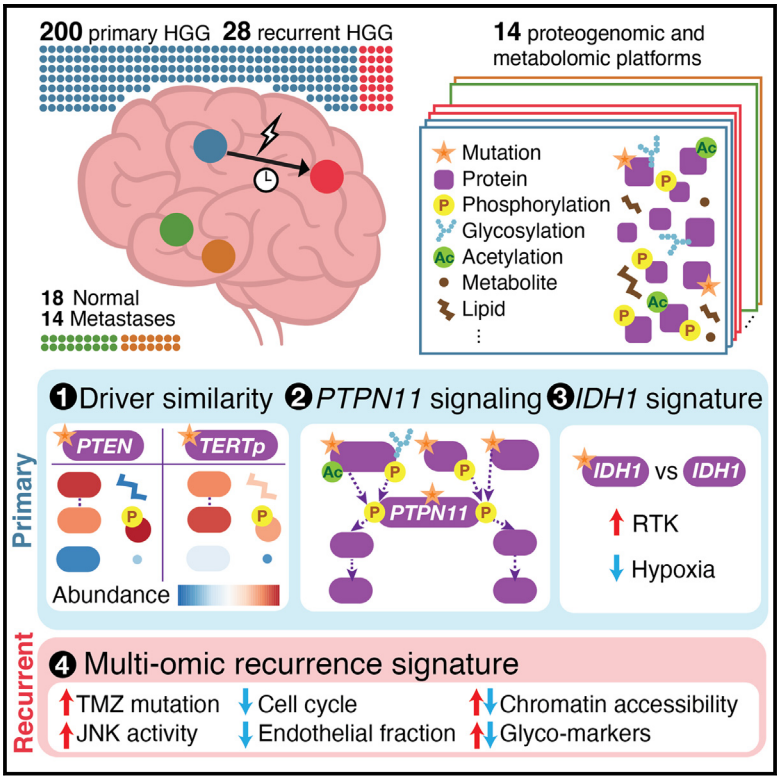
Cancer Cell

Article

Multi-scale signaling and tumor evolution in high-grade gliomas

Graphical abstract



Highlights

- Metabolome and glycoproteome data reveal driver interactions and recurrence markers
- Alterations in *TERTp*, *PTEN*, or *TERTp/EGFR* produce similar molecular features
- PTPN11 signaling links EGFR, PDGFR, and IDH1 to downstream effectors
- A low hypoxia signature and reduced AMPKA activities are found in IDH-mutant HGG

Authors

Jingxian Liu, Song Cao, Kathleen J. Imbach, ..., Li Ding, Philadelphia Coalition for a Cure, Clinical Proteomic Tumor Analysis Consortium

Correspondence

mchheda@wustl.edu (M.G.C.), rodland@ohsu.edu (K.D.R.), tao.liu@pnnl.gov (T.L.), lding@wustl.edu (L.D.)

In brief

Liu et al. integrate 14 proteogenomic, glycoproteomic, and metabolomic platforms on 228 high-grade gliomas composed of glioblastomas and grade 4 IDH-mutant astrocytomas, including tumors at recurrence. This study identifies multi-scale regulatory interactions governing tumor development and evolution, reveals heterogeneous upstream alterations converging on common downstream events, and suggests a central role for PTPN11 signaling across high-grade gliomas.

Liu et al., 2024, Cancer Cell *42*, 1217–1238
July 8, 2024
https://doi.org/10.1016/j.ccell.2024.06.004

CellPress

Cancer Cell

CellPress
OPEN ACCESS

Article

Multi-scale signaling and tumor evolution in high-grade gliomas

Jingxian Liu,[1,2,38] Song Cao,[1,2,38] Kathleen J. Imbach,[3,4,38] Marina A. Gritsenko,[5,38] Tung-Shing M. Lih,[6,38] Jennifer E. Kyle,[5,39] Tomer M. Yaron-Barir,[7,8,9,39] Zev A. Binder,[10,39] Yize Li,[1,2,39] Ilya Strunilin,[1,2,39] Yi-Ting Wang,[5,39] Chia-Feng Tsai,[5,39] Weiping Ma,[11,39] Lijun Chen,[6,39] Natalie M. Clark,[12] Andrew Shinkle,[1,2] Nataly Naser Al Deen,[1,2] Wagma Caravan,[1,2] Andrew Houston,[1,2] Faria Anjum Simin,[1,2] Matthew A. Wyczalkowski,[1,2] Liang-Bo Wang,[1,2]

(Author list continued on next page)

[1]Department of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA
[2]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA
[3]Josep Carreras Leukaemia Research Institute, Badalona, Spain
[4]Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
[5]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA
[6]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[7]Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA
[8]Englander Institute for Precision Medicine, Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA
[9]Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA
[10]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[11]Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[12]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[13]Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63130, USA
[14]Department of Neurology, Washington University in St. Louis, St. Louis, MO 63130, USA
[15]Department of Pathology, Spectrum Health and Helen DeVos Children's Hospital, Grand Rapids, MI, USA
[16]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA
[17]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA
[18]Department of Neurosurgery, Washington University School of Medicine, St. Louis, MO 63110, USA
[19]International Institute for Molecular Oncology, Poznań, Poland
[20]Poznan University of Medical Sciences, Poznań, Poland
[21]Van Andel Research Institute, Grand Rapids, MI, USA
[22]ICF, 530 Gaither Road Suite 500 Rockville, MD 20850, USA

(Affiliations continued on next page)

SUMMARY

Although genomic anomalies in glioblastoma (GBM) have been well studied for over a decade, its 5-year survival rate remains lower than 5%. We seek to expand the molecular landscape of high-grade glioma, composed of IDH-wildtype GBM and IDH-mutant grade 4 astrocytoma, by integrating proteomic, metabolomic, lipidomic, and post-translational modifications (PTMs) with genomic and transcriptomic measurements to uncover multi-scale regulatory interactions governing tumor development and evolution. Applying 14 proteogenomic and metabolomic platforms to 228 tumors (212 GBM and 16 grade 4 IDH-mutant astrocytoma), including 28 at recurrence, plus 18 normal brain samples and 14 brain metastases as comparators, reveals heterogeneous upstream alterations converging on common downstream events at the proteomic and metabolomic levels and changes in protein-protein interactions and glycosylation site occupancy at recurrence. Recurrent genetic alterations and phosphorylation events on PTPN11 map to important regulatory domains in three dimensions, suggesting a central role for PTPN11 signaling across high-grade gliomas.

INTRODUCTION

Glioblastoma (GBM) and other high-grade astrocytomas are among the most common and lethal brain tumors, with 5-year survival rates hovering below 5% despite decades of genomic analysis. Since the initial characterization of the GBM genome by The Cancer Genome Atlas in 2008, researchers have devoted significant effort to probing genomic and transcriptomic data for new therapeutic targets that will improve clinical outcomes. The contribution of deep genomic profiling to clinical classification

CellPress
OPEN ACCESS

Cancer Cell
Article

Erik Storrs,[1,2] Siqi Chen,[1,2] Ritvik Illindala,[1,13,14] Yuping D. Li,[1,13,14] Reyka G. Jayasinghe,[1,2] Dmitry Rykunov,[11] Sandra L. Cottingham,[15] Rosalie K. Chu,[16] Karl K. Weitz,[5] Ronald J. Moore,[5] Tyler Sagendorf,[5] Vladislav A. Petyuk,[5] Michael Nestor,[5] Lisa M. Bramer,[5] Kelly G. Stratton,[5] Athena A. Schepmoes,[5] Sneha P. Couvillion,[5] Josie Eder,[5] Young-Mo Kim,[5] Yuqian Gao,[5] Thomas L. Fillmore,[15] Rui Zhao,[5] Matthew E. Monroe,[5] Austin N. Southard-Smith,[1,2] Yang E. Li,[17,18] Rita Jui-Hsien Lu,[1,2] Jared L. Johnson,[7] Maciej Wiznerowicz,[19,20] Galen Hostetter,[21] Chelsea J. Newton,[21] Karen A. Ketchum,[22] Ratna R. Thangudu,[22] Jill S. Barnholtz-Sloan,[23] Pei Wang,[11] David Fenyö,[24,25] Eunkyung An,[26] Mathangi Thiagarajan,[27] Ana I. Robles,[26] D.R. Mani,[12] Richard D. Smith,[5] Eduard Porta-Pardo,[3] Lewis C. Cantley,[7,28,29] Antonio Iavarone,[30,31] Feng Chen,[1,13] Mehdi Mesri,[26] MacLean P. Nasrallah,[32] Hui Zhang,[6,33,34] Adam C. Resnick,[35,36] Milan G. Chheda,[1,13,14,*] Karin D. Rodland,[37,*] Tao Liu,[5,*] and Li Ding[1,2,13,17,40,*] Philadelphia Coalition for a Cure, and Clinical Proteomic Tumor Analysis Consortium

[23]Center for Biomedical Informatics and Information Technology & Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20850, USA
[24]Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY 10016, USA
[25]Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY 10016, USA
[26]Office of Cancer Clinical Proteomics Research, National Cancer Institute, Rockville, MD 20850, USA
[27]Frederick National Laboratory for Cancer Research, Frederick MD 21701, USA
[28]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[29]Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA
[30]Department of Neurological Surgery and Department of Biochemistry, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[31]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[32]Department of Pathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[33]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[34]Department of Urology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[35]Center for Data Driven Discovery in Biomedicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[36]Division of Neurosurgery, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[37]Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA
[38]These authors contributed equally
[39]These authors contributed equally
[40]Lead contact
*Correspondence: mchheda@wustl.edu (M.G.C.), rodland@ohsu.edu (K.D.R.), tao.liu@pnnl.gov (T.L.), lding@wustl.edu (L.D.)
https://doi.org/10.1016/j.ccell.2024.06.004

has been profound, culminating in the WHO re-classification of *IDH1* or 2 mutant tumors as grade 4 astrocytomas, rather than GBMs.[1] Despite this work, minimal progress has been made in treating high-grade gliomas. As targeted therapeutics rely heavily on modulation of protein function, we undertook an extensive study of proteins and post-translational modifications (PTMs) in 228 human GBM and grade 4 astrocytomas from 200 patients, to identify regulatory networks that impact the behavior of high-grade glioma to promote the development of small molecule inhibitors capable of changing disease trajectory.

Proteogenomics, as exemplified by the Clinical Proteomic Tumor Analysis Consortium analyses of adult GBM[2] and pediatric brain tumors,[3] focuses on applying high quality genomic, transcriptomic, and proteomic measurements to statistically significant numbers (low hundreds) of prospectively acquired tumor specimens to illuminate signal transduction and regulatory processes that may affect cell function and tumor progression. In this study, we have extended our prior results[2] to an independent cohort of 200 patients and added a limited set of paired primary-recurrent tumors from the same patient to provide an initial glimpse of proteogenomic changes associated with progression. Where feasible, we have added limited experiments directed at functional validation of key observations regarding the effects of *IDH1* variants and *PTPN11* variants on downstream signaling events, as well as providing orthogonal immunohistochemistry-based measurements using the CODEX platform, which provides spatial information relevant to the role of the tumor microenvironment (TME).

RESULTS

We characterized the proteogenomic and metabolic landscape of 228 tumor samples from 200 patients with high-grade glioma (HGG), as well as 18 unmatched normal brain samples, 14 brain metastases, and resections at multiple time points taken from a longitudinal cohort of 25 patients (Figures 1A and S1; Table S1). Overall, we generated a comprehensive dataset, including WXS (whole exome sequencing), WGS (whole genome sequencing), DNA methylation, miRNA, mRNA, snRNA (single-nuclei RNA sequencing), proteome, phosphoproteome, acetylome, glycoproteome, selected reaction monitoring (SRM), PRISM (high-pressure, high-resolution separations with intelligent selection and multiplexing)-SRM, IMAC (immobilized metal affinity chromatography)-SRM, metabolome, and lipidome data (Figure 1B). This enabled investigation of the proteogenomic changes at tumor recurrence using 53 longitudinal samples, and the shared and unique molecular and post-translational events that occur beyond genetic alterations using the full cohort (Figure 1C).

High-grade glioma evolution from diagnosis to recurrence is associated with genomic and proteomic drivers

We analyzed a longitudinal cohort of 53 tumor samples collected from 25 patients with grade 4 IDH-WT (GBM) and IDH-mutant astrocytoma (progression-free survival 1 month to four years; Figure 2A) to identify proteins differentially expressed in recurrent tumors (Figure 2B). Protein abundance changes largely

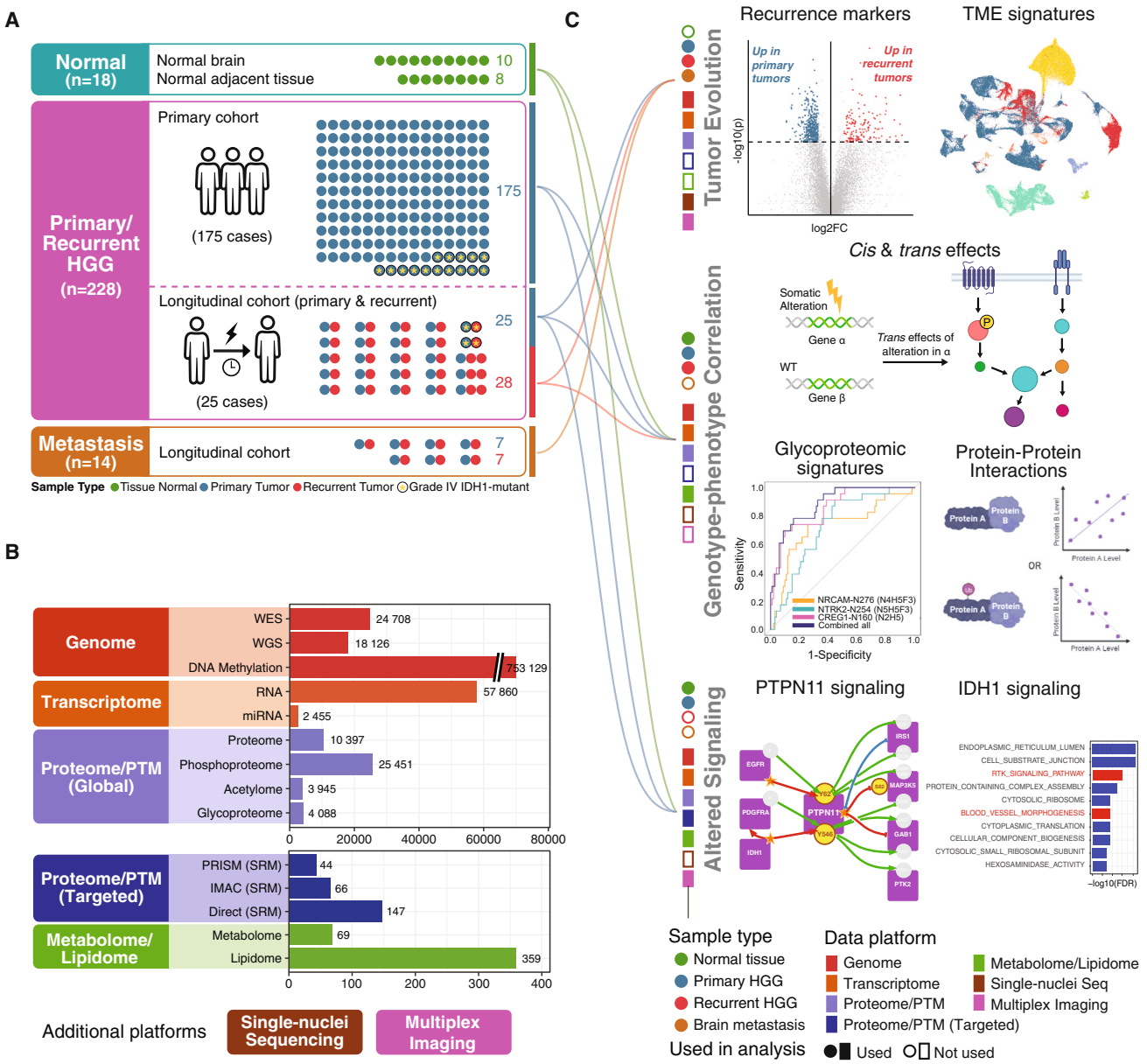


Figure 1. Overview of study cohort, technology platforms, and assays

(A) Cohort overview.

(B) Features quantified on 14 data platforms (excluding single-nuclei sequencing and multiplex imaging). Specific features quantified on each platform are: WXS-somatic variants, WGS-somatic variants, DNA methylation-CpG probes, RNA-RNA transcripts, miRNA-miRNA transcripts, proteome-proteins, phosphoproteome-phosphoproteins, acetylome-acetylproteins, glycoproteome-glycoproteins, PRISM SRM-proteins, IMAC SRM-phosphoproteins, direct SRM-proteins, metabolome-metabolites, lipidome-lipids. PRISM: high-pressure, high-resolution separations with intelligent selection and multiplexing. IMAC: immobilized metal affinity chromatography. SRM: selected reaction monitoring.

(C) Sub-cohorts used in each analysis.

See also Figure S1 and Table S1.

correlated with RNA (Figure S2A) and top differentially abundant proteins were shared by most cases (Figure S2B). Pathway analysis of proteins upregulated at recurrence revealed enriched biological processes related to oxidative phosphorylation, reactive oxygen species, and myogenesis. Proteins downregulated at recurrence were enriched in cell cycle, MYC targets, and DNA repair pathways.

We explored kinase activity in recurrent tumors compared to primary tumors based on differentially regulated phosphorylation sites using the Kinase Library, an atlas of experimentally validated substrate specificities for more than three hundred serine/threonine kinases (Figure 2D).[4] Splicing kinases such as CLK1-4 and SRPK1-3 are significantly more active in primary tumors, in concordance with our transcriptomic data (Figures 2E;

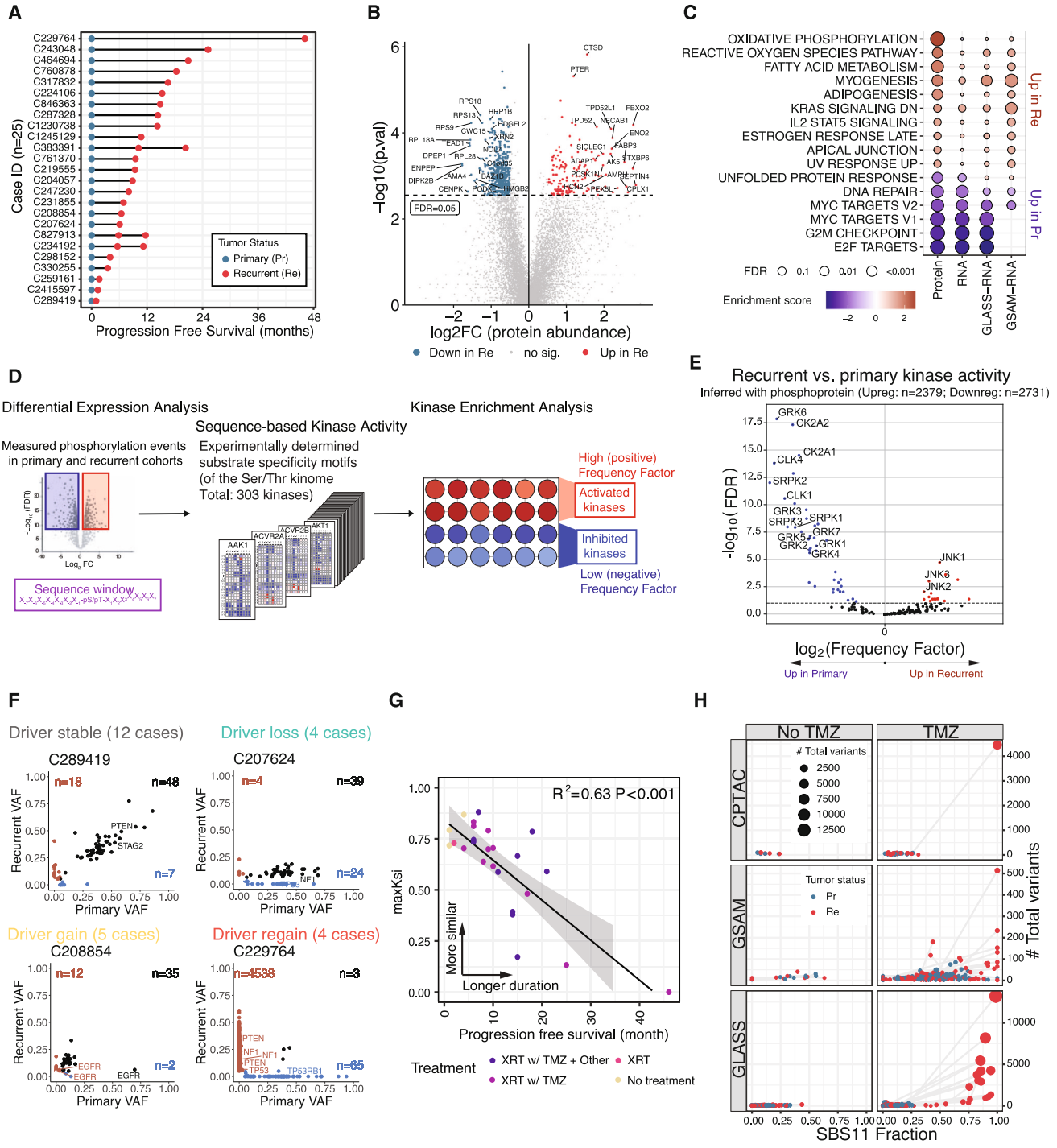


Figure 2. Proteogenomics of tumor evolution

(A) Primary and recurrent tumor sample collection for 25 patients with grade 4 astrocytoma.

(B) Differentially abundant proteins between recurrent and primary tumors (FDR < 0.05, log2FC > 0.5).

(C) Differentially enriched pathways in the current cohort, and GLASS and GSAM cohorts.

(D) Workflow of Kinase Library enrichment analysis.

(E) Kinase activity analysis comparing recurrent and primary tumors.

(F) Demonstrative variant allele frequency (VAF) correlation plots indicating 4 modes of driver mutation progression observed: stable (all mutations persist, 12 cases), loss (all unique mutations are in primary tumor, 4 cases), gain (all unique mutations are in recurrent tumor, 5 cases), and regain (both gain and loss of different mutations in the same gene, 4 cases). Number of mutations that are shared across timepoints, unique to either primary or recurrence are labeled. Demonstrative cases were randomly chosen.

(legend continued on next page)

Table S2). Moreover, casein kinase 2 (CK2) family (CK2A1 and CK2A2) and G protein-coupled receptor kinase (GRK) family (GRK1–7) members were more highly expressed in primary tumors, suggesting a role for G protein-coupled receptor (GPCR) signaling in tumor progression. Conversely, the c-Jun N-terminal kinase (JNK) family, typically activated under stressors like inflammation and hypoxia,[5] was upregulated in recurrent tumors.

To study tumor genetic evolution, we examined mutational shifts in each primary-recurrent pair (Figure 2A). We calculated variant allele frequency (VAF) for all mutations and observed shared and unique mutations in each pair. Since VAF can be biased by overall tumor purity, we compared the presence of variants (all variants in driver genes, other variants passing VAF > 0.05) instead of absolute VAF values. Focusing on the presence of driver gene mutations over time, we observed four patterns: stability, loss, gain, and regain of driver mutations (Figures 2F and S2C). No single pattern, nor any accumulation of a driver gene alteration, dominated the cohort, indicating heterogeneous mechanisms are associated with recurrence (Figure S2D). Expanding the focus to all somatic mutations, we found that patients with a shorter time to progression had more shared mutations between primary and recurrent tumors (Figure 2G). We quantified this trend by using clonal similarity statistics (maxKsi), which describe the possibility of observing shared mutations independently with a background GBM mutational frequency distribution. Clonal similarity negatively correlated with progression time, suggesting that recurrent tumors continuously evolved from their primary tumors (Pearson R^2 = 0.63, $p < 0.001$) (Figure 2G). We conducted the same analysis on two published longitudinal HGG cohorts, Glioma Longitudinal Analysis Consortium (GLASS)[6] and Glioblastoma, Stability of Actionable Mutations (GSAM).[7] While the GLASS cohort showed a similar negative correlation (Figure S2E), the GSAM cohort showed no correlation (Figure S2F) possibly due to the small number of mutations detected by targeted-panel sequencing.

Mutational signature analyses of primary and recurrent tumors identified two dominant signatures, SBS1 and SBS11, that associate with age and temozolomide treatment,[8] respectively (Figure S2F). SBS1 is positively associated with patient age (Pearson R = 0.56, p = 0.002) (Figure S2H). SBS11, a signature for temozolomide-induced DNA damage, was highly present in one recurrent tumor from a patient treated with eleven cycles of temozolomide (TMZ) that had recurrence at four years (Figure 2H). This patient's primary and recurrent tumors showed distinct mutational profiles, with 4,439 mutations in the recurrent tumor, compared to 68 in the primary tumor, including mutations in *POLE*, *MSH2*, and *MSH6*. Such hypermutation with a high-SBS11 signature was also observed in a subset of TMZ-treated patients in the GSAM and GLASS cohorts but not in patients who did not receive TMZ (Figure 2H). Previously, SBS11-associated hypermutation was reported with high-grade transformation from low-grade glioma.[9,10] We found the same in recurrent grade 4 astrocytomas.

