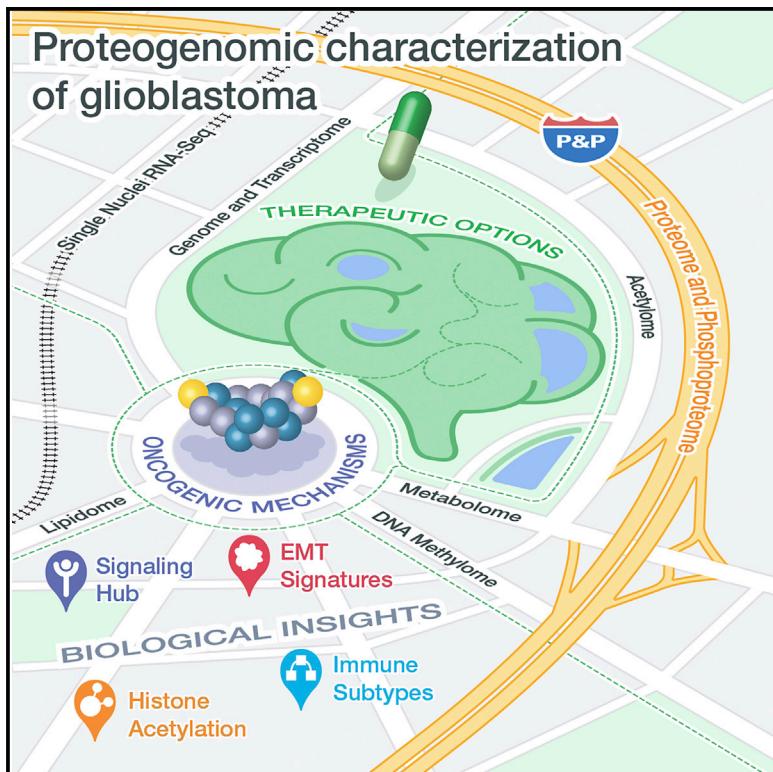


Proteogenomic and metabolomic characterization of human glioblastoma

Graphical Abstract



Authors

Liang-Bo Wang, Alla Karpova, Marina A. Gritsenko, ..., Tao Liu, Li Ding, Clinical Proteomic Tumor Analysis Consortium

Correspondence

karin.rodland@pnnl.gov (K.D.R.), tao.liu@pnnl.gov (T.L.), lding@wustl.edu (L.D.)

In Brief

Wang et al. perform integrated proteogenomic analysis of adult glioblastoma (GBM), including metabolomics, lipidomics, and single nuclei RNA-Seq, revealing insights into the immune landscape of GBM, cell-specific nature of EMT signatures, histone acetylation in classical GBMs, and the existence of signaling hubs which could provide therapeutic vulnerabilities.

Highlights

- Phosphorylated PTPN11 and PLCG1 represent a signaling hub in RTK-altered tumors
- Four immune GBM subtypes exist, characterized by distinct immune cell populations
- Mesenchymal subtype EMT signature is specific to tumor cells but not to stroma
- Histone H2B acetylation is enriched in classical GBMs with low macrophage content



Article

Proteogenomic and metabolomic characterization of human glioblastoma

Liang-Bo Wang,^{1,2,48} Alla Karpova,^{1,2,48} Marina A. Gritsenko,^{3,48} Jennifer E. Kyle,^{3,48} Song Cao,^{1,2,48} Yize Li,^{1,2,48} Dmitry Rykunov,^{4,49} Antonio Colaprico,^{5,6,49} Joseph H. Rothstein,^{4,49} Runyu Hong,^{7,8,49} Vasileios Stathias,^{5,9,10,49} MacIntosh Cornwell,^{7,11,49} Francesca Petralia,^{4,49} Yige Wu,^{1,2} Boris Reva,⁴ Karsten Krug,¹² Pietro Pugliese,¹³ Emily Kawaler,^{7,8} Lindsey K. Olsen,¹⁴ Wen-Wei Liang,^{1,2} Xiaoyu Song,^{15,16} Yongchao Dou,^{17,18} Michael C. Wendl,^{2,19,20} Wagma Caravan,^{1,2} Wenke Liu,^{7,8} Daniel Cui Zhou,^{1,2} Jiayi Ji,^{15,16} Chia-Feng Tsai,³ Vladislav A. Petyuk,³ Jamie Moon,³ Weiping Ma,⁴ Rosalie K. Chu,³ Karl K. Weitz,³ Ronald J. Moore,³ Matthew E. Monroe,³ Rui Zhao,³

(Author list continued on next page)

¹Department of Medicine, Washington University in St. Louis, St. Louis, MO 63130, USA

²McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63130, USA

³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

⁴Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁵Sylvester Comprehensive Cancer Center, University of Miami, FL 33136, USA

⁶Division of Biostatistics, Department of Public Health Science, University of Miami, FL 33136, USA

⁷Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY 10016, USA

⁸Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY 10016, USA

⁹Department of Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

¹⁰BD2K-LINCS Data Coordination and Integration Center, Miami, FL 33136, USA

¹¹Department of Medicine, NYU Grossman School of Medicine, New York, NY 10016, USA

¹²Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

¹³Department of Science and Technology, University of Sannio, 82100, Benevento, Italy

¹⁴Department of Biology, Brigham Young University, Provo, UT 84602, USA

¹⁵Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

¹⁶Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

¹⁷Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA

¹⁸Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹⁹Department of Genetics, Washington University in St. Louis, St. Louis, MO 63130, USA

²⁰Department of Mathematics, Washington University in St. Louis, St. Louis, MO 63130, USA

²¹Department of Neurology, Northwell Health System, Lake Success, NY 11042 USA

(Affiliations continued on next page)

SUMMARY

Glioblastoma (GBM) is the most aggressive nervous system cancer. Understanding its molecular pathogenesis is crucial to improving diagnosis and treatment. Integrated analysis of genomic, proteomic, post-translational modification and metabolomic data on 99 treatment-naive GBMs provides insights to GBM biology. We identify key phosphorylation events (e.g., phosphorylated PTPN11 and PLCG1) as potential switches mediating oncogenic pathway activation, as well as potential targets for EGFR-, TP53-, and RB1-altered tumors. Immune subtypes with distinct immune cell types are discovered using bulk omics methodologies, validated by snRNA-seq, and correlated with specific expression and histone acetylation patterns. Histone H2B acetylation in classical-like and immune-low GBM is driven largely by BRDs, CREBBP, and EP300. Integrated metabolomic and proteomic data identify specific lipid distributions across subtypes and distinct global metabolic changes in *IDH*-mutated tumors. This work highlights biological relationships that could contribute to stratification of GBM patients for more effective treatment.

INTRODUCTION

Glioblastoma (GBM) is the most common primary malignant brain tumor, with roughly 12,000 new cases annually in the United

States and median survival under 2 years (Delgado-López and Corrales-García, 2016; Ostrom et al., 2019). The Cancer Genome Atlas (TCGA) (Brennan et al., 2013; The Cancer Genome Atlas Research Network, 2008) and other studies (Yan et al., 2009)



Xiaolu Yang,^{1,27} Seungyeul Yoo,⁴ Azra Krek,⁴ Alexis Demopoulos,²¹ Houxiang Zhu,^{1,2} Matthew A. Wyczalkowski,^{1,2} Joshua F. McMichael,^{1,2} Brittany L. Henderson,¹⁴ Caleb M. Lindgren,¹⁴ Hannah Boekweg,¹⁴ Shuangjia Lu,^{1,2} Jessika Baral,^{1,2} Lijun Yao,^{1,2} Kelly G. Stratton,³ Lisa M. Bramer,³ Erika Zink,³ Sneha P. Couvillion,³ Kent J. Bloodsworth,³ Shankha Satpathy,¹² Weiva Sieh,^{4,15,16} Simina M. Boca,²² Stephan Schürer,^{5,9,10,23} Feng Chen,^{1,24,25} Maciej Wiznerowicz,^{26,27} Karen A. Ketchum,²⁸ Emily S. Boja,²⁹ Christopher R. Kinsinger,²⁹ Ana I. Robles,²⁹ Tara Hiltke,²⁹ Mathangi Thiagarajan,³⁰ Alexey I. Nesvizhskii,^{31,32} Bing Zhang,^{17,18} D.R. Mani,¹² Michele Ceccarelli,^{33,47} Xi S. Chen,^{5,6} Sandra L. Cottingham,³⁴ Qing Kay Li,³⁵ Albert H. Kim,³⁶ David Fenyö,^{7,8} Kelly V. Ruggles,^{7,11} Henry Rodriguez,²⁹ Mehdi Mesri,²⁹ Samuel H. Payne,¹⁴ Adam C. Resnick,^{37,38} Pei Wang,⁴ Richard D. Smith,³ Antonio Iavarone,^{39,40,41,42} Milan G. Chheda,^{1,25,43} Jill S. Barnholtz-Sloan,^{44,45} Karin D. Rodland,^{3,46,*} Tao Liu,^{3,*} and Li Ding^{1,2,19,25,50,*} Clinical Proteomic Tumor Analysis Consortium

²²Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington, DC 20007, USA

²³Institute for Data Science & Computing, University of Miami, FL 33136, USA

²⁴Department of Cell Biology and Physiology, Washington University in St. Louis, St. Louis, MO 63130, USA

²⁵Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63130, USA

²⁶International Institute for Molecular Oncology, 60-203 Poznań, Poland

²⁷Poznań University of Medical Sciences, 61-701 Poznań, Poland

²⁸ESAC, Inc., Rockville, MD 20850, USA

²⁹Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD 20892, USA

³⁰Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

³¹Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA

³²Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

³³Department of Electrical Engineering and Information Technology, University of Naples "Federico II", 80128, Naples, Italy

³⁴Department of Pathology, Spectrum Health and Helen DeVos Children's Hospital, Grand Rapids, MI 49503, USA

³⁵Department of Pathology, Johns Hopkins University, Baltimore, MD 21231, USA

³⁶Department of Neurological Surgery, Washington University in St. Louis, St. Louis, MO 63130, USA

³⁷Center for Data Driven Discovery in Biomedicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

³⁸Division of Neurosurgery, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

³⁹Institute for Cancer Genetics, Columbia University Medical Center, New York, NY 10032, USA

⁴⁰Department of Neurology, Columbia University Medical Center, New York, NY 10032, USA

⁴¹Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA

⁴²Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA

⁴³Department of Neurology, Washington University in St. Louis, St. Louis, MO 63130, USA

⁴⁴Case Comprehensive Cancer Center and Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

⁴⁵Research and Education, University Hospitals Health System, Cleveland, OH 44106, USA

⁴⁶Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA

⁴⁷BIOGEM, 83031 Ariano Irpino, Italy

⁴⁸These authors contributed equally

⁴⁹These authors contributed equally

⁵⁰Lead contact

*Correspondence: karin.rodland@pnnl.gov (K.D.R.), tao.liu@pnnl.gov (T.L.), lding@wustl.edu (L.D.)

<https://doi.org/10.1016/j.ccr.2021.01.006>

have reshaped the World Health Organization classification of nervous system tumors (Louis et al., 2016) to include molecular features (Brat et al., 2018; Louis et al., 2017). GBM is categorized as either *IDH*-wild type (*IDH*-WT; ~90%) or *IDH*-mutant (~10%). *IDH*-WT GBMs fall into three distinct subclasses (proneural, classical, and mesenchymal) based on genomic alterations and gene expression signatures (Verhaak et al., 2010; Wang et al., 2017). Methylome-based classification is being used to differentially diagnose brain tumors (Karimi et al., 2019; Nassiri et al., 2019) and may become clinically useful for GBM.

Surgical resection, chemotherapy, and radiotherapy remain the standard of care (Stupp et al., 2005; Perry et al., 2017), with the recent addition of tumor treating fields (Stupp et al., 2017). Promising immunotherapies have been proposed, including immune checkpoint inhibitors, vaccines, chimeric antigen receptor T cell (CAR-T) therapy, and viral therapy, though none have cleared Phase III trials (Lim et al., 2018; McGranahan et al., 2019). Despite different subtypes, no specific treatment works more effectively in

a pre-specified subset of patients based on transcriptomics, though those with MGMT promoter methylation respond better to temozolomide (Stupp et al., 2005).

Here, we integrated proteogenomic and metabolomic data from 10 platforms including whole genome sequencing (WGS), whole exome sequencing (WES), RNA sequencing (RNA-seq), microRNA-seq (miRNA-seq), single nuclei RNA-seq (snRNA-seq), DNA methylation arrays, proteome, phosphoproteome, acetylome, lipidome, and metabolome to investigate 99 treatment-naïve GBMs prospectively collected by the Clinical Proteomic Tumor Analysis Consortium (CPTAC). We report new immune-based subtypes, activation of DNA repair pathways via upregulated phosphosite levels of DNA repair genes in *TP53*-mutated tumors, an apparent phospho-signaling bottleneck in receptor tyrosine kinase (RTK)-altered tumors, and enrichment of histone H2B acetylation and low macrophage content in classical-like GBM tumors. We used single-cell data to investigate contributions of various cell types to

bulk tumor signatures and analyzed the mesenchymal subtype to discern epithelial-mesenchymal transition (EMT) signatures in tumor and infiltrating immune cells. The data presented here furnish a resource for future GBM studies.

RESULTS

Proteogenomic and metabolomic features delineate molecular subtypes of glioblastoma

We characterized the proteogenomic landscape of 99 GBMs and 10 unmatched GTEx normal brain samples. This cohort has diverse origins and clinical characteristics typical of adult GBM (Table S1). Six cases harbored *IDH1* R132H mutations and had earlier disease onset than those with *IDH1*-WT (median 47 vs. 59 years, t test p = 0.055). We detected one additional non-hotspot *IDH1* mutation (R222C).

All samples were homogenized and aliquoted for each of the ten different omics assays (Figures 1A and S1A–S1C; STAR methods). Mass spectrometry (MS) quantified protein, phosphorylation and acetylation, as previously described (Dou et al., 2020; Mertins et al., 2016) (Figures S1A–S1C). Metabolome and lipidome levels were respectively measured by label-free gas and liquid chromatography coupled to MS.

Genomic properties of our cohort were comparable to those of TCGA GBM cohort (Brennan et al., 2013) (Figure 1B). We identified many structural variants (SV) in oncogenes, including *EGFR* and *PDGFRA*, and tumor suppressors *PTEN* and *NF1*. *EGFR* mutations often co-occurred with *EGFR* SV and amplification events (p < 0.01). WES and WGS identified *TERT* promoter (*TERT*p) mutations with variant allele frequency (VAF) >5% (Figure 1B). Copy number analysis identified common focal and arm-level copy number variations (CNVs) (Figure S2A).

We added protein and phosphosite abundance to prior clustering studies of gene expression (Figures 1C and S2B–S2F; STAR methods). While our results are concordant with TCGA expression-based classification (Wang et al., 2017), 27 tumors (29%) were classified as a different subtype (Figures 1C and S2B). Based on similarities with gene expression subtypes, we designated the three clusters observed in *IDH1*-WT tumors as nmf1 (proneural-like; n = 29), nmf2 (mesenchymal-like; n = 37), and nmf3 (classical-like; n = 26). Pathway enrichment analysis of RNA, protein, and phosphosite abundances indicated that nmf1 was enriched for synaptic vesicle cycle and neurotransmission transport; nmf2 was enriched for innate immune response, including neutrophil degranulation, phagocytosis, and extracellular matrix organization; and nmf3 was enriched for mRNA splicing and RNA metabolism. Based on the known functional effects of protein acetylation (Narita et al., 2019), nmf1/proneural-like cluster had a higher abundance of acetylated proteins involved in the TCA cycle and metabolism of amino acids, whereas the nmf2/mesenchymal-like cluster was enriched for innate immune system activation, peroxisomal protein import and glycolysis. The nmf3/classical-like subtype was enriched for acetylation of chromatin modifiers and DNA repair proteins.

Clinical data associated with the three subtypes indicated that tumors with relatively low multi-omics membership scores for two or more subtypes, i.e. those of “mixed subtype” (n = 12) (Figure S2B) were associated with worse prognosis (log rank

test p = 1.7e-4; Figure S2C) compared with those of non-mixed subtype (excluding *IDH1*-mutant tumors). We identified three proteins associated with poor survival across all tumors: low expression of HIST3H2BB (log rank test p = 0.0034), high expression of MT-CYB (p = 0.03), and high expression of PRODH (p = 0.096). Genome-wide DNA methylation profiling identified six DNA methylation subtypes, including two distinct glioma CpG island methylator phenotype (G-CIMP) subtypes (dm2 and dm6). The dm6 subtype is *IDH*-mutant-specific with upregulation of chromatin organization pathways, while dm2 consists of *IDH1*-WT tumors with upregulation of transcription and mRNA splicing pathways (Figure S2D). Two subtypes showed elevated expression of different *de novo* DNA methylases (Figure S2E). We also examined *cis* associations of DNA methylation with RNA or protein abundances using iProFun (Song et al., 2019) (Table S3). For example, 38 of 90 tumors (42%) exhibited hypermethylation of the *MGMT* promoter region and significantly decreased *MGMT* RNA and protein levels (Welch's t test p = 4.9e-11 and 2.6e-6, respectively) (Figure S2G; Table S2).

Driver genetic alterations influence oncogenic protein abundance and phosphorylation

We associated genetic alterations (mutations, CNVs, fusions, and SVs) with RNA, protein expression and phosphorylation levels, observing 95 *cis-trans* phosphorylation events (Figures 2A and 2B). We found strong *cis* effects for *EGFR* and *PDGFRA*, with significant increases in RNA and protein expression and increased phosphorylation at S1166 and S1067/S1070, respectively. At the *trans* level, tumors with *EGFR* alterations presented elevated CTNNB1 (β -catenin) protein despite decreased mRNA, and increased phosphorylation of both PTPN11 (Shp2) at Y62 and PLCG1 at Y783 (Figures 2A and 2B). These observations illustrate the importance of protein measurements to study pathway activation.

Tumor suppressors *RB1*, *NF1*, *PTEN*, and *ATRX* demonstrated good concordance between genetic alterations and decreased RNA, protein, and phosphorylation levels of their respective gene products. Although the general effect of *TP53* mutations on increased protein stability is known, we identified specific phosphosites that correlate with increased stability. Phosphorylation of TP53 at S315 and TP53BP1 at S1099, S1106, and S1109 correlated with increased TP53 protein expression (Pearson r = 0.89 and 0.53, respectively) (Figures 2A, 2B, and S3A). TP53 alterations were largely missense mutations (Figure 2C) within its DNA binding domain (Figure S3B), including hotspot sites and several rare variants. TP53 protein expression was consistently elevated relative to *TP53* WT GBMs, an effect not seen at the RNA level. RNA-seq read alignment demonstrated that a recurrent splice-site mutation, X126_splice, results in an alternative splice site, producing an in-frame 7aa truncation in exon 5 (Figures 2C and S3C).

We assessed kinases known to phosphorylate TP53 and its downstream targets. In *TP53* mutants (Figures 2A, 2B, and 2D), we detected elevated protein and/or phosphorylation in ATR, MAPK3, CDK2, and CDK9, while *MDM2* was decreased at both RNA and protein levels. Tumors with *TP53* mutations showed upregulated phosphosites, but not increased protein levels, of DNA repair genes (Figure 2D), suggesting specific phosphosite regulation.



Figure 1. Proteogenomic summary of the cohort

(A) Summary of 10 data types generated in this study.

(B) Overview of significantly altered genes found in at least 5% of samples, showing tumor mutation burden (\log_2 WES mutation count) and structural variants, fusions, and CNVs. Subtypes are based on results in panel (C).

(C) Multi-omics clustering of tumor samples by NMF using CNV, expression, and protein and phosphoprotein abundances. Heatmaps show differential expression between subtypes, including DNA methylation, acetylome, metabolome, and lipidome, and characteristic features for each subtype. Pathway enrichment analysis highlights differences between subtypes. Neuron activity related pathways, immune response pathways, and cell cycle pathways were respectively enriched in the nmf1 (proneural-like), nmf2 (mesenchymal-like), and nmf3 (classical-like) subtypes.

See also Figures S1 and S2, and Tables S1, S2, and S3.

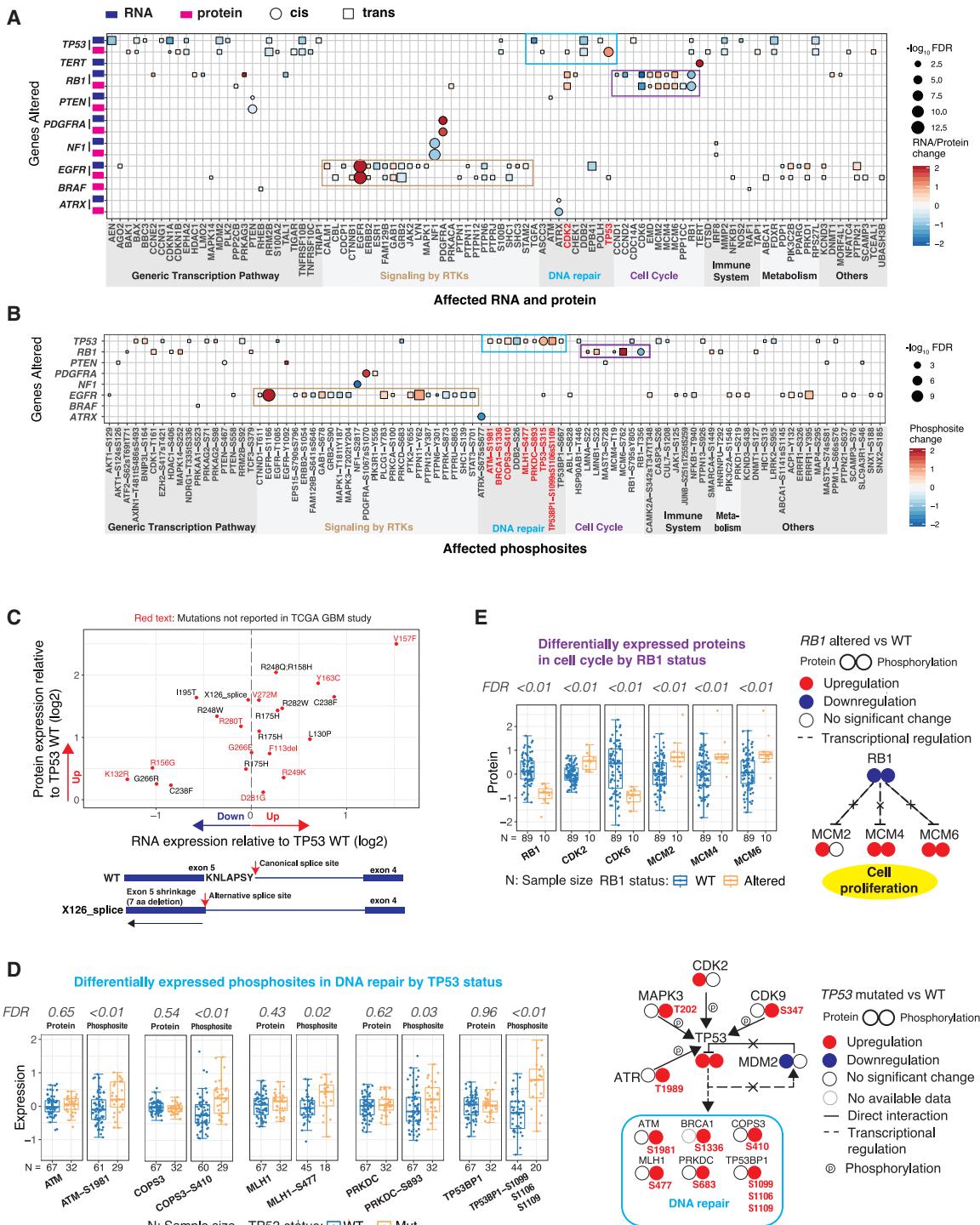


Figure 2. Cis and trans effects of SMGs and effects of TP53 regulations on DNA repair genes and RB1 on cell cycle genes

(A) Cis and trans effects of significantly mutated genes (y axis) on RNA and protein level (x axis) showing that effects are often similar.

(B) Cis and trans effects of significantly mutated genes (y axis) on protein phosphorylation status (x axis).

(C) Comparison of RNA and protein expression in TP53-mutated versus WT samples. Bottom: alternative TP53 splice site for the X126 mutation.

(D) Differentially expressed proteins and phosphoproteins in DNA repair genes for TP53-mutant (n = 32) versus TP53 WT (n = 67) samples. Right: a schematic of differences in expression and phosphorylation in the context of known pathway regulation.

(E) RB1 alterations associated with protein expression of CDK2, CDK6, MCM2, MCM4, and RB1. Right: A schematic of proposed interplay among RB1, MCM2, MCM4, and MCM6 in RB1-altered (n = 10) and WT (n = 89) samples.

See also Figure S3.

We observed negative feedback between *RB1* and downstream targets, *CDK2*, *CDK6*, *MCM2*, *MCM4*, and *MCM6*, while *NF1* had similar effects on *IRF8* (Figures 2A and 2B). *RB1*-altered samples (12% of the cohort) showed significantly downregulated *RB1* and upregulated *MCM2*, *MCM4*, and *MCM6* protein expression (Figure 2E). In addition, in samples with *NF1* alterations, we observed upregulation of protein and RNA of *IRF8*, a transcription factor that controls microglial motility (Masuda et al., 2014) (Figures 2A and 2B).

GBMs exploit various telomere lengthening mechanisms. WGS data identified *TERTp* hotspot mutations in 74% of primary GBMs (NM_198253.2:c.-124C>T [C228T] and c.-146C>T [C250T]) (Killela et al., 2013), resulting in increased *TERT* RNA expression (Figures S3D and S3E). We also found that *ATRX* mutations were mutually exclusive with *TERTp* hotspot mutations and co-occurred with *TP53* and *IDH1* R132H mutations (Figure S3F). Nine *ATRX* mutants had significantly diminished *ATRX* RNA and protein levels (Figure S3G). Immunohistochemistry (IHC) staining confirmed the loss of *ATRX* protein in tumor cells in *ATRX*-WT *IDH1* R132H mutant tumors (Figure S3H). Despite *ATRX* loss, expression of its complex partners was not affected at RNA or protein levels, raising questions about the function of the complex in the absence of *ATRX* protein (Figures S3I and S3J).

RTK signaling cascades are activated in GBM

Genomic loci associated with RTKs, such as EGFR, PDGFRA, and MET, are frequently amplified in GBM (Brennan et al., 2013). We identified 45 tumors with *EGFR* SVs, all having copy number amplifications, suggesting high concordance between SV and CNV (Figures 3A and S4A). All tumors with mutated *EGFR* and SV have correspondingly high RNA, protein, and Y1172 phosphorylation levels, indicating EGFR pathway activation.

In *EGFR*-altered samples, high EGFR autophosphorylation was observed, along with increased abundance and phosphorylation of pleckstrin homology-like domain family A member proteins (PHLDA1 and PHLDA3), transcription factor SOX9, cell adhesion protein CTNND2 (δ -catenin), and cell cycle proteins CDK6 and CDKN2C (Figures 3B and 3C). *PHLDA1* and *PHLDA3* were not differentially expressed in RNA, consistent with activation of downstream EGFR pathway components (Ling et al., 2011; Liu et al., 2019). SOX9 expression was validated in five of six tumors (three *EGFR*-altered and three *EGFR*-WT) using IHC (Figure S4D).

We also observed increased phosphorylation levels of PTPN11-Y62, PLCG1-Y783, RB1-S795Y805, MAP3K1-S1408, and specific EGFR sites in *EGFR*-altered samples (Figure 3B, bottom). Notably, the total protein level of PTPN11 was comparable between the two groups, suggesting its activity is regulated primarily by phosphorylation (Figure 3C). A similar pattern is observed with PLCG1 (PLC γ 1), where Y783 phosphorylation was significantly higher in *EGFR*-altered versus *EGFR*-WT samples (Wilcox false discovery rate [FDR] < 0.01; Figure S4C), despite no significant difference in PLC γ 1 protein expression (FDR = 0.11). Since phosphorylation of PLCG1 on Y783 is activating (Poulin et al., 2005), this could provide a mechanism for EGFR activation of PLCG1's known effects on proliferation, migration, and invasiveness (Kunze et al., 2014).

We performed a kinase-substrate study for EGFR and PDGFRA and identified high levels of GAB1 phosphorylation at

Y689 and Y657, consistent with high EGFR expression. In addition, PTPN11 phosphosites at Y546 and Y584 were associated with high PDGFRA expression (Figure 3D) and have been observed in lung cancers with ALK fusions (Voena et al., 2007). Activation of PTPN11 through either EGFR- or PDGFRA-related phosphorylation in GBM suggests it may represent a shared RTK signaling hub. PTPN11, GAB1, and GRB2 form a complex and are co-regulated by RTKs to activate the RAS pathway (Montagner et al., 2005). Figures 2A and S4C show that EGFR activation status is associated with upregulated GAB1 and downregulated GRB2 protein expression. We validated the elevated GAB1 expression in *EGFR*-altered tumors using IHC (Figure S4D).

Distinct immune marker expression and epigenetic events characterize GBM immune subtypes

We generated cell-type immune enrichment scores using single-sample GSEA by xCell (Aran et al., 2017), finding four distinct immune-based GBM subtypes (Figure 4A; STAR methods). Immune subtype 1 (im1) showed overall higher scores (Figure S5A), including elevated levels of microglia, macrophages, and lymphocytes. Immune subtypes 2 and 3 (im2 and im3) displayed reciprocal ratios of macrophages and lymphocytes, with im2 higher in macrophages and lower in lymphocytes and im3 having higher neuron score (Figures 4A and S5A). Immune subtype 4 (im4) is distinct from the others, with substantially lower enrichment for all immune cell types (Figures 4A and S5A). The mesenchymal subtype was enriched in im1 and im2 (Fisher test p = 1.65e-15), while *IDH* mutants were overrepresented in im3 (p = 1.02e-5). The DNA methylation dm3 subtype was strongly associated with im1 (Fisher test p = 1.26e-5), consistent with the association of dm3 with immune gene expression (Figure S2D). The four immune subtypes were confirmed in the TCGA GBM cohort using transcriptome data (Figures S5A and S5B) and protein deconvolution (Figures 4A and S5C).

We performed snRNA-seq on 18 GBM samples using the same cryopulverized material from previous analyses (7 im1, 5 im2, 1 im3, 5 im4 samples; Figure 4A). TAMs comprised the major non-neoplastic cell population in the GBM TME (Figure 4B). Im4 samples showed consistently low percentages of T cell and TAM infiltration (1.3% and 6% on average, respectively) compared with those of the other immune subtypes (7% and 19% on average, respectively). Im1 tumors showed higher scores for microglia and macrophages at the bulk level, but this was not observed in im1-3 in snRNA-seq (Figure S5D), which may be explained by the bulk data being driven by the percentage of TAMs and the expression of genes in the TAMs themselves. Supporting this hypothesis, we observed genes upregulated at the RNA and protein levels in TAMs in im1 (Figure 4C). We validated key immune markers (CD3, CD68, CD163, PD-1, and PD-L1) in five tumors from three immune subtypes using IHC (Figures S4E and S4F).

We identified differentially expressed proteins (DEPs) and phosphoproteins (DEPPs) in known immune targets (Chen and Hambardzumyan, 2018). Gene and protein expression levels of immune targets, including negative regulatory immune checkpoints (e.g., PD-1 and TIM-3), chemokines (e.g., CCL2), macrophage-specific cytokines (e.g., CSF-1), and their receptors (CSF-1R), were significantly higher in im1 (Figure 4A) (Butowski et al., 2016). We also identified overrepresented pathways

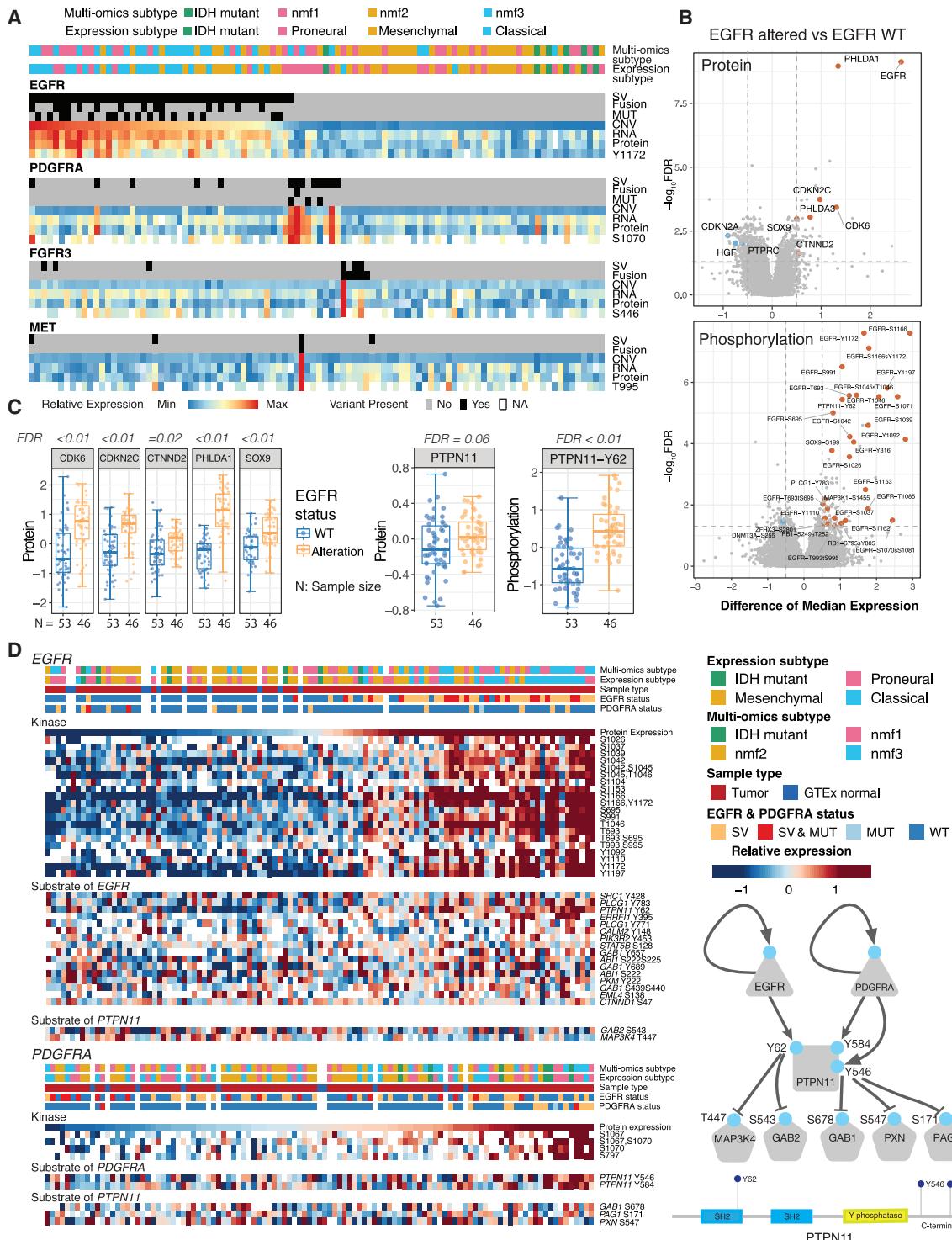


Figure 3. Alterations in RTKs and associations with expression, phosphosite status, and downstream targets

(A) Structural variations (SV), fusions, mutations (MUT), and copy number variations (CNV) in *EGFR*, *PDGFRA*, *FGFR3*, and *MET* and their *cis* effects.

(A) Structural variations (SV), frameshift mutations (WES), and copy number variations (CNV) in EGFR^{WT}, EGFR^{mut}, EGFR^{WT}, and EGFR^{mut}.

(C) Proteomic association of altered EGFR ($n = 46$) on protein expression of key genes, compared to samples with EGFR WT ($n = 53$) (left). PTPN11 level is not

(D) Heatmap showing significant ($FDR < 0.1$) *cis*- and *trans*-regulated sites of EGFR and PDGFRA kinases. Both EGFR and PDGFR regulate phosphorylation of PTPN11. The schematic (right) shows dual regulation of PTPN11 by EGFR and PDGFRA and the downstream substrates that PTPN11 may dephosphorylate.

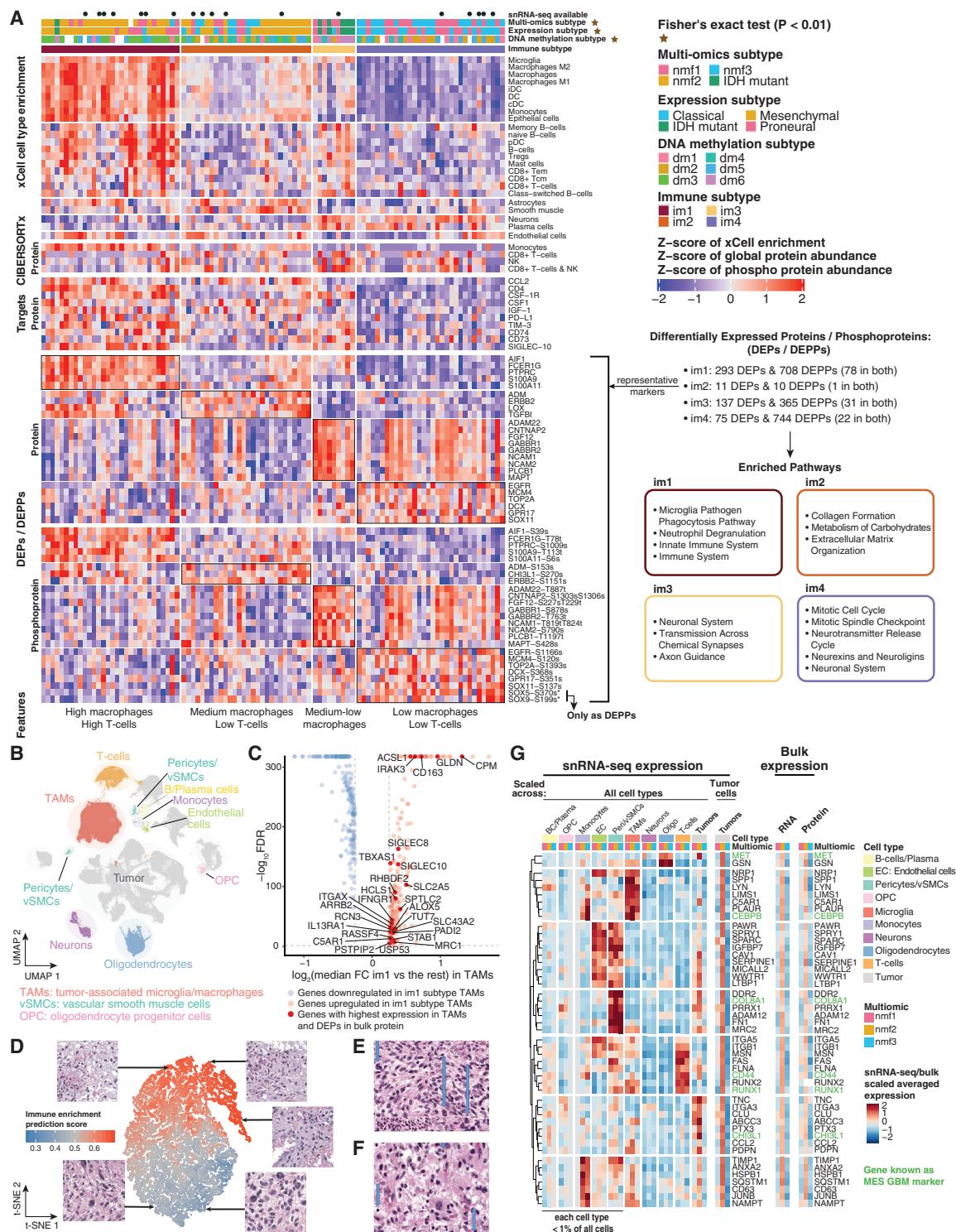


Figure 4. Cell-type enrichment, immune marker expression, and enrichment pathways among the four immune subtypes

(A) The four immune subtypes identified by consensus clustering showing cell-type features, immune checkpoints, and potential immunotherapy targets. Differential expression is between tumors of one immune subtype versus the rest based on global protein and phosphoprotein abundance (DEPs/DEPPs: FDR < 0.05 and $\log_2 FC > 0.8$) and the corresponding enriched pathways (FDR < 0.05 and $\log_2 FC \geq 3$ markers included in the pathway).

(B) snRNA-seq UMAP plot colored by cell types observed in 18 discovery cohort GBM samples. OPC, oligodendrocyte progenitor cells; TAM, tumor-associated microglia/macrophage; vSMC, vascular smooth muscle cell.

(C) Differentially expressed genes in TAMs between im1 subtype samples versus the remaining cohort. Figure shows genes with absolute value of average $\log_2 FC > 0.25$ and Wilcoxon test FDR-adjusted p values.

(legend continued on next page)

among DEPs and DEPPs (Table S4). Pathways in im1, including immune system and microglia pathogen phagocytosis, were mostly immune-related, whereas collagen formation and angiogenesis-related proteins were upregulated in im2, neuronal system pathways in im3, and cell cycle and gliogenesis pathways in im4 (Figure 4A, Table S4).

We analyzed morphologic differences between immune subtypes by applying a deep learning model using sampled tumor tiles from H&E-stained sections. Dimensional reduction (t-SNE) of im4 compared with other clusters (im1-3) (Figures 4D-4F) revealed a substantial number of large cells, some of which are giant cells, in im4 tiles, with few in im1-3 tiles. Biological relevance of these cells is unclear. Inflammatory cell fractions were noted in ~20% of im1-3 tiles compared to 5% in im4 tiles.

Mesenchymal tumor and microenvironment characteristics

Application of CausalPath (Babur et al., 2018) to the protein and phosphoprotein expression data (Figure S7A, Table S5) showed upregulation of the hypoxia pathway in mesenchymal tumors, evidenced by significant activation of multiple HIF-1 downstream targets (network permutation $p = 0.0012$). Increased angiogenesis was also evident in mesenchymal tumors, as demonstrated by upregulation of FLT1, MMP14, ENG, and SERPINE1. We observed complex regulation of macrophage activation and polarization through the upregulation of STAT3, ICAM1, SPI1, and CEBPB. In addition, the M1 polarization marker ARG1 showed increased expression (Arlauckas et al., 2018), along with SERPINE1 and HCK proteins, which promote M2 polarization (Kubala et al., 2018). The elevated inflammatory response in mesenchymal tumors may result in downstream activation of either hypoxia or macrophage polarization through multiple mediators, including LANE, IL18, and CD40 (Figure S7A).

While the origin of the mesenchymal signature in GBM has been controversial, snRNA-seq data enabled identifying mesenchymal features in tumor and immune cells. We found EMT-related genes upregulated in the tumor cells of nmf2 samples (Figures 4G and S5E). Among those, mesenchymal markers CHI3L1 and MET had the highest expression in tumor cells, along with other EMT-related genes, such as TNC, ITGA3, and PDPN. Some EMT-related genes were also expressed by nontumor cell types, including TAMs, T-cells, pericytes, endothelial cells, and oligodendrocytes. However, there were few subtype-specific differences in expression of these markers by nontumor cells. Accordingly, all EMT-related genes that were upregulated in tumor cells were also significantly increased in bulk RNA and protein data in nmf2 GBMs (Figure 4G). In addition to immune infiltration,

nmf2 GBMs evidently have intrinsic mesenchymal tumor cell-specific properties as measured by bulk proteomics.

Differential acetylation of histone proteins is associated with specific subtypes and pathways

Histone acetylation regulates gene expression but is frequently aberrant in cancer (Eberharter and Becker, 2002). We detected more than 30 acetylation sites on histones H1, H2A, H2B, H3.3, and H4. Unsupervised clustering of these sites across all samples identified subsets of tumors with differentially acetylated histones H1, H2B, H3.3, and H4 (Figure 5A). Histone acetylation was generally upregulated in tumors compared to normal samples, with a subset of tumors having elevated H1, H3, and H4 acetylation, while a different cluster exhibited significantly increased acetylation of H2B N-terminal sites (Figure 5A).

We performed Lasso linear regression between histone acetylation sites and the protein and acetylation abundances of histone acetyltransferases (HATs), bromodomain-containing proteins (BRDs), and deacetylases (HDACs). It revealed potential connections between HATs and BRDs and H2B acetylation sites, for example CREBBP and EP300, whose protein and acetylation levels showed substantial respective associations with H2B-K12, K13, K16, K17, and K21 and H2B-K21, and K24 sites (Figure 5B). These observations suggest that H2B hyperacetylation in some tumors may depend on CREBBP/EP300 activity. H2B acetylation sites also correlated with protein and acetylation abundance of BRD1, BRD3, and BRD4 proteins, which bind acetylated histones and mediate transcription (LeRoy et al., 2008). We observed significant negative correlation between H2B acetylation and the TME enrichment score (Figure 5D), while other histones showed positive associations. Our analysis identified that some pathways related to immune infiltration, such as ferroptosis, mast cells, and reactive oxygen species pathways had negative correlations with H2B acetylation, while spliceosome, nuclear receptors and SUMOylation pathways were positively correlated (Figure 5C). Two key proteins in the SUMOylation pathway, SUMO1 and UBE2l, were upregulated in samples with high H2B acetylation (Figure 5E). Interestingly, UBE2l correlated moderately with cell cycle regulator CDK6 (Pearson $r = 0.413$), a UBE2l target that is stabilized upon SUMOylation (Figure S6A) (Bellail et al., 2014). These observations suggest that H2B acetylation assists in distinguishing immune cells and other cell types.

Lipid composition and metabolomic features associated with GBM subtypes

We quantified 582 lipids in 75 tumors and 7 GTEx normal samples (Figure 6A; STAR methods), identifying more than 500 lipids that were differentially abundant across the four multi-omics

(D) Features captured by the deep learning model. Each dot represents a tile of H&E slides in the test set, colored according to prediction score (red: predicted immune-high; blue: predicted immune-low). The 20,000 sampled tiles from 99 patients were clustered by t-SNE to their activation maps (a 1,250-long vector for each tile) from the final layer of the model.

(E) H&E tile images from im4 tumors, with arrows indicating giant cells. The highlighted region contains multiple noncontinuous tiles clustered closely in t-SNE space.

(F) H&E tile images from non-im4 tumors, with arrows indicating the inflammatory cells. The highlighted region contains multiple noncontinuous tiles clustered closely in t-SNE space.

(G) Heatmaps showing snRNA-seq (left) and bulk protein (right) expression of genes upregulated in the nmf2 subtype in tumor cells. Expression values were scaled sample-wise across all cell types (or across tumor cells as labeled) and then averaged across multi-omics subtypes. Protein expression is shown for samples with snRNA-seq available.

See also Figure S5 and Table S4.

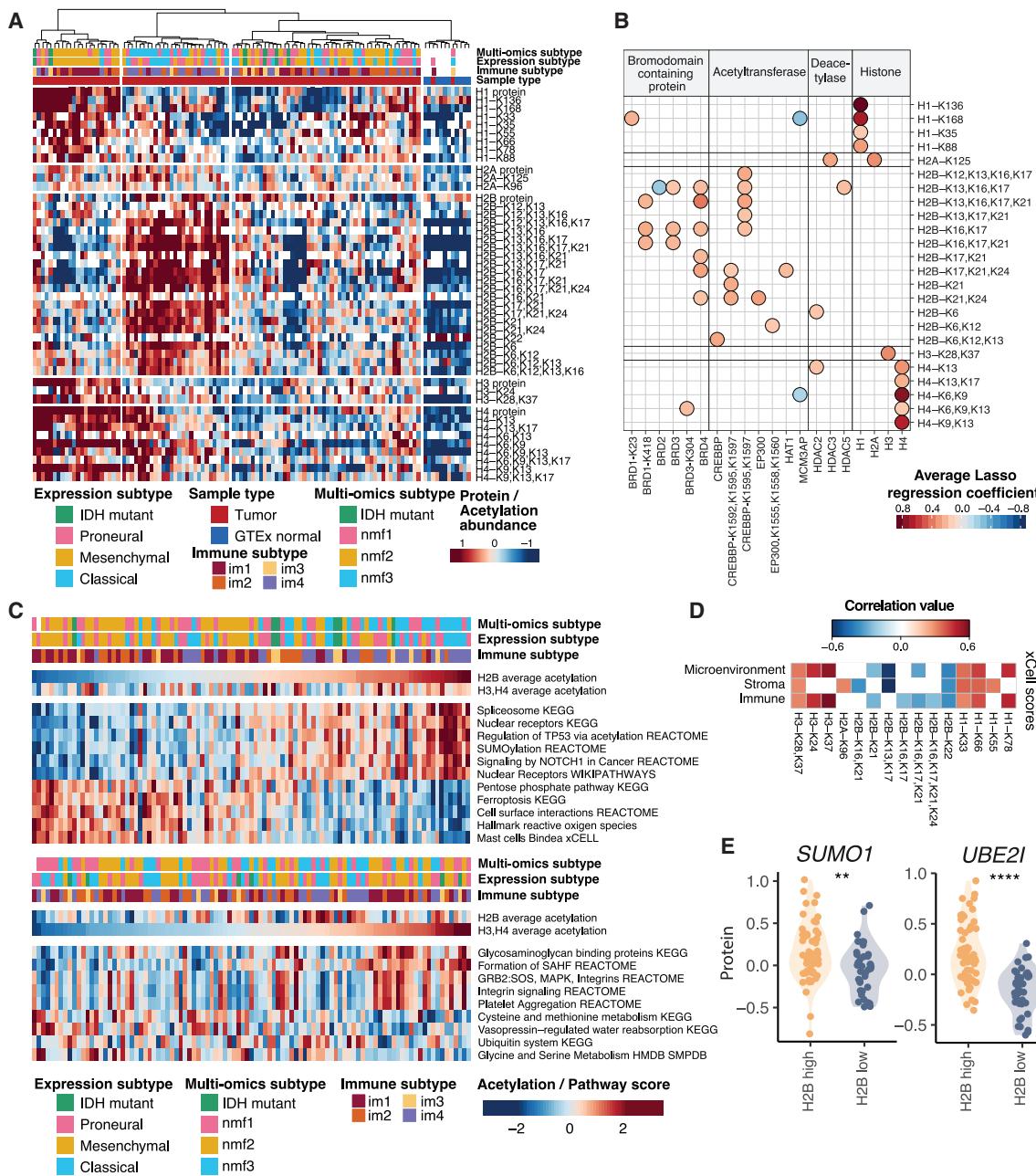


Figure 5. Histone acetylation associations with immune subtypes and pathways

- (A) Unsupervised clustering of histone protein and site-level acetylation reveals distinct clusters of tumors enriched for acetylation of histones H2B, H3, and H4.
 - (B) Significant associations between histone acetylation sites and histone acetyltransferase, deacetylases, and bromodomain-containing proteins.
 - (C) Pathways associated with levels of acetylation of histones H2B, H3, or H4 by multi-omics subtype.
 - (D) Significant Spearman correlation between xCell scores and acetylation of histone sites (FDR <0.05).
 - (E) SUMO1 and UBE2I protein expression across samples with high and low H2B acetylation.

subtypes (Wilcoxon; FDR < 0.05, Figure 1C). The mesenchymal subtype demonstrated elevated abundance of triacylglycerols (TGs), as well as depleted levels of phosphatidylcholines (PCs) and other types of phospholipids (Figures 6A and 6B). The proneural-like subtype was enriched for very long chain fatty acid lipids (VLCFAs) and glycerophospholipids with long-chain (LC)

polyunsaturated fatty acids (PUFAs) (Figure 6B). As for metabolites, the proneural-like cluster exhibited significantly increased levels of creatinine and homocysteine and reduced levels of L-cysteine and palatinitol (Wilcoxon; FDR ≤ 0.05).

We explored the connection between the differential abundance of 22:4- and 22:6-containing lipids and their metabolically

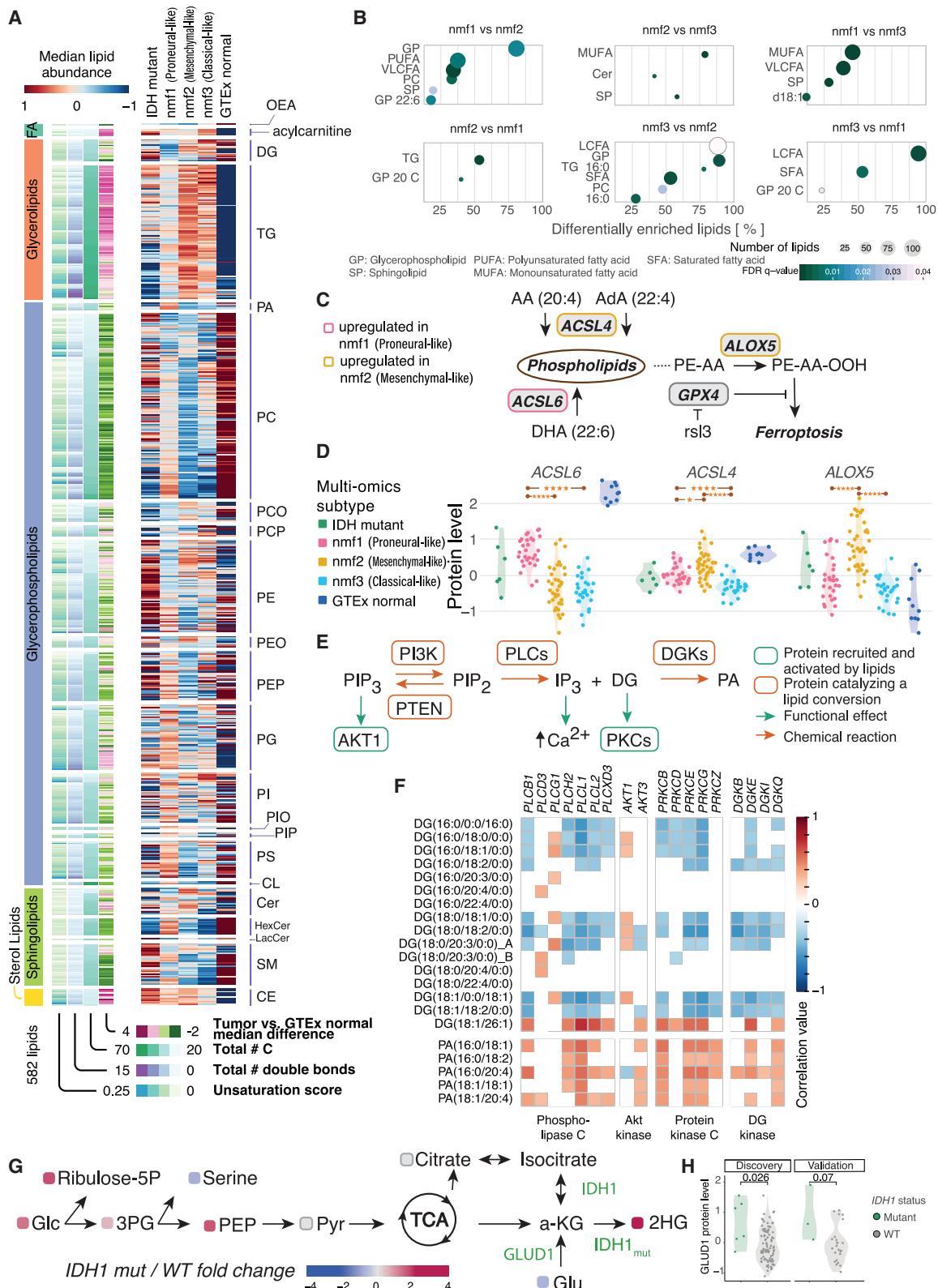


Figure 6. Lipidome and metabolome data map to major metabolic and signaling pathways

(A) Average abundance of all lipids detected across the four tumor subtypes and GTeX normal samples. Lipids are sorted by total number of double-chain bonds and total number of carbons in side chains.

(legend continued on next page)

related neuroprotective proteins (Bhagat and Das, 2015). A co-regulation example is shown with 22:6 (likely docosahexaenoic acid, DHA) and ACSL6 (an acyl-coA synthetase) (Figure 6C). Tumor samples had drastically diminished protein expression of ACSL6 (Figure 6D) and increased content of DHA-containing phosphatidylglycerols (PGs) (Figure S6C) and TGs (Figure S6D), while other phosphatidylethanolamine (PE), PC, and phosphatidylserine (PS) DHA-containing lipids were downregulated (Figure S6E). In addition, the proneural-like subtype demonstrated elevated expression of ACSL6 and phospholipids carrying DHA compared to the mesenchymal-like subtype. With respect to DHA metabolism, normal tissues were most similar to the proneural-like subtype and least similar to the mesenchymal-like subtype.

H2B acetylation-related pathway analysis identified upregulation of the ferroptosis pathway in mesenchymal-like GBMs (Figure 5C). For example, proteins ACSL4 and ALOX5 (arachidonate 5-lipoxygenase) were significantly upregulated only in the mesenchymal-like subtype (Figure 6D). ACSL4 incorporates arachidonic acid (AA - 20:4) and adrenic acid (AdA - 22:4) into PEs (Doll et al., 2017; Kagan et al., 2017) and ALOX5 catalyzes oxidation of PUFA (Gaschler and Stockwell, 2017). Their upregulation could indicate a higher content of oxidized PEs in this subtype. Downregulation of intact PE with PUFA was also observed in this subtype (Figure S6G). We also examined diacylglycerol (DG) levels in the context of enzymes related to DG production (Figure 6E). Figure 6F shows significant correlation between DGs and AKT1, PLCD3, and PLCG1 protein expression. PLCG1 phosphorylation was also affected by EGFR alterations (Figure 2B).

We compared metabolite abundances in *IDH*-mutant versus *IDH*-WT tumors. While 2-HG was the most highly abundant metabolite in *IDH*-mutant tumors (median log₂FC = 3.62, FDR < 0.05), we found other differentially present metabolites with p < 0.05, although they did not pass the FDR cutoff (Figure 6G). Several metabolites involved in glycolysis showed increased abundance in *IDH* mutants, while serine and glutamate levels were reduced. Glutamate may contribute to alpha-ketoglutarate levels and to 2-HG levels via GLUD1- and IDH1-catalyzed reactions. Supporting this hypothesis is the negative correlation between GLUD1 protein expression and glutamate abundance (Pearson's r = -0.29, p < 0.01) and significant upregulation of GLUD1 in *IDH* mutants (Wilcoxon p-value < 0.05). Our validation cohort confirmed this elevated expression of GLUD1 in *IDH*-mutant tumors (Wilcoxon p-value = 0.07, Figure 6H).

Key oncogenic pathways and therapeutic opportunities in GBM

We integrated genetic alterations and the RNA, protein, and phosphosite levels per expression subtype to examine three important oncogenic signaling pathways in GBM: RTK/RAS, PI3K/AKT, and p53/cell cycle (Figure 7A). We found that expression outlier percentage was much higher than genetic alteration rate in RTK pathways (Figure 3). Moreover, analyzed tumors harbored at least one genetic alteration or outlier expression in at least one of the three pathways.

In the RTK/RAS pathway, classical tumors predominantly showed amplified EGFR, while proneural and *IDH*-mutant tumors showed amplified PDGFRA, both resulting in higher RNA, protein, and phosphosite abundances of EGFR and PDGFRA, respectively, illustrating convergence at the functional level. For mesenchymal tumors, we observed upregulated MET and downregulated NF1 protein abundance.

In the PI3K pathway, proneural, mesenchymal, and classical tumors showed lower expression of PTEN due to mutations and deletions, which potentially activate AKT1 and AKT2 through phosphatidylinositol (3,4,5)-trisphosphate (PIP₃). In contrast, AKT3 expression was higher in *IDH*-mutant and proneural tumors, explained by active expression of AKT3 in adult brains (Easton et al., 2005). In the p53/cell cycle pathway, we observed subtype-specific amplification and increased expression of MDM2 in mesenchymal and MDM4 in proneural and classical tumors. We also observed differences between *IDH* WT and mutant tumors in CDKN2A/B.

In conjunction with druggability information from DGdb (Cotto et al., 2018) and DEPO (Sun et al., 2018), we conducted kinase-substrate and outlier analyses to identify druggable pairs (Figure 7B; Table S5). We found that GSK3B phosphorylation is positively associated with phosphorylation of its downstream proteins involved in mammalian target of rapamycin (mTOR) signaling (e.g., RPTOR and TSC1) and Wnt signaling (e.g., CTNNB1 and APC). Another player in mTOR signaling, AKT1S1, had many significant connections with AKT1, AKT2, and AKT3 kinases. EGFR was also found to phosphorylate CTNNB1 S33 at the N-terminal, known to mediate CTNNB1 proteasomal degradation (Park et al., 2004). Phosphosite outlier analysis corroborated these findings with interactions for GSK3B, AKT1, MAPK1, MAPK3, and EGFR (Figure S7B; Table S5). The MAP kinase cascade was associated with diverse proteins, including ABL1 kinase, which in turn had a strong

- (B) Lipid Mini-On enrichment analysis of lipid properties upregulated in subtype versus second subtype.
 - (C) Contribution of enzymes that activate PUFA (ACSL4 and ACSL6) to the phospholipid pool and the connection of PUFA-containing PE to ferroptosis.
 - (D) Protein expression of ACSL6, ACSL4, and ALOX5 across tumor multi-omics subtypes and GTEx normal tissues.
 - (E) Schematic diagram of lipid conversion reactions essential for cell signaling.
 - (F) Correlation among DG, phosphatidic acid (PA), and phospholipases C (PLCs; cleaves PIP₂ into DG and IP₃), Akt kinases (interact with PIP₃), protein kinases C (PKCs; interact with DG), and DG kinases (DGKs; phosphorylate DG to produce PA).
 - (G) *IDH1* mutants display elevated abundance of glucose, glycolytic intermediate metabolites, and 2-HG, along with reduced abundance of glutamate and serine.
 - (H) GLUD1 protein expression is upregulated in *IDH1* mutants in both discovery and validation cohorts.
- CE, Cholesteryl ester; CL, Cardiolipin; Cer, Ceramide; FA, Fatty acid; GP, Glycerophospholipid; Glc, Glucose; Glu, Glutamate; HexCer, Hexosylceramide; LCFA, Long chain fatty acid; LacCer, Lactosylceramides; MUFA, Monounsaturated fatty acid; OEA, Oleoylethanolamide; PCO, Phosphatidylcholine with an alkyl ether substituent; PCP, Phosphatidylcholine with a plasmalogen substituent; PEO, Phosphatidylethanolamine with an alkyl ether substituent; PEP (in panel A): Phosphatidylethanolamine with a plasmalogen substituent. PEP (G): Phosphoenolpyruvic acid. PI, Phosphatidylinositol; PIO, Phosphatidylinositol with an alkyl ether substituent; PIP, Phosphatidylinositol with a plasmalogen substituent; PLC, Phospholipase C; Pyr, Pyruvic acid; SM, Sphingomyelin; SP, Sphingolipid; 3PG, 3-Phosphoglyceric acid.
- See also Figure S6.

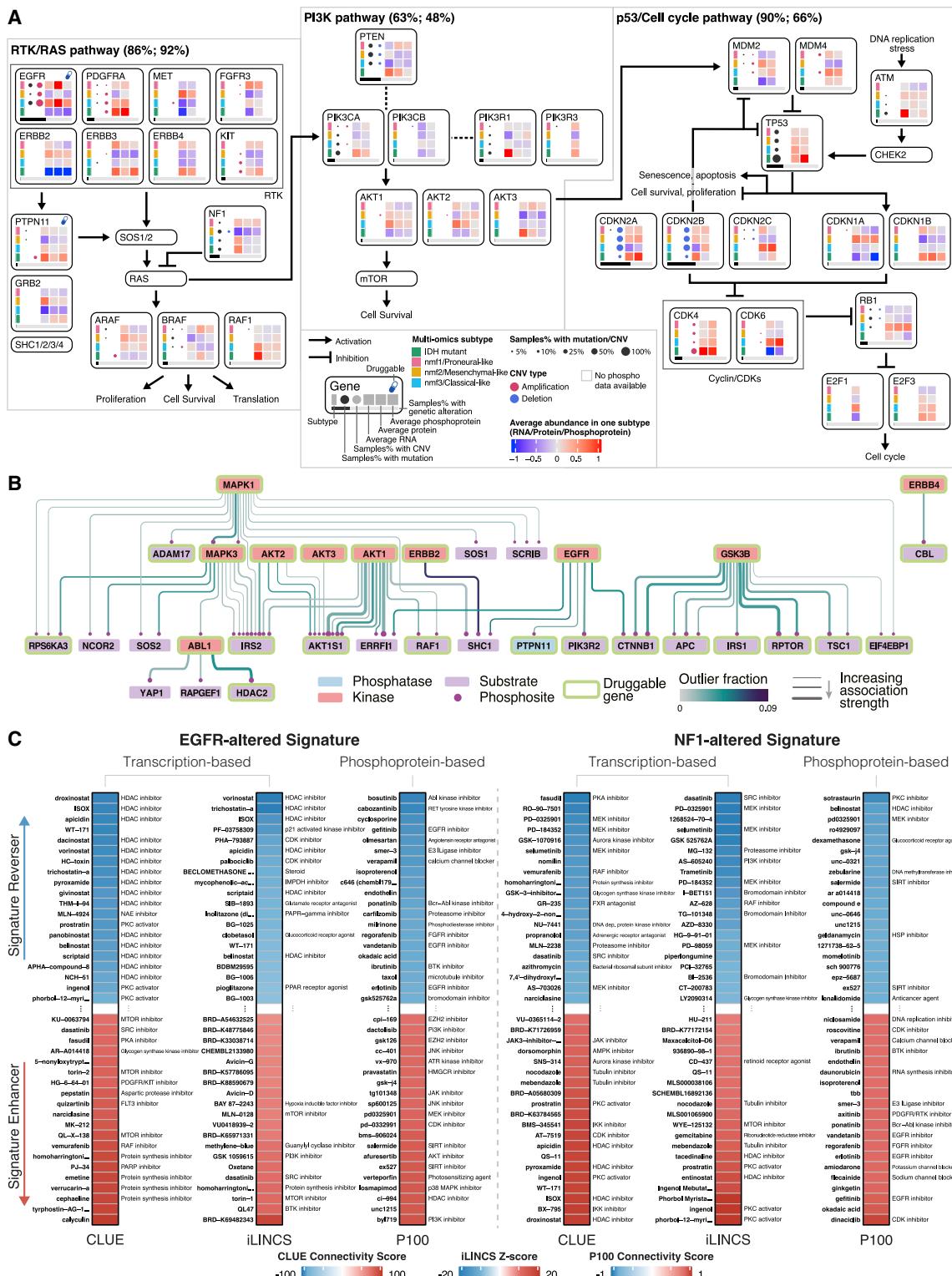


Figure 7. Summary of pathway alterations and potential therapeutic targets

(A) Three oncogenic pathways frequently altered in GBM. Each gene is annotated with mutational and CNV frequency, RNA, protein, and phosphoprotein abundance, by multi-omics subtype. Horizontal bar below gene box indicates frequency of alteration across all tumors. Also indicated are the proportion of tumors with genetic alterations (first percentage) and protein and phosphoprotein outlier expression (second percentage) for each pathway.

association with HDAC2 deacetylase with ~5% of outliers. We also found increased HDAC2 S422 phosphorylation, coinciding with its functional activity (Eom and Kook, 2015).

Using the Library of Integrated Network-Based Cellular Signatures (LINCS) (Keenan et al., 2018; Stathias et al., 2019), we calculated the similarity between alteration-specific RNA or phosphoprotein signatures from our study with corresponding transcriptional (L1000 assay) (Subramanian et al., 2017) and phosphoproteomic LINCS signatures (P100 assay) (Litichevskiy et al., 2018) to identify compounds predicted to reverse tumor signatures of the cohort. Phosphoproteomic data yielded more robust results than the transcriptional-response data. For example, while RNA signatures suggested that HDAC inhibitors might reverse the EGFR signature, phosphoproteomic data indicated kinase inhibitors, beyond the expected EGFR inhibitors, to be more highly connected with potential inhibitory effects (Figure 7C). In contrast, *NF1*-altered samples exhibited concordance between the transcriptional and phosphoproteomic perturbation-response analyses: both suggested MAPK inhibitors as top signature-reversing candidates (Figure 7C). RB1 alteration and immune subtypes agreed variously when comparing perturbations predicted to reverse the gene-altered cell states using transcriptomics versus phosphoproteomics readouts (Figure S7C, Table S6).

DISCUSSION

GBM was one of the earliest subjects of deep genomic and transcriptomic analysis (Brennan et al., 2013) and targeted MS studies (Gu et al., 2017; He et al., 2007). However, most patients are still treated with a standard of care developed almost two decades ago (Stupp et al., 2005), underscoring the need for deeper insights. Here, we extended classical sequencing approaches with comprehensive integration of MS-based proteome, phosphoproteome, acetylome, metabolome, and lipidome analyses and single-cell transcriptomics. Multi-omics analysis identified a subset of patients with mixed subtypes compared with traditional sequencing-based subtypes, who exhibit shortened overall survival. Phosphoproteomic data indicate that PLCG1 and PTPN11 act as a common signaling hub for multiple RTKs.

RNA and protein expression data from bulk tumors indicate that GBM subtypes differ in infiltrating macrophages and the distribution of specific immune cell types. In particular, we discovered an unexpected immune subtype im3 exhibiting a relative depletion of the macrophage-microglia immunosuppressive infiltration typical of GBM, but which contains significant enrichment of T lymphocytes and natural killer (NK) cells. It is also enriched for *IDH*-mutated tumors. We validated these findings using IHC and an independent patient cohort. Interestingly, im1 TAMs demonstrated upregulation of M2 polarization

markers, such as CD163 and MRC1, suggesting a role in tumor promotion (Pinto et al., 2019).

The mesenchymal subtype has high bulk-level RNA expression of EMT signatures (Behnan et al., 2019), but it was not known if this was due to the tumor cells themselves or the high number of infiltrating immune cells. Our data indicate that tumor cells in the mesenchymal subtype display an enhanced EMT signature, along with an increase in the relative proportion of immune cells in the tumor stroma. Cell-type-specific gene expression indicates that both the tumor cells and the stromal components contribute to the overall mesenchymal signatures. This study reveals an association between H2B acetylation and patterns of protein expression associated with immune cell functions in GBM. Furthermore, comparison of the lipid and metabolic signatures in GBM subtypes and normal brain tissues reveals shared characteristics that may be associated with neuronal phenotypes and IDH status. Similarly, from a metabolic view, mesenchymal-like GBMs differ substantially from other subtypes and have specific metabolic vulnerabilities not present in other GBMs.

The multidimensional analysis of patient specimens described in this investigation adds context to prior genomic and transcription-based investigations of GBM and suggests avenues for further mechanistic studies. Rapid advancement of single-cell genomics and proteomics technologies will facilitate deeper analyses of GBM heterogeneity and TME interactions. We hope these advances will improve patient stratification for clinical trials and lead, ultimately, to personalized treatments.

CONSORTIA

The members of the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium are Anupriya Agarwal, Mutual Amin, Eunkyung An, Matthew L. Anderson, David W. Andrews, Jessika Baral, Jill S. Barnholtz-Sloan, Thomas Bauer, Chet Birger, Michael J. Birrer, Kent J. Bloodsworth, Lili Blumenberg, Simina M. Boca, William E. Bocik, Hannah Boekweg, Emily S. Boja, Uma Borate, Melissa Borucki, Lisa M. Bramer, Meghan C. Burke, Shuang Cai, Anna Pamela Calinawan, Song Cao, Wagma Caravan, Steven A. Carr, Michele Ceccarelli, Sandra Cerda, Daniel W. Chan, Changcun Chang, Alyssa Charanum, Feng Chen, Lin S. Chen, Xi S. Chen, David Chesla, Milan G. Chheda, Arul M. Chinnaian, Shrabanti Chowdhury, Rosalie K. Chu, Marcin P. Cieślik, David J. Clark, Karl Clouser, Antonio Colaprico, MacIntosh Cornwell, Sandra L. Cottingham, Sneha P. Couvillion, Daniel Cui Zhou, Houston Culpepper, Andrzej Czechński, Tomasz Czernicki, Fulvio D'Angelo, Jacob Day, Stephanie De Young, Emek Demir, Alexis Demopoulos, Saravana Mohan Dhanasekaran, Rajiv Dhir, Li Ding, Marcin J. Domagalski, Yongchao Dou, Brian Druker, Elizabeth Duffy, Maureen Dyer, Nathan J. Edwards, Robert Edwards, Kimberly Elburn, Matthew J. Ellis,

(B) Dysregulated phospho-signaling in RTK, PI3K, WNT, and NOTCH pathways across all tumors. Thickness of a kinase-substrate connecting line indicates degree to which variation in kinase phosphorylation explains observed variation in the substrate phosphosite abundance. Line color indicates percentage of samples with outlier phosphorylation. Kinases governing multiple substrates with substantial phosphorylation outliers may be potential therapeutic targets.
(C) Drug connectivity analysis using alteration-specific transcriptional (CLUE and iLINCS) and phosphoproteomic (P100) signatures (altered tumors versus WT tumors). Twenty compounds that most strongly reverse or enhance the signature are highlighted along with their known mechanisms of action.
See also Figure S7 and Tables S5 and S6.

Jennifer Eschbacher, David Fenyö, Jayson B. Field, Steven Foltz, Alicia Francis, Stacey Gabriel, Nikolay Gabrovski, Luciano Garofano, Yifat Geffen, Gad Getz, Michael A. Gillette, Andrew K. Godwin, Denis Golbin, Charles A. Goldthwaite, Jr., Marina A. Gritsenko, Zeynep Gumus, Ziad Hanhan, Linda I. Hannick, Pushpa Hariharan, Dawn Hayward, David Heiman, Brittany L. Henderson, Tara Hiltke, Barbara Hindenach, Katherine A. Hoadley, Runyu Hong, Galen Hostetter, Chen Huang, Antonio Iavarone, Eric Jaehnig, Paweł Jarmużek, Scott D. Jewell, Jiayi Ji, Nan Ji, Corbin D. Jones, Marcin Jędryka, Bożena Kamińska, Renée Karabon, Alla Karpova, Alcida Karz, Wojciech Kaspera, Emily Kawaler, Selim Kaycee, Karen A. Ketchum, Albert H. Kim, Lyndon Kim, Christopher R. Kinsinger, Ramani B. Kothadia, Azra Krek, Karsten Krug, Chandan Kumar-Sinha, Jennifer E. Kyle, Jonathan Lei, Felipe D. Leprevost, Kai Li, Qin Li, Qing Kay Li, Tian Li, Yize Li, Zibo Li, Wen-Wei Liang, Yuxing Liao, Jena Lilly, Caleb M. Lindgren, Hongwei Liu, Tao Liu, Wenke Liu, Shuangjia Lu, Jan Lubinski, Weiping Ma, Melanie MacMillan, Rashna Madan, William Maggio, Piotr Makarewicz, Ewa Malc, Anna Malovannaya, D. R. Mani, Andrzej Marchel, Sailaja Mareedu, Sanford P. Markey, Annette Marrero-Oliveras, John Martignetti, Nina Martinez, Yosef Maruvka, Rafał Matkowski, Nicolle Maunganidze, Jason E. McDermott, Peter B. McGarvey, John McGee, Joshua F. McMichael, Mehdi Mesri, Piotr Mieczkowski, Simona Migliozzi, Francesmary Modugno, Matthew E. Monroe, Rebecca Montgomery, Jamie Moon, Ronald J. Moore, Sergiusz Nawrocki, Alexey I. Nesvizhskii, Chelsea J. Newton, Lindsey K. Olsen, Gilbert S. Omenn, Umut Ozbek, Oxana V. Paklina, Akash J. Patel, Amanda G. Paulovich, Samuel H. Payne, Amy M. Perou, Francesca Petralia, Vladislav A. Petryuk, Alexander R. Pico, Paul D. Piehowski, Dimitris G. Placanakis, Larisa Polonskaya, Olga Potapova, Barbara Pruetz, Pietro Pugliese, Liqun Qi, Shakti Ramkissoon, Adam C. Resnick, Boris Reva, Shannon Richey, Gregory Riggins, Anna Roberts-Pilgrim, Karna Robinson, Ana I. Robles, Nancy Roche, Karin D. Rodland, Henry Rodriguez, Daniel C. Rohrer, Brian R. Rood, Larissa Rossell, Joseph Rothstein, Kelly V. Ruggles, Dmitry Rykunov, Shankha Satpathy, Sara R. Savage, Eric E. Schadt, Stephan Schürer, Sohini Sengupta, Yan Shi, Zhiao Shi, Yvonne Shutack, Weiva Sieh, Shilpi Singh, Tara Skelly, Richard D. Smith, Lori J. Sokoll, Xiaoyu Song, Vasileios Stathias, Jakub Stawicki, Stephen E. Stein, Kelly G. Stratton, James Suh, Wojciech Szopa, Dave Tabor, Donghui Tan, Lauren Tang, Darlene Tansil, Ratna R. Thangudu, Mathangi Thiagarajan, Cristina Tognon, Elie Traer, Chia-Feng Tsai, Shirley Tsang, Jeffrey Tyner, Ki Sung Um, Dana R. Valley, Suhas Vasaikar, Negin Vatanian, Uma Velkulou, Michael Vernon, Weiqing Wan, Bing Wang, Junmei Wang, Liang-Bo Wang, Pei Wang, Alex Webster, Sun Wei, Karl K. Weitz, Bo Wen, Michael C. Wendl, Jeffrey R. Whiteaker, George D. Wilson, Maciej Wiznerowicz, Yige Wu, Matthew A. Wyczalkowski, Xiaolu Yang, Lijun Yao, Seungyeul Yoo, Yuriy Zakhartsev, Robert Zelt, Bing Zhang, Hui Zhang, Liwei Zhang, Zhen Zhang, Grace Zhao, Rui Zhao, Yuanli Zhao, Houxiang Zhu, Jun Zhu, Erika Zink.

STAR METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **ADDITIONAL RESOURCES**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Specimens and clinical data
 - Sample processing
 - Validation cohort sample collection and processing
- **METHOD DETAILS**
 - Sample processing for genomic DNA and total RNA extraction
 - Whole exome sequencing
 - PCR-free whole genome sequencing
 - Illumina Infinium methylationEPIC beadchip array
 - RNA sequencing
 - MS sample processing and data collection
 - Polar metabolites and lipid mass spectrometry
 - Immunohistochemistry (IHC) validation of genetic alteration downstream impact and immune cell compositions
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Tumor exclusion criteria
 - Genomic data analysis
 - RNA quantification and analysis
 - snRNA-seq quantification and analysis
 - MS data interpretation
 - Other proteogenomic analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.ccr.2021.01.006>.

ACKNOWLEDGMENTS

This work was supported by grants U24CA210972, U24CA210955, U24CA210954, U24CA210985, U24CA210993, U24CA210967, U24CA210986, U01CA214125, and U24CA210979 from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium, by grant R01HG009711 from National Human Genome Research Institute to L.D., and R01NS107833 from National Institutes of Health to M.G.C. The MS-based proteomics work was performed at the Environmental Molecular Sciences Laboratory (grid.436923.9), a U.S. Department of Energy National Scientific User Facility located at the Pacific Northwest National Laboratory operated under contract DE-AC05-76RL01830. We thank Dr. Charles A. Goldthwaite for manuscript editing.

AUTHOR CONTRIBUTIONS

Study Conception & Design: J.S.B., K.D.R., T.L., L.D.

Performed Experiment or Data Collection: M.A.G., W.C., C.T., R.K.C., K.W.K., R.J.M., M.E.M., R.Z., X.Y., E.Z., S.P.C., K.J.B., F.C.

Computation & Statistical Analysis: L.W., AKarpova, J.E.K., S.C., Y.L., D.R., A.C., J.H.R., R.H., V.S., MCornwell, F.P., Y.W., B.R., K.K., P.P., E.K., L.K.O., WLiang, X.S., Y.D., M.C.W., WLiu, D.C.Z., J.J., V.A.P., J.M., W.M., S.Y., AKrek, A.D., H.Z., B.L.H., C.M.L., H.B., S.L., J.B., L.Y., K.G.S., L.M.B., SSatpathy, W.S., M.W., A.I.N., B.Z., D.R.M., MCeccarelli, X.S.C., D.F., K.V.R., S.H.P., P.W., L.D.

Data Interpretation & Biological Analysis: L.W., AKarpova, J.E.K., S.C., Y.L., D.R., A.C., J.H.R., R.H., V.S., MCornwell, F.P., Y.W., B.R., K.K., P.P., E.K., L.K.O., WLiang, X.S., Y.D., M.C.W., WLiu, D.C.Z., J.J., S.Y., AKrek, A.D., B.L.H., C.M.L., H.B., S.L., J.B., L.Y., W.S., M.W., K.A.K., S.L.C., Q.K.L., J.S.B., L.D.

Writing – Original Drafts: L.W., AKarpova, M.A.G., J.E.K., S.C., Y.L., D.R., V.S., M.A.W., K.D.R., T.L., L.D.

Writing – Review & Editing: L.W., AKarpova, M.A.G., J.E.K., S.C., Y.L., A.C., V.S., M.C.W., J.F.M., S.M.B., E.S.B., A.H.K., M.M., S.H.P., A.I., M.G.C., J.S.B., K.D.R., T.L., L.D.

Supervision: SSchürer, F.C., A.I.N., B.Z., D.R.M., MCecarelli, X.S.C., D.F., K.V.R., H.R., S.H.P., A.C.R., P.W., R.D.S., J.S.B., K.D.R., T.L., L.D.

Administration: E.S.B., C.R.K., A.I.R., T.H., M.T., A.H.K., H.R., M.M., K.D.R., T.L., L.D.

DECLARATION OF INTERESTS

S.Y. is employed by Sema4. A.H.K. consults for Monteris Medical. P.W. is a statistical consultant for Sema4. M.G.C. receives research support from Orbus Therapeutics and NeoimmuneTech Inc, and royalties from UpToDate.

Received: February 3, 2020

Revised: June 2, 2020

Accepted: January 11, 2021

Published: February 11, 2021

REFERENCES

- 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169.
- Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220.
- Arlauckas, S.P., Garren, S.B., Garris, C.S., Kohler, R.H., Oh, J., Pittet, M.J., and Weissleder, R. (2018). Arg1 expression defines immunosuppressive subsets of tumor-associated macrophages. *Theranostics* **8**, 5842–5854.
- Babiceanu, M., Qin, F., Xie, Z., Jia, Y., Lopez, K., Janus, N., Facemire, L., Kumar, S., Pang, Y., Qi, Y., et al. (2016). Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* **44**, 2859–2872.
- Babur, Ö., Luna, A., Korkut, A., Durupinar, F., Siper, M.C., Dogrusoz, U., Aslan, J.E., Sander, C., and Demir, E. (2018). Causal interactions from proteomic profiles: molecular data meets pathway knowledge. *BioRxiv*, 258855.
- 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.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weersinghe, A., Colaprico, A., Wendt, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e18.
- Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112.
- 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.
- Behnau, J., Finocchiaro, G., and Hanna, G. (2019). The landscape of the mesenchymal signature in brain tumours. *Brain J. Neurol.* **142**, 847–866.
- Bellail, A.C., Olson, J.J., and Hao, C. (2014). SUMO1 modification stabilizes CDK6 protein and drives the cell cycle and glioblastoma progression. *Nat. Commun.* **5**, 4234.
- Benelli, M., Pescucci, C., Marseglia, G., Severgnini, M., Torricelli, F., and Magi, A. (2012). Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* **28**, 3232–3239.
- Bhagat, U., and Das, U.N. (2015). Potential role of dietary lipids in the prophylaxis of some clinical conditions. *Arch. Med. Sci.* **11**, 807–818.
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theor. Exp.* **2008**, P10008.
- Blumenberg, L., Kawaler, E., Cornwell, M., Smith, S., Ruggles, K., and Fenyö, D. (2019). BlackSheep: a bioconductor and bioconda package for differential extreme value analysis (Bioinformatics).
- Brat, D.J., Aldape, K., Colman, H., Holland, E.C., Louis, D.N., Jenkins, R.B., Kleinschmidt-DeMasters, B.K., Perry, A., Reifenberger, G., Stupp, R., et al. (2018). cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV". *Acta Neuropathol.* **136**, 805–810.
- 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.
- Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U S A* **101**, 4164–4169.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, Rahul (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420.
- Butowski, N., Colman, H., De Groot, J.F., Omuro, A.M., Nayak, L., Wen, P.Y., Cloughesy, T.F., Marimuthu, A., Haidar, S., Perry, A., et al. (2016). Orally administered colony stimulating factor 1 receptor inhibitor PLX3397 in recurrent glioblastoma: an Ivy Foundation Early Phase Clinical Trials Consortium phase II study. *Neuro-Oncol.* **18**, 557–564.
- Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474.
- Chen, Z., and Hambardzumyan, D. (2018). Immune microenvironment in glioblastoma subtypes. *Front. Immunol.* **9**, 1004.
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222.
- Chu, A., Robertson, G., Brooks, D., Mungall, A.J., Birol, I., Cope, R., Ma, Y., Jones, S., and Marra, M.A. (2016). Large-scale profiling of microRNAs for the cancer genome atlas. *Nucleic Acids Res.* **44**, e3.
- 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.
- Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.-S.M., Chang, H.-Y., et al. (2019). Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **179**, 964–983.e31.
- 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.
- Colaprico, A., Olsen, C., Bailey, M.H., Odom, G.J., Terkelsen, T., Silva, T.C., Olsen, A.V., Cantini, L., Zinovyev, A., Barillot, E., et al. (2020). Interpreting pathways to discover cancer driver genes with Moonlight. *Nat. Commun.* **11**, 69.
- Cotto, K.C., Wagner, A.H., Feng, Y.-Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073.
- Daily, K., Ho Sui, S.J., Schriml, L.M., Dexheimer, P.J., Salomonis, N., Schroll, R., Bush, S., Keddache, M., Mayhew, C., Lotia, S., et al. (2017). Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. *Sci. Data* **4**, 170030.

- Delgado-López, P.D., and Corrales-García, E.M. (2016). Survival in glioblastoma: a review on the impact of treatment modalities. *Clin. Transl. Oncol.* **18**, 1062–1071.
- Ding, Z., Mangino, M., Aviv, A., Spector, T., and Durbin, R. (2014). Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* **42**, e75.
- Doll, S., Proneth, B., Tyurina, Y.Y., Panzilius, E., Kobayashi, S., Ingold, I., Irmler, M., Beckers, J., Aichler, M., Walch, A., et al. (2017). ACSL4 dictates ferroptosis sensitivity by shaping cellular lipid composition. *Nat. Chem. Biol.* **13**, 91–98.
- Dou, Y., Kawaler, E.A., Zhou, D.C., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic characterization of endometrial carcinoma. *Cell* **180**, 729–748.e26.
- Duan, G., Li, X., and Köhn, M. (2015). The human DEPhosphorylation database DEPOD: a 2015 update. *Nucleic Acids Res.* **43**, D531–D535.
- Easton, R.M., Cho, H., Roovers, K., Shineman, D.W., Mizrahi, M., Forman, M.S., Lee, V.M.-Y., Szabolcs, M., de Jong, R., Oltersdorf, T., et al. (2005). Role for Akt3/protein kinase Bgamma in attainment of normal brain size. *Mol. Cell. Biol.* **25**, 1869–1878.
- Eberharter, A., and Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep.* **3**, 224–229.
- Eom, G.H., and Kook, H. (2015). Role of histone deacetylase 2 and its post-translational modifications in cardiac hypertrophy. *BMB Rep.* **48**, 131–138.
- 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.
- 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.
- 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.
- Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* **16**, 4.
- Gao, Q., Liang, W.-W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.-W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al. (2018). Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* **23**, 227–238.e3.
- Gaschler, M.M., and Stockwell, B.R. (2017). Lipid peroxidation in cell death. *Biochem. Biophys. Res. Commun.* **482**, 419–425.
- Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367.
- 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. *Bioinforma. Oxf. Engl.* **31**, 3838–3840.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849.
- Gu, X., Hua, Z., Dong, Y., Zhan, Y., Zhang, X., Tian, W., Liu, Z., Thiele, C.J., and Li, Z. (2017). Proteome and acetylome analysis identifies novel pathways and targets regulated by Perifosine in neuroblastoma. *Sci. Rep.* **7**, 42062.
- Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* **20**, 213.
- Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *BioRxiv*, 576827.
- He, H., Conrad, C.A., Nilsson, C.L., Ji, Y., Schaub, T.M., Marshall, A.G., and Emmett, M.R. (2007). Method for lipidomic analysis: p53 expression modulation of sulfatide, ganglioside, and phospholipid composition of U87 MG glioblastoma cells. *Anal. Chem.* **79**, 8423–8430.
- Hiller, K., Hangebrauk, J., Jäger, C., Spura, J., Schreiber, K., and Schomburg, D. (2009). MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal. Chem.* **81**, 3429–3439.
- 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.
- 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.
- Kagan, V.E., Mao, G., Qu, F., Angeli, J.P.F., Doll, S., Croix, C.S., Dar, H.H., Liu, B., Tyurin, V.A., Ritov, V.B., et al. (2017). Oxidized arachidonic and adrenic PEs navigate cells to ferroptosis. *Nat. Chem. Biol.* **13**, 81–90.
- 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.
- Karimi, S., Zuccato, J.A., Mamatjan, Y., Mansouri, S., Suppiah, S., Nassiri, F., Diamandis, P., Munoz, D.G., Aldape, K.D., and Zadeh, G. (2019). The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. *Clin. Epigenetics* **11**, 185.
- Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohman, A.B., Silverstein, M.C., Lachmann, A., et al. (2018). The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* **6**, 13–24.
- Killela, P.J., Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L.A., Friedman, A.H., Friedman, H., Gallia, G.L., Giovannella, B.C., et al. (2013). TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. U S A* **110**, 6021–6026.
- Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinforma. Oxf. Engl.* **23**, 1495–1502.
- Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277.
- Kim, S., Gupta, N., and Pevzner, P.A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **7**, 3354–3363.
- Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594.
- 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.
- 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.
- Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522.
- Kubala, M.H., Punj, V., Placencio-Hickok, V.R., Fang, H., Fernandez, G.E., Spoto, R., and DeClerck, Y.A. (2018). Plasminogen activator inhibitor-1 promotes the recruitment and polarization of macrophages in cancer. *Cell Rep* **25**, 2177–2191.e7.
- Kunze, K., Spieker, T., Gamerding, U., Nau, K., Berger, J., Dreyer, T., Sindermann, J.R., Hoffmeier, A., Gattenlöchner, S., and Bräuninger, A. (2014). A recurrent activating PLCG1 mutation in cardiac angiosarcomas increases apoptosis resistance and invasiveness of endothelial cells. *Cancer Res.* **74**, 6173–6183.
- 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. *Bioinforma. Oxf. Engl.* 33, 1744–1746.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- LeRoy, G., Rickards, B., and Flint, S.J. (2008). The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol. Cell* 30, 51–60.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079.
- Li, S., Shen, D., Shao, J., Crowder, R., Liu, W., Prat, A., He, X., Liu, S., Hoog, J., Lu, C., et al. (2013). Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* 4, 1116–1130.
- Li, M., Xie, X., Zhou, J., Sheng, M., Yin, X., Ko, E.-A., Zhou, T., and Gu, W. (2017). Quantifying circular RNA expression from RNA-seq data using model-based framework. *Bioinforma. Oxf. Engl.* 33, 2131–2139.
- 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.
- Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L., and Cesareni, G. (2019). SIGNOR 2.0, the SIGnaling network open resource 2.0: 2019 update. *Nucleic Acids Res.* 48, D504–D510.
- Lim, M., Xia, Y., Bettegowda, C., and Weller, M. (2018). Current state of immunotherapy for glioblastoma. *Nat. Rev. Clin. Oncol.* 15, 422.
- Ling, S., Chang, X., Schultz, L., Lee, T.K., Chaux, A., Marchionni, L., Netto, G.J., Sidransky, D., and Berman, D.M. (2011). An EGFR-ERK-SOX9 signaling cascade links urothelial development and regeneration to cancer. *Cancer Res.* 71, 3812–3821.
- Litichevskiy, L., Peckner, R., Abelin, J.G., Asiedu, J.K., Creech, A.L., Davis, J.F., Davison, D., Dunning, C.M., Egertson, J.D., Egri, S., et al. (2018). A library of phosphoproteomic and chromatin signatures for characterizing cellular responses to drug perturbations. *Cell Syst* 6, 424–443.e7.
- Liu, L., Shi, Y., Shi, J., Wang, H., Sheng, Y., Jiang, Q., Chen, H., Li, X., and Dong, J. (2019). The long non-coding RNA SNHG1 promotes glioma progression by competitively binding to miR-194 to regulate PHLDA1 expression. *Cell Death Dis.* 10, 463.
- Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820.
- Louis, D.N., Aldape, K., Brat, D.J., Capper, D., Ellison, D.W., Hawkins, C., Paulus, W., Perry, A., Reifenberger, G., Figarella-Branger, D., et al. (2017). Announcing cIMPACT-NOW: the consortium to inform molecular and practical approaches to CNS tumor taxonomy. *Acta Neuropathol.* 133, 1–3.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173, 338–354.e15.
- Masuda, T., Nishimoto, N., Tomiyama, D., Matsuda, T., Tozaki-Saitoh, H., Tamura, T., Kohsaka, S., Tsuda, M., and Inoue, K. (2014). IRF8 is a transcriptional determinant for microglial motility. *Purinergic Signal.* 10, 515–521.
- Matzke, M.M., Waters, K.M., Metz, T.O., Jacobs, J.M., Sims, A.C., Baric, R.S., Pounds, J.G., and Webb-Robertson, B.-J.M. (2011). Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinforma. Oxf. Engl.* 27, 2866–2872.
- McGranahan, T., Therkelsen, K.E., Ahmad, S., and Nagpal, S. (2019). Current state of immunotherapy for treatment of glioblastoma. *Curr. Treat. Options Oncol.* 20, 24.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- 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.
- Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Claußer, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534, 55–62.
- 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.
- Montagner, A., Yart, A., Dance, M., Perret, B., Salles, J.-P., and Raynal, P. (2005). A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation. *J. Biol. Chem.* 280, 5350–5360.
- 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.
- Narita, T., Weinert, B.T., and Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. *Nat. Rev. Mol. Cell Biol.* 20, 156–174.
- Nassiri, F., Mamatjan, Y., Suppiah, S., Badhiwala, J.H., Mansouri, S., Karimi, S., Saarela, O., Poisson, L., Gepfner-Tuma, I., Schittenhelm, J., et al. (2019). DNA methylation profiling to predict recurrence risk in meningioma: development and validation of a nomogram to optimize clinical management. *Neuro-Oncol.* 21, 901–910.
- Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782.
- Ostrom, Q.T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., and Barnholtz-Sloan, J.S. (2019). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro-Oncol.* 21, v1–v100.
- 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.
- Park, C.S., Kim, S.I., Lee, M.S., Youn, C.-Y., Kim, D.J., Jho, E.-H., and Song, W.K. (2004). Modulation of beta-catenin phosphorylation/degradation by cyclin-dependent kinase 2. *J. Biol. Chem.* 279, 19592–19599.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Perfetto, L., Briganti, L., Calderone, A., Cerquone Perpetuini, A., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T., et al. (2016). SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* 44, D548–D554.
- Perry, J.R., Laperriere, N., O'Callaghan, C.J., Brandes, A.A., Menten, J., Phillips, C., Fay, M., Nishikawa, R., Cairncross, J.G., Roa, W., et al. (2017). Short-course radiation plus temozolomide in elderly patients with glioblastoma. *N. Engl. J. Med.* 376, 1027–1037.
- Pilarczyk, M., Najafabadi, M.F., Kouril, M., Vasiliauskas, J., Niu, W., Shamsaei, B., Mahi, N., Zhang, L., Clark, N., Ren, Y., et al. (2019). Connecting Omics Signatures of Diseases, Drugs, and Mechanisms of Actions with iLINCs. *BioRxiv*, 826271.
- Pinto, M.L., Rios, E., Durães, C., Ribeiro, R., Machado, J.C., Mantovani, A., Barbosa, M.A., Carneiro, F., and Oliveira, M.J. (2019). The two faces of

- tumor-associated macrophages and their clinical significance in colorectal cancer. *Front. Immunol.* 10, 1875.
- Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11, 395.
- 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). DAReTE: a statistical tool for quantitative analysis of -omics data. *Bioinforma. Oxf. Engl.* 24, 1556–1558.
- Poulin, B., Sekiya, F., and Rhee, S.G. (2005). Intramolecular interaction between phosphorylated tyrosine-783 and the C-terminal Src homology 2 domain activates phospholipase C-gamma1. *Proc. Natl. Acad. Sci. U S A* 102, 4276–4281.
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
- Rieckmann, J.C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S.S., Lanzavecchia, A., Mann, M., et al. (2017). Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* 18, 583–593.
- 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.
- Ruepp, A., Waagstein, 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.
- Salomonis, N., Dexheimer, P.J., Omberg, L., Schroll, R., Bush, S., Huo, J., Schriml, L., Ho Sui, S., Kedda, M., Mayhew, C., et al. (2016). Integrated genomic analysis of diverse induced pluripotent stem cells from the progenitor cell biology consortium. *Stem Cell Rep.* 7, 110–125.
- Silva, T.C., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., and Noushmehr, H. (2016). TCGA Workflow: analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Res.* 5, 1542.
- Sokolov, A., Paull, E.O., and Stuart, J.M. (2016). One-class detection of cell states in tumor subtypes. *Pac. Symp. Biocomput.* 21, 405–416.
- 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.
- Stathias, V., Turner, J., Koleti, A., Vidovic, D., Cooper, D., Fazel-Najafabadi, M., Pilarczyk, M., Terryn, R., Chung, C., Umeano, A., et al. (2019). LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Res.* 48, D431–D439.
- Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* 352, 987–996.
- Stupp, R., Taillibert, S., Kanner, A., Read, W., Steinberg, D., Lhermitte, B., Toms, S., Idibaih, A., Ahluwalia, M.S., Fink, K., et al. (2017). Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA* 318, 2306–2316.
- Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.-A., Jones, D.T.W., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., et al. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22, 425–437.
- Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437–1452.e17.
- Sun, S.Q., Mashl, R.J., Sengupta, S., Scott, A.D., Wang, W., Batra, P., Wang, L.-B., Wyczalkowski, M.A., and Ding, L. (2018). Database of evidence for precision oncology portal. *Bioinforma. Oxf. Engl.* 34, 4315–4317.
- Tabb, D.L., Wang, X., Carr, S.A., Clouser, K.R., Mertins, P., Chambers, M.C., Holman, J.D., Wang, J., Zhang, B., Zimmerman, L.J., et al. (2016). Reproducibility of differential proteomic technologies in CPTAC fractionated xenografts. *J. Proteome Res.* 15, 691–706.
- The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- The Bioconda Team, Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valeris, R., and Köster, J. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* 15, 475–476.
- The Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068.
- Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967.
- 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.
- Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 177, 1035–1049.e19.
- Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110.
- Voena, C., Conte, C., Ambrogio, C., Boeri Erba, E., Boccalatte, F., Mohammed, S., Jensen, O.N., Palestro, G., Inghirami, G., and Chiarle, R. (2007). The tyrosine phosphatase Shp2 interacts with NPM-ALK and regulates anaplastic lymphoma cell growth and migration. *Cancer Res.* 67, 4278–4286.
- 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.
- Webb-Robertson, B.-J.M., Matzke, M.M., Metz, T.O., McDermott, J.E., Walker, H., Rodland, K.D., Pounds, J.G., and Waters, K.M. (2013). Sequential projection pursuit principal component analysis—dealing with missing data associated with new -omics technologies. *BioTechniques* 54, 165–168.
- 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.
- 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. *J. Open Source Softw.* 4, 1686.
- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinforma. Oxf. Engl.* 26, 1572–1573.
- 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.
- Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., et al. (2009). IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360, 765–773.
- 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.

- Yoo, S., Huang, T., Campbell, J.D., Lee, E., Tu, Z., Geraci, M.W., Powell, C.A., Schadt, E.E., Spira, A., and Zhu, J. (2014). MODMatcher: multi-omics data matcher for integrative genomic analysis. *PLoS Comput. Biol.* 10, e1003790.
- Zecha, J., Satpathy, S., Kanashova, T., Avanessian, S.C., Kane, M.H., Clouser, K.R., Mertins, P., Carr, S.A., and Kuster, B. (2019). TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. *Mol. Cell. Proteomics* 18, 1468–1478.
- Zhang, J., White, N.M., Schmidt, H.K., Fulton, R.S., Tomlinson, C., Warren, W.C., Wilson, R.K., and Maher, C.A. (2016). INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res.* 26, 108–118.
- Zhou, W., Laird, P.W., and Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 45, e22.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-IDH1-R132H (clone H09)	Dianova	Cat# DIA-H09; RRID: AB_2335716
Rabbit polyclonal anti-ATRX	Sigma	Cat# HPA001906; RRID: AB_1078249
Rabbit monoclonal anti-SOX9 (clone D8G8H)	Cell Signaling	Cat# 82630S; RRID: AB_2665492
Mouse monoclonal anti-GAB1 (clone H-7)	Santa Cruz	Cat# sc-13319; RRID: AB_2107855
Rabbit polyclonal anti-CD3 (clone A0452)	Dako	Cat# A00452; RRID: AB_2335677
Mouse monoclonal anti-CD68 (clone KP1)	Ventana	Cat# 790-2931; RRID: AB_2335972
Mouse monoclonal anti-CD163 (clone 10D6)	Novacastra	Cat# NCL-L-CD163; RRID: AB_2756375
Mouse monoclonal anti-PD-1 (clone NAT105)	Cell Marque	Cat# 315M-96; RRID: AB_1160829
Mouse monoclonal anti-PD-L1 (clone 22C3)	Dako	Cat# M3653; RRID: AB_2833074
Bacterial and virus strains		
Biological samples		
Primary tumor and normal tissue samples	This paper	See Methods: Experimental Model and Subject Details
Patient-derived xenograft tissue samples	Washington University in St. Louis	See Methods: Method Details
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
Ni-NTA Superflow Agarose Beads	Qiagen	Catalog: 30410
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

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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
Tris (hydroxymethyl)aminomethane hydrochloride pH 8.0	Sigma	Catalog: T2694
Trifluoroacetic acid	Sigma	Catalog: 91707
Urea	Sigma	Catalog: U0631
Critical commercial assays		
BCA Protein Assay Kit	ThermoFisher Scientific	Catalog: A53225
Infinium MethylationEPIC Kit	Illumina	Catalog: WG-317-1003
TMT-11 reagent kit	ThermoFisher Scientific	Catalog: A34808
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold	Illumina	Catalog: RS-122-2301
PTMScan® Acetyl-Lysine Motif [Ac-K] Kit	Cell Signaling	Catalog: 13416
Deposited data		
CPTAC GBM proteomic data	this study	https://cptac-data-portal.georgetown.edu/cptac/s/S048 ; https://cptac-data-portal.georgetown.edu/study-summary/S057 ; https://pdc.cancer.gov/
CPTAC GBM genomic and snRNA-seq data	this study	https://portal.gdc.cancer.gov/projects/CPTAC-3
Clinical and genomic data of the validation cohort	this study	https://pedcbiportal.kidsfirstdrc.org/study/summary?id=phgg_cbtc ; https://cavatica.sbggenomics.com/u/cavatica/pbta-cbtc/
Software and algorithms		
Ascore v1.0.6858	(Beausoleil et al., 2006)	https://github.com/PNNL-Comp-Mass-Spec/AScore
MASIC	(Monroe et al., 2008)	https://github.com/PNNL-Comp-Mass-Spec/MASIC
MS-GF+ v9981	(Kim and Pevzner, 2014)	https://github.com/MSGFPlus/msgfplus
mzRefinery	(Gibbons et al., 2015)	https://omics.pnl.gov/software/mzrefinery
BIC-Seq2	(Xi et al., 2016)	http://compbio.med.harvard.edu/BIC-seq/
GISTIC2 v2.0.22	(Mermel et al., 2011)	https://github.com/broadinstitute/gistic2
Strelka v2.9.2	(Kim et al., 2018)	https://github.com/Illumina/strelka
VarScan v2.3.8	(Koboldt et al., 2012)	https://dkoboldt.github.io/varscan/
Pindel v0.2.5	(Ye et al., 2009)	https://github.com/genome/pindel
MuTect v1.1.7	(Cibulskis et al., 2013)	https://github.com/broadinstitute/mutect
somaticwrapper v1.3 and v1.5	Li Ding Lab	https://github.com/ding-lab/somaticwrapper
Samtools v1.2	(Li et al., 2009)	https://www.htslib.org/
GATK v4.0.0.0	(McKenna et al., 2010)	https://github.com/broadgsa/gatk
bam-readcount v0.8	McDonnell Genome Institute	https://github.com/genome/bam-readcount
germlinewrapper v1.1	Li Ding Lab	https://github.com/ding-lab/germlinewrapper
Manta v1.6.0	(Chen et al., 2016)	https://github.com/Illumina/manta
DELLY v0.8.1	(Rausch et al., 2012)	https://github.com/dellytools/delly

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Telseq v0.0.1	(Ding et al., 2014)	https://github.com/zd1/telseq
HTSeq v0.11.2	(Anders et al., 2015)	https://github.com/simon-anders/htseq
EricScript v0.5.5	(Benelli et al., 2012)	https://sites.google.com/site/bioericscript/
INTEGRATE v0.2.6	(Zhang et al., 2016)	https://sourceforge.net/projects/integrate-fusion/
STAR-Fusion v1.5.0	(Haas et al., 2019)	https://github.com/STAR-Fusion/STAR-Fusion
BWA v0.7.17-r1188	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
CIRI v2.0.6	(Gao et al., 2015)	https://sourceforge.net/projects/ciri/
RSEM v1.3.1	(Li and Dewey, 2011)	https://deweylab.github.io/RSEM/
Bowtie2 v2.3.3	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
R-rollup	(Polipitiya et al., 2008)	https://omics.pnl.gov/software/danter
MODMatcher	(Yoo et al., 2014)	https://github.com/integrativenetworkbiology/Modmatcher
ConsensusClusterPlus v1.48.0	(Wilkerson and Hayes, 2010)	https://bioconductor.org/packages/ConsensusClusterPlus/
louvain-igraph v0.6.1	(Blondel et al., 2008)	https://doi.org/10.5281/zenodo.1054103
TCGAbiolinks v2.11.1	(Colaprico et al., 2016)	http://bioconductor.org/packages/TCGAbiolinks/
iProFun	(Song et al., 2019)	https://github.com/songxiaoyu/iProFun
BlackSheep	(Blumenberg et al., 2019)	https://github.com/ruggleslab/blackSheep
xCell v1.2	(Aran et al., 2017)	http://xcell.ucsf.edu/
CIBERSORTx	(Newman et al., 2019)	https://cibersortx.stanford.edu/
MoonlightR v1.12.0	(Colaprico et al., 2020)	http://bioconductor.org/packages/MoonlightR
CausalPath v.7c5b934	(Babur et al., 2018)	https://github.com/PathwayAndDataAnalysis/causalpath
Seurat v3.1.2	(Butler et al., 2018)	https://cran.r-project.org/web/packages/Seurat
edgeR v3.28.1	(Robinson et al., 2010)	https://www.bioconductor.org/packages/edgeR/
CLUE (data v1.1.1.2 and software v1.1.1.43)	(Subramanian et al., 2017)	https://clue.io/
iLINCs v2.7.0	(Pilarczyk et al., 2019)	https://www.ilincs.org/
R v3.6	R Development Core Team	https://www.R-project.org/
Bioconductor v3.9	(Huber et al., 2015)	https://bioconductor.org/
Tidyverse	(Wickham et al., 2019)	https://www.tidyverse.org/
Python v3.7	Python Software Foundation	https://www.python.org/
Bioconda	(The Bioconda Team et al., 2018)	https://bioconda.github.io/
Snakemake v5.6	(Köster and Rahmann, 2012)	https://snakemake.readthedocs.io/
ComplexHeatmap	(Gu et al., 2016)	https://www.bioconductor.org/packages/ComplexHeatmap/
scikit-learn v0.23.2	(Pedregosa et al., 2011)	https://scikit-learn.org/
Other		
RefSeq (downloaded from UCSC Genome Browser on 2018-06-29)	(O'Leary et al., 2016)	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., 2019)	https://www.gencodegenes.org/ ; https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
gnomAD v2.1	(Karczewski et al., 2020)	https://gnomad.broadinstitute.org/
The 1000 genomes project (final phase release on 2013-05-02)	(The 1000 Genomes Project Consortium, 2015)	https://www.internationalgenome.org/
OmniPath (downloaded on 2018-03-29)	(Türei et al., 2016)	http://omnipathdb.org/
DEPOD (downloaded on 2018-03-29)	(Duan et al., 2015)	http://depod.bioss.uni-freiburg.de/
CORUM (downloaded on 2018-06-29)	(Ruepp et al., 2010)	https://mips.helmholtz-muenchen.de/corum/
SIGNOR v2.0 (downloaded on 2018-10-29)	(Licata et al., 2019)	https://signor.uniroma2.it/
Reactome (downloaded on 2018-11-01)	(Fabregat et al., 2018)	https://reactome.org/
NetworkIN 3.0	(Horn et al., 2014)	https://networkin.info/
LINCS data portal (P100 Level 4)	(Stathias et al., 2019)	http://lincsportal.ccs.miami.edu/dcic-portal/

RESOURCE 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 raw proteomic data reported in this paper can be accessed via the CPTAC Data Portal at: <https://cptac-data-portal.georgetown.edu/cptac/s/S048>. Genomic, transcriptomic, and snRNA-seq data files can be accessed via Genomic Data Commons (GDC) at: <https://portal.gdc.cancer.gov/projects/CPTAC-3>. Clinical and processed genomic data of our validation cohort from Children's Brain Tumor Tissue Consortium (CBTTC) can be accessed via PedcBioPortal at: https://pedcbioportal.kidsfirstdrc.org/study-summary?id=phgg_cbttc and via CAVATICA at: <https://cavatica.sbggenomics.com/u/cavatica/pbta-cbttc/>. Proteomic data files can be accessed via Proteomic Data Commons (PDC) at: <https://pdc.cancer.gov/>. Processed data used in this publication can be found the CPTAC Data Portal at: <https://cptac-data-portal.georgetown.edu/study-summary/S057>, at Table S2, the cptac Python package, and LinkedOmics ([Vasaikar et al., 2018](#)).

ADDITIONAL RESOURCES

The CPTAC program website, which includes details about program initiatives, investigators, and datasets, can be accessed at: <https://proteomics.cancer.gov/programs/cptac>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Specimens and clinical data

Tumor and germline blood samples from 99 qualified cases were collected from 10 tissue source sites in strict accordance with the CPTAC-3 protocol with informed consent from the patients. No adjacent tissue was collected as part of this study, however, 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 = 55) and females (n = 44) from 6 different countries. Histopathologically defined adult glioblastoma tumors were only considered for analysis, with an age range of 24-88. Clinical data were obtained from the tissue source sites and reviewed for correctness and completeness of data.

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 (LN_2) then transferred to a liquid nitrogen freezer for storage until approval for shipment to the BCR.

Specimens were shipped using a cryoport that maintained an average temperature of under -140°C to the BCR 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 LN_2 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 percent 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 LN_2 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 LN_2 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.

Validation cohort sample collection and processing

High grade glioma adolescent and young adults (AYA) cohort was used in validation studies were collected through Children's Brain Tumor Tissue Consortium (CBTTC) sites including Children's Hospital of Philadelphia (CHOP), Seattle Children's Hospital, Meyer Children's Hospital, UCSF Benioff Children's Hospital, University of Pittsburgh, Lurie Children's Hospital, Children's National Medical Center, and through the HUP-CHOP Neurosurgery Tumor Tissue Bank Collaborative at the Hospital of University of Pennsylvania. All samples were fresh frozen collected 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.

METHOD DETAILS

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 QIAasympo DNA Mini Kit and QIAasympo 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 enough DNA yield to go through 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 (Fisher et al., 2011), 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 capturing > 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 RNA library preparation and sequencing

About 20–30 mg of cryopulverized powder from GBM specimens was resuspended in Lysis buffer (10 mM Tris-HCl (pH 7.4); 10 mM NaCl; 3 mM MgCl₂; and 0.1% NP-40). This suspension was pipetted gently for 6–8 times, incubated on ice for 30 seconds, and pipetted again for 4–6 times. The lysate containing free nuclei was filtered through a 40 µm cell strainer. We washed the filter with 1 mL Wash and Resuspension buffer (1X PBS + 2% BSA + 0.2 U/µL RNase inhibitor) and combined the flow through with the original filtrate. After a 6-minute centrifugation at 500 x g and 4°C, the nuclei pellet was resuspended in 500 µL of Wash and Resuspension buffer. After staining by DRAQ5, the nuclei were further purified by Fluorescence Activated Cell Sorting (FACS). FACS-purified nuclei were centrifuged again and resuspended in a small volume (about 30 µL). After counting and microscopic inspection of nuclei quality, the nuclei preparation was diluted to about 1,000 nuclei/µL. About 20,000 nuclei were used for single-nuclei RNA sequencing (snRNA-seq) by the 10X Chromium platform. We loaded the single nuclei onto a Chromium Chip B Single Cell Kit, 48 rxns (10x Genomics, PN-1000073) and processed them through the Chromium Controller to generate GEMs (Gel Beads in Emulsion). We then prepared the sequencing libraries with the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3, 16 rxns (10x Genomics, PN-1000075) following the manufacturer's protocol. Sequencing was performed on an Illumina NovaSeq 6000 S4 flow cell. The libraries were pooled and sequenced using the XP workflow according to the manufacturer's protocol with a 28x8x98bp sequencing recipe. The resulting sequencing files were available as FASTQs per sample after demultiplexing.

MS sample processing and data collection

Protein extraction and Lys-C/Trypsin tandem digestion

Approximately 50 mg of each of the cryopulverized tumor and normal 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 x 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, an aliquot of the same amount of sequencing-grade modified trypsin (Promega, V5117) 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 the final concentration of formic acid and centrifuged for 15 min at 1,500 x g at 4°C before transferring samples into new tubes leaving the resulting pellet behind. After 3 fold dilution with 0.1% formic acid, tryptic peptides were desalting on C18 SPE (Waters tC18 SepPak, WAT054925) and dried using Speed-Vac.

TMT-11 labeling of peptides

Desalted peptides from each sample were labeled with 11-plex TMT reagents (ThermoFisher Scientific). Peptides (400 µg) from each of the samples 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 according to the optimized TMT labeling protocol described previously ([Zecha et al., 2019](#)). 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, 60 µL 50 mM HEPES pH8.5, 20% ACN solution was added to dilute the samples, and 12 µL of 5% hydroxylamine was added and incubated for 15 min at RT to quench the labeling reaction. Peptides labeled by different TMT reagents were then mixed, dried using Speed-Vac, reconstituted with 3% acetonitrile, 0.1% formic acid and desalting on tC18 SepPak SPE columns.

Peptide fractionation by basic reversed-phase liquid chromatography (bRPLC)

Approximately 3.5 mg of 11-plex TMT labeled sample was separated on a reversed phase Agilent Zorbax 300 Extend-C18 column (250 mm x 4.6 mm column containing 3.5-µm particles) using the 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

Fe³⁺-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 in 500 µL IMAC binding/wash buffer (80% acetonitrile, 0.1%

trifluoroacetic acid) and incubated with 20 µL of the 50% bead suspension for 30 min at RT. After incubation, the beads were sequentially washed with 50 µL of wash buffer (1X), 50 µL of 50% acetonitrile, 0.1% trifluoroacetic acid (1X), 50 µL of wash buffer (1X), and 50 µL of 1% formic acid (1X) 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 K₂HPO₄, pH 7.0). Sixty microliter of 50% acetonitrile, 0.1% formic acid was used for elution of phosphopeptides from the C18 stage tips after two washes with 100 µL of 1% formic acid. 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 combined into four samples follow concatenation scheme by combining 3 fractions that were 4 fractions apart (i.e., combining fractions #1, #5 and #9 as a new fraction) and dried down 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 Na₂HPO₄ and 50 mM NaCl). After dissolving the peptide, the pH of the peptide solution was checked using pH indicator paper. 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 buffers 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, half of the antibody in each tube was transferred to the peptide solution and incubated on a rotator overnight at 4°C. After removing the supernatant, the reacted beads were washed with 1 mL of PBS buffer 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 two 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 13 µL of 2% acetonitrile, 0.1% formic acid contained 0.01% DDM (n-Dodecyl β-D-maltoside) right before the LC-MS/MS analysis.

The acetylated peptides prepared by IP from the IMAC flow-through may very well miss those peptides that are both phosphorylated and acetylated. Splitting the samples for independent IP and IMAC may improve the chance of recovering such peptides, assuming having both PTMs on the same peptide does not impact the affinity of either the IP or IMAC process. However, acetylated peptides are estimated to be 10 times lower in abundance than the phosphopeptides, hence much larger input may be needed to recover the dual-modified peptides. Given the extremely low stoichiometry of these dual-modified peptides and the sample size limitations, it was not pursued in this work.

LC-MS/MS analysis

Fractionated samples prepared for global proteome, phosphoproteome, and acetylome analysis were separated 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 profile (min: %MP B): 0:2, 1:6, 85:30, 94:60, 95:90, 100:90, 101:50, 110:50 (for global proteome and phosphoproteome analysis); 0:2, 1:6, 235:30, 244:60, 245:90, 250:90, 251:50, 260:50 (for acetylome analysis). 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 4 × 10⁵) 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 1 × 10⁵) 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 except that the max ion time was 125 ms.

Construction and utilization of the comparative reference samples

As a quality control measure, two different types of “Comparative Reference” (“CompRef”) patient-derived xenograft (PDX) samples were generated as previously described (Li et al., 2013; Tabb et al., 2016) and used to monitor the longitudinal performance of the proteomics workflow throughout the course of this study. Briefly, the PDX tumors from established basal and luminal breast cancer intrinsic subtypes were raised subcutaneously in 8-week old NOD.Cg-Prkdc^{scid} II2rg^{tm1Wjl}/SzJ mice (Jackson Laboratories, Bar Harbor, ME) using procedures reviewed and approved by the Institutional Animal Care and Use Committee at Washington University in St. Louis. Xenografts were grown in multiple mice, pooled, and cryopulverized to provide a sufficient amount of uniform material for the duration of the study. Full proteome, phosphoproteome and acetylome process replicates of each of the two types of CompRef samples were prepared and analyzed as standalone 11-plex TMT experiments alongside every 4 TMT-11 experiments of the study samples, using the same analysis protocol as the patient samples. These interstitially analyzed CompRef samples were evaluated for depth of proteome, phosphoproteome, and acetylome coverage and for consistency in quantitative comparison between the basal and luminal models.

Global proteome and phosphoproteome analysis of high grade brain tumor samples

An independent cohort of 39 high grade (age 14–39 years old) brain tumors obtained from the Children’s Brain Tumor Tissue Consortium (CBTC) were analyzed using the same procedures for TMT-based quantitative global proteome and phosphoproteome analysis of the adult GBM and normal brain tissue samples, with modifications in front-end protein extraction and digestion described as follows: Approximately 50 mg of each of brain tumor tissues were cryopulverized and lysed separately in 800 μ L of lysis buffer (6 M urea, 25 mM Tris, pH 8.0, 1 mM EDTA, 1 mM EGTA, 1:100 v/v Sigma protease inhibitor, 1:100 v/v Sigma phosphatase inhibitor cocktail 2, and 1:100 v/v Sigma phosphatase inhibitor cocktail 3). Lysates were precleared by centrifugation at 20,000 $\times g$ for 10 min at 4°C. After adjusting the protein concentration to approximately 1.5 μ g/ μ L, 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 to 2 M urea concentration with 25 mM Tris, pH 8.0 and digested with Lys-C at 1:50 enzyme-to-substrate ratio. After 2 h of digestion at 25°C, aliquot of sequencing grade modified trypsin at 1:25 enzyme-to-substrate ratio was added to the samples and further incubated at 25°C for 14 h.

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 30 mg using a modified Folch extraction (Nakayasu et al., 2016). 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. Sample 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 $\times g$ 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 μ L 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 (Webb-Robertson et al., 2014). To protect carbonyl groups and reduce the number of tautomeric isomers, 20 μ L 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 μ L 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 analysed 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 (Webb-Robertson et al., 2014). Briefly, 1 μ L of each sample was injected onto a HP-5MS column (30 m \times 0.25 mm \times 0.25 μ m; 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 μ L chloroform plus 95 μ L of methanol. The TLEs were analyzed as outlined in the previous study (Kyle et al., 2017). 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 μ L of reconstituted sample was injected onto a Waters CSH column (3.0 mm \times 150 mm \times 1.7 μ m 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 μ L/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 (Webb-Robertson et al., 2014). GC-MS raw data files were processed using Metabolite Detector software v2.0.6 beta (Hiller et al., 2009). 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 (Kind et al., 2009), 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) (Kyle et al., 2017). 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 (Pluskal et al., 2010). Aligned features were manually verified and peak apex intensity values were exported for subsequent statistical analysis.

Immunohistochemistry (IHC) validation of genetic alteration downstream impact and immune cell compositions

IHC stains for IDH1, ATRX, SOX9, GAB1, CD3, CD68, CD163, PD-1, and PD-L1 were performed at the Johns Hopkins Hospital clinical IHC laboratory using the autostainers (Ventana XT and Dako). Briefly, tissue blocks were cut into 5-micron thickness sections prior to incubation with primary antibodies. Heat antigen retrieval was performed to enhance signal detection. Primary antibodies were diluted according to standard protocols and/or manufacturer suggestions. A mouse-HRP and/or rabbit-AP polymer detection systems were used to develop immunostaining. Slides were counterstained with hematoxylin and dehydrated for permanent mounting. Appropriate positive and negative controls were also included during the assay.

QUANTIFICATION AND STATISTICAL ANALYSIS

Tumor exclusion criteria

One sample (C3L-03747) was excluded from the downstream analysis since it failed the expert pathology review (high necrosis) and had low correlation of RNA and protein or phosphoprotein.

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

Copy Number Variant (CNV) were detected using BIC-Seq2 (NBICseq-norm v0.2.4 and NBICseq-seg v0.7.2) (Xi et al., 2016) from WGS tumor and normal paired BAMs using Li Ding Lab's BIC-Seq2 pipeline v2.0 <https://github.com/ding-lab/BICSEQ2>. We used a bin size of 100bp and a lambda of 3 (smoothing parameter for CNV segmentation). To further summarize the arm-level copy number change, we used a weighted sum approach (Vasalikar et al., 2019), in which the segment-level \log_2 copy ratios for all the segments located in the given arm were added up with the length of each segment being weighted. We then used GISTIC2 v2.0.22 (Mermel et al., 2011) to integrate results from individual patients and identify genomic regions recurrently amplified or deleted in our samples. The threshold of arm-level CNV was 0.3 for gain and -0.3 for loss.

We defined a tumor with chr7 amplification or chr10 deletion when the GISTIC results of at least one chromosome arm exceeded the threshold (± 0.3). For samples that both chromosome arms were within the GISTIC threshold or its GISTIC result was not available, we considered the tumor with the chromosome amplification or deletion if the averaged CNV values exceeded the threshold (± 0.2).

Somatic variant calling

Somatic variants were called from WES tumor and normal paired BAMs using somaticwrapper v1.5, 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 (Kim et al., 2018), VarScan v2.3.8 (Koboldt et al., 2012), Pindel v0.2.5 (Ye et al., 2009), and MuTect v1.1.7 (Cibulskis et al., 2013). 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 ≤ 0.02 and tumor VAF ≥ 0.05
- Read depth in tumor ≥ 14 and normal ≥ 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

We additionally called somatic whole-genome variants using WGS tumor and normal paired BAMs using somaticwrapper v1.3 identical to the exome version except that we kept non-exonic variants.

Germline variant calling and annotation

Germline variant calling was performed using Li Ding Lab's pipeline germlinewrapper v1.1, which implements multiple tools for the detection of germline INDELS and SNVs. Germline SNVs were identified using VarScan v2.3.8 (with parameters: `-min-var-freq 0.10 -p-value 0.10, -min-coverage 3 -strand-filter 1`) operating on a mpileup stream produced by samtools v1.2 (with parameters: `-q 1 -Q 13`) and GATK v4.0.0.0 (McKenna et al., 2010) using its haplotype caller in single-sample mode with duplicate and unmapped reads removed and retaining calls with a minimum quality threshold of 10. All resulting variants were limited to the coding region of the full-length transcripts obtained from Ensembl release 95 plus additional two base pairs flanking each exon to cover splice donor/acceptor sites. We required variants to have allelic depth ≥ 5 reads for the alternative allele in both tumor and normal samples. We used bam-readcount v0.8 for reference and alternative alleles quantification (with parameters: `-q 10 -b 15`) in both normal and tumor samples. Additionally, we filtered all variants with $\geq 0.05\%$ frequency in gnomAD v2.1 (Karczewski et al., 2020) and The 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015).

TERT promoter mutation calling

We used bam-readcount to count reads in WGS tumor and blood normal BAMs at the known hotspot positions at hg38 chr5:1295113 and chr5:1295135. We called a mutation if it was not observed in matching blood normal BAM and VAF > 5%. For all tumor samples lacking a *TERT*p hotspot mutation, we performed the readcount across the entire *TERT* promoter region from chr5:1294200 to chr5:1295601 (hg38). In these cases, we applied a more stringent VAF cutoff of 10%.

Structural variant calling

Structural variants (SVs) were called by Manta v1.6.0 (Chen et al., 2016) and DELLY v0.8.1 (Rausch et al., 2012) from WGS tumor and normal paired BAMs. We ran Manta on canonical chromosomes with the default record- and sample-level filters. For DELLY, we followed somatic SV workflow Only SV calls with PASS filter status were kept for downstream analysis. Lastly, we manually reviewed all the SV calls in the genes of interest (e.g. EGFR and PDGFRA).

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 (Bady et al., 2012) 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 $M = \log_2(\text{beta} / (1 - \text{beta}))$. 9 tumors with low data quality of DNA methylation array were excluded from the prediction. 38 out of the remaining 90 tumors (42%) were predicted to be MGMT promoter DNA hypermethylated.

Telomere length quantification and telomere genotyping

We used Telseq v0.0.1 (Ding et al., 2014) to estimate the telomere length using WXS and WGS tumor and blood normal paired BAMs. We defined telomere length ratio as the ratio between the estimated tumor telomere length and the estimated blood normal telomere length. While WXS and WGS-based telomere length ratios were well correlated, we used WGS based lengths for the downstream analysis. We defined long telomere phenotype as tumors with WGS telomere length ratio > 1.2 , and short telomere phenotype as WGS telomere length ratio < 0.8 .

We identified telomere genotypes as the following:

- *TERT*p hotspot if tumor has *TERT*p hotspot mutation
- *ATRX*mut for all remaining tumors with only *ATRX* mutation
- *ATRX*mut *IDH1*mut for all remaining tumors with both *ATRX* and *IDH1* mutated
- *IDH1*mut for all remaining tumors with only *IDH1* mutation
- *TERT*p not hotspot for all remaining tumors without hotspot mutation in *TERT* promoter and expressing *TERT*
- WT for the remaining tumors that do not fall into any category

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 (Anders et al., 2015) 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 downloaded 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.

RNA fusion detection

We used three callers, STAR-Fusion v1.5.0 (Haas et al., 2019), INTEGRATE v0.2.6 (Zhang et al., 2016), and EricScript v0.5.5 (Benelli et al., 2012), to call consensus fusion/chimeric events in our samples. Calls by each tool using tumor and normal RNA-Seq data

were then merged into a single file and extensive filtering is done. As STAR-Fusion has higher sensitivity, calls made by this tool with higher supporting evidence (defined by fusion fragments per million total reads, or FFPM > 0.1) were required, or a given fusion must be reported by at least 2 callers. We then removed fusions present in our panel of blacklisted or normal fusions, which included uncharacterized genes, immunoglobulin genes, mitochondrial genes, and others, as well as fusions from the same gene or paralog genes and fusions reported in TCGA normal samples (Gao et al., 2018), GTEx tissues (reported in STAR-Fusion output), and non-cancer cell studies (Babiceanu et al., 2016). Finally, we removed normal fusions from the tumor fusions to curate the final set.

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 pipeline 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.

Circular RNA prediction and quantification

The hg38 reference genome and GDC's annotations were used for the circRNA analysis. First, CIRI v2.0.6 (Gao et al., 2015) was used to call circular RNA with default parameters and BWA v0.7.17-r1188 (Li and Durbin, 2009) was used as a mapping tool. The cutoff of supporting reads for circRNA was set to 10. Then a pseudo-linear transcript strategy was used to quantify circular RNA expression (Li et al., 2017). In brief, for each sample, linear transcripts of circular RNAs were extracted and 75bp (read length) from the 3' end was copied to the 5' end. The modified transcripts were called pseudo-linear transcripts. Transcripts of linear genes were also extracted and mixed with pseudo-linear transcripts. RSEM v1.3.1 (Li and Dewey, 2011) with Bowtie2 v2.3.3 (Langmead and Salzberg, 2012) as the mapping tool was used to quantify circular RNA expression based on the mixed transcripts. After quantification, the upper quantile method was applied for normalization and the normalized matrix was log₂-transformed.

snRNA-seq quantification and analysis

snRNA-seq data preprocessing

For each sample, we obtained the unfiltered feature-barcode matrix per sample by passing the demultiplexed FASTQs to Cell Ranger v3.1.0 '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 pre-built human genome reference 3.0.0 (GRCh38 and Ensembl 93).

Seurat v3.1.2 (Butler et al., 2018; Hafemeister and Satija, 2019) was used for all subsequent analysis. 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%).

Each sample was scaled and normalized using Seurat's 'SCTransform' function to correct for batch effects (with parameters: var.s.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 (Blondel et al., 2008) and the top 30 PCA dimensions via Seurat's 'FindNeighbors' and 'FindClusters' (with parameters: resolution = 0.5) functions. The resulting merged and normalized matrix was used for the subsequent analysis.

snRNA-seq cell type annotation

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, GABA2, 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).

snRNA-seq analysis

Differentially expressed genes within TAM cells, T-cells and Tumor cells were identified by FindMarkers function comparing cells belonging to one subtype (immune subtype or multi-omics subtype) to the rest. Wilcoxon statistical test was used. log₂FC > 0.25 and FDR < 0.05 was used to filter DEGs.

MS data interpretation

Quantification of TMT global proteomics data

LC-MS/MS analysis of the TMT11-labeled, bRPLC fractionated samples generated a total of 264 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 (Gibbons et al., 2015; Kim and Pevzner, 2014; Kim et al., 2008) 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 \pm 10 ppm parent ion tolerance, allowed for isotopic error in precursor ion selection, and searched a decoy database 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. Inference of parsimonious protein set at gene level resulted in the identification of protein groups covering 11,141 genes.

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 \log_2 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, \log_2 relative protein abundance for each sample and re-centered to achieve a common median of 0.

Quantification of phosphopeptides

Phosphopeptide identification for the 132 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 (Beausoleil et al., 2006) 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 44 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.

Preprocessing of proteomics tables

Due to the quantification of small values close to 0 on spectrum level, some extreme positive or negative values were generated after \log_2 transform of relative protein/phosphopeptide/acetyl peptide abundance, which may have negative impact on the downstream analysis of the data sets. To identify TMT outliers with extreme values, we perform inter-TMT *t*-test for each individual protein/phosphopeptide/acetyl peptide. For a specific protein/phosphopeptide/acetyl peptide, relative abundance level of each TMT value was compared against all the other TMT values using Spearman two-sample test. Outlier was defined if the p-value passed a certain threshold. In the global proteome data, 153 TMT values were identified as outliers with inter-TMT *t*-test p-value lower than 10e-6, as a result 1,530 data points (0.14% of all observations) were removed from the data sets. In the phosphoproteome data, 379 TMT values were identified as outliers with inter-TMT *t*-test p-value lower than 10e-10, resulting in 3,790 data points (0.09% of all observations) removed from the data sets. In the acetylome data, 12 TMT values were identified as outliers with inter-TMT *t*-test p-value lower than 10e-14, and 120 data points (0.015% of all observations) were removed from the data sets.

Batch effects were checked using the \log_2 relative protein/phosphopeptide abundance or protein/acetyl peptide abundance, and removed using Combat algorithm (Beausoleil et al., 2006) after TMT outlier filtering. Imputation was performed after batch effect correction to produce a different version of the data tables for some of the data analysis tools that are sensitive to missing values. The proteins/phosphopeptide/acetyl peptide with missing rate less than 50% were selected and imputed with the DreamAI algorithm <https://github.com/WangLab-MSSM/DreamAI> tailored for proteomics data.

Sample outlier identification of metabolome and lipidome

A robust Mahalanobis distance based on biomolecule abundance vectors (rMd-PAV) was calculated to identify potential sample outliers in the data (Matzke et al., 2011). For proteomics data, this distance was calculated based on four metrics: average correlation with samples in the same group, skewness of biomolecule abundance distribution, the proportion of missing data, and

median absolute deviation of abundances. These metrics, minus the proportion missing, were used for the metabolomics and lipidomics datasets. To confirm any sample outliers identified by rMd-PAV, a correlation heatmap was generated and sequential projection pursuit principal component analysis (PCA) was run (Webb-Robertson et al., 2013). No sample outliers were identified in the proteomics dataset. One outlier, C3N-01366, was removed from the metabolomics dataset; C3N-01370 was removed from the positive lipid dataset and C3L-03968 from the negative lipid dataset.

Normalization and protein quantification 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 (Polpitiya et al., 2008), in which peptides were scaled by a reference peptide and the protein abundance was set as the median of the scaled peptides.

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 as described previously (Clark et al., 2019). Using MODMatcher (Yoo et al., 2014), we confirmed that all samples were well aligned among RNA-Seq, proteomics and CNV (WGS) data.

Ancestry prediction using SNPs from 1000 genomes project

We used a reference panel of genotypes and a clustering based on principal components to identify likely ancestry. We selected 107,765 coding SNPs with a minor allele frequency > 0.02 from the final phase release of The 1000 Genomes Project (1000 Genomes Project Consortium et al., 2010). From this set of loci, we measured the depth and allele counts of each sample in our cohort using bam-readcount v0.8.0. Genotypes were then called for each sample based on the following criteria: 0/0 if reference count ≥ 8 and alternate count < 4; 0/1 if reference count ≥ 4 and alternate count ≥ 4 ; 1/1 if reference count < 4 and alternate count ≥ 8 ; and ./ (missing) otherwise. After excluding markers with missingness > 5%, 70,968 markers were kept for analysis. We performed PCA on the 1000 Genomes samples to identify the top 20 principal components. We then projected our cohort onto the 20-dimensional space representing the 1000 Genomes data. We then trained a random forest classifier with the 1000 Genomes dataset using these 20 principal components. The 1000 Genomes dataset was split 80/20 for training and validation respectively. On the validation dataset our classifier achieved 99.6% accuracy. We then used the fitted classifier to predict the likely ancestry of our cohort.

Multi-omics subtyping using non-negative matrix factorization (NMF)

We selected the following proteogenomic features to the sample availability: CNV, bulk RNA, protein, and phosphoprotein expression. Due to limited sample amounts, not all tumors were analyzed for DNA methylome, metabolome, and miRNA. We used non-negative matrix factorization (NMF) implemented in the NMF R-package (Gaujoux and Seoighe, 2010) to perform unsupervised clustering of tumor samples and to identify proteogenomic features that show characteristic expression patterns for each cluster. Briefly, given a factorization rank k (where k is the number of clusters), NMF decomposes a $p \times n$ data matrix V into two matrices W and H such that multiplication of W and H approximates V . Matrix H is a $k \times n$ matrix whose entries represent weights for each sample (1 to N) to contribute to each cluster (1 to k), whereas matrix W is a $p \times k$ matrix representing weights for each feature (1 to p) to contribute to each cluster (1 to k). Matrix H was used to assign samples to clusters by choosing the k with maximum score in each column of H . For each sample, we calculated a cluster membership score as the maximal fractional score of the corresponding column in matrix H . We defined a “cluster core” as the set of samples with cluster membership score > 0.5 . Matrix W containing the weights of each feature to a certain cluster was used to derive a list of representative features separating the clusters using the method proposed in (Kim and Park, 2007).

To enable integrative multi-omics clustering we enforced all data types (and converted if necessary) to represent log₂-ratios to either a common reference measured in each TMT plex (proteome, phosphoproteome), an *in silico* common reference calculated as the median abundance across all samples (RNA gene expression) or to gene copy numbers relative to matching normal blood sample (CNV). All data tables were then concatenated and all rows containing missing values were removed. To remove uninformative features from the dataset prior to NMF clustering, we removed features with the lowest standard deviation (bottom 5th percentile) across all samples. Each row in the data matrix was further scaled and standardized such that all features from different data types were represented as z-scores.

Since NMF requires a non-negative input matrix we converted the z-scores in the data matrix into a non-negative matrix as follows:

- 1) Create one data matrix with all negative numbers zeroed
- 2) Create another data matrix with all positive numbers zeroed and the signs of all negative numbers removed
- 3) Concatenate both matrices resulting in a data matrix twice as large as the original, but with positive values only and zeros and hence appropriate for NMF

The resulting matrix was then passed to NMF analysis in R using the factorization method described in (Brunet et al., 2004). To determine the optimal factorization rank k (number of clusters) for the multi-omic data matrix, we tested a range of clusters between $k = 2$ and 8. For each k , we factorized matrix V using 50 iterations with random initializations of W and H . To determine the optimal

factorization rank, we calculated cophenetic correlation coefficients to measure how well the intrinsic structure of the data is recapitulated after clustering. Finally, we picked the k with maximal cophenetic correlation for cluster numbers between $k = 3$ and 8.

To achieve robust factorization of the multi-omics data matrix V , we took the optimal factorization rank k , repeated the NMF analysis for 200 iterations with random initializations of W and H , and partitioned the samples into clusters as described above. Due to the non-negative transformation of the z-scored data matrix, feature weight matrix W contained two separate weights for positive and negative z-scores of each feature, respectively. To revert the non-negative transformation and to derive a single signed weight for each feature, we first normalized each row in matrix W by dividing by the sum of feature weights in each row, aggregated both weights per feature and cluster by keeping the maximal normalized weight and multiplication with the sign of the z-score the initial data matrix. Thus, the resulting transformed matrix W_{signed} contained signed cluster weights for each feature in the input matrix.

For each cluster, we calculated normalized enrichment scores (NES) of cancer-relevant gene sets by projecting the matrix of signed multi-omic feature weights W_{signed} onto hallmark pathway gene sets (Liberzon et al., 2015) using ssgSEA (Barbie et al., 2009) available on <https://github.com/broadinstitute/ssGSEA2.0> (parameters: gene.set.database="h.all.v6.2.symbols.gmt" sample.norm.type="rank" weight=1 statistic="area.under.RES" output.score.type="NES" nperm=1000 global.fdr=TRUE min.overlap=5 correl.type="z.score"). To derive a single weight for each gene measured across all omics data types, we retained the weight with maximal absolute amplitude. We then associated the resulting clusters to sample-level variables by testing for overrepresentation in the cluster core sample sets using Fisher's exact test. The following clinical variables were used: expression subtype, sex, vital status, and smoking history.

The entire NMF workflow has been implemented as a module on Broad's Cloud platform Terra (<https://app.terra.bio/>). The docker containers encapsulating the source code and the required R packages for NMF clustering and ssGSEA were available on Docker-hub (broadcptac/pgdac_mo_nmf:9, broadcptac/pgdac_ssgsea:5).

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 (Wang et al., 2017), which contained 50 highly expressed genes in classical, proneural, and mesenchymal IDH WT tumors. Tumors with recurrent mutations in IDH1/2 (IDH1 R132H specifically in our cohort) were assigned to be IDH mutant tumors. We then performed consensus clustering on all tumors based on the selected gene expression in $\log_2(\text{FPKM-UQ} + 1)$ 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 $k = 3$.

Unsupervised clustering of DNA methylation

Methylation subtypes were segregated based on the top 8,000 most variable probes using k-means consensus clustering as previously described (Sturm et al., 2012). We first removed underperforming probes (Zhou et al., 2017), and then the samples with more than 30% missing values. Remaining missing values were imputed using the mean of the corresponding probe value. We then performed clustering 1000 times using the ConsensusClusterPlus R package (parameters: maxK = 10 reps = 1000 pItem = 0.8 pFeature = 1 clusterAlg = "km" distance = "euclidean"). We choose $k = 6$ based on the delta area plot of consensus CDF.

MolecularNeuroPathology (MNP) DNA Methylation Classification of Central Nervous System (CNS) Tumors

We applied the existing DNA methylation classification of CNS tumors developed by the MolecularNeuroPathology group (Capper et al., 2018) to our cohort. The processed microarray beta values and the classification of the MNP cohort (v11b2) were downloaded from the GEO dataset GSE90496 (the MNP reference set) and the supplemental tables of the original publication, consisting of total 600 GBM tumors and control samples. We included probes from CPTAC samples with < 20% of missing values across all samples. Remaining missing values were imputed using the mean of the corresponding probe value. Top 10,000 variable probes out of the total 375,969 shared probes across MNP and CPTAC GBM samples were selected to construct the shared DNA methylome space. We used PCA (parameters: random_state = 202012) to extract the top 30 principal components and assigned the MNP classification to CPTAC samples using nearest neighbor implemented by scikit-learn (KNeighborsClassifier with parameters: k = 9, algorithm='brute'). While an official MNP DNA methylation classifier exists online (<https://www.molecularneuropathology.org/mnp>), we were not able to access it since our registration was not approved by the site at the time of writing.

Unsupervised clustering of miRNA expression

Unsupervised miRNA expression subtype identification was performed on mature miRNAs expression ($\log_2 \text{TPM}$) from 98 tumors with miRNA-seq available using Louvain clustering (Blondel et al., 2008) implemented in louvain-igraph v0.6.1. Top 50 differentially expressed miRNAs from each miRNA-based subtype were selected (Table S3).

Unsupervised clustering of individual data type

We also applied clustering across all tumors using individual data types including RNA, protein, phosphoprotein, acetylprotein, lipids, and metabolites. The DreamAI-imputed expression values were used for protein, phosphoprotein and acetylprotein. For other data types, features with too many missing values were discarded (RNA: 40%; metabolites: 80% of all tumors). We then selected the top 95% most variable remaining features to perform consensus clustering using ConsensusClusterPlus R package (parameters: maxK = 10 reps = 1000 pItem = 0.8 pFeature = 1 clusterAlg = "km" distance = "euclidean" seed = 201909). Optimal number of k was selected based on the delta area plot of consensus CDF. The resulting clusters per data type were shown in Figure S2F.

Determination of stemness score

Stemness scores were calculated as previously described (Malta et al., 2018). Firstly, we used MoonlightR (Colaprico et al., 2020) to query, download, and preprocess the pluripotent stem cell samples (ESC and iPSC) from the Progenitor Cell Biology Consortium (PCBC) dataset (Daily et al., 2017; Salomonis et al., 2016). Secondly, to calculate the stemness scores based on mRNA expression, we built a predictive model using one-class logistic regression (OCLR) (Sokolov et al., 2016) on Progenitor Cell Biology Consortium (PCBC) dataset. For mRNA expression-based signatures, to ensure compatibility with our cohort, we first mapped the Ensembl IDs to Human Genome Organization (HUGO) gene names and dropped any genes that had no such mapping. The resulting training matrix contained 12,945 mRNA expression values measured across all available PCBC samples. To calculate mRNA-based sternness index (mRNAsI), we used FPKM-UQ mRNA expression values for all CPTAC GBM tumors and GTEx samples. We used TCGAanalyze_Stemness function from the R package TCGAbiolinks (Colaprico et al., 2016) and following our previously described workflow (Silva et al., 2016), with "stemSig" argument set to PCBC stemSig.

Multi-omics *cis* association analysis using iProFun

We integrated somatic mutation, CNV, DNA methylation, RNA, protein, phosphorylation (phospho) and acetylation (acetyl) levels via iProFun (Song et al., 2019) to investigate the functional impacts of DNA alterations in GBM. All data types were preprocessed to eliminate potential issues for analysis such as batch effects, missing data and major unmeasured confounding effects before the iProFun analysis. As phosphoprotein and acetylprotein were measured in a small subset of the genes in comparison with RNA and protein, we considered three sets of iProFun analysis using different combination functional outcomes (mRNA/protein, mRNA/protein/phospho, and mRNA/protein/acetyl) to include as many as possible genes and omics for investigation. For each set of outcomes (e.g. RNA and protein), we considered their levels perturbed jointly by three DNA alterations (somatic mutation, CNV, and DNA methylation). The effects of DNA methylation on molecular traits are usually smaller than mutation and CNV, and thus adjusting their effects in analysis is critical to obtain unconfounded associations for methylation. In addition, we adjusted age, sex, and tumor purity in the analysis. Tumor purity was determined using xCell (Aran et al., 2017) from RNA-Seq data.

The iProFun procedure was applied to a total of 7,464 genes with measured RNA/protein, 4,433 genes with measured RNA/protein/phospho, and 1,315 genes with measured RNA/protein/acetyl data, respectively, for their *cis* regulatory patterns in tumors. For example, when we considered DNA methylation for its effects on RNA/protein/phospho, we started with the traditional linear regression for each of the three outcomes separately:

$$\text{RNA} \sim \text{methylation} + \text{covariates}$$

$$\text{protein} \sim \text{methylation} + \text{covariates}$$

$$\text{phospho} \sim \text{methylation} + \text{covariates}$$

The covariates here include CNV, somatic mutations (genes with mutation rate $\geq 10\%$), age, sex, and tumor purity. Then iProFun took the association summary statistics from these three regressions as input to call posterior probabilities of belonging to each of the eight possible configurations (e.g., “None”, “RNA only”, “protein only”, “phospho only” “RNA & protein”, “RNA & phospho”, “protein & phospho” and “all three”) and to determine significance associations.

A gene was identified to present significant and biologically meaningful association if the association passes three criteria: (1) the satisfaction of biological filtering procedure, (2) posterior probabilities $> 75\%$, and (3) empirical false discovery rate (eFDR) $< 10\%$. Specifically, the biological filtering criterion requires that CNV presents positive associations with all the types of molecular quantitative traits (QTs), DNA methylation presents negative associations with all the types of molecular QTs, and mutation requires the association across all outcome platforms preserve consistent directions (either positive or negative). Secondly, a significance was called only if the posterior probabilities $> 75\%$ of a predictor being associated with a molecular QT, by summing over all configurations that are consistent with the association of interest. For example, the posterior probability of a methylation being associated with mRNA expression levels was obtained by summing up the posterior probabilities in the following four association patterns – “RNA only”, “RNA & protein”, “RNA & phospho” and “all three”, all of which were consistent with methylation being associated with mRNA expression. Lastly, we calculated the empirical FDR (eFDR) via 100 permutations per molecular QTs by shuffling the label of the molecular QTs and required eFDR $< 10\%$ by selecting a minimal cutoff value of α that $75\% < \alpha < 100\%$. The eFDR is calculated by:

$$\text{eFDR} = \frac{\text{Average } \# \text{genes with posterior probability } > \alpha \text{ in permuted data}}{\text{Average } \# \text{genes with posterior probability } > \alpha \text{ in original data}}$$

Table S3 presents the results of whether the DNA methylation/CNV/mutation of a gene has perturbed any of its cis QTs (mRNA, protein, phosphoprotein and acetylprotein).

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

We aggregated a set of interacting proteins (e.g. kinase/phosphatase-substrate or complex partners) from OmniPath (downloaded on 2018-03-29) (Türel et al., 2016), DEPOD (downloaded on 2018-03-29) (Duan et al., 2015), CORUM (downloaded on 2018-06-29) (Ruepp et al., 2010), Signor2 (downloaded on 2018-10-29) (Perfetto et al., 2016), and Reactome (downloaded on 2018-11-01) (Fabregat et al.,

2018). We focused our analyses on 18 GBM SMGs previously reported in the literature: PIK3R1, PIK3CA, PTEN, RB1, TP53, EGFR, IDH1, BRAF, NF1, PDGFRA, ATRX, and TERTp (Bailey et al., 2018; Brennan et al., 2013).

For each interacting protein pair, we split samples with and without mutations in partner A and compare expression levels (RNA, protein and phosphosites) both in *cis* (partner A) and in *trans* (partner B), calculating a median difference in expression and testing for significance with the Wilcoxon rank sum test, with the Benjamini-Hochberg multiple test correction. For mutational impact analysis on lipidome or metabolome, all possible pairs between SMGs and metabolites/lipids were tested.

Protein and RNA marker identification for multi-omics mixed subtype

By comparing the tumors of multi-omics mixed subtypes (nmf_cluster_membership score ≤ 0.55) to other (non-mixed) tumors, we identified 276 differentially expressed genes and 690 differentially expressed proteins. For each differentially expressed gene/protein, we binned all the tumors into three groups based on their expression level: high, medium, and low. We then compared the survival outcome (log-rank test) in the high expression group to the low expression group using the functions TCGAanalyze_divideGroups(), TCGAanalyze_SurvivalKM(), and TCGAanalyze_survival() from TCGAbiolinks (Colaprico et al., 2016), and the function surv_median() from the R package survminer. We identified 19 genes and 40 proteins with significantly differential survival outcome (Table S3).

Kinase-substrate pairs 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 20 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 A_i depends on kinase protein expression S_i and error E_i ,

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

For the i -th trial for kinase phosphosite abundance in the *trans* associations, substrate phosphosite abundance A_i depends on kinase phosphosite expression K_i , substrate protein expression S_i and error E_i ,

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

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.

For the broader investigation of signaling cascades, we included total 214 kinases and 43 phosphatases if they satisfied either of the genetic alteration criteria or at least three criteria below:

- 5% and more tumors with copy number alterations
- 2 and more tumors with somatic mutations
- Top 20% variable gene expression
- Top 35% variable protein expression
- Significantly different RNA or protein expression between tumor and normal (FDR ≤ 0.01)

Differential proteomic, phosphosite, metabolome and lipidome analysis

TMT-based global proteomic, phosphoproteomic, and acetylation, as well as metabolome and lipidome data were used to perform pairwise differential analysis between groups of samples. A Wilcoxon rank-sum test was performed to determine differential abundance of proteins, PTMs and metabolites. At least four samples in both groups were required to have non-missing values and the p-value was adjusted using the Benjamini-Hochberg procedure. For phosphorylation markers in each genomic subtype, the adjusted p-value for the protein change was required to be ≥ 0.05 .

Phosphoproteome outlier analysis

Outlier Analysis was done using BlackSheep's DEVA analysis (Blumenberg et al., 2019). Phosphopeptide analysis was done on data that was aggregated per protein, summing together outlier values across all phosphosites. Protein analysis was performed using TMT-based global proteomic data, RNA analysis was done using FPKM-UQ normalized transcript data. The DEVA method calculates interquartile range (IQR) and median values for the given dataset, and then defines outliers as values greater than the median plus 1.5x IQR. Features were prefiltered to include an outlier value in at least 30% of samples in the group of interest and for features that had a higher proportion of features in the group of interest compared to the rest of the population. Statistics were calculated using a Fisher's exact test and p-values were corrected using the Benjamini-Hochberg procedure. Druggability of a gene/protein was performed using DGIdbR (Cotto et al., 2018).

Copy number impact on transcriptome and proteome

To evaluate copy number impact on RNA and protein expression, we applied gene-wise correlation analysis on CNV versus RNA expression and on CNV versus protein expression. Correlation was performed by Pearson's correlation method. Both correlation coefficient and p-value were computed and adjusted by the Benjamini-Hochberg procedure.

Cell type enrichment deconvolution using gene expression

The abundance of each cell type was inferred by the xCell web tool (Aran et al., 2017), which performed the cell type enrichment analysis from gene expression data for 64 immune and stromal cell types (default xCell signature). 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.

Immune clustering using cell type enrichment scores

Immune subtypes of the GBM tumors were generated on the consensus clustering of the cell type enrichment scores by xCell ([Wilkerson and Hayes, 2010](#)). Among the 64 cell types tested in xCell, we selected 40 cell types with at least 2 samples with xCell enrichment $p < 0.01$, which filtered out the cell types not typical in the brain. xCell generated an immune score per sample that integrates the enrichment scores B cells, CD4+ T-cells, CD8+ T-cells, DC, eosinophils, macrophages, monocytes, mast cells, neutrophils, and NK cells. In addition, we included microglia using the scores by ssGSEA based on its marker genes: *P2RY12*, *TMEM119*, *SLC2A5*, *TGFBR1*, *GPR34*, *SALL1*, *GAS6*, *MERTK*, *C1QA*, *PROS1*, *CD68*, *ADGRE1*, *AIF1*, *CX3CR1*, *TREM2*, and *ITGAM*. The microglia ssGSEA score was computed using the R package GSVA (gsva function with method='ssgsea'). We performed consensus immune clustering based on the z-score normalized xCell and microglia scores. The consensus clustering was determined by the R package ConsensusClusterPlus (parameters: clusterAlg='kmdist' distance='spearman').

Cell type enrichment deconvolution using protein abundance

We applied CIBERSORTx ([Newman et al., 2019](#)) to compute immune cell fractions from bulk protein abundance. To characterize the immune context using proteomics, we generated a signature matrix based on the dataset from Rieckmann et al. ([Rieckmann et al., 2017](#)). Briefly, 28 distinct human hematopoietic cell types from peripheral blood of healthy donors were sorted by flow cytometry. Erythrocytes and platelets were excluded from subsequent analyses. Cellular proteomes were analyzed in single runs by high-resolution MS using a quadrupole Orbitrap instrument. Each cell type proteome was measured from four donors. The proteomic dataset included 10,134 proteins and 104 steady state samples. The samples were first scaled to have mean zero and standard deviation equal to one. We grouped the 26 subtypes into nine cell types: B cells, basophils, dendritic cells, eosinophils, monocytes, natural killer cells (NKs), neutrophils, CD4+ T-cells and CD8+ T-cells. We took imputed values from [Table S3](#) of the Rieckmann et al. paper to generate a signature matrix of these nine cell types. CIBERSORT was applied to the GBM imputed protein abundance matrix using a batch correction and relative values. CD8+ T-cells cells and NKs were merged by summing their relative values. Z-scores of relative values were used for boxplots and heatmap.

Deep learning histopathology image analysis

We trained deep learning models for 3 different prediction tasks based on histopathology images, including the G-CIMP phenotype (positive and negative), immune response (im4 subtype as immune low and the rest of the tumors as immune high), and telomere length (short, normal, and long). Digital histopathology slides and associated quantified features (cellularity, necrosis, tumor nuclei, age, tumor weight) of samples used in proteomics analysis were downloaded from The Cancer Imaging Archive (TCIA) database. Labels were at per-case (patient) level. The images and their corresponding labels were then divided into 3 datasets at per-case level with 70% of cases in training set, 15% of cases in validation set, and 15% of cases in testing set. Due to the large size of the scanned histopathology slides, they were tiled into 299-by-299-pixel pieces with overlapping area of 49 pixels from each edge at 20X, 10X, and 5X resolution. In this process, tiles with over 30% of background pixels were removed. Qualified tiles, quantified features, and labels of each set were then loaded into a designated TFrecords file. After the data preparation, convolutional neural network (CNN) architectures, including InceptionV1 to V4, InceptionResNetV1 and V2, and self-designed simple CNNs, were trained from scratch. Statistical metrics, such as area under ROC, area under PRC, and top-1 accuracy, were used to evaluate the performance. The best model for each task was picked at the minimum validation loss point. Trained models were tested on the testing set and the statistical metrics of the testing set were used to compare the performance of different models on the same tasks.

A visualization method designed to unveil the features learned by the models was applied to discover histological features associated with G-CIMP phenotype, telomere length, and immune response in the cohort. Firstly, the activation score vectors of each tile from the fully connected layer immediately before the output layer in the testing set were extracted as representation of the input samples. A randomly sampled subset of these activation score vectors was dimensionally reduced into 2-dimensional space by t-SNE with each point representing an image tile. