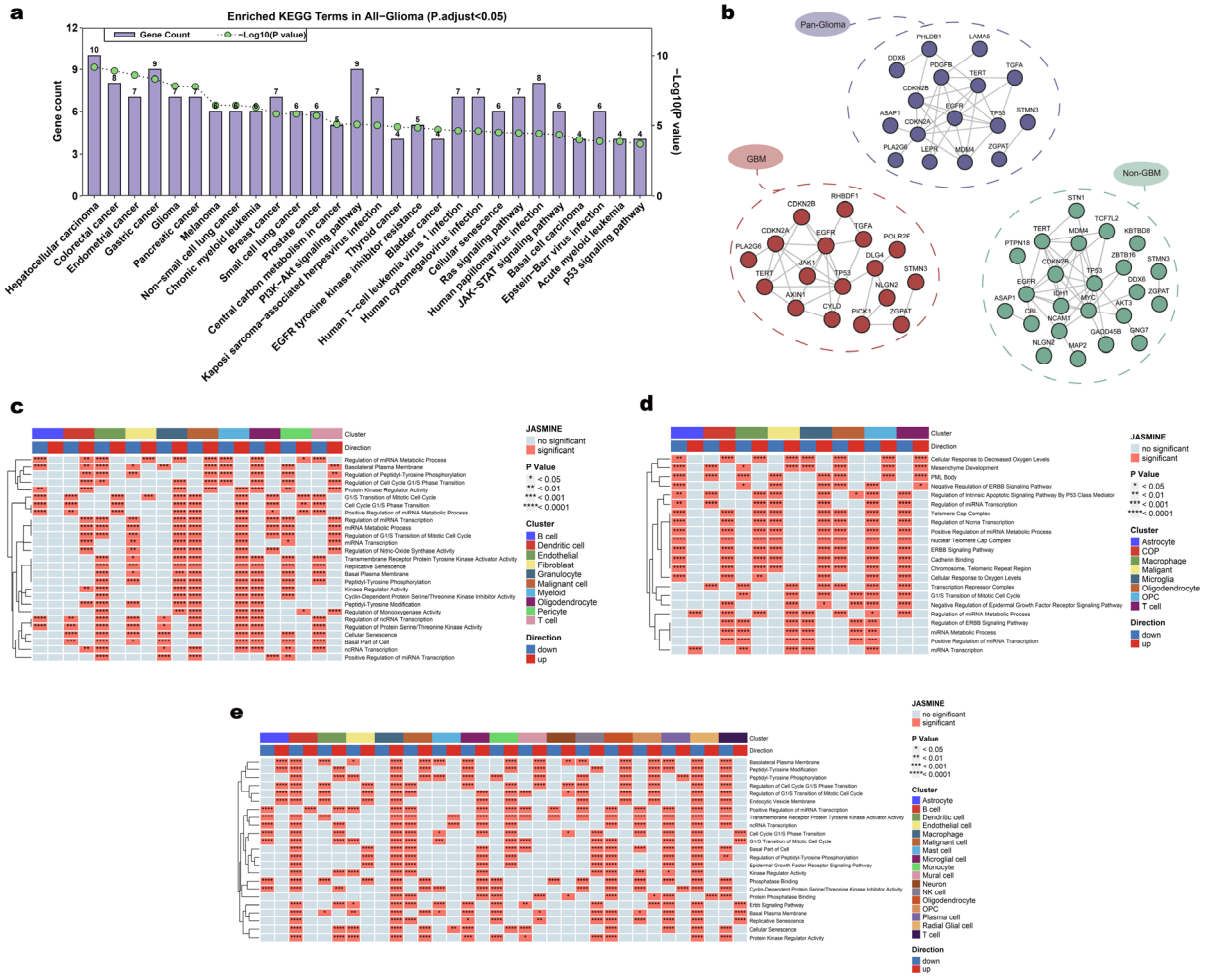

Fig2. Enrichment of identified genes. (a) The bar plot represents the number of genes associated with each KEGG pathway (left Y-axis), while the green dots indicate the -log10(P value) for pathway enrichment (right Y-axis). **(b)** Protein–protein interaction network for identified potential causal genes. **(c-e)** Heatmap of significant gene set distributions across cell clusters. Asterisks indicate P-values: * <0.05, ** <0.01, *** <0.001, **** <0.0001. The left dendrogram shows clustering based on gene set similarity. "Direction" indicates upregulation (red) or downregulation (blue) of gene sets in each cluster. Cell clusters are color-coded as shown.

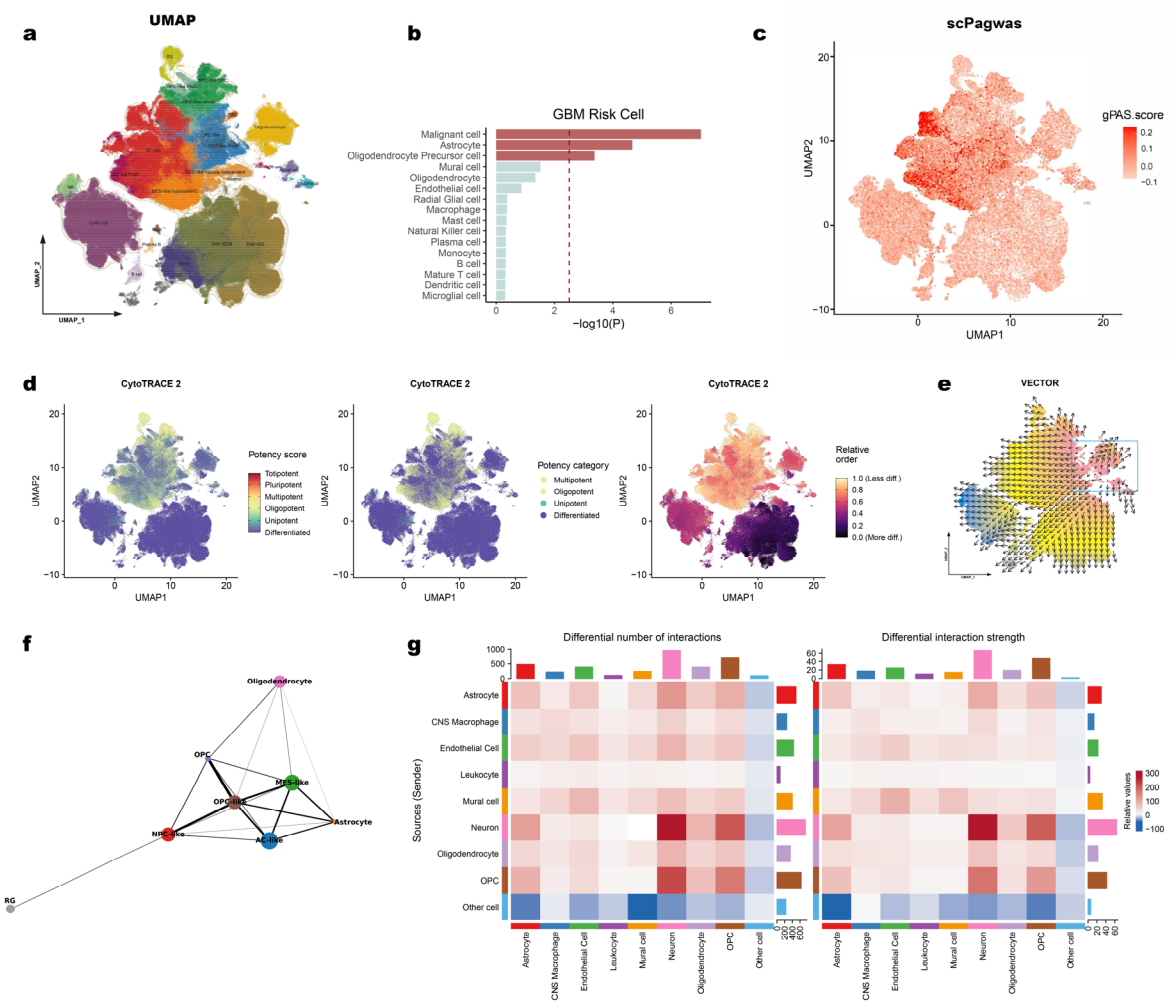

Fig3. Cell-of-origin and TME cell interactions in GBM. (a) UMAP visualization of GBmap (core) **(b)** Bar plot showing genetic risk cell types identified in scPagwas, which ranked by p value. Dashed red line indicates the threshold after multiple testing correction. **(c)** UMAP visualization of gPAS scores derived from scPagwas analysis. Cells are colored according to their gPAS score, ranging from low (white) to high (red), indicating regions with higher pathway activity. **(d)** UMAP visualization of differentiation ability colored by CytoTRACE 2 predictions. (Left) UMAP colored by predicted potency score, with values ranging from totipotent (red) to differentiated (blue). (Center) UMAP colored by predicted potency category. (Right) UMAP colored by predicted relative differentiation order, transitioning from less differentiated (yellow) to more differentiated (purple). **(e)** UMAP visualization of pseudotime by VECTOR, the pseudotime order of cells is depicted by color, transitioning from red to blue. The blue rectangle highlights the region corresponding to the starting cells. Vector representation of inferred developmental directions for cells in GBM. **(f)** Graph plot illustrating differentiate relationship identified by PAGA, Nodes represent cell types, and edge thickness reflects connectivity strength. **(g)** Heatmaps showing the differences in brain cell interactions between GBM non-malignant cells and healthy controls in GBM high-incidence brain regions. interaction number (left) and interaction strength (right) among cell populations across two datasets. Rows represent source (sender) cell types, and columns represent target cell types. The top bar plot shows the total incoming signaling for each target cell type, while the side bar plot indicates outgoing signaling for each source cell type. Red indicates increased signaling in the GBM patient dataset compared to the healthy dataset, and blue indicates decreased.

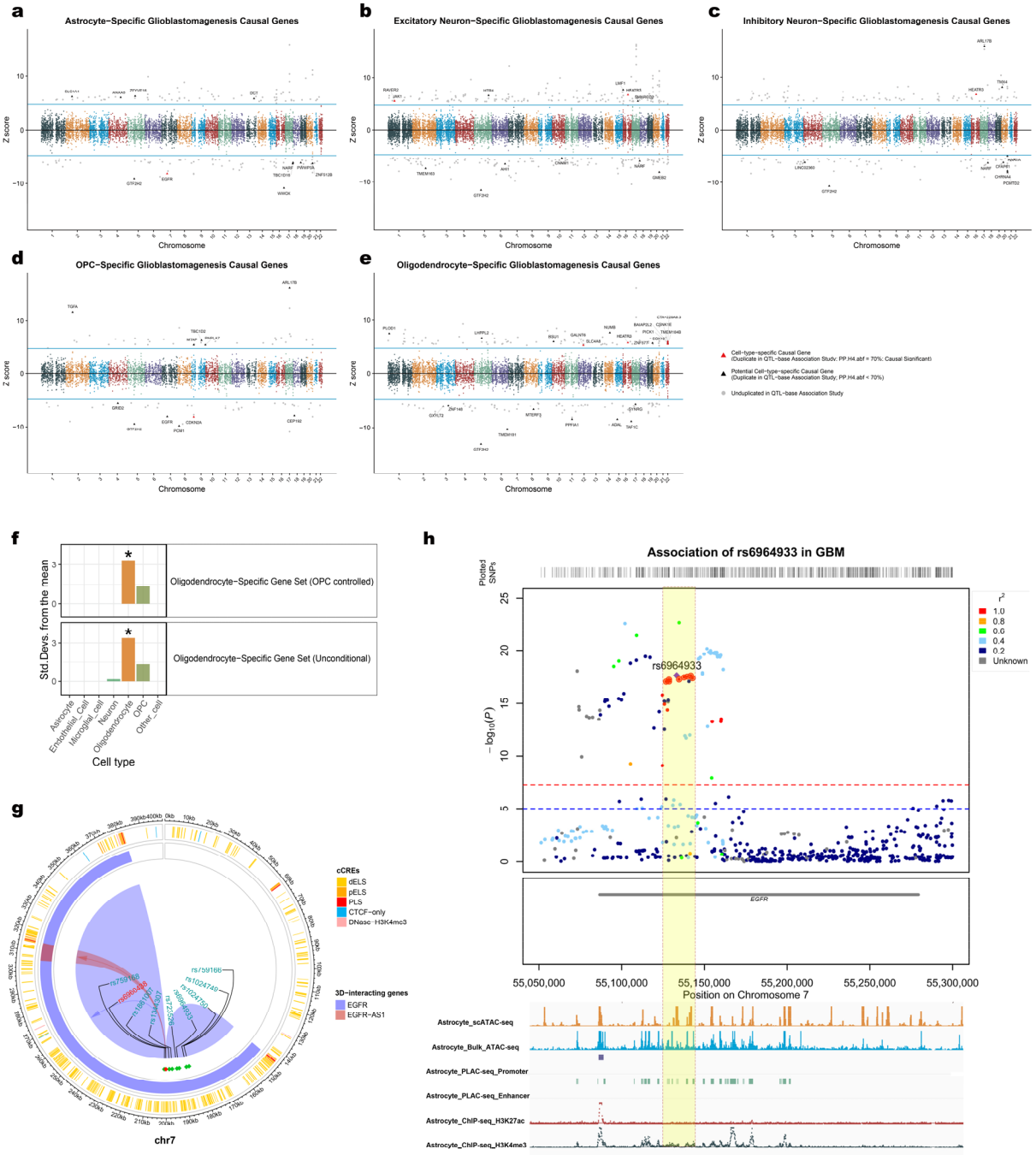

Fig 4. Cell-type-specific causal effect in glioblastomagenesis. (a-e) Manhattan-style plots displaying cell-type-specific glioblastomagenesis causal genes across chromosomes for astrocytes, excitatory neurons, inhibitory neurons, OPCs, and oligodendrocytes. Z-scores represent gene effect size and direction in QTL-based association study, with significant cell-type-specific causal genes highlighted as red triangles (Duplicated in QTL-based association study, PP.H4.abf ≥ 70%, and with causal significance) and potential causal genes as black triangles (Duplicated in QTL- based association study, and PP.H4.abf < 70%). Grey dots indicate unduplicated genes from QTL-based association studies. Horizontal blue lines denote significant thresholds (P= 2.78e-6; 0.05/17,964). **(f)** Bar plots showing the cell type enrichment test in significant oligodendrocyte-specific gene sets. (Top) Results for OPC-controlled analysis. (Bottom) Results for unconditional analysis. Asterisks (*) highlight significant deviations, indicating overexpression in oligodendrocytes compared to the bootstrapped mean. **(g)** Circular plot illustrating cis-regulatory elements (cCREs) and 3D genomic interactions in rs69600438 and its high LD credit set. Outer tracks show cCREs, categorized as dELS (yellow), pELS (orange), PLS (red), CTCF-only

(blue), and DNase-H3K4me3 (cyan). Inner arcs represent 3D interactions between regulated genes. SNPs of interest are labeled and linked to their associated regions. Purple and red shading indicates chromatin loops connecting regulatory elements to target genes. Visualization by the ViSNP package. **(h)** Regional Manhattan plot showing rs6964933 and its high LD credit set in GBM GWAS. The y-axis displays -log10(p-values), and points are colored based on linkage disequilibrium (LD) with rs6964933. Horizontal dashed lines indicate genome-wide significance (5×10^-8) and suggestive significance (1×10^-5) thresholds. Peak plot generated by IGV v2.17.2 displayed with additional tracks for astrocyte-specific epigenomic peaks, including snATAC-seq, bulk ATAC-seq, regulatory elements from PLAC-seq, and histone modifications (H3K27ac and H3K4me3) from ChIP-seq. The highlighted yellow region marks the position of rs6964933 and its high-LD credible set.

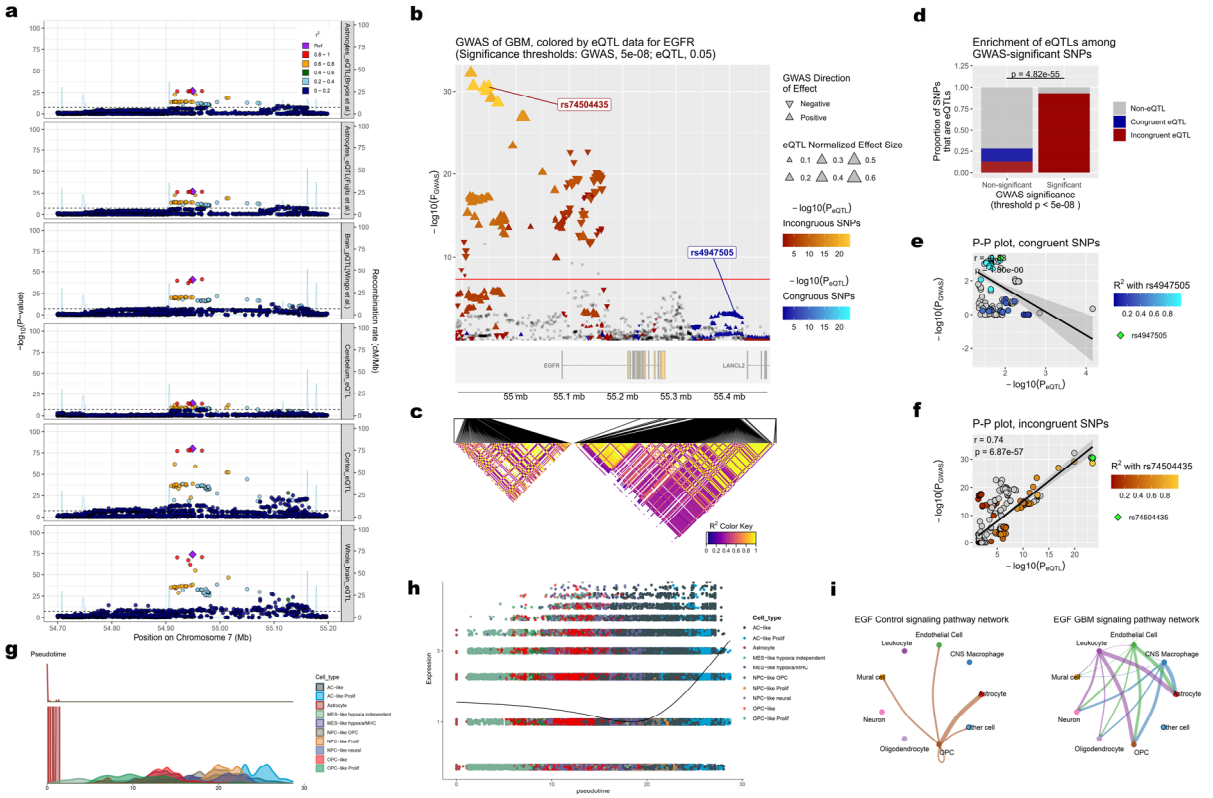

Fig 5. EGFR effect in glioblastomagenesis (a) Regional association plots for distinct brain QTL datasets showing genetic signals in the EGFR cis region. Points represent SNPs, colored by LD (R2) with the lead SNP (purple diamond). The y-axis displays -log10(p-values) for associations, and the light blue line indicates recombination rates across the region. The dashed horizontal line denotes the significance threshold (5×10^-8). (b) eQTpLot visualization of the colocalization and driving effect between GBM GWAS and astrocyte-specific cis-eQTL data (Fujita et al.) in the EGFR cis region. Manhattan plot showing GWAS signal, with SNPs colored by eQTL P value and direction (triangles for positive and negative effects). Genomic region plot showing the genomic structure. (c) LD heatmap depicting pairwise LD across the locus. (d) Bar plot quantifying the enrichment of eQTLs among GWAS-significant SNPs (p<5e−8), distinguishing congruent and incongruent eQTLs. (e) P-P plot correlating GWAS and eQTL p-values for congruent SNPs, (f) P-P plot correlating GWAS and eQTL p-values for incongruent SNPs.(g) Density plot demonstrates cell type distributions along pseudotime inferred by Monocle3, with astrocytes designated as the starting cell population. Each curve represents the relative proportion of a specific cell type as pseudotime progresses, highlighting differentiation trajectories.(h) Pseudotime trajectory plot showing expression dynamics of EGFR along pseudotime, with astrocytes set as the starting cell population in Monocle3. Points represent individual cells, colored by cell type, and the black curve indicates the smoothed trend of EGFR expression across pseudotime. (i) Circle plot comparison of EGF signaling pathway networks between healthy controls and GBM non-malignant cells. The edge color represents the source cell group as the sender, and the edge weight reflects interaction strength, with thicker lines indicating stronger signals.