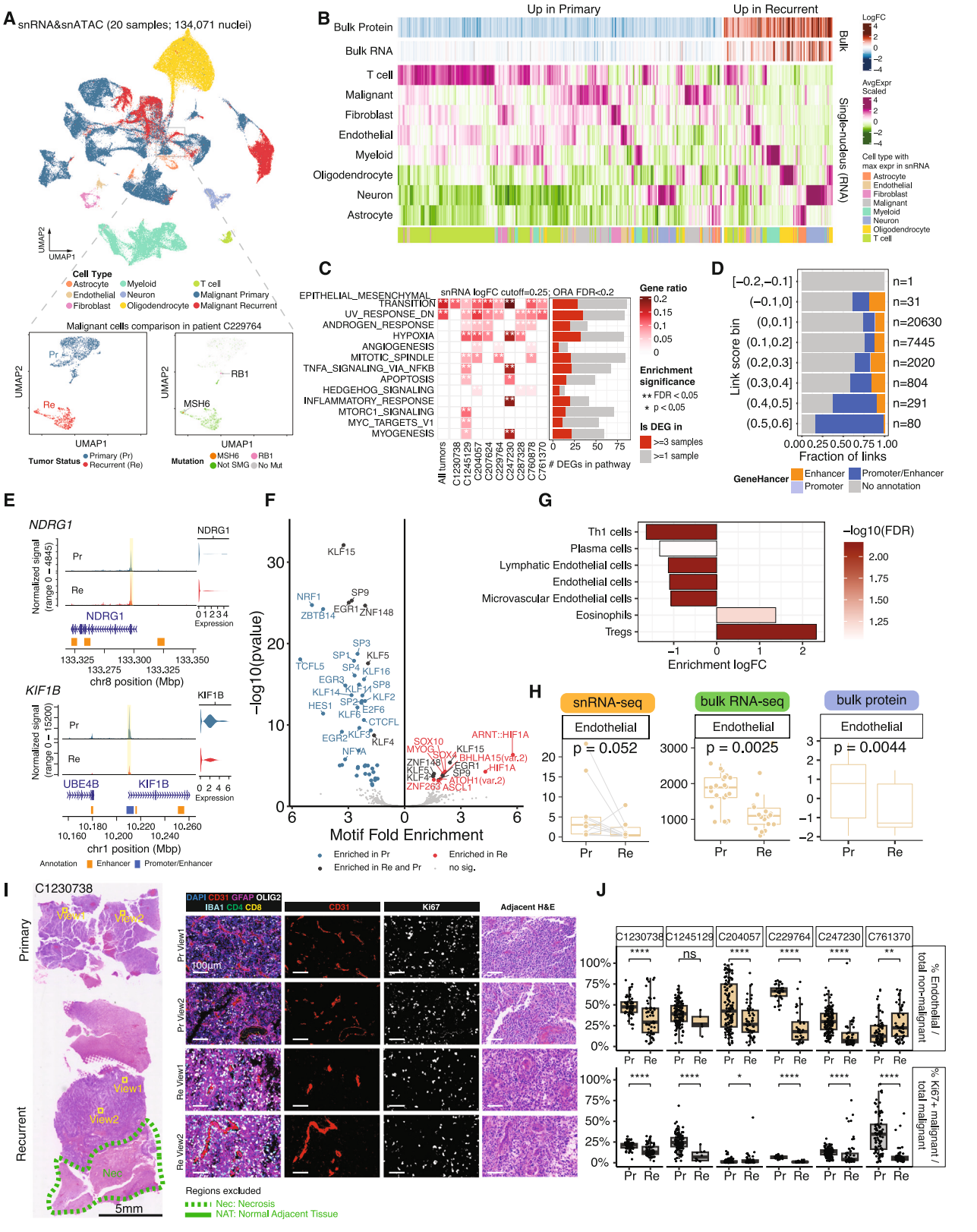
Tumor-intrinsic and TME-associated features in primary and recurrent high-grade glioma

To delineate tumor intrinsic and TME-associated features at recurrence, we performed snRNA analysis on 10 primary and recurrent HGG pairs, represented by 134K nuclei. The overlap between snRNA profiles of primary and recurrent gliomas varied considerably, with one extreme case (patient C229764) exhibiting distinct tumor cell clusters from primary and recurrent samples (Figure 3A), consistent with the individual's high mutational burden from bulk WXS. A shift in driver mutation is further supported by *RB1* and *MSH6* mutations detected from snRNA transcripts, mutations unique to the patient's primary and recurrent tumor cells, respectively (Figure 2F).

To understand how specific cell types contribute to differentially abundant proteins (Figure 2B), we calculated the average transcriptional expression from each cell type using snRNA (Figure 3B). Diverse cell types, including malignant cells, immune cells, and stroma cells, contributed to up- and downregulated proteins (Figure 3B). To investigate transcriptional changes intrinsic to malignant cells, we performed cohort-level and patient-level differential expression analysis between malignant cells in primary and recurrent tumors (one sample with fewer than 50 malignant cells was excluded). Differentially expressed genes (DEGs) between primary and recurrent malignant cells were enriched in epithelial-mesenchymal transition, UV response, androgen response, and hypoxia pathways (Figure 3C) with shared DEGs across patients (Figure 3D). Using matching snATAC-seq data, we identified 78,787 accessible chromatin regions (ACRs) overlapping with known gene regulatory elements such as promoters and enhancers annotated in the GeneHancer database[11] (Table S3, Methods). Correlation between ACR accessibility and gene expression data identified 14,280 ACRs as regulatory elements with 31,302 ACR-gene correlations (Table S3). 31% of such ACRs have been previously described in the GeneHancer database and show a stronger link score, whereas 69% are previously unreported (Figure 3D). Differentially accessible chromatin regions (DACRs) between primary and recurrent malignant cells at the promoter and enhancer regions of genes such as NDRG1 and KIF1B indicate potential epigenetic regulation in recurrent malignant cells (Figure 3E). Motif enrichment analysis indicates differential transcription factor activity at recurrence (Figure 3F). Primary malignant cells show higher activity in transcription factors NRF1, Sp/KLF, and the E2F family previously described in GBM related to a metabolic-active and proliferative state.[12–14] Recurrent malignant cells show activity in SOX4, SOX10, and HIF1A, suggesting a switch toward a stem-like state post-treatment[15,16] (Table S3).

Recurrence was also associated with changes in the TME. Cell type enrichment analysis with bulk RNA-seq showed the

(G) Progression-free survival negatively correlates with tumor clonal similarity statistics (MaxKsi), based on somatic mutations (R^2 = 0.63, p value < 0.001).
(H) Relationship between fraction of somatic mutations with SBS11 signature and total number of somatic mutations in each tumor sample, categorized by whether patient received TMZ treatment prior to recurrent tumor collection. GLASS and GSAM cohort shown for comparison. Gray line connects tumor samples from same patient.
See also Figure S2 and Table S2.



(legend on next page)

expected high neuron content in normal and normal-adjacent tissues compared to HGGs and high epithelial content in brain metastases of epithelial origin compared to HGGs (Figure S3A). Comparing recurrent and primary tumors showed a significantly decreased enrichment score of endothelial cells (paired Wilcoxon test, $p = 0.0048$) and type 1 T helper (Th1) cells ($p = 0.0012$) and a significant increase in regulatory T (Treg) cells ($p = 0.0032$) in recurrent tumors (Figure 3G). A decrease in endothelial cells was also supported by snRNA-seq and protein deconvolution data performed on the same longitudinal tumor samples (Figures 3H and S3B) and corroborated by the GLASS and GSAM cohorts (Figure S3C). Multiplex imaging with codetection by indexing (CODEX) on 12 samples indicated major cell types in the brain (Figures 3I and S3D; Table S4). Whole-slide quantification showed a decrease in fraction of endothelial cells over all non-malignant cells in 5 out of 6 patients at recurrence (Figure 3J). However, recurrent tumors exhibited highly heterogeneous vascular morphology, such as large blood vessels embedded in dense malignant cells (C1230738), engorged vessels surrounded by necrosis and normal adjacent tissue (C1245129), and sparse microvasculature with overall low cellularity (C761370) (Figure S3E). Additionally, we observed a decreased proliferative rate in glial-lineage cells in dense tumor ($GFAP^+$ and/or $OLIG2^+$), i.e., areas enriched for malignant cells, in 5 out of 6 patients (Figure 3J), concordant with a decrease in the cell-cycle pathway observed from bulk RNA-seq and proteomics (Figure 2C). Taken together, recurrent high-grade gliomas demonstrated transcriptional and epigenetic changes in malignant cells and TME remodeling.

Impact and association of genetic alterations on proteomics and metabolomics in HGG

After evaluating progression-associated events, we conducted an alteration association analysis in primary tumors. A snapshot of the driver gene profile in the cohort demonstrates which gene alterations significantly co-occur or are mutually exclusive. *Cis* analyses highlight the effects of each highly altered driver gene on its own associated RNA, protein, and PTM levels. *Trans* effects analyses profile the alterations' effects on distal molecular events. Finally, evaluating drivers in a pairwise manner identified drivers that are individually associated with similar downstream effects.

We identified 13 highly altered driver genes in this cohort (Table S5). Conducting an alteration co-occurrence analysis for all possible pairs of the 13 driver genes confirmed several observations from TCGA,[17] such as a high co-occurrence of *RB1* and *TP53* alterations. We also observed mutual exclusivity of *EGFR* and *PTPN11* alterations, which in retrospect was also present in the TCGA cohort (Figure 4A). In *cis* analyses, tumor suppressors, such as *ATRX* and *RB1*, exhibited relatively lower RNA and protein expression when the gene was altered, whereas oncogenes such as *EGFR* and *PDGFRA* displayed elevated levels of their RNA and protein (Figure 4B). Analysis of PTMs showed that when altered, ATRX exhibited increased phosphorylation at S784, while mutant IDH1 demonstrated a decrease in acetylation at K224 (Figure 4B). While alteration of epidermal growth factor receiver (EGFR) was associated with a slight increase in phosphorylation at Y316, N2H8F0S0G0 glycosylation at N603 was the strongest *cis* PTM effect observed for EGFR (Figure 4B).

To understand common downstream signaling cascades within tumors, we asked which drivers produce similar downstream effects when either gene was altered. We ran *trans* analysis in a pairwise manner for all possible driver combinations. The impact of a given driver's alteration on *trans* effects was evaluated in samples lacking alteration in the comparator driver before conducting the reverse assessment for the other driver in the pair (Figure 4C). For comparative assessments, we grouped two main classes of *trans* effects. The first group comprised effects on proteins and PTMs; the second group comprised effects on the lipidome and metabolome. Driver pairs that exhibited strong concordance in their protein/PTM consequences exhibited the same positive association in their effects on lipids and metabolites (Figures 4D; Table S5).

Of all investigated driver pairs, *TERTp* and *PTEN* alterations correlated most strongly in their *trans* effects on proteins and PTMs (Figure 4E, left, Pearson R = 0.62, p < 1e-100). Among the large number of downstream similarities in protein and

Figure 3. Single-nuclei sequencing reveals malignant cell-intrinsic and tumor microenvironment (TME)-associated signatures in recurrent grade 4 astrocytomas

(A) UMAP of snRNA-sequencing data from 10 primary-recurrent pairs.

(B) Heatmap showing scaled average snRNA expression of differentially abundant proteins between recurrent and primary tumors. Top two rows: log-fold change of bulk protein and RNA. Middle eight rows: average expression of same genes from snRNA in indicated cell type, scaled by both row and column. Bottom row: cell type having highest snRNA expression of given gene product.

(C) Malignant cell-intrinsic dysregulated pathways in recurrent tumors detected from snRNA-based differentially expressed gene (DEG) between each case's recurrent versus primary tumor. Red in barplot: number of genes differentially expressed in at least 3 cases.

(D) Correlation (link score) distribution between accessible chromatin region (ACR) accessibility and gene expression in malignant cells. ACRs overlapping with GeneHancer are annotated. Number of links shown on right.

(E) Examples of chromatin accessibility and snRNA expression in primary and recurrent malignant cells for two genes.

(F) Motif enrichment in differential ACRs. Motifs significantly enriched (hypergeometric test with Benjamini-Hochberg multiple test correction, FDR < 0.05) in red (DACRs in primary tumors), blue (DACRs in recurrent tumors), and gray (both sets of DACRs).

(G) Cell-type enrichment score comparison from bulk RNA-seq representing the TME cell-type composition between recurrent and primary tumors (differentially enriched, FDR < 0.1).

(H) Endothelial cell composition of total tumor microenvironment population is decreased in recurrent tumor, from snRNA cell counts and bulk RNA and bulk protein deconvolution fractions. Boxplots show interquartile range (IQR; boxes span first to third quartiles, middle line indicates median and whiskers show closest values within 1.5x IQR). Paired samples from the same patient are connected by lines. Wilcoxon *p* values were calculated.

(I) Multiplexed immunofluorescence results (CODEX) of three primary-recurrent tumor pairs.

(J) Proportion of endothelial cells and Ki67+ malignant cells from 6 primary-recurrent matched tumor sections are quantified in 1 mm × 1 mm tiles respectively. Boxplots show interquartile range. Wilcoxon *p* values were calculated (ns: $p > 0.05$, $*p \leq 0.05$, $**p \leq 0.01$, $****p \leq 0.0001$). See also Figure S3 and Tables S3 and S4.

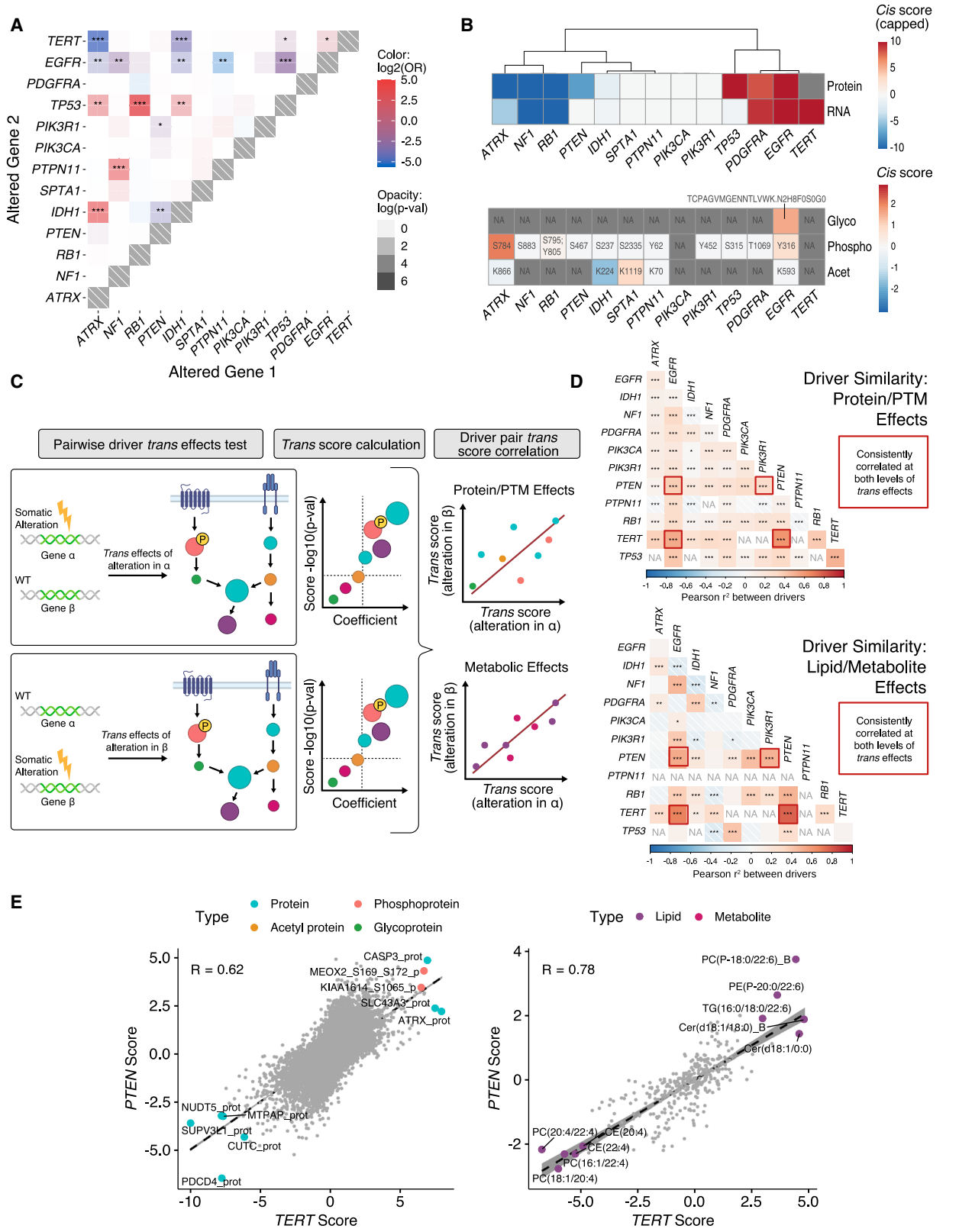


Figure 4. *Cis* and *trans* results highlight similarities between primary tumors at different omics levels

(A) Somatic alteration mutual exclusivity analysis. Significance of co-occurrence is indicated by both opacity and asterisks. Fisher's exact test with Benjamini & Hochberg adjusted *p* values were calculated (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

(legend continued on next page)

PTM profiles, the two genes were associated with diminished PDCD4 expression and elevated CASP3 expression, suggesting each had a similar influence on cell death pathways (Figure S4A). Effects of *PTEN* and *TERTp* alterations on the metabolome and lipidome were similarly strongly positively associated (Figure 4E, right, Pearson R = 0.78, $p < 8.96e-81$). *EGFR* and *TERTp* alterations demonstrated a similar scenario. We observed a positive association in the protein and PTM effects when either of the two genes were altered (Figure S4B, left, Pearson R = 0.53, $p = 0$); among these were lower SUPV3L1 and increased MTARC2 protein levels relative to samples WT for both genes (Figure S4C). The positive correlation between *EGFR* and *TERTp* was similarly strong when compared on the metabolomic and lipidomic scale, (Figure S4B, right, Pearson R = 0.68, $p < 3.10e-57$).

Associations between glycosylation and phosphorylation in EGFR activities and tumor recurrence

As protein glycosylation is involved in numerous biological functions, and membrane-bound or secreted glycoproteins are potential therapeutic targets and/or biomarkers,[18–21] we performed quantitative glycoproteomic analysis on our entire cohort, revealing 671 upregulated and 674 downregulated N-linked intact glycopeptides (IGPs) in grade 4 astrocytomas relative to normal brain (Figure 5A). The majority of upregulated IGPs originated from glycoproteins that were associated with ECM-receptor interactions or complement and coagulation cascades (Figures 5A; Table S6).

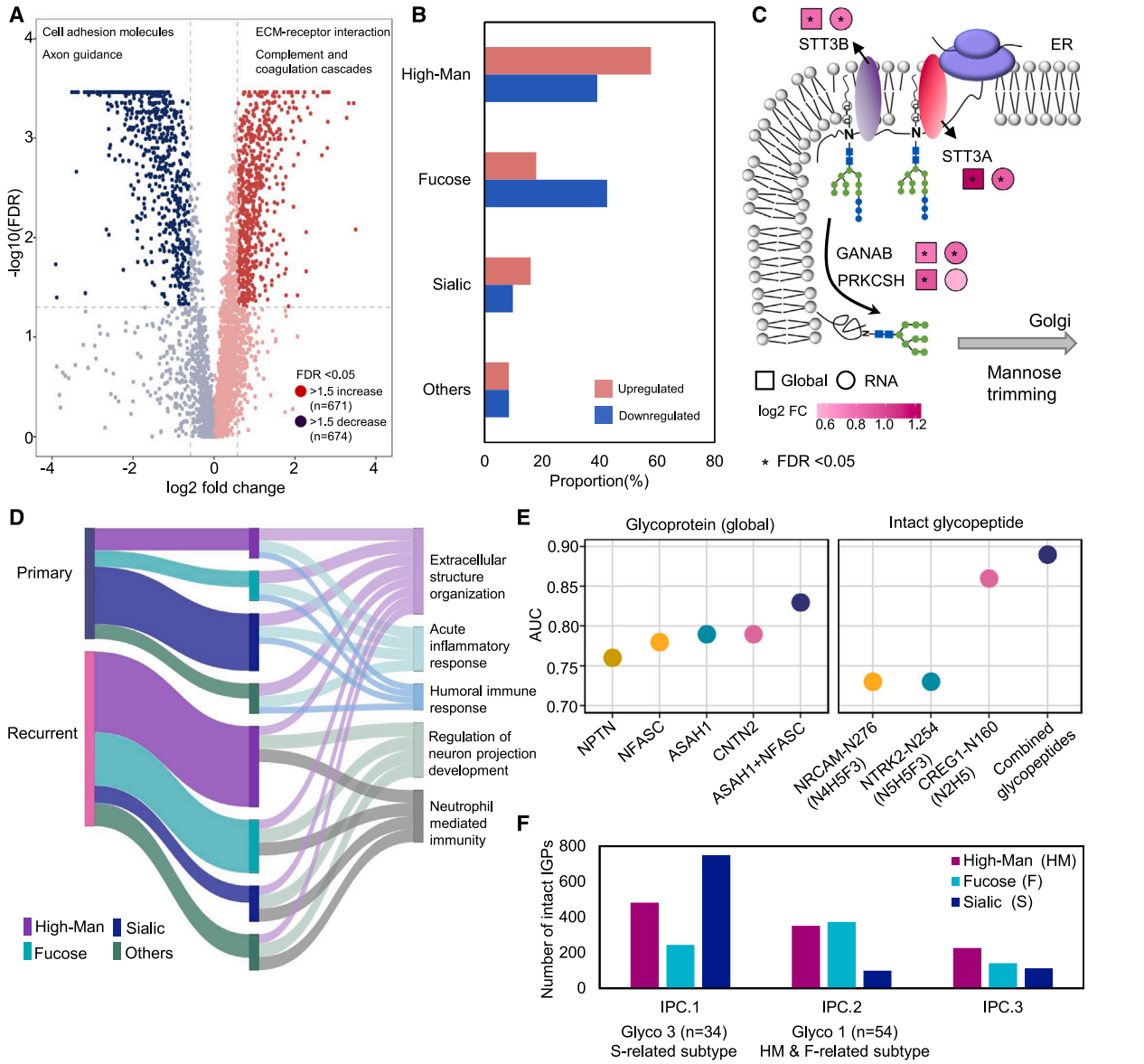
We investigated glycan types according to the monosaccharide compositions of the identified glycopeptides: oligomannose only (High-Man), fucosylated glycans only (Fucose), sialylated glycans including sialylation only and sialylation with fucosylation (Sialic), and glycans that did not fit into the first three categories (Others). A majority of upregulated IGPs (primary vs. normal) contained High-Man glycans, while downregulated IGPs were mostly fucosylated (Figure 5B). We found that the expression profile of N-linked IGPs was mostly positively associated with global protein expression of the corresponding glycoproteins (Figure S5A). However, site-specific heterogeneity in glycan occupancy can contribute to variation in IGP abundances observed from the same proteins.[22,23] Moreover, glycan synthesis and conjugation to glycoproteins are regulated by glycosylation enzymes,[24] implicating enzymes with aberrant expression profiles as potential treatment targets. In this particular cohort, overexpression of STT3A, STT3B, GANAB, and PRKCSH was associated with primary tumors compared to normal tissues (Figures 5C; Table S6). STT3A and STT3B are catalytic subunits of the OST complex, which is involved in transferring the Glc3Man9GlcNAc2 precursor from lipid-linked oligosaccharides to the NXS/T motif on nascent polypeptides.[25] GANAB and PRKCSH are responsible for trimming the two innermost α1,2-Glc residues from the precursor to produce Man9GlcNAc2 and releasing N-linked glycoproteins from the ER.[26,27] The upregulation of these glycosylation enzymes indicates an increase in N-linked glycosylation in high-grade gliomas. Most of the upregulated IGPs with High-Man glycans of N2H8 or N2H9 (N=HexNAc and H=Hex, Figure S5B) further suggests incomplete/premature termination of N-glycan biosynthesis. The elevation in High-Man may impact the functions of the glycoproteins enhancing cancer progression.[28,29]

We used glycoproteomic data to identify potential recurrent HGG-associated glycoproteins. Compared to primary tumors, overexpressed IGPs of recurrent tumors were mostly High-Man or Fucose glycans from glycoproteins involved in biologic processes such as regulation of neuron projection development and neutrophil mediated immunity (Figures 5D; Table S6). We identified four glycoproteins (CNTN2, NPTN, ASAH1, and NFASC) with elevated glycosylation activities that may relate to global expression in recurrent tumors (Figure S5C). Abundances of these glycoproteins successfully differentiated recurrent and primary tumors via receiver operating characteristic (ROC) analysis (Figures 5E and S5D). A panel combining ASAH1 and CNTN2 had an area under the curve (AUC) of 0.83. Intriguingly, both glycoproteins were associated with neutrophil mediated immunity (Table S6). In addition, we examined glycoproteins with increased glycosylation activities in recurrent tumors (vs. primary tumors) at the IGP level but not at the global level. Three IGPs, NRCAM-N276 (N4H5F3), NTRK2-N254 (N5H5F3), and CREG1-N160 (N2H5) discriminated recurrent from primary tumors, achieving an AUC of 0.89 when combined in a panel (Figures 5E and S5D).

Focusing on primary tumors, we defined three glycoproteomic subtypes (Glyco 1–3) using unbiased glycoproteomic grouping of tumors with three intact glycopeptide clusters (IPC 1–3) (Figure S5E). Glyco 1 was associated with IPC 2, characterized by axon development and neuron projection guidance (Table S6; Figures 5F, S5E, and S5F). Additionally, tumors in Glyco 1 were associated with the proneural subtype ($p < 0.05$, hypergeometric test, Table S6). Glyco 2 was associated with the classical subtype, whereas Glyco 3 was related to the mesenchymal subtype. Furthermore, Glyco 3 was associated with IPC 1, which primarily contained Sialic glycans and was characterized by extracellular structure organization and acute inflammatory response (Table S6; Figures 5F, S5E, and S5F).

As a case study of the value of integrating multiple data types, we analyzed the spatial relationship between the most common *EGFR* somatic mutations observed in this cohort (A289 and G598), the significantly upregulated phosphorylation event

(B) *Cis* effects for 13 common drivers demonstrating how gene alterations impact their own RNA and protein (top), and post-translational changes (bottom), relative to tumors WT for the given gene. Heatmaps reflect *cis* result score (capped at +/−10 for RNA and protein scores, gray where measure was N/A).
(C) *Trans* analysis: effects of somatically altered genes on other proteins, post-translational modifications (PTMs), metabolites, and lipid levels. *Trans* effects were measured pairwise for each somatically altered gene. Each *trans* effect was scored for each pair and correlation between pairs was calculated.
(D) Driver gene similarity scores were evaluated at the protein/PTM and lipid/metabolite level. Colors indicate the Pearson correlation coefficient (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).
(E) Pearson correlation in *trans* scores between *TERTp* and *PTEN* effects at protein/PTM levels (p < 1e-100) and lipid/metabolite levels ($p < 8.96e-81$). Top 5 positively and negatively scored events are labeled and colored according to data type.
See also Figure S4 and Table S5.



(legend on next page)

implicated by the *cis* analysis (Y316), and the glycosylation of N352 and N603 using an EGFR active dimer structure in the UCSF Chimera software.[30,31] Visualizing the modifications on the EGFR dimer demonstrates that the Y316 residue is located at an interface between the two EGFR molecules' extracellular domains (Figure 5G). Moreover, the G598 mutation colocalized with the N603 glycosylation, both of which occur approximately 47 amino acids before the transmembrane alpha helix of the protein, and N603 is related to the autoinhibitory tether interactions of EGFR.[32,33] In contrast, when mapped onto the structure, the N352 glycosylation site falls 58 amino acids from the A289 mutation and is adjacent to the binding site of both EGF and cetuximab[30,31] (Figures 5G and S5G). The glycosite N352 is essential for EGFR to maintain its functional conformation to allow EGF binding.[32,33] Although the impact of glycosylated N352 and N603 on the phosphorylation in *EGFR*-altered tumors is not yet understood, we found positive associations among Y316 phosphorylation and N352 ($p = 0.043$) and N603 ($p = 0.0069$) glycosylations in *EGFR*-altered but not *EGFR*-WT, tumors (Table S6). Thus, mapping mutations and PTMs identified in this cohort suggests opportunities to understand steric changes underlying the biology of driver events. These data may be useful for future PTM crosstalk/cross-correlation studies.

Protein-protein interaction analysis assisted by targeted proteomics and the Kinase Library identifies pathway disruptions

We next leveraged global proteomic data to detect indications of disrupted protein-protein interactions (PPIs) beyond the initial EGFR focus, using protein abundance correlations as a mathematical proxy to estimate possible interactions[34] (Figure 6A, left). For this analysis, we prioritized proteins known to interact as supported by more than three databases. Baseline calculations in primary tumors demonstrated strong positive associations in protein subunits and other proteins well-known to associate, such as PIK3R1 and PIK3CA, and negative associations between other proteins, such as PKN2 and PDPK1 (Figure 6A, right). To identify the possible role of PPI aberrations in HGG, we examined differences in correlations between proteins encoded by genetically altered and WT primary tumors before conducting a similar assessment in matched primary and recurrent tumors. Overall, we found many putative disruptions between interactors within the EGFR and platelet derived growth factor receptor alpha (PDGFRA) signaling pathways (Figure 6B).

Somatic alterations, particularly mutations, interrupt PPIs and in turn contribute to oncogenic phenotypes.[35] We examined all database-supported PPIs involving at least one highly altered gene in our dataset, comparing protein partner correlations between samples with or without a somatic alteration in the gene encoding that protein. Fourteen protein relationships exhibited significant aberrations in correlation when in somatically altered samples ($p < 0.01$), and half of these significant hits involved interactions with EGFR (Figure 6C). Closer examination of RB1-MDM4 and PTPN11-IRS1 showed that *RB1* and *PTPN11* alterations were associated with strong negative protein abundance correlations between interaction partners in primary tumors, respectively, suggesting that mutations in these proteins disrupt their interactions (Figure 6D). In contrast, wild-type EGFR-ERBB4 and EGFR-CBL interactions showed the expected positive correlations in abundance, while samples with *EGFR* alterations exhibited loss of these associations (Figure S6A). In summary, we identified disruption of PPIs as a common effect of somatic gene alterations, providing potential insight into biology and/or therapy.

Using a similar PPI analysis, we examined whether interacting proteins exhibited altered correlations between primary and recurrent samples. Focusing specifically on PPIs that involved at least one well-known cancer driver gene, we selected three protein interaction compartments: PPIs correlated in primary tumors but not in recurrent tumors, PPIs correlated in both primary and recurrent tumors, and PPIs correlated in recurrent but not primary samples (Figure 6E). This undertaking identified 85 significant oncogenic-related PPI correlations related to recurrence status, including PPIs highly correlated in both primary and recurrent tumor groups ($p < 0.01$, Figure S6C). Examination of some examples showed that while RB1 and CDK2 protein abundance did not significantly correlate in primary tumor samples, they corresponded strongly in recurrent tumors (Figure 6F, left). Although RB1 and CDK6 protein abundance strongly correlated in both primary and recurrent tumor samples, CDK2 and CDK6 exhibited an association only in recurrent samples (Figure S6C). Similarly, AKT1 and XIAP protein abundance is highly correlated in recurrent, but not primary, tumors (Figure 6F, right). In contrast, other protein pairs, such as TP53 and UBE2I or EGFR and NUMBL, significantly correlated in primary samples while their association was absent in recurrent tumors (Figure S6D). These findings suggest that recurrent tumors may diverge from their primary counterparts through differences in protein interactions, possibly reflecting a treatment-resistant phenotype.

We next investigated how a gene alteration affects signaling pathways beyond its own cognate protein by comparing phosphorylation sites of five altered genes in mutated and WT tumors (Figure 6G). While some mutations activated similar kinases, probing the full landscape of the human kinome revealed distinct kinase activity patterns across the different mutation contexts. For example, tumors with both *RB1* and *TP53* mutations

Figure 5. Altered glycosylation programs in grade 4 astrocytomas and topologic relation between somatic mutations, glycosites, and phosphosite on EGFR

(A) Differential analysis of primary tumors vs. normal brain in four predominantly tumor-enriched pathways.
(B) Glycan type distributions based on upregulated/downregulated intact glycopeptides in primary tumors (vs. normal brain).
(C) Glycosylation enzymes associated with altered glycosylation activities in primary tumors.
(D) Glycan type distribution and biological processes associated with recurrent grade 4 astrocytomas vs. primary tumors.
(E) The dots (matching ROC curve colors in Figure S5D) correspond to the AUCs of the glycoproteins and IGPs that may serve as potential biomarkers for recurrent grade 4 astrocytoma (vs. primary tumors).
(F) Glycan type distribution for intact glycopeptide clusters (IPCs) associated with glycoproteomic subtypes.
(G) EGFR dimer, demonstrating spatial relationships between somatic mutations, phosphosite, and glycosites.
See also Figure S5 and Table S6.

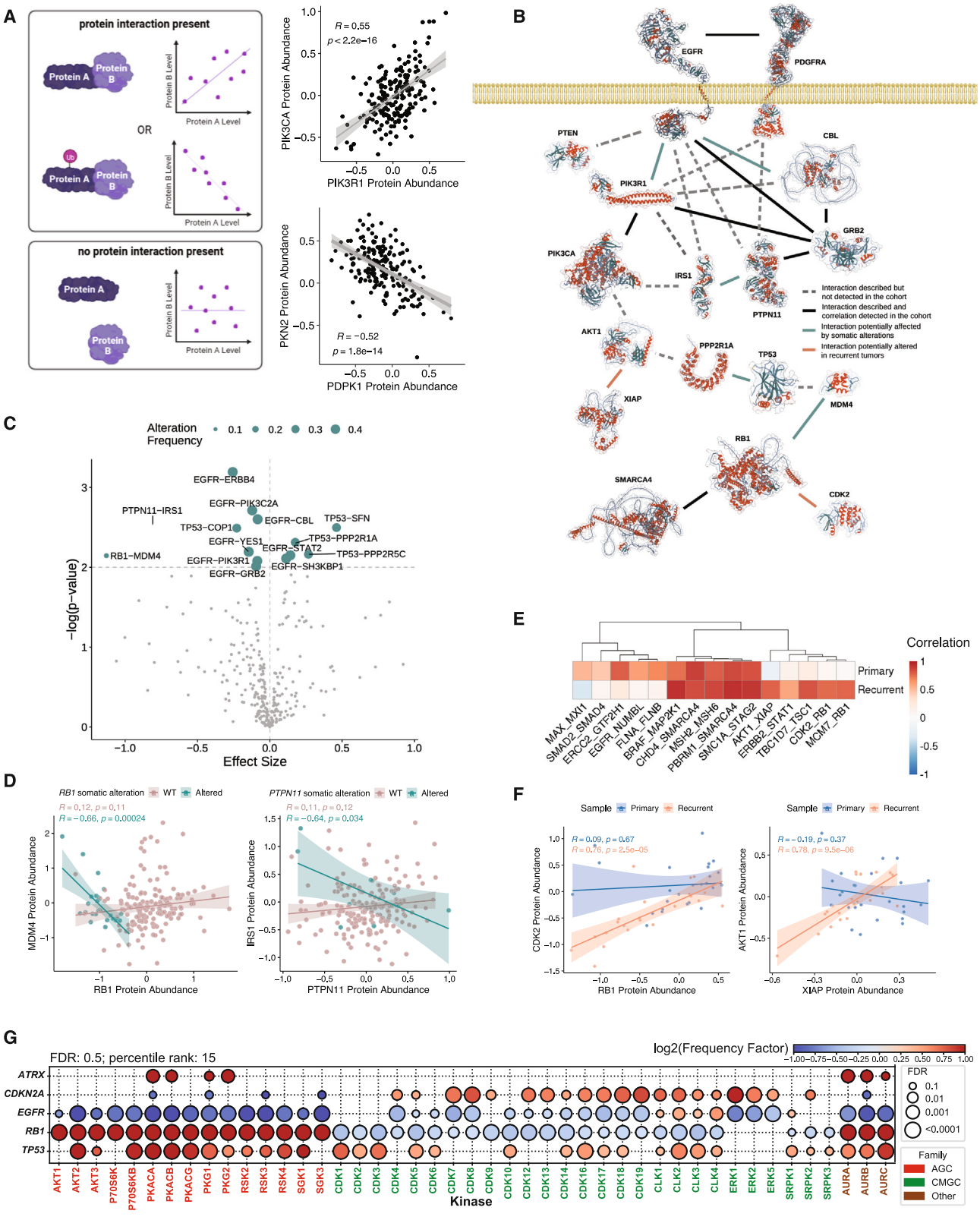


Figure 6. Protein-protein interaction (PPI) differences are associated with somatic alterations and recurrence status

(A) Left, schematic of protein interaction correlations. Right, examples of high correlation in abundance and negative correlation in abundance.

(B) Inferred PPIs in EGFR and PDGFRA signaling pathways in indicated contexts.

(legend continued on next page)

exhibited activation of the PI3K-AKT pathway. *TP53*-mutant tumors showed activation of cell-cycle-related kinases (CDK1-6), whereas *RB1*-mutated tumors showed less activation of these kinases when compared to their respective WT tumors (Figure 6G). Similarly, the PI3K-AKT pathway was less active in *EGFR*-altered tumors. Finally, while *CDKN2A*-altered tumors manifested high activity of both cell-cycle-related kinases (CDK1-6) and ERK kinases, *EGFR*-altered tumors showed less activation of the ERK kinases (Figure 6G). Additional common and distinct activity patterns can be detected across the kinome when comparing the effect of different mutations on cellular signaling (Figure S6E).

PTPN11 may serve as a network hub for EGFR, PDGFRA, and IDH1 signaling

After analyzing the putative effects of somatic drivers in HGG, we focused on one particular driver: *PTPN11*. The importance of *PTPN11* mutation in grade 4 astrocytoma has been underappreciated, although its functional relevance has been documented in diseases such as Noonan syndrome and juvenile myelomonocytic leukemia.[36,37] In this study, we identified *PTPN11* mutations in eleven patients (~6%) and *PTPN11* amplification in one patient (Figure 7A). *PTPN11* mutations, which were mutually exclusive with *EGFR* alteration (Figure 7A), were located at the N-terminal src homology 2 (N-SH2) and kinase phosphatase (PTP) domains (Figure 7B). When mapped onto a partial 3D structure,[38] the Y62 phosphosite and A72T, E76K, R498W, G503A, G506P, T507K, and Q510L mutations are located at the intramolecular interaction site between the PTP and N-SH2 domains (Figure 7B), known to regulate PTP catalytic sites.[38] Mutations may disrupt this interaction and switch PTPN11 from an inhibitory state to an active state. Furthermore, *cis* and *trans* analyses revealed high PTPN11-Y546 phosphorylation in *IDH1*-mutant tumors, which is distinct from the high PTPN11-Y62 phosphorylation observed in *EGFR*-altered tumors (false discovery rate [FDR] < 0.05 and $p < 0.001$; Figure 7C). We validated these findings using an IMAC-SRM target assay (Figure S7A). IDH-mutant tumors exhibited high PDGFRA protein abundance,[39] and increased levels of PDGFRA protein were associated with high PTPN11-Y546 phosphorylation (Figure S7B), thus resulting in high PTPN11-Y546 phosphorylation in *IDH1*-mutant tumors. Interestingly, mutations in *PTPN11* did not affect its Y62 and Y546 phosphorylation levels (Figure 7C).

As the effect of Y62 and Y546 phosphorylation on PTPN11 activity is not well understood, we investigated the effect of these two phosphorylation sites on the signaling pathways downstream of PTPN11 by applying the Kinase Library motif analysis to phosphorylation sites that are positively or negatively correlated with Y62 and Y546 to infer kinase activity patterns associated with the phosphorylation of each site. Strikingly, we found opposite kinase activity patterns associated with the phosphorylation of these two sites. Y62 phosphorylation was associated with decreased activity of the PI3K/AKT pathway and less resultant activity of the ERK kinases, while Y546 was strongly associated with ERK activity, without strong activity of the PI3K/AKT kinases, suggesting alternative activation of the MAPK pathway (Figure 7D). Phosphorylation levels of known activating sites of the corresponding kinases further supported the inferred kinase activity (Figure S7D). Our analysis revealed that the phosphorylation of Y546 but not Y62 was significantly associated with decreased activity of hypoxia-related kinases (AMPKAs and AMPK-related kinases) (Figures 7D and S7E). The observed activity patterns of the PI3K-AKT kinases associated with Y62 and Y546 phosphorylation in patients were also observed in cell lines with corresponding phosphomimetics compared to Y-to-F mutations (Figure S7F). Interestingly, while AMPKA and AMPK-related kinases were inhibited in association with Y546 phosphorylation in tumors, these kinases were activated in Y546E cell lines, potentially demonstrating a negative feedback loop in response to the constitutive negative charge at amino acid 546. One possible explanation of these phenomena is that PTPN11-Y546 phosphorylation is associated with *IDH1* mutation (Figure 7C), and *IDH1*-mutant tumors exhibit lesser hypoxia and cellular stress responses compared to *IDH1-WT* tumors (Figures 7C–7F). Since PTPN11 acts upstream of the PI3K-AKT and MAPK pathways,[40,41] our data suggest that the phosphorylation of Y62 is inhibitory, while the phosphorylation of Y546 is activating.

We examined the downstream phosphorylation sites regulated by *PTPN11* mutations and two phosphorylation sites, Y62 and Y546, through focused *cis* and *trans*, PPI, and kinase/phosphatase-substrate analyses. To discover putative PTPN11 downstream targets, we focused on phosphosites on PTPN11 partners identified in known interaction databases, including OmniPath,[42] NetworKIN,[43] DEPOD,[44] and SIGNOR.[45] We observed distinct phosphorylation sites regulated by *PTPN11* mutations, Y62 phosphorylation, and Y546 phosphorylation, respectively (Figures 7C–7E). We found low MAP3K5 S82 phosphorylation in *PTPN11*-mutant tumors despite no significant difference in MAP3K5 protein abundance (Figure 7C). Phosphorylation levels between the adjacent S82 and S83 sites were positively correlated (Figure S7C). However, *cis* and *trans* analyses did not return S83 as a significant hit, likely because only two patients with *PTPN11* mutation have S83 phosphorylation data available. High PTPN11 Y62 and Y546 phosphorylation was associated with low IRS1 phosphorylation at multiple sites and low MAP3K5 S958 phosphorylation based on kinase substrate analyses, respectively (Figure 7E). Furthermore, PPI

(C) 14 significantly altered PPIs (linear regression with interaction term p value < 0.01) in primary tumors when one protein partner is somatically altered.
(D) Pearson correlations between PPIs according to *RB1* and *PTPN11* alterations, respectively (Pearson R p value <0.05).
(E) Subset of protein pairs exhibiting correlation in primary and/or recurrent tumors for all matched samples with protein data available (Pearson R > 0 and Pearson p value <0.01 in indicated group). Related to Figure S6B.
(F) RB1 and CDK2 protein abundance Pearson correlations in recurrent and matched grade 4 astrocytomas (left). XIAP and AKT1 protein abundance Pearson correlations in recurrent and primary samples (right).
(G) Effects of somatic mutations in genes (y axis), on activity of kinases (x axis), using Kinase Library applied to differentially abundant phosphoproteins compared to WT tumors.
See also Figure S6.

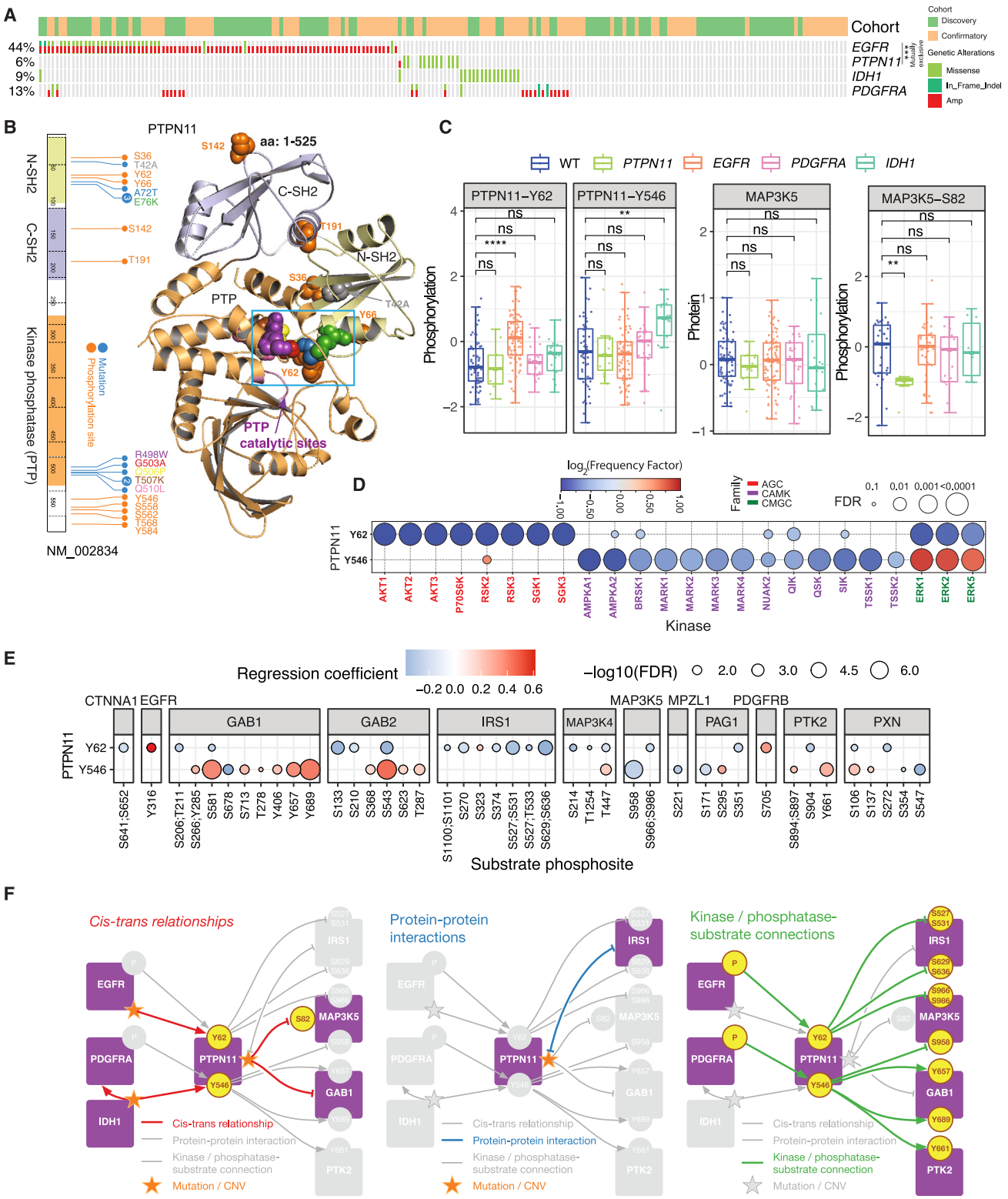


Figure 7. The PTPN11 signaling hub in grade 4 astrocytoma

(A) *PTPN11*, *EGFR*, *IDH1*, *and PDGFRA* alteration status in primary HGGs. The *p* value was calculated with Fisher's exact test (***$p < 0.001$).

(B) *PTPN11* mutations and observed phosphorylations mapped onto 3D protein structure.

(C) *cis/trans* analyses of phosphorylated residues and driver mutations. Boxplots show interquartile ranges of normalized protein and phosphoprotein abundance. Wilcoxon *p* values were calculated (ns: $p > 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$).

(legend continued on next page)

analyses showed that *PTPN11* mutations alter the protein's interaction with IRS1 and PTK2 (Figures 6C and 7F). Collectively, our integrated analyses demonstrated that more than 60% of grade 4 astrocytomas have a PTPN11-centered signaling hub through *PTPN11* mutation or altered PTPN11 phosphorylation, suggesting different regulatory roles for distinct phosphorylation sites on PTPN11.

Metabolic enzymes/pathways connected to RTK activation and dysregulated hypoxia signaling in IDH-mutant HGG

Given the contrasting molecular pathways involved in the pathogenesis and evolution of GBM versus IDH-mutant astrocytoma, an understanding of the underlying biology of each tumor type is important for effective patient care. We identified fifteen p.R132H and one p.R132C mutation events in *IDH1* in our cohort (Figure 8A). The increased number of IDH-mutant tumors compared to findings from our previous study[2] enabled us to carry out comparative analyses between GBMs and IDH-mutant astrocytomas. Differential methylation and metabolite abundance analyses identified three metabolites upregulated in abundance in IDH-mutant tumors: 2-hydroxyglutarate (2-HG), glycerol 3-phosphate, and myo-inositol (Figure 8A). In addition, among 737,419 experimentally measured CpG sites, IDH-mutant tumors had 7,914 upregulated and 134 downregulated methylation sites compared to GBM (absolute change > 0.5, FDR < 0.05, Figure 8A). We observed differentially abundant proteins in IDH-mutant tumors compared to GBM involved in dysregulated pathways including cancer-related RTKs and blood vessel morphogenesis (hypoxia) pathways (Figure 8A). RTK signaling proteins, including CSF1R, MAPK1, PDGFRA, PLCB1, PRKCB, PTPN11, and SMOC2, were significantly upregulated in IDH-mutant samples (Figures 8B and S8A). Many of these were also upregulated at the RNA level, including *PDGFRA*, *PLCB1*, *PRKCB*, and *PTPN11.* We performed a focused DNA methylation analysis of the regions near RTK genes (+/− 500K window) between the two tumor types. We identified upregulated DNA methylation sites in IDH-mutant astrocytomas near several RTK genes, including *PDGFRA* (∼2K bps), *PLCB1* (∼287K bps), and *PTPN11* (∼235K bps) (Figure S8B).

Proteogenomic integration revealed a lower hypoxia signature in IDH-mutant tumors compared to GBM (Figure 8C). The hypoxia signature of IDH-mutant tumors is comparable to normal and normal-adjacent tissues (NATs). HIF1A pathway genes, including *ANXA1*, *COL5A1*, *FN1*, *SERPINE1*, *SLC13A3*, and *VEGFA,* show consistently low protein expression in IDH-mutant tumors (Figure 8C). We observed concordant low RNA expression for these genes in IDH-mutant astrocytomas similar to the TCGA cohort[46] (Figures S8C and S8D). Protein interaction analysis using STRING[47] supports the positive correlations among these *HIF1A* pathway members (Figure 8D). Note that there is no detected expression of the HIF1A protein. *HIF1AN* (FIH-1), an inhibitor of *HIF1A*,[48] showed upregulated expression in IDH-mutant astrocytomas. By using median protein expression among genes involved in the H1F1A pathway as a hypoxia score, we separated tumors, regardless of IDH mutational status, into high-hypoxia (positive score) and low-hypoxia groups (negative score). Patients in our cohort with low hypoxia scores exhibited significantly better survival (p value < 0.01; Figure S8E). We further considered age and sex as covariates for survival analyses and observed the same conclusion (p = 0.05, Figure S8E). In addition, we examined the survival difference in GBM and IDH-mutant tumors, respectively. We observed the same trend in patients with GBM (Figure S8G). However, for the IDH-mutant cohort, these analyses did not achieve statistical significance as only one IDH-mutant tumor had a high hypoxia score.

To follow up on the finding that IDH-WT tumors have greater activation of the HIF1A pathway (Figures 8C and 8D), we conducted CODEX analyses on IDH-WT and IDH-mutant astrocytoma and NAT sections. We found high SERPINE1 expression in IDH-WT but not IDH-mutant tumors (Figure 8E) We further examined regions exhibiting robust SERPINE1 staining in two IDH-WT samples (C3L-01046 and C3*N*-03448), with additional markers for cell-type identification (Table S4). By inferring the cell boundaries with segmentation, we found SERPINE1 was expressed in some, but not all, $OLIG2^+$ and/or $GFAP^+$ cells, as well as in a subset of $IBA1^+$ cells, again almost exclusively in the IDH-WT tumors. We then leveraged scRNA data from our previously reported 19 CPTAC GBM samples,[2] to probe SERPINE1 transcriptional expression across different cell types. Consistent with our CODEX findings, macrophages and malignant cells emerged as the top two expressors of SERPINE1 among the five most abundant cell types observed across samples (Figures S8G and S8H). While there is no single specific marker that definitively marks malignant cells, our prior scRNA data suggest OLIG2 and GFAP were most highly expressed in malignant cells (Figure S8H). In summary, this data is consistent with greater HIF1A pathway activation in IDH-WT tumors supported by higher SERPINE1 levels in tumor cells and macrophages.

Differentially abundant phosphorylation sites between IDH-mutant astrocytomas and GBM were enriched in actin-related and cell-cell junction pathways (Figure S8I). We further compared miRNA expression between IDH-mutant astrocytomas and GBM and identified several dysregulated miRNAs associated with *IDH1* mutation (Figure S8J). For instance, miR-504-5p is upregulated in IDH-mutant astrocytomas; prior work has demonstrated miR-504 overexpression extended survival of mice hosting xenografts derived from GBM stem cells.[49]

We next explored the effect of IDH mutation on intracellular signaling through phosphorylation and kinase activity, leveraging the Kinase Library. IDH-mutant tumors showed significantly lower activity of AMPKAs and AMPK-related kinases compared to GBM (Figure 8F). Other interesting observations from this analysis include the low activity of immune-related kinases (TBK1 and IKKE) in IDH-mutant astrocytomas and high

(D) Kinase activity in samples with PTPN11-Y62 and PTPN11-Y546 phosphorylation events.
(E) Kinase substrate analysis for PTPN11-Y62 and PTPN11-Y546 phosphorylation.
(F) Analyses of *cis* and *trans* effects, PPIs, and kinase/phosphatase-substrate measurements reveal multiple proteins likely regulated by PTPN11. More than 60% of tumors share the PTPN11-centered signaling hub.
See also Figure S7.

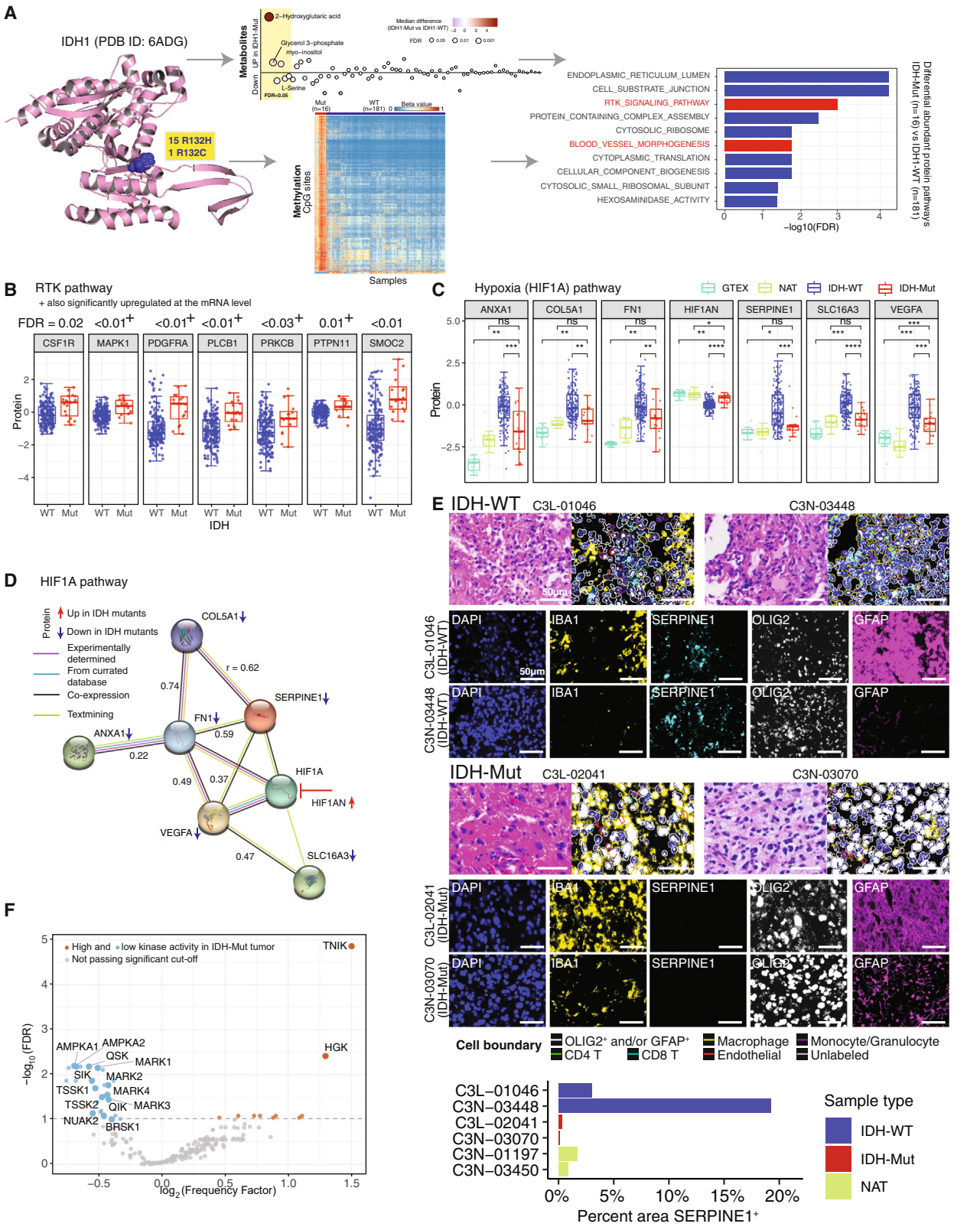


Figure 8. Dysregulated pathways in grade 4 *IDH*-mutant astrocytomas

(A) Hotspot *IDH1* R132H/R132C mutations result in 2-HG production and hypermethylation phenotype. Proteogenomics identified dysregulated pathways in IDH-mutant tumors compared to GBMs.

(legend continued on next page)

activation of TNIK, a kinase reported to be involved in regulation of neuronal dendrite extension and neuronal signaling (Figures 8F and S8K).[50,51]

DISCUSSION

In this report, we compared 181 IDH-WT primary GBMs with 16 IDH-mutant grade 4 astrocytomas, previously classified as GBMs at the time of surgery, with an emphasis on understanding the molecular basis of the significantly improved prognosis associated with IDH mutation. This work adds to the prior CPTAC study of GBM,[2] which had a smaller number of samples and proteogenomic platforms and no matched longitudinal samples. That study identified the origin of EMT signatures in mesenchymal GBM as being the tumor cells and not stroma; identified four GBM immune subtypes driven by distinct immune populations; histone H2B acetylation enrichment in classical GBM; and pointed to potential importance of PTPN11. Here, our key findings include: (a) single-cell resolution of chromatin accessibility and RNA expression in longitudinal samples from the same patient; (b) similar glycoproteomic, post-translational, and metabolomic features in tumors with distinct mutations (TERTp/PTEN and TERTp/EGFR); (c) convergence of pathways downstream of EGFR, PDGFR, and IDH1 on PTPN11; (d) association of a low hypoxia signature and reduced AMPKA activity in IDH-mutant tumors; and (e) metabolomic and glycoproteomic markers of tumor recurrence. In addition to the data suggesting significant convergence downstream of individual genetic alterations, we believe that the high-quality multi-omic data presented here provides a valuable resource to the neuro-oncology community. These data can be mined to generate testable hypotheses using experimental model systems such as transgenic mice and genetically manipulated cell lines to test mechanisms driving tumor progression or to identify exploitable vulnerabilities that can improve clinical outcomes.

Tumor recurrence is nearly universal in HGG, and improved understanding of how tumors change over time and in response to treatment may improve therapeutic strategies. By studying a set of 25 primary and recurrent tumors at the single-cell level, as well as their metabolome, lipidome, glycoproteome, and phosphoproteome, we found that the tumors transition over time and treatment toward a more mesenchymal state and show decreased cell cycle and mismatch repair gene expression. A subset of proteins increased their glycosylation in recurrent tumors and we observed an enrichment in features associated with neuronal projections and neutrophil-mediated immunity. These results are generally consistent with the observations of Kim et al.[52] who, using a larger number of longitudinal samples, observed a decrease in proliferation and an increase in neuronal characteristics at the level of protein abundance. Our investigation provides more mechanistic detail, in the observation of appropriate changes in transcription factor expression potentially driving the transition to a more neuronal state, and the concomitant expression of a neuronal-like glycosylation pattern. Our comparison of driver gene mutations between paired primary and recurrent tumors largely showed heterogeneous patterns of evolution. These findings highlight the array of genetic and post-translational differences between aggressive primary and recurrent tumors and suggest that targeting protein interactions, PTMs, and metabolites may be effective against recurrence.

Limitations of this study include a relatively modest sample size of limited ethnic and racial diversity and reliance on a bulk sampling strategy. Small subsets of the cohort including brain metastases and secondary recurrent GBM were not included in major analyses and are provided as a resource. Study conclusions are based primarily on statistical associations between molecular observations and available clinical parameters, with limited mechanistic validation.

In summary, a major result of this study was the identification of common features among tumors harboring disparate driver gene alterations, highlighting the complexity of tumor-promoting signaling events. Tracking the molecular landscape of initial and recurrent tumors from the same patient across time on these same platforms demonstrated the heterogeneous changes that transpire as tumors evolve under treatment, while also highlighting pathways that are frequently enriched at the time of recurrence. We hope the focus on identifying cooperating events and common regulatory networks may provide additional targets for therapeutic intervention, pending functional validation studies beyond the scope of this work.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCES AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human participants
- METHOD DETAILS
  - Sample processing
  - Longitudinal cohort sample collection and processing
  - Sample processing for genomic DNA and total RNA extraction
  - Whole exome sequencing
  - PCR-free whole genome sequencing
  - Illumina infinium methylationEPIC beadchip array
  - RNA sequencing
  - Single-nuclei multiome library construction and sequencing

(B) Specific RTK pathway genes are upregulated in IDH-mutant astrocytomas. Boxplots show interquartile ranges of normalized protein abundance. FDR was calculated with Wilcoxon test and Benjamini & Hochberg adjustment.
(C) IDH-mutant tumors have less protein expression of hypoxia pathway members compared to GBMs. Boxplots show interquartile ranges of normalized protein abundance. Wilcoxon p values were calculated (ns: $p > 0.05$, *$p < 0.05$, **$p \leq 0.01$, ***$p < 0.001$, ****$p \leq 0.0001$).
(D) PPIs from STRING[46] among HIF1A pathway genes.
(E) CODEX images of IDH-WT (top) and mutant patient samples (middle). Whole-slide percent cellular area with SERPINE1 staining is quantified for each sample (bottom).
(F) Kinase activity differences between IDH-mutant astrocytomas and GBMs, based on phosphoproteomic measurements. See also Figure S8 and Table S4.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2024.06.004.

CONSORTIA

The members of the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium are Anupriya Agarwal, Iyad Alnahhas, Mitual Amin, Eunkyung An, Matthew L Anderson, David W Andrews, Meenakshi Anurag, Jill S. Barnholtz-Sloan, Thomas Bauer, Jasmin Bavarva, Zev A. Binder, Michael J Birrer, Uma Borate, Melissa Borucki, Lisa M. Bramer, Steven Brem, Shuang Cai, Lewis C. Cantley, Song Cao, Wagma Caravan, Steven A Carr, Daniel W Chan, Lijun Chen, Lin S Chen, Xi S Chen, David Chesla, Milan G. Chheda, Arul M Chinnaiyan, Shrabanti Chowdhury, Rosalie K. Chu, Marcin P Cieślik, Natalie M. Clark, Karl Clauser, Antonio Colaprico, Sandra L. Cottingham, Sneha P. Couvillion, Reese Crispen, Tomasz Czernicki, Fulvio D'Angelo, Emek Demir, Saravana Mohan Dhanasekaran, Li Ding, Marcin J Domagalski, Yongchao Dou, Brian Druker, Josie Eder, Nathan J Edwards, Jennifer Eschbacher, David Fenyö, Thomas H. Fillmore, Nikolay Gabrovski, Yuqian Gao, Luciano Garofano, Yifat Geffen, Gad Getz, Michael A Gillette, Denis Golbin, Charles A. Goldthwaite Jr., Marina A. Gritsenko, Zeynep Gumus, Ziad Hanhan, Beth Hermes, Tara Hiltke, Galen Hostetter, Andrew Houston, Chen Huang, Antonio Iavarone, Kathleen J. Imbach, Eric Jaehnig, Scott D. Jewell, Jared L. Johnson, Kevin D Judy, Alicia Karz, Wojciech Kaspera, Karen A. Ketchum, Lyndon Kim, Young-Mo Kim, Mateusz P Koptyra, Ramani B Kothadia, Chandan Kumar-Sinha, Jennifer E. Kyle, John YK Lee, Jonathan Lei, Felipe D Leprevost, Kai Li, Qing Kay Li, Yize Li, Zibo Li, Yuxing Liao, Tung-Shing M. Lih, Jena Lilly, Jingxian Liu, Tao Liu, Rita Jui-Hsien Lu, Yin Lu, Jan Lubínski, Weiping Ma, William Maggio, Anna Malovannaya, D. R. Mani, Nina Martinez, Jason E McDermott, Peter B McGarvey, Mehdi Mesri, Simona Migliozzi, Matthew E. Monroe, Ronald J. Moore, Nataly Naser Al Deen, MacLean P. Nasrallah, Michael Nestor, Alexey I Nesvizhskii, Chelsea J. Newton, Donald M O'Rourke, Gilbert S Omenn, Umut Ozbek, Oxana V Paklina, Amanda G Paulovich, Samuel H Payne, Francesca Petralia, Vladislav A. Petyuk, Alexander R Pico, Paul D Piehowski, Alexander Pilozzi, Dmitris Placantonakis, Eduard Porta-Pardo, Olga Potapova, Barbara Pruetz, Adam C. Resnick, Boris Reva, Ana I. Robles, Karin D. Rodland, Henry Rodriguez, Daniel C. Rohrer, Brian Rood, Larissa Rossell, Kelly V Ruggles, Dmitry Rykunov, Tyler Sagendorf, Atom Sarkar, Shankha Satpathy, Sara R Savage, Eric E Schadt, Athena A Schepmoes, Scott R Shepard, Zhiao Shi, Yvonne Shutack, Richard D. Smith, Lori J Sokoll, Yingduo Song, Yizhe Song, Austin N. Southard-Smith, Jakub Stawicki, Phillip B Storm, Kelly G. Stratton, Ilya Strunilin, Wojciech Szopa, Ratna R. Thangudu, Mathangi Thiagarajan, Cristina Tognon, Elie Traer, Chia-Feng Tsai, Shirley Tsang, Alan R Turtz, Jeffrey Tyner, Negin Vatanian, Weiqing Wan, Liang-Bo Wang, Pei Wang, Yi-Ting Wang, Michael W Weaver, Karl K. Weitz, Bo Wen, Jeffrey R Whiteaker, Bart Williams, George D Wilson, Maciej Wiznerowicz, Yige Wu, Matthew A. Wyczalkowski, Lijun Yao, Tomer M. Yaron, Eric L Zager, Bing Zhang, Hui Zhang, Xu Zhang, Zhen Zhang, Rui Zhao, Jun Zhu.

ACKNOWLEDGMENTS

This work was supported by grants U24CA210972, U24CA210955, U24CA210954, U24CA210985, U24CA210993, U24CA210967, U24CA210986, U01CA214125, U24CA210979, U01CA214114, U01CA214116, U24CA271012, and U24CA271079 from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC), by grant R01HG009711 from National Human Genome Research Institute (NHGRI) to L.D., R01NS107833 and R01NS117149 from National Institutes of Health to M.G.C., PID2019-107043RA-I00 and RYC2019-026415-I from the Spanish Science Ministry and LABAE20038PORT from the AECC Scientific Foundation to E.P.-P. The MS-based proteomics work described herein was performed at the Environmental Molecular Sciences Laboratory (grid.436923.9), a U.S. Department of Energy (DOE) National Scientific User Facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, WA. PNNL is a multi-program national laboratory operated by the Battelle Memorial Institute for the DOE under contract DE-AC05-76RL01830. Three-dimensional protein modeling and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. We thank the Philadelphia Coalition for a Cure (PC4C) for support in procurement and coordination of longitudinal surgical samples.

AUTHOR CONTRIBUTIONS

Study conception and design: J.S.B.-S., A.I., F.C., M.M., M.P.N., H.Z., A.C.R., M.G.C., K.D.R., T.L., L.D.; performed experiment or data collection: M.A.G., T.-S.M.L., J.E.K., Y.-T.W., C.-F.T., L.C., A.S., F.A.S., S.C. Y.D.L., R.I., R.G.J., N.N.A.D., W.C., A.H., R.K.C., K.K.W., R.J.M., A.A.S., S.P.C., J.E., Y.-M.K., Y.G., T.L.F., R.Z., M.E.M., C.J.N., K.A.K., R.R.T.; computation and statistical analysis: J.L., S.C., K.J.I., T.-S.M.L., J.E.K., T.M.Y., Y.L., I.S., Z.A.B., Y.-T.W., C.-F.T., W.M., N.M.C., M.A.W., L.-B.W., E.S., D.R., T.S., V.A.P., M.N., L.M.B., K.G.S., A.N.S., R.J.-H.L., D.F., E.P.-P.; data interpretation and biological analysis: J.L., S.C., K.J.I., T.-S.M.L., J.E.K., T.M.Y., Y.L., I.S., Z.A.B., Y.-T.W., C.-F.T., W.M., N.M.C., D.R., S.L.C., A.N.S., R.J.-H.L., J.L.J., M.W., G.H., E.P.-P., F.C., M.P.N., A.C.R., M.G.C., K.D.R., T.L., L.D.; writing – original drafts: J.L., S.C., K.J.I., T.-S.M.L., T.M.Y., Y.L., Y.E.L., F.C. M.G.C., K.D.R., T.L., L.D.; writing – review and editing: J.L., S.C., K.J.I., T.-S.M.L., T.M.Y., Y.L., M.G.C., K.D.R., T.L., L.D., M.P.N.; supervision: P.W., D.F., D.R.M., R.D.S., E.P.-P., L.C.C., A.I., F.C. M.M., H.Z., A.C.R., M.G.C., K.D.R., T.L., L.D.; administration: E.A., M.T., A.I.R., M.M., M.G.C., K.D.R., T.L., L.D.

DECLARATION OF INTERESTS

T.M.Y. is a co-founder, stockholder, and member of the board of directors of DESTROKE, Inc., an early stage start-up developing mobile technology for automated clinical stroke detection. J.L.J. has received consulting fees from Scorpion Therapeutics and Volastra Therapeutics. L.C.C. is a founder and member of the board of directors of Agios Pharmaceuticals and is a founder of Petra Pharmaceuticals. L.C.C. is an inventor on patents (pending) for Combination Therapy for PI3K-associated Disease or Disorder, and The Identification of Therapeutic Interventions to Improve Response to PI3K Inhibitors for Cancer Treatment. L.C.C. is a co-founder and shareholder in Faeth Therapeutics. P.W. is a statistical consultant for Sema4. M.G.C. receives research support from Merck, Orbus Therapeutics, and NeoimmuneTech Inc, and royalties from UpToDate.

Received: November 20, 2023
Revised: March 12, 2024
Accepted: June 10, 2024
Published: July 8, 2024

REFERENCES

1. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. *23*, 1231–1251. https://doi.org/10.1093/neuonc/noab106.
2. Wang, L.B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and metabolomic characterization of human

glioblastoma. Cancer Cell *39*, 509–528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.

3. Petralia, F., Tignor, N., Reva, B., Koptyra, M., Chowdhury, S., Rykunov, D., Krek, A., Ma, W., Zhu, Y., Ji, J., et al. (2020). Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. Cell *183*, 1962–1985.e31. https://doi.org/10.1016/j.cell.2020.10.044.
4. Johnson, J.L., Yaron, T.M., Huntsman, E.M., Kerelsky, A., Song, J., Regev, A., Lin, T.-Y., Liberatore, K., Cizin, D.M., Cohen, B.M., et al. (2023). An atlas of substrate specificities for the human serine/threonine kinome. Nature *613*, 759–766. https://doi.org/10.1038/s41586-022-05575-3.
5. Kiernan, E.A., Smith, S.M.C., Mitchell, G.S., and Watters, J.J. (2016). Mechanisms of microglial activation in models of inflammation and hypoxia: Implications for chronic intermittent hypoxia. J. Physiol. *594*, 1563–1577. https://doi.org/10.1113/JP271502.
6. Varn, F.S., Johnson, K.C., Martinek, J., Huse, J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., et al. (2022). Glioma progression is shaped by genetic evolution and microenvironment interactions. Cell *185*, 2184–2199.e16. https://doi.org/10.1016/j.cell.2022.04.038.
7. Hoogstrate, Y., Draaisma, K., Ghisai, S.A., van Hijfte, L., Barin, N., de Heer, I., Coppieters, W., van den Bosch, T.P.P., Bolleboom, A., Gao, Z., et al. (2023). Transcriptome analysis reveals tumor microenvironment changes in glioblastoma. Cancer Cell *41*, 678–692.e7. https://doi.org/10.1016/j.ccell.2023.02.019.
8. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.
9. Zhang, B., Wan, W., Li, Z., Gao, Z., Ji, N., Xie, J., Wang, J., Wang, B., Lai-Wan Kwong, D., Guan, X., et al. (2022). A prognostic risk model for glioma patients by systematic evaluation of genomic variations. iScience *25*, 105681. https://doi.org/10.1016/j.isci.2022.105681.
10. Daniel, P., Meehan, B., Sabri, S., Jamali, F., Sarkaria, J.N., Choi, D., Garnier, D., Kitange, G., Glennon, K.I., Paccard, A., et al. (2022). Detection of temozolomide-induced hypermutation and response to PD-1 checkpoint inhibitor in recurrent glioblastoma. Neurooncol. Adv. *4*, vdac076. https://doi.org/10.1093/noajnl/vdac076.
11. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database *2017*, bax028. https://doi.org/10.1093/database/bax028.
12. Terekhanova, N.V., Karpova, A., Liang, W.-W., Strzalkowski, A., Chen, S., Li, Y., Southard-Smith, A.N., Iglesia, M.D., Wendl, M.C., Jayasinghe, R.G., et al. (2023). Epigenetic regulation during cancer transitions across 11 tumour types. Nature *623*, 432–441. https://doi.org/10.1038/s41586-023-06682-5.
13. Yu, H., Li, Z., and Wang, M. (2020). Expression and prognostic role of E2F transcription factors in high-grade glioma. CNS Neurosci. Ther. *26*, 741–753. https://doi.org/10.1111/cns.13295.
14. Luo, J., Wang, X., Xia, Z., Yang, L., Ding, Z., Chen, S., Lai, B., and Zhang, N. (2015). Transcriptional factor specificity protein 1 (SP1) promotes the proliferation of glioma cells by up-regulating midkine (MDK). Mol. Biol. Cell *26*, 430–439. https://doi.org/10.1091/mbc.E14-10-1443.
15. Wu, Y., Fletcher, M., Gu, Z., Wang, Q., Costa, B., Bertoni, A., Man, K.H., Schlotter, M., Felsberg, J., Mangei, J., et al. (2020). Glioblastoma epigenome profiling identifies SOX10 as a master regulator of molecular tumour subtype. Nat. Commun. *11*, 6434. https://doi.org/10.1038/s41467-020-20225-w.
16. Lo Dico, A., Martelli, C., Diceglie, C., Lucignani, G., and Ottobrini, L. (2018). Hypoxia-Inducible Factor-1α Activity as a Switch for Glioblastoma Responsiveness to Temozolomide. Front. Oncol. *8*, 249. https://doi.org/10.3389/fonc.2018.00249.
17. Ding, L., Bailey, M.H., Porta-Pardo, E., Thorsson, V., Colaprico, A., Bertrand, D., Gibbs, D.L., Weerasinghe, A., Huang, K.-l., Tokheim, C., et al. (2018). Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. Cell *173*, 305–320.e10. https://doi.org/10.1016/j.cell.2018.03.033.
18. Zhang, H., Li, X.-j., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. *21*, 660–666. https://doi.org/10.1038/nbt827.
19. Chandler, K., and Goldman, R. (2013). Glycoprotein Disease Markers and Single Protein-omics. Mol. Cell. Proteomics *12*, 836–845. https://doi.org/10.1074/mcp.R112.026930.
20. Varki, A. (2017). Biological roles of glycans. Glycobiology *27*, 3–49. https://doi.org/10.1093/glycob/cww086.
21. Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. Nat. Rev. Nephrol. *15*, 346–366. https://doi.org/10.1038/s41581-019-0129-4.
22. Cao, L., Huang, C., Cui Zhou, D., Hu, Y., Lih, T.M., Savage, S.R., Krug, K., Clark, D.J., Schnaubelt, M., Chen, L., et al. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell *184*, 5031–5052.e26. https://doi.org/10.1016/j.cell.2021.08.023.
23. Thaysen-Andersen, M., Packer, N.H., and Schulz, B.L. (2016). Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease. Mol. Cell. Proteomics *15*, 1773–1790. https://doi.org/10.1074/mcp.O115.057638.
24. Bieberich, E. (2014). Synthesis, Processing, and Function of N-glycans in N-glycoproteins. Adv. Neurobiol. *9*, 47–70. https://doi.org/10.1007/978-1-4939-1154-7_3.
25. Wen, P., Chen, J., Zuo, C., Gao, X., Fujita, M., and Yang, G. (2022). Proteome and Glycoproteome Analyses Reveal the Protein N-Linked Glycosylation Specificity of STT3A and STT3B. Cells *11*, 2775. https://doi.org/10.3390/cells11182775.
26. Strous, G.J., Van Kerkhof, P., Brok, R., Roth, J., and Brada, D. (1987). Glucosidase II, a protein of the endoplasmic reticulum with high mannose oligosaccharide chains and a rapid turnover. J. Biol. Chem. *262*, 3620–3625.
27. De Masi, R., and Orlando, S. (2022). GANAB and N-Glycans Substrates Are Relevant in Human Physiology, Polycystic Pathology and Multiple Sclerosis: A Review. Int. J. Mol. Sci. *23*, 7373. https://doi.org/10.3390/ijms23137373.
28. De Leoz, M.L.A., Young, L.J.T., An, H.J., Kronewitter, S.R., Kim, J., Miyamoto, S., Borowsky, A.D., Chew, H.K., and Lebrilla, C.B. (2011). High-Mannose Glycans are Elevated during Breast Cancer Progression. Mol. Cell. Proteomics *10*, M110.002717. https://doi.org/10.1074/mcp.M110.002717.
29. Park, D.D., Phoomak, C., Xu, G., Olney, L.P., Tran, K.A., Park, S.S., Haigh, N.E., Luxardi, G., Lert-itthiporn, W., Shimoda, M., et al. (2020). Metastasis of cholangiocarcinoma is promoted by extended high-mannose glycans. Proc. Natl. Acad. Sci. USA *117*, 7633–7644. https://doi.org/10.1073/pnas.1916498117.
30. Arkhipov, A., Shan, Y., Das, R., Endres, N.F., Eastwood, M.P., Wemmer, D.E., Kuriyan, J., and Shaw, D.E. (2013). Architecture and Membrane Interactions of the EGF Receptor. Cell *152*, 557–569. https://doi.org/10.1016/j.cell.2012.12.030.
31. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.
32. Azimzadeh Irani, M., Kannan, S., and Verma, C. (2017). Role of N-glycosylation in EGFR ectodomain ligand binding. Proteins *85*, 1529–1549. https://doi.org/10.1002/prot.25314.
33. Chen, Y.-J., Yen, T.-C., Lin, Y.-H., Chen, Y.-L., Khoo, K.-H., and Chen, Y.-J. (2021). ZIC-cHILIC-Based StageTip for Simultaneous

Glycopeptide Enrichment and Fractionation toward Large-Scale N-Sialoglycoproteomics. Anal. Chem. *93*, 15931–15940. https://doi.org/10.1021/acs.analchem.1c03224.

34. Li, Y., Porta-Pardo, E., Tokheim, C., Bailey, M.H., Yaron, T.M., Stathias, V., Geffen, Y., Imbach, K.J., Cao, S., Anand, S., et al. (2023). Pan-cancer proteogenomics connects oncogenic drivers to functional states. Cell *186*, 3921–3944.e25. https://doi.org/10.1016/j.cell.2023.07.014.
35. Ruffalo, M., and Bar-Joseph, Z. (2019). Protein interaction disruption in cancer. BMC Cancer *19*, 370. https://doi.org/10.1186/s12885-019-5532-5.
36. Tartaglia, M., Mehler, E.L., Goldberg, R., Zampino, G., Brunner, H.G., Kremer, H., Van Der Burgt, I., Crosby, A.H., Ion, A., Jeffery, S., et al. (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat. Genet. *29*, 465–468. https://doi.org/10.1038/ng772.
37. Loh, M.L., Vattikuti, S., Schubbert, S., Reynolds, M.G., Carlson, E., Lieuw, K.H., Cheng, J.W., Lee, C.M., Stokoe, D., Bonifas, J.M., et al. (2004). Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. Blood *103*, 2325–2331. https://doi.org/10.1182/blood-2003-09-3287.
38. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J., and Shoelson, S.E. (1998). Crystal structure of the tyrosine phosphatase SHP-2. Cell *92*, 441–450. https://doi.org/10.1016/s0092-8674(00)80938-1.
39. Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114. https://doi.org/10.1038/nature16490.
40. Neel, B.G., Gu, H., and Pao, L. (2003). The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. Trends Biochem. Sci. *28*, 284–293. https://doi.org/10.1016/S0968-0004(03)00091-4.
41. Liu, K.-W., Feng, H., Bachoo, R., Kazlauskas, A., Smith, E.M., Symes, K., Hamilton, R.L., Nagane, M., Nishikawa, R., Hu, B., and Cheng, S.-Y. (2011). SHP-2/PTPN11 mediates gliomagenesis driven by PDGFRA and INK4A/ARF aberrations in mice and humans. J. Clin. Invest. *121*, 905–917. https://doi.org/10.1172/JCI43690.
42. Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods *13*, 966–967. https://doi.org/10.1038/nmeth.4077.
43. Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M.B., and Pawson, T. (2008). NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res. *36*, D695–D699. https://doi.org/10.1093/nar/gkm902.
44. Damle, N.P., and Köhn, M. (2019). The human DEPhOsphorylation Database DEPOD: 2019 update. Database *2019*, baz133. https://doi.org/10.1093/database/baz133.
45. Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L., and Cesareni, G. (2020). SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. Nucleic Acids Res. *48*, D504–D510. https://doi.org/10.1093/nar/gkz949.
46. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The Somatic Genomic Landscape of Glioblastoma. Cell *155*, 462–477. https://doi.org/10.1016/j.cell.2013.09.034.
47. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613. https://doi.org/10.1093/nar/gky1131.
48. Mahon, P.C., Hirota, K., and Semenza, G.L. (2001). FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. Genes Dev. *15*, 2675–2686. https://doi.org/10.1101/gad.924501.
49. Bier, A., Hong, X., Cazacu, S., Goldstein, H., Rand, D., Xiang, C., Jiang, W., Ben-Asher, H.W., Attia, M., Brodie, A., et al. (2020). miR-504 modulates the stemness and mesenchymal transition of glioma stem cells and their interaction with microglia via delivery by extracellular vesicles. Cell Death Dis. *11*, 899. https://doi.org/10.1038/s41419-020-03088-3.
50. Burette, A.C., Phend, K.D., Burette, S., Lin, Q., Liang, M., Foltz, G., Taylor, N., Wang, Q., Brandon, N.J., Bates, B., et al. (2015). Organization of TNIK in dendritic spines. J. Comp. Neurol. *523*, 1913–1924. https://doi.org/10.1002/cne.23770.
51. Coba, M.P., Komiyama, N.H., Nithianantharajah, J., Kopanitsa, M.V., Indersmitten, T., Skene, N.G., Tuck, E.J., Fricker, D.G., Elsegood, K.A., Stanford, L.E., et al. (2012). TNiK Is Required for Postsynaptic and Nuclear Signaling Pathways and Cognitive Function. J. Neurosci. *32*, 13987–13999. https://doi.org/10.1523/JNEUROSCI.2433-12.2012.
52. Kim, K.H., Migliozzi, S., Koo, H., Hong, J.H., Park, S.M., Kim, S., Kwon, H.J., Ha, S., Garofano, L., Oh, Y.T., et al. (2024). Integrated proteogenomic characterization of glioblastoma evolution. Cancer Cell *42*, 358–377.e8. https://doi.org/10.1016/j.ccell.2023.12.015.
53. Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J., and Gygi, S.P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. *24*, 1285–1292. https://doi.org/10.1038/nbt1240.
54. Monroe, M.E., Shaw, J.L., Daly, D.S., Adkins, J.N., and Smith, R.D. (2008). MASIC: A software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC–MS(/MS) features. Comput. Biol. Chem. *32*, 215–217. https://doi.org/10.1016/j.compbiolchem.2008.02.006.
55. Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277. https://doi.org/10.1038/ncomms6277.
56. Gibbons, B.C., Chambers, M.C., Monroe, M.E., Tabb, D.L., and Payne, S.H. (2015). Correcting systematic bias and instrument measurement drift with mzRefinery. Bioinformatics *31*, 3838–3840. https://doi.org/10.1093/bioinformatics/btv437.
57. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121. https://doi.org/10.1038/nmeth.3252.
58. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. JOSS *4*, 1686. https://doi.org/10.21105/joss.01686.
59. The Bioconda, T., Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., and Köster, J. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat. Methods *15*, 475–476. https://doi.org/10.1038/s41592-018-0046-7.
60. Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522. https://doi.org/10.1093/bioinformatics/bts480.
61. Xi, R., Lee, S., Xia, Y., Kim, T.-M., and Park, P.J. (2016). Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. Nucleic Acids Res. *44*, 6274–6286. https://doi.org/10.1093/nar/gkw491.
62. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.
63. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594. https://doi.org/10.1038/s41592-018-0051-x.
64. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan

2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. https://doi.org/10.1101/gr.129684.111.

65. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871. https://doi.org/10.1093/bioinformatics/btp394.

66. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.

67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

68. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

69. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

70. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

71. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

72. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

73. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

74. Polpitiya, A.D., Qian, W.-J., Jaitly, N., Petyuk, V.A., Adkins, J.N., Camp, D.G., Anderson, G.A., and Smith, R.D. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics *24*, 1556–1558. https://doi.org/10.1093/bioinformatics/btn217.

75. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170.

76. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

77. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71. https://doi.org/10.1093/nar/gkv1507.

78. Song, X., Ji, J., Gleason, K.J., Yang, F., Martignetti, J.A., Chen, L.S., and Wang, P. (2019). Insights into Impact of DNA Copy Number Alteration and Methylation on the Proteogenomic Landscape of Human Ovarian Cancer via a Multi-omics Integrative Analysis. Mol. Cell. Proteomics *18*, S52–S65. https://doi.org/10.1074/mcp.RA118.001220.

79. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220–314. https://doi.org/10.1186/s13059-017-1349-1.

80. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.

81. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

82. Hu, Y., Shah, P., Clark, D.J., Ao, M., and Zhang, H. (2018). Reanalysis of Global Proteomic and Phosphoproteomic Data Identified a Large Number of Glycopeptides. Anal. Chem. *90*, 8065–8071. https://doi.org/10.1021/acs.analchem.8b01137.

83. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.

84. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. *12*, 77. https://doi.org/10.1186/1471-2105-12-77.

85. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

86. Xu, T., Le, T.D., Liu, L., Su, N., Wang, R., Sun, B., Colaprico, A., Bontempi, G., and Li, J. (2017). CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics *33*, 3131–3133. https://doi.org/10.1093/bioinformatics/btx378.

87. Mauguen, A., Seshan, V.E., Begg, C.B., and Ostrovnaya, I. (2019). Testing clonal relatedness of two tumors from the same patient based on their mutational profiles: update of the Clonality R package. Bioinformatics *35*, 4776–4778. https://doi.org/10.1093/bioinformatics/btz486.

88. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189.

89. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773. https://doi.org/10.1093/nar/gky955.

90. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

91. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.

92. Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res. *38*, D497–D501. https://doi.org/10.1093/nar/gkp914.

93. Calderone, A., Iannuccelli, M., Peluso, D., and Licata, L. (2020). Using the MINT Database to Search Protein Interactions. CP in Bioinformatics *69*, e93. https://doi.org/10.1002/cpbi.93.

94. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., et al. (2019). The BioGRID interaction database: 2019 update. Nucleic Acids Res. *47*, D529–D541. https://doi.org/10.1093/nar/gky1079.

95. del Toro, N., Shrivastava, A., Ragueneau, E., Meldal, B., Combe, C., Barrera, E., Perfetto, L., How, K., Ratan, P., Shirodkar, G., et al. (2022). The IntAct database: efficient access to fine-grained molecular interaction data. Nucleic Acids Res. *50*, D648–D653. https://doi.org/10.1093/nar/gkab1006.

96. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655. https://doi.org/10.1093/nar/gkx1132.

97. Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., and Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. Nat. Methods *11*, 603–604. https://doi.org/10.1038/nmeth.2968.

98. Vasaikar, S.V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. *46*, D956–D963. https://doi.org/10.1093/nar/gkx1090.

99. National Cancer Institute Clinical Proteomic Tumor Analysis, C (2023). The Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme Collection (CPTAC-GBM). The Cancer Imaging Archive.

100. Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. *12*, R1. https://doi.org/10.1186/gb-2011-12-1-r1.

101. Nakayasu, E.S., Nicora, C.D., Sims, A.C., Burnum-Johnson, K.E., Kim, Y.-M., Kyle, J.E., Matzke, M.M., Shukla, A.K., Chu, R.K., Schepmoes, A.A., et al. (2016). MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. mSystems *1*, e00043-16. https://doi.org/10.1128/mSystems.00043-16.

102. Webb-Robertson, B.-J., Kim, Y.-M., Zink, E.M., Hallaian, K.A., Zhang, Q., Madupu, R., Waters, K.M., and Metz, T.O. (2014). A statistical analysis of the effects of urease pre-treatment on the measurement of the urinary metabolome by gas chromatography–mass spectrometry. Metabolomics *10*, 897–908. https://doi.org/10.1007/s11306-014-0642-1.

103. Kyle, J.E., Crowell, K.L., Casey, C.P., Fujimoto, G.M., Kim, S., Dautel, S.E., Smith, R.D., Payne, S.H., and Metz, T.O. (2017). LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. Bioinformatics *33*, 1744–1746. https://doi.org/10.1093/bioinformatics/btx046.

104. Kind, T., Wohlgemuth, G., Lee, D.Y., Lu, Y., Palazoglu, M., Shahbaz, S., and Fiehn, O. (2009). FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. Anal. Chem. *81*, 10038–10048. https://doi.org/10.1021/ac9019522.

105. Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf. *11*, 395. https://doi.org/10.1186/1471-2105-11-395.

106. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. *4*, 222. https://doi.org/10.1038/msb.2008.61.

107. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968. https://doi.org/10.1093/bioinformatics/btq054.

108. Shi, T., Fillmore, T.L., Sun, X., Zhao, R., Schepmoes, A.A., Hossain, M., Xie, F., Wu, S., Kim, J.-S., Jones, N., et al. (2012). Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. Proc. Natl. Acad. Sci. USA *109*, 15395–15400. https://doi.org/10.1073/pnas.1204366109.

109. Gelman, S.J., Naser, F., Mahieu, N.G., McKenzie, L.D., Dunn, G.P., Chheda, M.G., and Patti, G.J. (2018). Consumption of NADPH for 2-HG Synthesis Increases Pentose Phosphate Pathway Flux and Sensitizes Cells to Oxidative Stress. Cell Rep. *22*, 512–522. https://doi.org/10.1016/j.celrep.2017.12.050.

110. Cremer, A., Ellegast, J.M., Alexe, G., Frank, E.S., Ross, L., Chu, S.H., Pikman, Y., Robichaud, A., Goodale, A., Häupl, B., et al. (2020). Resistance Mechanisms to SYK Inhibition in Acute Myeloid Leukemia. Cancer Discov. *10*, 214–231. https://doi.org/10.1158/2159-8290.CD-19-0209.

111. Bady, P., Sciuscio, D., Diserens, A.-C., Bloch, J., Van Den Bent, M.J., Marosi, C., Dietrich, P.-Y., Weller, M., Mariani, L., Heppner, F.L., et al. (2012). MGMT methylation analysis of glioblastoma on the Infinium methylation BeadChip identifies two distinct CpG regions associated with gene silencing and outcome, yielding a prediction model for comparisons across datasets, tumor grades, and CIMP-status. Acta Neuropathol. *124*, 547–560. https://doi.org/10.1007/s00401-012-1016-2.

112. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898. https://doi.org/10.1126/science.aav1898.

113. Sharma, V., Eckels, J., Schilling, B., Ludwig, C., Jaffe, J.D., MacCoss, M.J., and MacLean, B. (2018). Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline. Mol. Cell. Proteomics *17*, 1239–1244. https://doi.org/10.1074/mcp.RA117.000543.

114. Toghi Eshghi, S., Shah, P., Yang, W., Li, X., and Zhang, H. (2015). GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides. Anal. Chem. *87*, 5181–5188. https://doi.org/10.1021/acs.analchem.5b00024.

115. Toghi Eshghi, S., Yang, W., Hu, Y., Shah, P., Sun, S., Li, X., and Zhang, H. (2016). Classification of Tandem Mass Spectra for Identification of N- and O-linked Glycopeptides. Sci. Rep. *6*, 37189. https://doi.org/10.1038/srep37189.

116. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.

117. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. Bioinformatics *17*, 520–525. https://doi.org/10.1093/bioinformatics/17.6.520.

118. Pengelly, R.J., Gibson, J., Andreoletti, G., Collins, A., Mattocks, C.J., and Ennis, S. (2013). A SNP profiling panel for sample tracking in whole-exome sequencing studies. Genome Med *5*, 89. https://doi.org/10.1186/gm492.

119. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell *32*, 42–56.e6. https://doi.org/10.1016/j.ccell.2017.06.003.

120. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

121. Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C.C., McIntosh, B.J., Leow, K.X., Schwartz, M.S., et al. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat. Biotechnol. *40*, 555–565. https://doi.org/10.1038/s41587-021-01094-0.

122. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 1–11. https://doi.org/10.1038/ncomms3612.

123. Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408. https://doi.org/10.1038/s41586-020-2188-x.

124. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. J. Digit. Imag. *26*, 1045–1057. https://doi.org/10.1007/s10278-013-9622-7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD45	Novus Biologicals	Cat# NBP234287A
CD3e	Akoya	Cat# 4550119; RRID: AB_2936080
CD4	Akoya	Cat# 4550112; RRID: AB_3094499
CD8	Akoya	Cat# 4250012; RRID: AB_2915960
FOXP3	Akoya	Cat# 4550071; RRID: AB_2927679
CD20	Akoya	Cat# 4450094; RRID: AB_3094498
CD68	Akoya	Cat# 4550113; RRID: AB_2935894
CD163	Akoya	Cat# 4250079; RRID: AB_2935895
IBA1	Thermo Fisher Scientific	Cat# MA5-27726; RRID: AB_2735228
CD11b	Abcam	Cat# ab187537
HLA-DR	Abcam	Cat# ab209968; RRID: AB_3095318
PTPRZ1	Abcam	Cat# ab290651
OLIG2	Abcam	Cat# ab220796; AB_2923001
PanCytokeratin	Akoya	Cat# 4450020; RRID: AB_3083456
GFAP	Invitrogen	Cat# 13-0300
CD31	Abcam	Cat# 226157; RRID: AB_3094823
Vimentin	Biolegend	Cat# 677802; RRID: AB_2565982
Ki67	Akoya	Cat# 4450096; RRID: AB_3094497
GLUT1	Abcam	Cat# ab196357; RRID: AB_2832207
VEGFA	Abcam	Cat# ab185238
PAI1	Abcam	Cat# ab237780
HIF1A	Abcam	Cat# ab210073
FN1	Thermo Fisher Scientific	Cat# 2335-MSM8-P1ABX
PTPN11-Y546	Abcam	Cat# ab225574
Biological samples		
Primary tumor and normal tissue samples	This paper	
Chemicals, peptides, and recombinant proteins		
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma	Catalog: H3375
Acetic Acid, glacial	Sigma	Catalog: AX0074-6
Acetonitrile, HPLC grade	J.T. Baker	Catalog: 9829-03
Acetonitrile anhydrous	Sigma	Catalog: 271004
Ammonium hydroxide solution	Sigma	Catalog: 338818
Aprotinin	Sigma	Catalog: A6103
Dithiothreitol	Thermo Scientific	Catalog: 20291
Ethylenediaminetetraacetic acid	Sigma	Catalog: E7889
Formic acid	Sigma	Catalog: 0507
Iodoacetamide	Sigma	Catalog: A3221
Iron (III) chloride	Sigma	Catalog: 451649
HPLC Grade Water	J.T. Baker	Catalog: 4218-03
Hydroxylamine Solution 50%	Sigma	Catalog: 467804
Leupeptin	Roche	Catalog: 11017101001
Lysyl Endopeptidase	Wako Chemicals	Catalog: 129-02541
Methanol, HPLC grade	Fluka	Catalog: 34966

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ni-NTA Superflow Agarose Beads	Qiagen	Catalog: 30410
Oasis MAX 1 cc Vac Cartridge, 30 mg Sorbent per Cartridge	Waters	Catalog: 186000366
Phenylmethylsulfonyl fluoride	Sigma	Catalog: 93482
Phosphatase Inhibitor Cocktail 2	Sigma	Catalog: P5726
Phosphatase Inhibitor Cocktail 3	Sigma	Catalog: P0044
Potassium phosphate dibasic	Sigma	Catalog: P3786
Potassium phosphate monobasic	Sigma	Catalog: P9791
PUGNAc	Sigma	Catalog: A7229
Reversed-phase tC18 SepPak	Waters	Catalog: WAT054925
Sequencing grade modified trypsin	Promega	Catalog: V517
Sodium butyrate	Sigma	Catalog: 303410
Sodium chloride	Sigma	Catalog: S7653
Sodium fluoride	Sigma	Catalog: S7920
Triethylammonium acetate buffer	Sigma	Catalog: 90358
Trifluoroacetic acid	Sigma	Catalog: 91707
Tris (hydroxymethyl)aminomethane hydrochloride pH 8.0	Sigma	Catalog: T2694
Trypsin	Promega	Catalog: V5117
Urea	Sigma	Catalog: U0631
Critical commercial assays		
BCA Protein Assay Kit	ThermoFisher Scientific	Catalog: A53225
Infinium MethylationEPIC Kit	Illumina	Catalog: WG-317-1003
PTMScan® Acetyl-Lysine Motif [Ac-K] Kit	Cell Signaling	Catalog: 13416
TMT-11 reagent kit	ThermoFisher Scientific	Catalog: A34808
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold	Illumina	Catalog: RS-122-2301
Q5 site-directed mutagenesis kit	New England Biolabs	Catalog: E0554S
Deposited data		
CPTAC WES, WGS, DNA-methylation, bulk RNA, miRNA, multiome snRNA seq data	This paper	Genomic Data Commons (GDC): https://portal.gdc.cancer.gov/projects/CPTAC-3
CPTAC proteome, phosphoproteome, acetylome, glycoproteome, metabolome and lipidome data	This paper	Proteomic Data Commons (PDC): https://pdc.cancer.gov/
CPTAC multiome snATAC seq data	This paper	Cancer Data Service (CDS): https://dataservice.datacommons.cancer.gov/
CODEX and histopathology images	This paper	The Cancer Imaging Archive (TCIA): https://doi.org/10.7937/K9/TCIA.2018.3RJE41Q1 Imaging Data Commons: https://portal.imaging.datacommons.cancer.gov/explore/filters/?collection_id=CPTAC&collection_id=cptac_gbm
RNA expression data from The Cancer Genome Atlas Research Network (TCGA) cohort	Brennan et al.[46]	https://portal.gdc.cancer.gov/
Processed DNA somatic alteration and RNA-seq data from Glioma Longitudinal Analysis Consortium (GLASS) cohort	Varn et al.[6]	https://www.synapse.org/glass
Targeted exome-seq and RNA-seq data from Glioblastoma, Stability of Actionable Mutations (GSAM) cohort	Hoogstrate et al.[7]	European Genome Phenome Archive (EGA: EGAD00001007860, EGA: EGAD00001004593)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pLEX307 PTPN11	Addgene	Plasmid #140941; RRID: Addgene_140941
Software and algorithms		
Ascore v1.0.6858	Beausoleil et al.[53]	https://github.com/PNNL-Comp-Mass-Spec/AScore
MASIC	Monroe et al.[54]	https://github.com/PNNL-Comp-Mass-Spec/MASIC
MS-GF+ v9981	Kim et al.[55]	https://github.com/MSGFPlus/msgfplus
mzRefinery	Gibbons et al.[56]	https://omics.pnl.gov/software/mzrefinery
Bioconductor v3.9	Huber et al.[57]	https://bioconductor.org/
Tidyverse	Wickham et al.[58]	https://www.tidyverse.org/
Bioconda	Grüning et al.[59]	https://bioconda.github.io/
Snakemake v5.6	Köster et al.[60]	https://snakemake.readthedocs.io/
BIC-Seq2	Xi et al.[61]	http://compbio.med.harvard.edu/BIC-seq/
GISTIC2 v2.0.22	Mermel et al.[62]	https://github.com/broadinstitute/gistic2
Strelka v2.9.2	Kim et al.[63]	https://github.com/Illumina/strelka
VarScan v2.3.8	Koboldt et al.[64]	https://dkoboldt.github.io/varscan/
Pindel v0.2.5	Ye et al.[65]	https://github.com/genome/pindel
MuTect v1.1.7	Cibulskis et al.[66]	https://github.com/broadinstitute/mutect
somaticwrapper v1.6	Li Ding Lab	https://github.com/ding-lab/somaticwrapper
Samtools v1.2	Li et al.[67]	https://www.htslib.org/
Bedtools v2.26.0	Quinlan et al.[68]	https://bedtools.readthedocs.io/en/latest/
GATK v4.1.9.0	McKenna et al.[69]	https://github.com/broadgsa/gatk
bam-readcount v0.7.4	McDonnell Genome Institute	https://github.com/genome/bam-readcount
Rcwrapper	Li Ding Lab	https://github.com/ding-lab/rcwrapper
10Xmapping	Li Ding Lab	https://github.com/ding-lab/10Xmapping
HTSeq v0.11.2	Anders et al.[70]	https://github.com/simon-anders/htseq
BWA v0.7.17-r1188	Li et al.[71]	http://bio-bwa.sourceforge.net/
RSEM v1.3.1	Li et al.[72]	https://deweylab.github.io/RSEM/
Bowtie2 v2.3.3	Langmead et al.[73]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
R-rollup	Polpitiya et al.[74]	https://omics.pnl.gov/software/danter
ConsensusClusterPlus v1.48.0	Wilkerson et al.[75]	https://bioconductor.org/packages/ConsensusClusterPlus/
louvain-igraph v0.6.1	Blondel et al.[76]	https://doi.org/10.5281/zenodo.1054103
TCGAbiolinks v2.11.1	Colaprico et al.[77]	http://bioconductor.org/packages/TCGAbiolinks/
iProFun	Song et al.[78]	https://github.com/songxiaoyu/iProFun
xCell	Aran et al.[79]	http://xcell.ucsf.edu/
Seurat v4.1.1	Butler et al.[80]	https://cran.r-project.org/web/packages/Seurat
edgeR v3.28.1	Robinson et al.[81]	https://www.bioconductor.org/packages/edgeR/
MS-PyCloud	https://www.biorxiv.org/content/10.1101/320887v1.full	https://bitbucket.org/mschnau1/ms-pycloud/src/main/
GPQuest2.1	Hu et al.[82]	https://github.com/huizhanglab-jhu/GPQuest
WebGestalt	Liao et al.[83]	http://www.webgestalt.org/
pROC	Robin et al.[84]	https://www.expasy.org/resources/proc
Caret	https://doi.org/10.18637/jss.v028.i05	https://cran.r-project.org/web/packages/caret/index.html
ComplexHeatmap	Gu et al.[85]	https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
CancerSubtypes	Xu et al.[86]	https://bioconductor.org/packages/release/bioc/html/CancerSubtypes.html
Clonality	Mauguen et al.[87]	http://bioconductor.org/packages/release/bioc/html/Clonality.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
RefSeq (downloaded from UCSC Genome Browser on 2018-06-29)	O'Leary et al.[88]	https://www.ncbi.nlm.nih.gov/refseq/; https://genome.ucsc.edu/cgi-bin/hgTables; RRID:SCR_003496
GENCODE v22 (download from GDC Reference Files)	Frankish et al.[89]	https://www.gencodegenes.org/; https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files
gnomAD v2.1	Karczewski et al.[90]	https://gnomad.broadinstitute.org/
The 1000 Genomes Project (final phase release on 2013-05-02)	1000 Genomes Project Consortium[91]	https://www.internationalgenome.org/
OmniPath (downloaded on 2018-03-29)	Türei et al.[42]	http://omnipathdb.org/
DEPOD (downloaded on 2018-03-29)	Damle et al.[44]	http://depod.bioss.uni-freiburg.de/
CORUM (downloaded on 2018-06-29)	Ruepp et al.[92]	https://mips.helmholtz-muenchen.de/corum/
SIGNOR v2.0 (downloaded on 2018-10-29)	Licata et al.[45]	https://signor.uniroma2.it/
STRING	Szklarczyk et al.[47]	https://string-db.org/
MINT	Calderone et al.[93]	https://mint.bio.uniroma2.it/
BioGrid	Oughtred et al.[94]	https://thebiogrid.org/
IntAct	del Toro et al.[95]	https://www.ebi.ac.uk/intact/
Reactome (downloaded on 2018-11-01)	Fabregat et al.[96]	https://reactome.org/
NetworKIN 3.0	Horn et al.[97]	https://networkin.info/
UCSF Chimera	Park et al.[29]	https://www.cgl.ucsf.edu/chimera

RESOURCES AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Li Ding (lding@wustl.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Clinical data and proteomic data (raw MS files and processed data files of global proteomics and PTMs) and metabolomic data reported in this paper can be accessed via the Proteomic Data Commons (PDC) at: https://pdc.cancer.gov/. Genomic, transcriptomic, and multiome snRNA-seq data files can be accessed via Genomic Data Commons (GDC) at: https://portal.gdc.cancer.gov/projects/CPTAC-3. We recommend using a sample ID manifest to ensure exploration of the complete dataset including metastatic samples. Multiome snATAC-seq data files can be accessed via the Cancer Data Service (CDS) at Cancer Data Service (CDS) https://dataservice.datacommons.cancer.gov/. Processed data used in this publication can also be found in the PDC, the Python package and LinkedOmics.[98] Pathology and CODEX images can be accessed via The Cancer Imaging Archive (TCIA)[99] at https://doi.org/10.7937/K9/TCIA.2018.3RJE41Q1, and the Imaging Data Commons at https://portal.imaging.datacommons.cancer.gov/explore/filters/?collection_id=CPTAC&collection_id=cptac_gbm.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human participants

Tumor and germline blood samples from 207 qualified cases (200 primary tumor samples, 28 matched recurrent tumor samples) were collected from 10 tissue source sites in strict accordance with the CPTAC-3 protocol with informed consent from the patients. Eight normal adjacent tissue samples were collected as part of this study, and 10 normal samples from the frontal cortex were used in the analysis from the GTEx project (https://gtexportal.org/). This study contained both males (n = 124) and females (n = 83) from 9 different countries. Only histopathologically defined adult GBMs, grade 4 *IDH*-mutant astrocytomas and brain metastases were considered for analysis, with an age range of 24-88 years. Clinical data were obtained from the tissue source sites and reviewed for correctness and completeness of data. Institutional review boards at each Tissue Source Site (TSS) reviewed protocols and consent documentation, in adherence to Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines.

METHOD DETAILS

Sample processing

The CPTAC Biospecimen Core Resource (BCR) at the Pathology and Biorepository Core of the Van Andel Research Institute in Grand Rapids, Michigan manufactured and distributed biospecimen kits to the Tissue Source Sites (TSS) located in the US, Europe, and Asia. Each kit contains a set of pre-manufactured labels for unique tracking of every specimen respective to TSS location, disease, and sample type, used to track the specimens through the BCR to the CPTAC proteomic and genomic characterization centers.

Tissue specimens averaging 200 mg were snap-frozen by the TSS within a 30 min cold ischemic time (CIT) (CIT average = 13 min), and an adjacent segment was formalin-fixed paraffin-embedded (FFPE) and H&E stained by the TSS for quality assessment to meet the CPTAC GBM requirements. Routinely, several tissue segments for each case were collected. Tissues were flash frozen in liquid nitrogen (LN2) and transferred to a liquid nitrogen freezer for storage until approval for shipment to the BCR.

Specimens were shipped to the BCR using a cryoport that maintained an average temperature of under −140°C with a time and temperature tracker to monitor the shipment. Receipt of specimens at the BCR included a physical inspection and review of the time and temperature tracker data for specimen integrity, followed by barcode entry into a biospecimen tracking database. Specimens were again placed in LN2 storage until further processing. Acceptable GBM tumor tissue segments were determined by TSS pathologists based on the percent viable tumor nuclei (>60%), total cellularity (>50%), and necrosis (<50%). Segments received at the BCR were verified by BCR and Leidos Biomedical Research (LBR) pathologists, and the percentage of total area of tumor in the segment was also documented. Additionally, disease-specific working group pathology experts reviewed the morphology to clarify or standardize specific disease classifications and correlation to the proteomic and genomic data.

Specimens selected for the discovery set were determined on the maximal percent in the pathology criteria and best weight. Specimens were pulled from the biorepository using an LN2 cryocart to maintain specimen integrity and then cryopulverized. The cryopulverized specimen was divided into aliquots for DNA (30 mg) and RNA (30 mg) isolation and proteomics (50 mg) for molecular characterization. Nucleic acids were isolated and stored at −80°C until further processing and distribution; cryopulverized protein material was returned to the LN2 freezer until distribution. Shipment of the cryopulverized segments used cryoports for distribution to the proteomic characterization centers and shipment of the nucleic acids used dry ice shippers for distribution to the genomic characterization centers; a shipment manifest accompanied all distributions for the receipt and integrity inspection of the specimens at the destination. The DNA sequencing was performed at the Broad Institute, Cambridge, MA, and RNA sequencing was performed at the University of North Carolina, Chapel Hill, NC. Material for proteomic analyses was sent to the Proteomic Characterization Center (PCC) at Pacific Northwest National Laboratory (PNNL), Richland, Washington.

Longitudinal cohort sample collection and processing

The longitudinal cohort was collected through the HUP-CHOP Neurosurgery Tumor Tissue Bank Collaborative at the Hospital of the University of Pennsylvania. All samples were collected fresh-frozen at the time of surgery, shipped, and stored in BioRC (Biorepository Resource Center) at Children's Hospital of Philadelphia. 30 mg tissue pieces were cut/chipped off using disposable scalpels on dry ice and delivered to Fred Hutchinson Cancer Research Center for sample preparation and proteomic analysis.

Sample processing for genomic DNA and total RNA extraction

Our study sampled a single site of the primary tumor from surgical resections, due to the internal requirement to process a minimum of 125 mg of tumor issue and 50 mg of adjacent normal tissue. DNA and RNA were extracted from tumor and blood normal specimens in a co-isolation protocol using Qiagen's QIAsymphony DNA Mini Kit and QIAsymphony RNA Kit. Genomic DNA was also isolated from peripheral blood (3-5 mL) to serve as matched normal reference material. The Qubit™ dsDNA BR Assay Kit was used with the Qubit® 2.0 Fluorometer to determine the concentration of dsDNA in an aqueous solution. Any sample that passed quality control and produced sufficient DNA yield to undergo various genomic assays was sent for genomic characterization. RNA quality was quantified using both the NanoDrop 8000 and quality assessed using Agilent Bioanalyzer. A sample that passed RNA quality control and had a minimum RIN (RNA integrity number) score of 7 was subjected to RNA sequencing. Identity match for germline, normal adjacent tissue, and tumor tissue was assayed at the BCR using the Illumina Infinium QC array. This beadchip contains 15,949 markers designed to prioritize sample tracking, quality control, and stratification.

Whole exome sequencing

Library construction

Library construction was performed as described in,[100] with the following modifications: initial genomic DNA input into shearing was reduced from 3 μg to 20-250 ng in 50 μL of solution. For adapter ligation, Illumina paired-end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adapter ligation, and library enrichment PCR. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30 μL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

In-solution hybrid selection

After library construction, libraries were pooled into groups of up to 96 samples. Hybridization and capture were performed using the relevant components of Illumina's Nextera Exome Kit and following the manufacturer's suggested protocol, with the following

exceptions. First, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's Nextera Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo) using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries were normalized to 2 nM.

Cluster amplification and sequencing

Cluster amplification of DNA libraries was performed according to the manufacturer's protocol (Illumina) using exclusion amplification chemistry and flowcells. Flowcells were sequenced utilizing sequencing-by-synthesis chemistry. The flowcells were then analyzed using RTA v.2.7.3 or later. Each pool of whole exome libraries was sequenced on paired 76 cycle runs with two 8 cycle index reads across the number of lanes needed to meet coverage for all libraries in the pool. Pooled libraries were run on HiSeq 4000 paired-end runs to achieve a minimum of 150x on target coverage per each sample library. The raw Illumina sequence data were demultiplexed and converted to fastq files; adapter and low-quality sequences were trimmed. The raw reads were mapped to the hg38 human reference genome and the validated BAMs were used for downstream analysis and variant calling.

PCR-free whole genome sequencing

Preparation of libraries for cluster amplification and sequencing

An aliquot of genomic DNA (350 ng in 50 μL) was used as the input into DNA fragmentation (aka shearing). Shearing was performed acoustically using a Covaris focused-ultrasonicator, targeting 385bp fragments. Following fragmentation, additional size selection was performed using a SPRI cleanup. Library preparation was performed using a commercially available kit provided by KAPA Biosystems (KAPA Hyper Prep without amplification module) and with palindromic forked adapters with unique 8-base index sequences embedded within the adapter (purchased from IDT). Following sample preparation, libraries were quantified using quantitative PCR (kit purchased from KAPA Biosystems), with probes specific to the ends of the adapters. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 1.7 nM and pooled into 24-plexes.

Cluster amplification and sequencing (HiSeq X)

Sample pools were combined with HiSeq X Cluster Amp Reagents EPX1, EPX2, and EPX3 into single wells on a strip tube using the Hamilton Starlet Liquid Handling system. Cluster amplification of the templates was performed according to the manufacturer's protocol (Illumina) with the Illumina cBot. Flowcells were sequenced to a minimum of 15x on HiSeq X utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads. Output from Illumina software was processed by the Picard data processing pipeline to yield BAMs containing demultiplexed, aggregated, aligned reads. All sample information tracking was performed by automated LIMS messaging.

Illumina infinium methylationEPIC beadchip array

The MethylationEPIC array uses an 8-sample version of the Illumina Beadchip that captures > 850,000 DNA methylation sites per sample. 250 ng of DNA was used for the bisulfite conversation using Infinium MethylationEPIC BeadChip Kit. The EPIC array includes sample plating, bisulfite conversion, and methylation array processing. After scanning, the data was processed through an automated genotype calling pipeline. Data generated consisted of raw idats and a sample sheet.

RNA sequencing

Quality assurance and quality control of RNA analytes

All RNA analytes were assayed for RNA integrity, concentration, and fragment size. Samples for total RNA-seq were quantified on a TapeStation system (Agilent, Inc. Santa Clara, CA). Samples with RINs > 8.0 were considered high quality.

Total RNA-seq library construction

Total RNA-seq library construction was performed from the RNA samples using the TruSeq Stranded RNA Sample Preparation Kit and bar-coded with individual tags following the manufacturer's instructions (Illumina, Inc. San Diego, CA). Libraries were prepared on an Agilent Bravo Automated Liquid Handling System. Quality control was performed at every step and the libraries were quantified using the TapeStation system.

Total RNA sequencing

Indexed libraries were prepared and run on HiSeq 4000 paired end 75 base pairs to generate a minimum of 120 million reads per sample library with a target of greater than 90% mapped reads. Typically, these were pools of four samples. The raw Illumina sequence data were demultiplexed and converted to FASTQ files, and adapter and low-quality sequences were quantified. Samples were then assessed for quality by mapping reads to the hg38 human genome reference, estimating the total number of reads that mapped, amount of RNA mapping to coding regions, amount of rRNA in sample, number of genes expressed, and relative expression of housekeeping genes. Samples passing this QA/QC were then clustered with other expression data from similar and distinct tumor types to confirm expected expression patterns. Atypical samples were then SNP typed from the RNA data to confirm source analyte. FASTQ files of all reads were then uploaded to the GDC repository.

miRNA-seq library construction

miRNA-seq library construction was performed from the RNA samples using the NEXTflex Small RNA-Seq Kit (v3, PerkinElmer, Waltham, MA) and bar-coded with individual tags following the manufacturer's instructions. Libraries were prepared on Sciclone Liquid Handling Workstation Quality control was performed at every step, and the libraries were quantified using a TapeStation system and an Agilent Bioanalyzer using the Small RNA analysis kit. Pooled libraries were then size selected according to NEXTflex Kit specifications using a Pippin Prep system (Sage Science, Beverly, MA).

miRNA sequencing

Indexed libraries were loaded on the Hiseq 4000 to generate a minimum of 10 million reads per library with a minimum of 90% reads mapped. The raw Illumina sequence data were demultiplexed and converted to FASTQ files for downstream analysis. Resultant data were analyzed using a variant of the small RNA quantification pipeline developed for TCGA (Chu et al., 2016). Samples were assessed for the number of miRNAs called, species diversity, and total abundance. Samples passing quality control were uploaded to the GDC repository.

Single-nuclei multiome library construction and sequencing

First, 15–25 mg of pulverized tissue was placed in a 5-ml Eppendorf tube on ice. Using a wide-bore pipette tip (Rainin), a lysis buffer prepared from the Nuclei Isolation protocol (10x Genomics) and SuperRNase inhibitor (Invitrogen) was added to the tube. The tissue solution was gently pipetted until the lysis liquid turned a slightly cloudy color. (The number of pipetting iterations depended on the specific tissue.) The tissue homogenate was then filtered through a 40-μm strainer (pluriSelect) and washed with a BSA wash buffer (2% BSA + 1 × PBS + RNase inhibitor). The filtrate was collected, centrifuged at 500g for 6 min at 4°C and resuspended with a BSA wash buffer. Then, 100 μL of cell lysis solution was set aside for unstained reference, while the rest was stained with 7AAD or DRAQ5 depending on ATAC or RNA protocol. Nuclei underwent FACS and sorting gates were based on size, granularity and dye staining signal. The final suspension was spun down at 500g for 6 min at 4°C and resuspended with a BSA wash buffer. More specific details about the RNA protocol can be found on protocols.io (RNA protocol: dx.doi.org/10.17504/protocols.io.14egn7w6zv5d/v1; for ATAC protocol 7AAD is substituted for dye).

Nuclei and barcoded beads were isolated in oil droplets via the 10x Genomics Chromium instrument. Single nuclei suspensions were counted and adjusted to a range of 500 to 1800 nuclei/μL using a hemocytometer. Reverse transcription was subsequently performed to incorporate cell and transcript specific barcodes. All snRNA-seq samples were run using the Chromium Next GEM Single Cell 3' Library and Gel Bead Kit v3.2 or v3.3 (10x Genomics). For the multiome kit, Chromium Next GEM Single Cell Multiome ATAC + Gene Expression was used. Barcoded libraries were then pooled and sequenced on the Illumina NovaSeq 6000 system with associated flow cells.

MS sample processing and data collection

Protein extraction and tandem Lys-C/trypsin digestion

Approximately 50 mg of each of the fresh frozen GBM tumor and GTEx normal brain tissues were homogenized separately in 200 μL of lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris, pH 8.0, 1 mM EDTA, 2 μg/mL aprotinin, 10 μg/mL leupeptin, 1 mM PMSF, 10 mM NaF, 1:100 v/v Sigma phosphatase inhibitor cocktail 2, 1:100 v/v Sigma phosphatase inhibitor cocktail 3, 20 μM PUGNAc, and 5 mM sodium butyrate). Lysates were precleared by centrifugation at 20,000 g for 10 min at 4°C and protein concentrations were determined by BCA assay (ThermoFisher Scientific) and adjusted to 8 μg/μL with lysis buffer. Proteins were reduced with 5 mM dithiothreitol for 1 h at 37°C, and subsequently alkylated with 10 mM iodoacetamide for 45 min at 25°C in the dark. Samples were diluted 1:3 with 50 mM Tris, pH 8.0 and digested with Lys-C (Wako) at 1:50 enzyme-to-substrate ratio. After 2 h of digestion at 25°C, sequencing grade modified trypsin (Promega, V5117) at 1:50 enzyme-to-substrate ratio was added to the samples and further incubated at 25°C for 14 h. The digested samples were then acidified with 100% formic acid to 1% of final concentration of formic acid and centrifuged for 15 min at 1,500 g to clear digest from precipitation. Tryptic peptides were desalted on C18 SPE (Waters tC18 SepPak) and dried using Speed-Vac.

TMT-11 labeling

Desalted peptides were labeled with 11-plex TMT reagents (ThermoFisher Scientific) using conditions modified from the manufacturer's instructions. Peptides (400 μg) from each of the tumors were dissolved in 80 μL of 50 mM HEPES, pH 8.5 solution, and mixed with 400 μg of TMT reagent that was dissolved freshly in 20 μL of anhydrous acetonitrile. Channel 126 was used for labeling the internal reference sample (pooled from all tumor and normal samples) throughout the sample analysis. After 1 h incubation at RT, each sample was diluted with 60 μL of 50 mM HEPES, pH 8.5 in 20% acetonitrile. Twelve μL of 5% hydroxylamine was then added and the samples were incubated for 15 min at RT to quench the reaction. Peptides labeled by different TMT reagents were then mixed, dried using Speed-Vac, reconstituted with 3% acetonitrile, 0.1% formic acid and were desalted on Waters tC18 SepPak SPE columns.

Peptide fractionation

Approximately 3.5 mg of 11-plex TMT labeled sample was separated on a reversed phase Agilent Zorbax 300 Extend-C18 column (250 mm × 4.6 mm column containing 3.5-μm particles) using Agilent 1200 HPLC System. Solvent A was 4.5 mM ammonium formate, pH 10, 2% acetonitrile and solvent B was 4.5 mM ammonium formate, pH 10, 90% acetonitrile. The flow rate was 1 mL/min and the injection volume was 900 μL. The LC gradient started with a linear increase of solvent B to 16% in 6 min, then linearly increased to 40% B in 60 min, 4 min to 44% B, 5 min to 60% B and another 14 of 60% solvent B. A total of 96 fractions were collected into a 96 well plate throughout the LC gradient. These fractions were concatenated into 24 fractions by combining 4 fractions that are

24 fractions apart (i.e., combining fractions #1, #25, #49, and #73; #2, #26, #50, and #74; and so on). For proteome analysis, 5% of each concatenated fraction was dried down and re-suspended in 2% acetonitrile, 0.1% formic acid to a peptide concentration of 0.1 mg/mL for LC-MS/MS analysis. The rest of the fractions (95%) were further concatenated into 12 fractions (i.e., by combining fractions #1 and #13; #3 and #15; and so on), dried down, and subjected to immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment.

Phosphopeptide enrichment using IMAC

Fe^{3+}-NTA-agarose beads were freshly prepared using the Ni-NTA Superflow agarose beads (QIAGEN, #30410) for phosphopeptide enrichment. For each of the 12 fractions, peptides were reconstituted to 0.5 g/μL in IMAC binding/wash buffer (80% acetonitrile, 0.1% trifluoroacetic acid) and incubated with 10 μL of the Fe^{3+}-NTA-agarose beads for 30 min at RT. After incubation, the beads were washed 2 times each with 50 μL of wash buffer and once with 50 μL of 1% formic acid on the stage tip packed with 2 discs of Empore C18 material (Empore Octadecyl C18, 47 mm; Supleco, 66883-U). Phosphopeptides were eluted from the beads on C18 using 70 μL of Elution Buffer (500 mM potassium phosphate buffer). 50% acetonitrile, 0.1% formic acid was used for elution of phosphopeptides from the C18 stage tips. Samples were dried using Speed-Vac, and later reconstituted with 12 μL of 3% acetonitrile, 0.1% formic acid for LC-MS/MS analysis.

Immunoaffinity purification of acetylated peptides

Tryptic peptides from the flow-through of IMAC were dried using Speed-Vac. The dried peptides were reconstituted in 1.4 mL of the immunoaffinity purification (IAP) buffer (50 mM MOPS/NaOH pH 7.2, 10 mM Na_2HPO_4 and 50 mM NaCl). After dissolving the peptide, the pH of the peptide solution was checked using pH indicator paper. The amount of reconstituted peptides was quantified via BCA assay and concatenated into 4 fractions by combining 3 fractions that are 4 fractions apart (i.e., combining fractions #1, #5 and #9 as a new fraction). The antibody beads from PTMScan® Acetyl-Lysine Motif [Ac-K] Kit (Cell Signaling, #13416) were freshly prepared. Briefly, the antibody beads were centrifuged at 2,000 x g for 30 sec and all buffer from the beads were removed; the antibody beads were then washed with 1 mL of IAP buffer for four times and finally resuspend in 40 μL of IAP buffer. For each fraction, 1/2 amounts of the antibody in each tube were transferred to the peptide solution and incubated on a rotator overnight at 4°C. After removing the supernatant, the reacted beads were washed with 1mL of PBS buffer for five times. For the elution of acetylated peptides, the antibody beads were incubated 2 times each with 50 μL of 0.15% TFA at room temperature for 10 min. The eluted peptides were transferred to the stage tip packed with 2 discs of Empore C18 material. The C18 stage tips were washed by 1% formic acid and 50% acetonitrile, and 0.1% formic acid was used for elution of peptides from the C18 stage tips. The eluted peptides were dried using Speed-Vac, and reconstituted with 12 μL of 3% acetonitrile, 0.1% formic acid right before the LC-MS/MS analysis.

Enrichment of intact glycopeptides

In this study, the intact glycopeptides were enriched from the flow-through of acetylated peptide enrichment. The flow-throughs were first desalted on reversed phase C18 SPE column (Waters), and then the intact glycopeptides were enriched using the mixed anion exchange (MAX) solid phase extraction column (Waters). MAX cartridges were conditioned with 3 x 1 mL of ACN, 3 x 1 mL of 100 mM triethylammonium acetate, 3 x 1 mL of water (Fisher chemical, LC/MS grade), and 3 x 1 mL of 95% ACN (v/v) 1% TFA (v/v). Samples were dissolved in 1 mL of 95 % ACN (v/v) 1 % TFA (v/v) and loaded onto cartridge twice. The non-glycosylated peptides were washed out by 95% ACN (v/v) 1% TFA (v/v). Finally, glycopeptides were eluted by 500 μL of 50% ACN (v/v) 0.1% TFA (v/v), dried and stored at -80°C prior to LC-MS/MS analysis.

LC-MS/MS analysis

Fractionated samples prepared for global proteome, phosphoproteome and acetylome analysis were separated under identical conditions using a nanoACQUITY UPLC system (Waters) by reversed-phase HPLC. The analytical column was manufactured in-house using ReproSil-Pur 120 C18-AQ 1.9 μm stationary phase (Dr. Maisch GmbH) and slurry packed into a 25-cm length of 360 μm o.d. x 75 μm i.d. fused silica picofrit capillary tubing (New Objective). The analytical column was heated to 50°C using an AgileSLEEVE column heater (Analytical Sales and Services). The analytical column was equilibrated to 98 % Mobile Phase A (MP A, 0.1% formic acid/3% acetonitrile) and 2% Mobile Phase B (MP B, 0.1% formic acid/90% acetonitrile) and maintained at a constant column flow of 200 nL/min. The sample was injected into a 5-μL loop placed in-line with the analytical column which initiated the gradient profiles (min:% MP B): 0:2, 1:6, 85:30, 94:60, 95:90, 100:90, 101:50, 110:50 (global proteome and phosphoproteome); 0:2, 1:6, 235:40, 244:60, 245:90, 250:90, 251:50, 260:50 (acetylome). The column was allowed to equilibrate at start conditions for 30 minutes between analytical runs.

MS analysis was performed using an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific). The global proteome and phosphoproteome samples were analyzed under identical conditions. Electrospray voltage (1.8 kV) was applied at a carbon composite union (Valco Instruments) coupling a 360 μm o.d. x 20 μm i.d. fused silica extension from the LC gradient pump to the analytical column and the ion transfer tube was set at 250°C. Following a 25 min delay from the time of sample injection, Orbitrap precursor spectra (AGC $4x10^5$) were collected from 350-1800 *m/z* for 110 min at a resolution of 60K along with data dependent Orbitrap HCD MS/MS spectra (centroid) at a resolution of 50K (AGC $1x10^5$) and max ion time of 105 ms for a total duty cycle of 2 seconds. Masses selected for MS/MS were isolated (quadrupole) at a width of 0.7 *m/z* and fragmented using a collision energy of 30%. Peptide mode was selected for monoisotopic precursor scan and charge state screening was enabled to reject unassigned 1+, 7+, 8+, and >8+ ions with a dynamic exclusion time of 45 seconds to discriminate against previously analyzed ions between +/- 10 ppm. The acetylome samples were analyzed under similar conditions.

Intact glycopeptides were analyzed by Orbitrap Fusion Lumos Tribrid (Thermo Scientific) combined with Easy nLC 1200 UPLC system (Thermo Scientific). Samples were reconstituted with 3 % ACN in 0.1 % formic acid (solvent A) before loading onto an in-house

packed column (0.75 μm I.D. x 27.5 cm length packed with ReproSil-Pur 120 C18-AQ, 1.9 μm). Loaded peptides were subjected to the gradient with 200 nL/min of flow rate as follows: 2 to 6 % B (90 % ACN 0.1 % F.A) for 1 min, 6 to 30 % B for 84 min, 30 to 60 % B for 9 min, 60 to 90 % B for 1 min, isocratic 90 % B for 5 min, 90 to 50 % B for 1 min, and isocratic 50 % B for 9 min. Parameters for glycoproteomic samples were set as follows: MS1 resolution - 60,000, mass range – 500 to 2000 *m/z*, RF Lens – 30%, AGC Target – 5.0e5, Max injection time – 50 ms, charge state include – 2-6, dynamic exclusion – 45 s. The cycle time was set to 2 s, and within this 2 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 50,000, high-energy collision dissociation activation energy (HCD) – 35, isolation width (*m/z*) – 0.7, AGC Target – 1.0e5, Max injection time – 100 ms.

Polar metabolites and lipid mass spectrometry

Metabolite and lipid extraction

Lipids and metabolite extracts were generated from the same pulverized tissue with a minimum of 20 mg using a modified Folch extraction.[101] Additional solvent was added such that the final volume was proportionate to the mass of the sample ensuring the solvent ratio is 3:8:4 H2O:CHCl3:MeOH. Samples were vortexed for 30 sec, chilled in an ice block for 5 min, and vortexed again for 30 sec. The samples were then centrifuged at 10,000 x g for 10 min at 4°C. The polar metabolite extract was transferred into a glass vial, dried in a speedvac, and stored at -20°C until chemical derivatization for gas chromatography mass spectrometry (GC-MS) analysis. The total lipid extract (TLE) was transferred into a glass vial, dried in a speedvac, and then reconstituted in 500 μL 1:1 chloroform/methanol for storage at -20°C until analysis.

Chemical derivatization of polar metabolites

Polar metabolites along with 50% of the TLE were chemically derivatized prior to metabolomics analysis. Chemical derivatization of metabolites was previously detailed.[102] To protect carbonyl groups and reduce the number of tautomeric isomers, 20 μL of methoxyamine in pyridine (30 mg/mL) was added to each sample, followed by vortexing for 30 seconds and incubation at 37°C with generous shaking for 90 minutes. To derivatize hydroxyl and amine groups to trimethylsilylated (TMS) forms, 80 μL of N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) was added to each vial, followed by vortexing for 10 seconds and incubation at 37°C with shaking for 30 minutes. The samples were allowed to cool to room temperature and were analyzed on the GC-MS the same day.

GC-MS analysis

An Agilent GC 7890A coupled with a single quadrupole MSD 5975C was used to analyze chemically derivatized metabolites. GC-MS analysis was previously detailed.[102] Briefly, 1 μL of each sample was injected onto a HP-5MS column (30 m × 0.25 mm × 0.25 μm; Agilent Technologies, Inc). The injection port temperature was held at 250°C throughout the analysis. The GC oven was held at 60°C for 1 minute after injection then increased to 325°C by 10°C/min, followed by a 5-minute hold at 325°C. Total analysis time was 34 minutes per injection. The helium gas flow rates were determined by the Agilent Retention Time Locking function based on analysis of deuterated myristic acid. Data were collected over the mass range 50 - 550 *m/z*. A mixture of fatty acid methyl esters (C8-C28) was analyzed once per day at the beginning of each batch together with the samples for retention index alignment purposes during subsequent data analysis.

LC-MS analysis

Stored plasma TLEs were dried in vacuo (45 min) and reconstituted in 5 μL chloroform plus 95 μL of methanol. The TLEs were analyzed as outlined in the previous study.[103] A Waters Acquity UPLC H class system interfaced with a Velos-ETD Orbitrap mass spectrometer was used for liquid chromatography tandem mass spectrometry (LC-MS/MS) analyses. 10 μL of reconstituted sample was injected onto a Waters CSH column (3.0 mm x 150 mm x 1.7 μm particle size) and separated over a 34-minute gradient (mobile phase A: ACN/H2O (40:60) containing 10 mM ammonium acetate; mobile phase B: ACN/IPA (10:90) containing 10 mM ammonium acetate) at a flow rate of 250 μL/min. Eluting lipids were introduced to the MS via electrospray ionization in both positive and negative modes, and lipids were fragmented using higher-energy collision dissociation (HCD) and collision-induced dissociation (CID).

Metabolite identification and data processing

Metabolite identifications and data processing were conducted as previously detailed.[102] GC-MS raw data files were processed using Metabolite Detector software v2.0.6 beta.[104] Retention indices (RI) of detected metabolites were calculated based on the analysis of the FAMEs mixture, followed by their chromatographic alignment across all analyses after deconvolution. Metabolites were identified by matching experimental spectra to an augmented version of the Agilent Fiehn Metabolomics Retention Time Locked (RTL) Library,[104] containing spectra and validated retention indices. All metabolite identifications were manually validated. The NIST 08 GC-MS library was also used to cross validate the spectral matching scores obtained using the Agilent library and to provide identifications for metabolites that were initially unidentified. The three most abundant fragment ions in the spectra of each identified metabolite were automatically determined by Metabolite Detector, and their summed abundances were integrated across the GC elution profile. A matrix of identified metabolites, unidentified metabolite features, and their corresponding abundances for each sample in the batch were exported for statistics.

Lipid identification and data processing

LC-MS/MS lipidomics data were analyzed using LIQUID (Lipid Informed Quantitation and Identification).[103] Confident identifications were selected by manually evaluating the MS/MS spectra for diagnostic and corresponding acyl chain fragments of the identified lipid. In addition, the precursor isotopic profile, extracted ion chromatogram, and mass measurement error along with the elution time were evaluated. To facilitate quantification of lipids, a reference database for lipids identified from the MS/MS data was created and features from each analysis were then aligned to the reference database based on their identification, *m/z* and retention time

using MZmine 2.[105] Aligned features were manually verified and peak apex intensity values were exported for subsequent statistical analysis.

SRM sample processing and data collection

Stable isotope-labeled peptides

Proteotypic peptides for the target proteins were selected for targeted proteomics analysis based on the TMT identification results and well-accepted criteria,[106] and the corresponding crude heavy stable isotope-labeled peptides were synthesized with $^{13}C/^{15}N$ on C-terminal lysine or arginine (New England Peptide, Gardner, MA). The heavy peptides were dissolved individually in 15% acetonitrile (ACN) and 0.1% formic acid (FA) at a concentration of 2 mM and used for creating a peptide internal standard (IS) mixture with a final concentration of 5 μM for each peptide.

SRM assay development

The heavy peptides in the peptide mixture were evaluated for peptide response and fragmentation pattern using LC-SRM. For each peptide, transition settings were as follows: (1) precursor charges: 2, 3 and 4; (2) fragment ion charges: 1, 2 and 3; (3) ion type: y ions; and (4) *m/z* window: 250-1500. Transition lists were generated with optimal collision energy values by Skyline software (Version 22.2).[107] LC-SRM was then used to evaluate all heavy peptides for stability of peptide retention time, reliable heavy peptides identification, transition interferences, and endogenous peptide detectability. In the end, 3 transitions per peptide were selected for the final assay configuration for targeted quantitation of a total of 194 peptides of the 105 target proteins and 75 phosphopeptides of 62 proteins. The tryptic digests of the patient samples were analyzed directly for 148 peptides; they were subjected to further chromatographic and affinity enrichment for detecting an additional 46 peptides and 75 phosphopeptides, respectively (see below).

PRISM fractionation

High-pressure, high-resolution separations coupled with intelligent selection and multiplexing (PRISM) fractionation[108] was performed for 46 peptides to achieve higher sensitivity quantitation by SRM. A nanoACQUITY UPLC system (Waters Corporation, Milford, MA) equipped with a reversed-phase capillary LC column (30-μm Jupiter C18 bonded particles packed in 200 μm i.d. × 50 cm capillary) was used. Separations were performed by reversed-phase LC fractionation at mobile phase flow rates of 2.2 μL/min on the binary pump systems using 10 mM ammonium formate (pH 7) in water as mobile phase A and 10 mM ammonium formate (pH 7) in 90% ACN as mobile phase B. Forty five microliter of sample with a peptide concentration of 0.5 μg/μL and 12 fmol/μL of heavy peptide standards was loaded onto the reversed-phase capillary column and separated into 96 fractions using a 190-min gradient of (min:%B): 35:1, 37:10, 52:15, 87:25, 112:35, 125:45, 150:90, 156:1. The eluent was automatically deposited every minute and concatenated onto 6 vials. Prior to peptide fraction collection, 20 μL of 0.1% FA was added to each well of the 96-well plate to avoid the loss of peptides. All the eluates were then dried under vacuum.

Phosphopeptides enrichment by IMAC

The in-house-made IMAC tip was capped in a tip-end with a 20-μm polypropylene frits disk followed by packing with Ni-NTA silica resin (QIAGEN, Hilden, Germany). First, Ni^{2+} ions were removed by adding 50 mM EDTA in 1 M NaCl. The tip was then activated with 100 mM $FeCl_3$ and equilibrated with 1% (v/v) acetic acid at pH 3.0 prior to sample loading. One microliter of 2000 fmol/μL of each of the crude IS phosphopeptides spiked into 200 μL of 1 μg/μL tryptic peptides or into 0.1% (v/v) trifluoroacetic acid, 80% ACN were subjected to IMAC. Followed by 1% (v/v) trifluoroacetic acid, 80% ACN, and 1% (v/v) acetic acid washing steps, the bound phosphopeptides were eluted by 200 mM $NH_4H_2PO_4$ onto the activated desalting SDB-XC StageTips for desalting and directly eluted to sample vials of LC-SRM then dried under vacuum.

LC-SRM data acquisition

For direct analysis, the digested tissue samples were reconstituted in 2% ACN/0.1% FA and spiked with 50 fmol/μL heavy peptides for a final concentration of 0.25 μg/μL, and 2 μL of the resulting samples were analyzed by LC-SRM using a Waters nanoACQUITY UPLC system coupled to a Thermo Scientific TSQ Altis triple quadrupole mass spectrometer (Thermo Scientific, Waltham, MA). A 100 μm i.d. × 10 cm, BEH 1.7-μm C18 capillary column (Waters) was operated at a temperature of 44 °C. The mobile phases were (A) 0.1% FA in water and (B) 0.1% FA in ACN. The peptide samples were separated at a flow rate of 400 nL/min using a 110-min gradient profile as follows (min:%B): 7:1, 9:6, 40:13, 70:22, 80:40, 85:95, 93:50, 94:95 and 95:1. Data were acquired in time-scheduled SRM mode (retention time window: 15 min). The PRISM fractions were reconstituted in 20 μL 2% ACN/0.1% FA and 9 μL were loaded onto the column and separated at a flow rate of 400 nL/min using a 70-min gradient profile as follows (min:%B): 11:0.5, 14: 5, 30:20, 37:25, 48:60, 31:95, 54:0.5. Data were acquired in unscheduled SRM mode. For enriched phosphopeptide samples, all the eluent from IMAC was dissolved in 12 μL and 9 μL were loaded onto the column and separated at a flow rate of 400 nL/min using a 42-min gradient profile as follows (min:%B): 11:0.5, 13: 5, 18:20, 23:25, 53:95, 36:0.5. Data were acquired in unscheduled SRM mode. The parameters of the triple quadruple instrument were set with 0.7 fwhm Q1 and Q3 resolution, and 1.2-s cycle time.

Cell line construction

We used immortalized human astrocytes, which we previously generated and described,[109] using hTERT, a dominant negative allele of TP53, and an shRNA targeting p16. Generation of IDH1 wild type and IDH1-R132H mutant expressing cells were previously described.[109] For PTPN11 constructs, we obtained pLEX307 PTPN11[110] (Addgene plasmid #140941). Note that this construct has a stop codon prior to the V5 tag. We then used Q5 site-directed mutagenesis kit (New England Biolabs) to create PTPN11 mutant alleles, which were lentivirally transduced in the immortalized astrocytes, separate from the IDH1 cell lines. We tested cell lines and

confirmed they were free of mycoplasma. Global proteome and phosphoproteome data generated from these cell lines are included in Table S8.

Co-detection by indexing (CODEX) multiplexed tissue imaging

A panel of 25 CODEX antibodies (Table S4) were designed for the human brain. Carrier-free antibodies (Table S4) were verified for their specificity by using immunofluorescence (IF) staining in multiple channels. Once verified, antibodies were conjugated using Akoya Antibody Conjugation Kit (Akoya Biosciences, SKU 7000009) with a barcode (Akoya Biosciences) assigned based on the IF staining results. Several common markers were directly purchased through Akoya Biosciences (Table S4). CODEX staining and imaging were performed according to the manufacturer's instruction (CODEX User Manual - Rev C). Briefly, 5 μm FFPE sections were placed on APTES (Sigma, #440140) coated coverslips and baked at 60°C overnight before deparaffinization. The next day, tissues were incubated in xylene, rehydrated in ethanol, and washed in ddH2O before antigen retrieval with TE buffer, pH 9 (Genemed, #10-0046) in boiling water for 10 min in a pressure cooker. Sections were then blocked using the blocking buffer (CODEX staining kit, SKU 7000008) and stained with the 25-marker antibody panel (Table S4) to a volume of 200 μl for 3 hours at room temperature in a humidified chamber. Imaging of the CODEX multicycle experiment was performed using Keyence fluorescence microscope (model BZ-X810) equipped with a Nikon CFI Plan Apo λ 20x/0.75 objective, the CODEX instrument (Akoya Biosciences, USA), and CODEX Instrument Manager (CIM) (Akoya Biosciences, USA). The raw images were then stitched and processed using the CODEX processor (Akoya Biosciences, USA). After multiplex imaging was completed, hematoxylin and eosin (H&E) staining was performed on the adjacent section from the same tissue.

Genomic data analysis

Harmonized genome alignment

WGS, WES, RNA-Seq sequence data were harmonized by NCI Genomic Data Commons (GDC) https://gdc.cancer.gov/about-data/gdc-data-harmonization, which included alignment to GDC's hg38 human reference genome (GRCh38.d1.vd1) and additional quality checks. All the downstream genomic processing was based on the GDC aligned BAMs to ensure reproducibility. However, RNA-Seq of 9 GTEx and 4 CPTAC samples didn't have the GDC harmonized BAMs available at the time of the analysis. We followed GDC's pipeline (same tool and parameters) to align those RNA-Seq samples. To ensure our alignment pipeline is identical to GDC, we randomly selected 10 samples with GDC BAMs available to apply our pipeline and obtain their gene level read count. All selected samples had identical gene counts using GDC or our BAMs.

Copy number variant calling

Somatic copy number variants were called using GATK v4.1.9.0.[69] GDC's hg38 human reference genome was binned into target intervals using the PreprocessIntervals function with bin-length 1000bp and interval-merging-rule of OVERLAPPING_ONLY. A panel of normals (PON) was then generated using each normal sample as input and the GATK functions CollectReadCounts with –interval-merging-rule OVERLAPPING_ONLY followed by CreateReadCountPanelOfNormals with –minimum-interval-median-percentile 5.0. For tumor samples, reads that overlap the target interval were counted using the GATK function CollectReadCounts. Tumor Read counts were then standardized and denoised utilizing the GATK function DenoiseReadCounts with the PON specified by –count-panel-of-normals. Allelic counts for tumor were generated for variants present in the af-only-gnomad.hg38.vcf from the GATK best practices (variants further filtered to 0.2 > af > 0.01 and entries marked with 'PASS') using the GATK function CollectAllelicCounts. Segments were then modeled using GATK function ModelSegments with the denoised copy ratio and tumor allelic counts used as input. Copy ratios for the segments were then called on the segment regions using the GATK function CallCopyRatioSegments. Bedtools intersect was used to map copy number ratios from segments to genes and assign the called amplification or deletions.[68] For genes overlapping multiple segments a custom python script was utilized to call that gene as amplified, neutral, or deleted based on a weighted copy number ratio calculated from copy ratio of each segment overlapped and the length of the overlap as well as the z-score threshold used by the CallCopyRatioSegments function.

Somatic variant calling

Somatic variants were called from WES tumor and normal paired BAMs using somaticwrapper v1.6.1, a pipeline designed for detection of somatic variants from tumor and normal exome data. The pipeline merges and filters variant calls from four callers: Strelka v2.9.2,[63] VarScan v2.3.8,[64] Pindel v0.2.5,[65] and MuTect v1.1.7.[66] SNV calls were obtained from Strelka, Varscan, and Mutect. Indel calls were obtained from Stralka2, Varscan, and Pindel. The following filters were applied to get variant calls of high confidence:

- Normal VAF $\leq$ 0.02 and tumor VAF $\geq$ 0.05
- Rescue low-vaf variants in SMG gene list by removing tumor VAF $\geq$ 0.05 requirement
- Read depth in tumor $\geq$ 14 and normal $\geq$ 8
- Indel length < 100 bp
- All variants must be called by 2 or more callers
- All variants must be exonic
- Exclude variants in dbSNP but not in COSMIC

TERT promoter hotspot mutation calling

We used rcwrapper v1.0 to count reads in WGS tumor and blood normal BAMs at the known hotspot positions at hg38 chr5:1295113 and chr5:1295135. Rcwrapper is an automatic pipeline which uses bamreadcount for counting reads for reference and variant allele for each variant position, then parses the VAF information from it. We called a mutation if it was not observed in matching blood normal BAM and VAF > 5%.

DNA methylation microarray processing

Raw methylation idat files were downloaded from CPTAC DCC and GDC. Beta values of CpG loci were reported after functional normalization, quality check, common SNP filtering, and probe annotation using Li Ding Lab's methylation pipeline v1.1 https://github.com/ding-lab/cptac_methylation. Resulting beta values of methylation were used for downstream analysis.

Classification of MGMT promoter DNA methylation status

We applied the MGMT-STP27 model[111] to determine the MGMT promoter DNA methylation status, which is a logistic regression prediction model based on the M values of the two probes in the MGMT promoter region, cg12434587 and cg12981137. M-values were converted from the beta values of the processed microarray data by M = log2(beta / (1 - beta)). 7 tumors with low data quality of DNA methylation array were excluded from the prediction. 86 out of the remaining 209 tumors (41%) were predicted to be MGMT promoter DNA hypermethylated.

RNA quantification and analysis

RNA quantification

We obtained the gene-level readcount, Fragments Per Kilobase of transcript per Million mapped reads (FPKM) and FPKM Upper Quartile (FPKM-UQ) values by following the GDC's RNA-Seq pipeline (Expression mRNA Pipeline) https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/, with the exception of running the quantification tools in the stranded mode. We used HTSeq v0.11.2[70] to calculate the gene-level stranded readcount (parameters: -r pos -f bam -a 10 -s reverse -t exon -i gene_id -m intersection-nonempty –nonunique=none) using GENCODE v22 (Ensembl v79) annotation down- loaded from GDC (gencode.gene.info.v22.tsv). The readcount was then converted to FPKM and FPKM-UQ using the same formula described in GDC's Expression mRNA Pipeline documentation.

miRNA quantification

miRNA-Seq FASTQ files were downloaded from GDC. We reported the mature miRNA and precursor miRNA expression in TPM (Transcripts Per Million) after adapter trimming, quality check, alignment, annotation, reads counting using Li Ding Lab's miRNA pipe- line https://github.com/ding-lab/CPTAC_miRNA. The mature miRNA expression was calculated irrespective of its gene of origin by summing the expression from its precursor miRNAs.

Multiome snRNA-seq/snATAC-seq quantification and analysis

snRNA/snATAC-seq data preprocessing

For each sample, we obtained the unfiltered feature-barcode matrix per sample by passing the demultiplexed FASTQs to Cell Ranger Arc v2.0.0 'count' command using default parameters and a customized pre-mRNA GRCh38 genome reference was built to capture both exonic and intronic reads. The customized genome reference modified the transcript annotation from the 10x Genomics prebuilt human genome reference 3.0.0 (GRCh38 and Ensembl 93).

Seurat v4.1.1[80] and Signac v1.7.0 were used for all subsequent analyses. We constructed a Seurat object using the unfiltered feature-barcode matrix for each sample. A series of quality filters were applied to the data to remove those cell barcodes which fell into any one of these categories recommended by Seurat: too few total transcript counts (< 300); possible debris with too few genes expressed (< 200) and too few UMIs (< 1,000); possible more than one cell with too many genes expressed (> 10,000) and too many UMIs (> 10,000); possible dead cell or a sign of cellular stress and apoptosis with too high proportion of mitochondrial gene expression over the total transcript counts (> 10%). Quality-control filtering of the ATAC-seq measurements from the same nuclei was performed using functions from the Signac package. Filters that were applied for the cell calling include: 1,000 < number of fragments in peaks < 20,000; percentage of reads in peaks > 15; ENCODE blacklist regions percentage < 0.05 (https://www.encodeproject.org/annotations/ENCSR636HFF/); nucleosome banding pattern score < 5; and enrichment-score for Tn5-integration events at transcriptional start sites > 2.

Each sample was scaled and normalized using Seurat's 'SCTransform' function to correct for batch effects (with parameters: vars.to.regress = c("nCount_RNA", "percent.mito"), variable.features.n = 3000). We then merged all samples and repeated the same scaling and normalization method. All cells in the merged Seurat object were then clustered using the original Louvain algorithm and the top 30 PCA dimensions via Seurat's 'FindNeighbors' and 'FindClusters' (with default parameters: resolution = 0.8) functions. The resulting merged and normalized matrix was used for the subsequent analysis.

Peak calling for snATAC-seq data

To call peaks on multiome snRNA/snATAC-seq data, we followed the workflow previously described.[12] Specifically, we used the MACS2 tool (v.2.2.7.1) through the CallPeaks function of the Signac package (v.1.3.0, https://github.com/timoast/signac). We further removed peaks from the Y chromosome, as well as those overlapping genomic regions containing 'N'. All peaks were resized to 501 bp centered at the peak summit defined by MACS2. We next performed the iterative removal procedure described previously[112] to get the set of non-overlapping peaks. In brief, we start with retaining the most significant peak by MACS2 peak score (−log10[q]), removing all peaks that have direct overlap with it. We repeat this procedure for the remaining peaks, until we have the set of

non-overlapping peaks. The resulting sample peak set was used to calculate peak-count matrix using FeatureMatrix from the Signac package, which was also used for downstream analysis. Peaks overlapping with known gene regulatory elements such as promoters and enhancers annotated in the GeneHancer database[11] were identified with function findOverlaps (package IRanges v2.28.0, parameters: minoverlap=250).

Single-nuclei sequencing cell type annotation

Putative cell types were assigned to each cluster by manually reviewing the expression of marker genes. The marker genes used were TMEM119, P2RY12, SLC2A5, TGFBR1, GPR34, SALL1, GAS6, MERTK, C1QA, C3, PROS1, CD68, ADGRE1, AIF1, CX3CR1, TREM2, ITGAM, SPI1, CSF1R, LAPTM5, RGS1, PTPRC (Microglia); LGALS2, FCER1G, FCN1, CSTA, S100A8, S100A9, S100A12, LYZ, CD68, CD14 (Monocytes); AIF1, CD68, LST1, IFITM2 (Macrophages); Microglia, macrophages were named together as tumor-associated microglia/macrophages. CD8A, CD8B, CD3E, CD3D, PRF1, GZMA, GZMB, GZMK, GZMH, CD4, IL7R, LTB, LDHB, CD69, FAS, KLRG1, CD28, DPP4 (CD4/CD8 T-cells); CD19, CD79A, CD79B, MS4A1, SDC1, IGHG1, IGHG3, IGH4 (B-cells/Plasma); MBP, PLP1, CLDN11, MOG, KLK6, CNDP1, GJB1, MAG, NKX6-2, OPALIN, FOLH1, CARNS1, MOBP, ERMN, TMEM125, CNTN2, ENPP2, SH3GL3, MAL, TF, ST18, TPPP (Oligodendrocytes); PPP1R1B, CPNE6, NTSR2, GJB6, SLC39A12, GABRA2, WIF1, GABRG1, HHATL, C16orf89, ACSBG1, FBXO2, MMP28, SNCG, RANBP3L, IQCA1, SLC14A1 (Astrocytes); FSTL5, GAD2, GRIN1, SYNPR, GABRG2, DLX5, SULT4A1, RBFOX3, CALY, SLC6A17, SLC32A1, CCK, GABRA1, CDH9, DLX6-AS1, KCNC2, MIR7-3HG, FRMPD4, CAMKV, PCP4L1 (Neurons); EMCN, FLT1, PECAM1, KDR, PLVAP, PLVAP, TEK, VWF, ACTA2, ANGPT2, COL1A1, COL3A1, COL5A1, COL12A1, EMILIN1, LUM (Stroma). Malignant cells were further confirmed by copy number calling with inferCNV using putative non-malignant cell types as reference (parameters: analysis_mode='subclusters', cutoff=0.1, denoise=T, HMM=T).

snRNA-seq analysis

Differentially expressed genes between malignant cells from the primary and recurrent tumors from the same patient were identified by FindMarkers function. Samples with fewer than 50 malignant cells are excluded in the snRNA malignant cell comparisons. Patient C247230 was excluded for this reason. Wilcoxon statistical test was used. log2FC > 0.25 and FDR < 0.05 was used to filter DEGs. Average gene expression in each cell type was calculated with function AverageExpression (parameters: assay=''RNA'').

snATAC-seq malignant peak-expression link

Correlation between accessible chromatin regions (ACRs) and gene expression in malignant cells were calculate by LinkPeaks function. Default parameters (method = "pearson", distance = 5e+05, pvalue_cutoff = 0.05, score_cutoff = 0.05) were used.

snATAC-seq differential accessibility analysis

Differentially accessible chromatin regions (DACRs) within malignant cells were identified by FindMarkers function (parameters: test.use = 'LR', min.pct = 0.05, latent.vars = c('nCount_peaksinters','case_id')) comparing cells belonging to recurrent tumor to paired primary tumor from the same patient. Wilcoxon statistical test was used. FDR < 0.05 was used to filter DACRs. To avoid single-sample driven differences, we calculated ACR accessibility changes in each case and filtered out cohort-level DACRs with change in only 1 case.

snATAC-seq motif enrichment analysis

Motif enrichment analysis from filtered DACRs (FDR<0.05, shared changes in at least 2 cases) were conducted with function FindMotifs. To balance up- and down-regulated DACRs, top 150 DACRs by log2FC in each direction were used in the analysis. Reference motif position weight matrix was obtained from JASPAR2020 matrix (species 9606). Motif enrichment with FDR<0.05 (hypergeometric test with Benjamini-Hochberg multiple test correction) was considered significant.

Mapping mutations to snRNA data

We applied our in-house tool 10Xmapping. It utilized the cell and molecular barcode information from an snRNA bam file to identify reads supporting the reference and variant alleles covering the variant site in each individual cell. These variants were adopted from high-confidence WXS-based somatic mutations.

MS data interpretation

Quantification of TMT global proteomics data

LC-MS/MS analysis of the TMT11-labeled, bRPLC fractionated samples generated a total of 360 global proteomics data files. The Thermo RAW files were processed with mzRefinery to characterize and correct for any instrument calibration errors, and then with MS-GF+ v9881[55] to match against the RefSeq human protein sequence database downloaded on June 29, 2018 (hg38; 41,734 proteins), combined with 264 contaminants (e.g., trypsin, keratin). The partially tryptic search used a ± 10 ppm parent ion tolerance, allowed for isotopic error in precursor ion selection, and searched a decoy data- base composed of the forward and reversed protein sequences. MS-GF+ considered static carbamidomethylation (+57.0215 Da) on Cys residues and TMT modification (+229.1629 Da) on the peptide N-terminus and Lys residues, and dynamic oxidation (+15.9949 Da) on Met residues for searching the global proteome data. Peptide identification stringency was set at a maximum 1% FDR at peptide level using PepQValue < 0.005 and parent ion mass deviation < 7 ppm criteria. A minimum of 6 unique peptides per 1000 amino acids of protein length was then required for achieving 1% at the protein level within the full data set.

The intensities of all 11 TMT reporter ions were extracted using MASIC software (Monroe et al., 2008). Next, PSMs passing the confidence thresholds described above were linked to the extracted reporter ion intensities by scan number. The reporter ion intensities from different scans and different bRPLC fractions corresponding to the same gene were grouped. Relative protein abundance was calculated as the ratio of sample abundance to reference abundance using the summed reporter ion intensities from peptides

that could be uniquely mapped to a gene. The pooled reference sample was labeled with TMT 126 reagent, allowing comparison of relative protein abundances across different TMT-11 plexes. The relative abundances were log2 transformed and zero-centered for each gene to obtain final relative abundance values.

Small differences in laboratory conditions and sample handling can result in systematic, sample-specific bias in the quantification of protein levels. In order to mitigate these effects, we computed the median, log2 relative protein abundance for each sample and re-centered to achieve a common median of 0.

Quantification of phosphopeptides

Phosphopeptide identification for the 180 phosphoproteomics data files were performed as in the global proteome data analysis described above (e.g., peptide level FDR < 1%), with an additional dynamic phosphorylation (+79.9663 Da) on Ser, Thr, or Tyr residues. The phosphoproteome data were further processed by the Ascore algorithm[53] for phosphorylation site localization, and the top-scoring sequences were reported. For phosphoproteomic datasets, the TMT-11 quantitative data were not summarized by protein but left at the phosphopeptide level. All peptides (phosphopeptides and global peptides) were labeled with TMT-11 reagent simultaneously. Separation into phospho- and non-phosphopeptides using IMAC was performed after the labeling. Thus, all the biases upstream of labeling are assumed to be identical between global and phosphoproteomic datasets. Therefore, to account for sample-specific biases in the phosphoproteome analysis, we applied the correction factors derived from median- centering the global proteomic dataset.

Quantification of acetylated peptides

Acetylated peptide identification for the 60 acetylome data files were performed as in the global proteome data analysis described above, with additional dynamic acetylation (+42.0105 Da) and carbamylation (+43.0058 Da) on Lys residues. The acetylation site localization, protein inference, and quantification of the acetylome data were performed in identical fashion as in the phosphoproteome data.

SRM quantification

SRM data were analyzed using the Skyline software (Version 22.2).[107] The total peak area ratios of endogenous light peptides and their heavy isotope-labeled internal standards (i.e., L/H peak area ratios) were exported for quantitation. Peak detection and integration were carried out according to two criteria: (1) same retention time and (2) similar peak area ratios for the transitions. All data were manually inspected to ensure correct retention time, peak detection and accurate integration. All SRM results including the assay characterization data are organized as Skyline files on the Panorama server[113] and can be accessed via https://panoramaweb.org/CPTAC_GBM.url (the account for reviewer is: Email: panorama+reviewer154@proteinms.net; Password: eJvuMZMw).

Normalization of metabolome and lipidome

Global median centering, where each sample is normalized to the median of its observed values, was used to normalize all datasets. Protein quantification was accomplished via R-rollup,[74] in which peptides were scaled by a reference peptide and the protein abundance was set as the median of the scaled peptides.

Quantification of intact glycopeptides

The DDA raw files of intact glycopeptides were converted to mzXML format ProteoWizard 3.0 with the Peak Picking option selected for all MS levels prior to the database searching against a customized N-linked glycopeptide database containing over 30,000 known glycosite-containing peptide sequences of human species3 and a glycan database containing 253 glycan compositions (GlycomeDB, http://www.glycome-db.org/) via GPQuest.[82,89,114,115,122,123] The theoretical b-/y-ions of N-linked glycopeptides (targeted/decoy) in the customized database were calculated and used as fragment ion index during database search. MS/MS spectra were preprocessed in a series of procedures including spectrum de-noising, oxonium ion evaluation, and glycan composition prediction. For each preprocessed potential MS/MS spectrum of an intact glycopeptide, the top 100 peaks were matched to the fragment ion index of candidate peptides in the database. A candidate peptide was retained if at least 6 peaks were matched. Further evaluations were performed on the spectrum and the matched candidate peptides by assessing the isotopic peaks and fragment ions (b-/y-ions and Y-type ions) to compute Morpheus score, where the candidate peptide with the highest Morpheus score was assigned to that particular MS/MS spectrum. The glycan composition of the intact glycopeptide was determined by computing the mass difference between the peptide sequence and the precursor mass followed by finding the glycan matching that mass difference in the glycan database. PSMs with summed b-/y-ion intensity less than 20% of total intensity of the spectrum were filtered out. FDR at PSM level was set at < 1% to ensure the precise identification of the intact glycopeptides. The intact glycopeptides were quantified using the report ions of TMT-11plex. The missing values were imputed (only for intact glycopeptides quantified in > 50% of the samples) using DreamAI (https://github.com/WangLab-MSSM/DreamAI).

Preprocessing of harmonized protein/phosphoproteome/acetylome data

Log2 transformed sample-to-reference ratio of MS2 (log2-ratio) intensity was produced for 149 confirmatory samples and 110 discovery samples. Numbers of proteins/phosphosites/acetyl-sites detected in the study are listed in Figure 1.

To preprocess the log2-ratio intensity data matrix, we followed the procedure applied on the proteomics data in GBM discovery study.[2] To remove the technical variation among the sample distribution globally, we first performed sample median alignment to the log2-ratio intensity matrix.

We then applied an 'Inter-TMTplex T-test' to detect and remove outlier TMT plexes for each protein. For each TMT plex, we performed a t-test between the log2-ratio of samples inside the plex and the log2-ratio of samples outside the plex. After double log

transformation, p-values of 'Inter-TMTplex T-test' falling beyond 4-standard deviations from the median of the entire data sets were flagged as outliers, and corresponding TMT plexes were labeled with NA in the data set. A number of outlier TMT plexes removed from the data sets are listed in Table S7.

After the outlier removal, we removed the batch effect from technical variations across TMT plexes of the discovery and confirmatory data matrix separately. We applied a correction using the ComBat algorithm.[116] Due to the sample requirement across all batches in the ComBat algorithm, we performed KNN imputation on the data using an 'impute' R package.[117] After we corrected the imputed data using ComBat, we replaced the missing data structure from before KNN imputation. Additionally, we applied the ComBat method with an adjustment on sample groups: primary / recurrent / metastasis / normal adjacent tissue.

To further align the data sets from discovery and confirmatory samples, we applied additional correction to features detected in both studies to remove batch effects from 3 cohorts: the Longitudinal cohort, the Primary cohort, and the Primary cohort expanded. ComBat was used to remove the batch effect among these cohorts, with adjustment on the sample group differences among primary / recurrent / metastasis / normal adjacent tissue / GTEX normal.

After aligning the two studies, we applied DreamAI[93] (https://github.com/WangLab-MSSM/DreamAI) to impute missing values on the aligned data from discovery and confirmatory samples separately. Imputation was performed on features that appeared on at least 50% of samples in the corresponding data sets. The number of imputed features was listed in Table S7.

Alignment of harmonized RNA/miRNA/SRM data from discovery and confirmatory samples

To align the RNA/ miRNA/ SRM (Direct-SRM and IMAC-Ratio table) data from discovery and confirmatory samples, we applied the same procedure as for proteome alignment, removing the batch effect from the longitudinal, primary, and primary expanded cohorts. using the ComBat algorithm with adjustment on the sample group: primary / recurrent / metastasis / normal adjacent tissue / GTEX normal.

Other proteogenomic analysis

Sample labeling check across data types

While multiple omics data enhance our understanding of complex molecular mechanisms underlying GBM, it is sometimes inevitable to have sample errors including sample swapping, shifting, or data contamination. Working on error-containing data is dangerous since it could lead to a wrong scientific location. Therefore, it is required to confirm whether different types of molecular data are pertained from the same individuals prior to data integration or public sharing. For the GBM dataset, we checked sample labeling across different types of data with the SeqQEst pipeline (https://github.com/ding-lab/SeqQEst). We obtained the variant allele frequency for 1,391 germline target SNP markers from WES, WGS, RNA-seq, and snRNA-seq. The target SNP markers were selected from ~5 million common missense SNPs (dbNSFP v3.5a; https://sites.google.com/site/jpopgen/dbNSFP)[118] across chromosomes 1–22 in the human genome based on WES and RNA-Seq data testing. We then calculated the Pearson correlation between each sequencing pair both within one case and between different cases with R function cor (parameter: use="pairwise.complete.obs"). Correlation coefficients between different cases were used to build a null distribution. All correlation coefficients between different platforms within one case rejected the null hypothesis with $p = 0.000$, indicating no sample swaps detected.

Multiomics subtyping using non-negative matrix factorization (NMF)

Expression based TCGA subtyping. Gene expression based subtypes were based on the 150 genes created by Wang et al., the most recent TCGA subtyping effort,[119] which contained 50 highly expressed genes in classical, proneural, and mesenchymal GBM. Tumors with recurrent mutations in *IDH1* were assigned to be IDH-mutant tumors. We then performed consensus clustering on all tumors based on the selected gene expression in log2(FPKM-UQ + 1) using ConsensusClusterPlus R package (parameters: maxK = 10 reps = 2000 pItem = 0.8 pFeature = 1 clusterAlg = "hc" distance = "pearson" seed = 201909). We chose the total number of clusters k = 5 based on the delta area plot of consensus CDF. The clusters were annotated with the TCGA subtypes based on their gene expression profiles. Three clusters (r1, r4, and r5) were merged due to their similar expression signature, which was identical to the clustering result while choosing k = 3.

Mutation impact on the RNA, proteome, phosphoproteome, lipidome and metabolome

We focused our analyses on 12 GBM SMGs previously reported in the literature: PIK3R1, PIK3CA, PTEN, RB1, TP53, EGFR, IDH1, BRAF, NF1, PDGFRA, ATRX, and TERTp.[46] For each feature, we split samples with and without any genetic alterations (including copy number alterations and mutation), and compared expression levels (RNA, protein, phosphosite, lipid and metabolite), calculating a median difference in expression and testing for significance with the Wilcoxon rank sum test, with the Benjamini-Hochberg multiple test correction.

Clonal similarity analysis

We calculated a clonal similarity statistics for each recurrent tumor with its paired primary tumor using somatic mutation data with Clonality R package.[87] We first lifted genome coordinates of our readcount data from hg38 to hg19 with UCSC online.

Lift Genome Annotations tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). We then obtained background mutation frequency with function get.mutation.frequencies (parameter: tcga.cancer.type = "GBM"). Finally, we ran SNVtest for each tumor pair and calculated Maximum likelihood estimate of Ksi, a parameter of the likelihood representing clonality strength.

Primary/recurrent tumor differential analysis

We compared RNA transcript and protein abundance levels between the primary and recurrent tumors with the Wilcoxon rank sum test, with the Benjamini-Hochberg multiple test correction. All genes (RNA and protein separately) were then ranked by log2 fold

change and pathway enrichment scores were calculated in R with function GSEA from package clusterProfiler (parameters: pvalueCutoff=1). In this analysis, we used the 50 cancer hallmark pathway genesets.[120]

Mutational signature analysis

To infer mutational signatures from somatic variants, we used package signatureanalyzer (parameters: -n 100 –cosmic cosmic3_exome –objective poisson).[8]

Cell type enrichment analysis

The abundance of each cell type was inferred by the xCell package,[79] which performed the cell type enrichment analysis from gene expression data for 64 immune and stromal cell types. xCell is a gene signatures-based method learned from thousands of pure cell types from various sources. We input the FPKM-UQ expression matrix of this study in xCell using the expression levels ranking.

CODEX whole-slide quantification

Multiplex images were segmented using the Mesmer pretrained nuclei + membrane segmentation model in the Deepcell[121] cell segmentation library. The DAPI channel was used as the nuclei segmentation image, and CD45, CD8, CD31, CD4, HLA-DR, GFAP, CD68 channels were merged and used as the membrane segmentation image. Following segmentation, cells were classified with gating strategy (Endothelial cells: $CD31^+$, macrophages: IBA1+ and/or CD163+, malignant cells: OLIG2+ and/or GFAP+, CD4 T cell: $CD4^+$ and IBA1-, CD8 T cell: $CD8^+$ and IBA-, monocyte/granulocyte: CD11b+). To eliminate batch effects, marker thresholds were set manually for each image. Non-overlapping tiles of resolution 2k by 2k pixels (1mm by 1mm) were cropped from each CODEX sample. Number of cells, area of cells and cellular pixel fraction of positive staining for each marker were then quantified per tile to account for sections of different sizes.

Kinase-substrate regression analysis

For each kinase-substrate protein pair supported by previous experimental evidence (OmniPath, NetworKIN, DEPOD, and SIGNOR), we tested the associations between all sufficiently detected phosphosites on the substrate and the kinase. For a kinase-substrate pair to be tested, we required both kinase protein/phosphoprotein expression and phosphosite phosphorylation to be observed in at least 30 samples in the respective datasets and the overlapped dataset. We then applied the linear regression model using lm function in R to test for the relation between kinase and substrate phosphosite. For the i-th trial for kinase phosphosite abundance in the cis associations, kinase phosphosite abundance Ai depends on kinase protein expression Si and error Ei,

$$A_i = M_1 S_i + B + E_i$$

For the i-th trial for kinase phosphosite abundance in the trans associations, substrate phosphosite abundance Ai depends on kinase phosphosite expression Ki substrate protein expression Si and error Ei,

$$A_i = M_1 S_i + M_2 K_i + B + E_i$$

where the regression slope M coefficients are determined by least-square calculation. The resulting p-values were adjusted for multiple testing using the Benjamini-Hochberg procedure.

Survival analysis

We used R package "survival" to perform survival analysis. The Kaplan-Meier curve of overall survival was used to compare the prognosis among different groups (function survfit). Log-rank test (from the R package survminer) was used to test the differential survival outcomes between categorical variables. The standard multivariate Cox-proportional hazard modeling was applied to estimate the hazard ratio of HIF1A score status (function coxph). Age and gender, as the covariates, were included in the model.

Driver gene exclusivity analysis

For each possible pair of all 13 highly altered genes in this dataset, a 2x2 contingency table was constructed according to the number of primary GBM samples exhibiting or lacking somatic alteration in the two driver genes. Fisher tests were then conducted to determine if any two of the 13 highly altered driver genes in this cohort exhibited significant co-occurrence or mutual exclusivity. The resulting odds ratio (OR) and Benjamini & Hochberg FDR adjusted p-values were noted for each pair. Results are displayed in Figure 4A, which depicts the log2-transformed OR values and FDR adjusted p-values.

Cis impact of somatically altered driver genes

To investigate the effects of a somatically altered gene on its corresponding transcript, protein, and PTM, we calculated *cis*-effect scores at each level. We implemented a linear regression modeling the effects of a particular driver gene alteration status on the protein abundance, RNA expression, or PTM corresponding to the gene.

$$Y = \beta_0 + \beta_1 A_g + \beta_2 P + \epsilon$$

where Y denotes an (n x 1) vector containing either the protein abundance, the RNA expression, phosphoproteomic, acetylation, or glycosylation level at a residue pertaining to the driver gene, A is a binary vector indicating the driver alteration status (1 if a patient has a predicted oncogenic alteration in the gene, 0 if not) for that particular driver gene (g) in a tumor sample. Tumor purity (P) was also included as a covariate in the model. The error ϵ is assumed to be normally distributed with a constant variance σ.

After running the model for each of the 13 drivers and determining the associated correlation and Benjamini & Hochberg FDR adjusted p-values at each data level, the *cis*-effect score was determined by multiplying the -log10(p-adj) by the sign of the correlation coefficient between the altered driver and the effect on the omics. This retains the directional impact of the altered driver on the

target of interest, relative to samples lacking any somatic alteration for the cancer driver gene. The scores were capped at an absolute value of 10 for visualization purposes. The *cis*-effect results are shown in Figure 4B. In the case of the PTM results, we show the event with the highest absolute score for each gene.

Driver similarity estimated by trans effects on proteome, PTMs, and metabolome

To investigate the wider effects of somatic alterations on other proteins, PTMs, metabolites and lipids, we conducted a pair-wise *trans* analysis. We ran a linear regression for each cancer driver, similar to the process described above for the *cis*-effect score determination, but considering samples in which another driver lacked any somatic alteration before conducting the reverse assessment (Figure 4C). For each possible driver gene pair, all *trans* events were tested with the following model:

$$Y = \beta_0 + \beta_1 A_g + \beta_2 P + \epsilon$$

where Y is a (n x 1) vector representing the abundance of a given protein, phosphoprotein, glycoprotein, acetylated protein, metabolite or lipid. A denotes a binary vector indicating the driver alteration status for that particular driver gene (g) in a tumor sample, and the tumor purity (P) calculated by ESTIMATE[122] was also included as a covariate. Lastly, the error ϵ is assumed to be normally distributed with a constant variance σ. The samples tested required no somatic alteration present in the other driver in the pair. We tested only those events for which there was data available for more than 30 samples overall, of which more than 10 needed to correspond to samples harboring a genetic alteration in the driver of interest.

For each driver, scores were assigned to all corresponding *trans* events by multiplying the -log10(p-value) of the association with driver alteration by the sign of the coefficient, as was conducted for the *cis*-effect scores. To determine cancer drivers that associate with similar changes when altered, a ''driverness similarity'' calculation was conducted for each pair of drivers. This was done for the protein and PTM events, before doing the same just in the context of metabolomic effects. The scores of all the drivers' events were compared to determine a holistic correlation between the two drivers at the protein/PTM and metabolomic scale (Figure 4D). Driver pairs highly correlated at both the protein/PTM and lipid/metabolite levels are highlighted in Figure 4D.

Protein-protein interaction databases

We integrated PPIs from the following sources: STRING[47] (minimum score above 700), MINT,[90] BioGrid,[94] IntAct,[95] CORUM[92] and the unbiased experimental interactome.[123] For the purposes of the rest of the analyses, we only kept those interactions described in at least 3 different databases (n = 14,768).

Regression model estimating effect of gene alterations on protein-protein interactions

As aforementioned, we prioritized PPIs documented in multiple databases (see above). To statistically test whether somatic alterations may alter protein co-expression of the driver gene (g) with a described PPI partner in primary GBM, we used a linear regression model with a non-linear interaction term. This was investigated using the following regression model to learn the β coefficients:

$$Y = \beta_0 + \beta_1 X_g + \beta_2 A_g + \beta_3 (X_g * A_g) + \beta_4 P + \epsilon$$

where Y is a (n x 1) vector representing the protein abundance of the PPI partner, X is a vector representing the protein abundance of the driver gene of interest (g), A is a binary vector indicating the driver alteration status for that particular driver gene (g) in a tumor sample, (X*A) is the interaction term composed of the element-wise product of X and A, and the tumor purity (P) was also included as a covariate. Lastly, the error ϵ is assumed to be normally distributed with a constant variance σ. Data for both protein pairs were required in more than 30 patients to be tested.

The PPI affected by somatic alteration are highlighted in Figure 6C, displayed in a volcano plot of the coefficient and corresponding unadjusted p-value resulting from the interaction term in the model (X*A). All events with p-value less than 0.01 were considered significant and highlighted.

Protein-protein interactions in primary versus recurrent GBM

In addition to estimating how somatic alterations may influence PPI, it was also of interest how protein abundance correlations may persist or change between matched primary and recurrent samples. Unlike the somatic alteration analysis, we also wanted to maintain PPI that remained associated in recurrent GBM, so a linear regression model with a non-linear interaction term was not appropriate in this case. Instead, we ran a linear model to evaluate the baseline PPI separately in the matched primary and recurrent samples:

$$Y = \beta_0 + \beta_1 X_g + \beta_2 P + \epsilon$$

where Y is a (n x 1) vector representing the protein abundance of the PPI partner in either primary or recurrent samples, X is a vector representing the protein abundance of the other interacting protein in the pair (g), and the tumor purity (P) was also included as a covariate. Lastly, the error ϵ is assumed to be normally distributed with a constant variance σ.

We used the outputs from this linear model to prioritize protein pairs associated in recurrent but not primary samples (recurrent p value < 0.01, primary p-value > 0.01), those present in primary but not recurrent samples (recurrent p-value > 0.01, primary p-value < 0.01), and PPI associated in both primary and recurrent samples (recurrent p value < 0.01, primary p-value < 0.01). Given the prioritized list, we then ran Pearson correlations between the protein abundances of each protein pair separately in the matched primary and recurrent samples. We further refined only those PPI with recurrent Pearson p value > 0.1 and primary Pearson p-value < 0.01 (recurrent negative), recurrent Pearson p value < 0.01 and primary Pearson $p > 0.1$ (recurrent positive), and those with both recurrent and primary Pearson $p < 0.01$ (double positive). Finally, to more specifically evaluate cancer-relevant genes (Table S5), we filtered the results to require at least one cancer driver partner in the recurrent positive/negative results and both

protein partners to be drivers in the double positive group. The full results are summarized in Figure S6B, while a subset of the more striking results is displayed in Figure 6E.

Kinase Library enrichment analysis

Full description of the substrate specificities atlas of the Ser/Thr kinome can be found in.[4] The phosphorylation sites detected in this study were scored by all the characterized kinases (303 S/T kinases), and their ranks in the known phosphoproteome score distribution were determined as described previously[4] (percentile score). For every non-duplicate, singly phosphorylated site, kinases that ranked within the top-15 kinases for the S/T kinases were considered as biochemically predicted kinases for that phosphorylation site. Towards assessing the activity patterns of the kinome in differential expression data (gene alteration, *IDH1* mutation, and recurrent vs. primary tumors), we compared the percentage of phosphorylation sites for which each kinase was predicted among the upregulated/downregulated phosphorylation sites (FDR of 0.1 for *IDH1* mutation and recurrent vs. primary tumors, FDR of 0.5 for gene alteration analysis), versus the percentage of biochemically favored phosphorylation sites for that kinase within the set of unregulated sites (sites with FDR above the threshold mentioned above). The ratio between these two percentages is defined as the ''Frequency Factor'' of this kinase. For PTPN11 analysis, we compared the percentage of phosphorylation sites for which each kinase was predicted among the positively/negatively correlated phosphorylation sites (FDR of 0.1 for *IDH1* mutation and recurrent vs. primary tumors, FDR of 0.5 for gene alteration analysis), versus the percentage of biochemically favored phosphorylation sites for that kinase within the set of uncorrelated sites (sites with FDR above the threshold mentioned above). Contingency tables were corrected using Haldane correction (adding 0.5 to the cases with zero in one of the counts). Statistical significance was determined using one-sided Fisher's exact test, and the corresponding *p* values were adjusted using the Benjamini-Hochberg procedure. Then, for every kinase, the most significant enrichment side (upregulated or downregulated) was selected based on the adjusted p-value and presented in the volcano plots and bubblemaps. In the volcano plots, kinases that were significant (adjusted *p* value <= 0.1) for both upregulated and downregulated analysis were excluded from downstream analysis. Bubblemaps were generated with size and color strength representing the adjusted p-values and frequency factors respectively, only displaying significant kinases (adjusted *p* value <= 0.1).

Mapping mutation, phosphorylation, and glycosylation events onto the EGFR protein

A previously published active dimer EGFR structure was used to visualize the 3-dimensional protein.[30] The most common EGFR point mutations in the cohort were selected for mapping: A289 and G598 (A265 and G574 in the structure, respectively). In addition, the top EGFR phosphorylation site implicated by the *cis* analysis, Y316 (Y292 on structure) was also mapped. Two glycosylation sites were selected for mapping onto the structure: N352 DSLSINATNIK (amino acids 323-333 on structure), and N603 TCPAGVMGENNTLVWK (amino acids 570-585 on structure). To ensure that the amino acid coordinates in this data were correctly mapped onto the corresponding coordinates in the structure, each event was matched by the sequence to which it corresponded (for point mutations, the surrounding sequence was used to map the particular amino acid affected). To assess the proximity to ligand binding sites, the N352 sequence was mapped onto PBD structures 1IVO for EGF and 6ARU for Cetuximab. All rendering and selection were done using UCSF Chimera software.[31] The final structures are depicted in Figures 5G and S5G.

Glycoproteomic characterization of grade 4 astrocytomas with IDH-mutation or IDH WT

Glycoproteomic (intact glycopeptides, IGPs) and global proteomic data were used to conduct pairwise differential analysis between primary tumors and normal tissues as well as between primary tumors and recurrent tumors. Median log2 fold changes were computed and Wilcoxon rank-sum test was performed that the p-values were adjusted via Benjamini-Hochberg method. We performed KEGG pathway and GO (biological processes) enrichment analysis via WebGestalt.[83] Upregulated/downregulated IGPs/glycoproteins referred to > 1.5-fold increase/decrease with FDR <0.05. A Sankey diagram was used to show differences in glycan types and biological processes between primary and recurrent grade 4 astrocytomas. The thickness of lines from primary/recurrent to the glycan types is proportional to the numbers of overexpressed IGPs (relative to primary or recurrent, depending on which one is considered). The thickness of lines from the glycan types to biological processes represents numbers of glycoproteins mapped to a particular biological process, listed on the right. The abundances of IGPs and phosphopeptides of EGFR were subjected to linear regression analysis using lm function in R (global protein expression was considered in the models as well) to determine interplays between glycosylation and phosphorylation in EGFR-altered tumors and EGFR-wt tumors. Receiver operating characteristic (ROC) curves were generated using pROC[84] based on the predictive models of glycoproteins and IGPs that were potential biomarkers of recurrent grade 4 astrocytoma compared to primary tumors. In brief, for each glyco-signature (glycoprotein or IGP) or a panel of multiple glyco-signatures, its discriminatory power through logistic regression was evaluated using ROC analysis. The data were log-transformed followed by z-score prior to ROC analysis. To ensure statistical stability of the results, we used bootstrap resampling (n=500) of the data to construct and evaluate the predictive model of a glyco-signature/panel. The mean ROC curves were depicted based on bootstrap resampling results and an area under the curve (AUC) was computed for the mean ROC curve. Any glyco-signature with AUC > 0.5 demonstrated its potential as a marker for recurrent grade 4 astrocytoma. IGPs (primary tumors only) with CVs in the > 25% quartile used to derive glycoproteomic subtypes via CancerSubtypes.[86] Specifically, 80% of the original sample pool was randomly subsampled without replacement and partitioned into three major clusters (glyco) using hierarchical clustering, which was repeated 2000 times. The IGPs were grouped into three clusters (IPC 1 to 3) using unsupervised hierarchical clustering in ComplexHeatmap.[85] Enrichment scores between protein subtypes in relation to the glycoproteomic subtypes were calculated according to the hypergeometric distribution. A hypergeometric test p-value<0.05 was required for the consideration of overlap between different omic subtypes.

Histopathology image analysis

Digital histopathology slides of samples used in proteomics analysis were downloaded from the CPTAC-GBM collection[99] in The Cancer Imaging Archive (TCIA).[124] Histopathological reviews were done on all samples for percent total cellularity, percent tumor nuclei and percent necrotic surface area. Histopathologically defined adult glioblastoma tumors, astrocytoma that progressed to glioblastoma and brain metastases were considered for analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

All data analyses were conducted in the R environment. Details of specific functions and libraries are provided in the STAR Methods sections. Statistical significance is determined by Wilcoxon rank-sum test, Fisher's exact test, hypergeometric test and Pearson correlation test. p values < 0.05 is considered statistically significant. Details of statistical tests are provided in the figure legends and the STAR Methods sections.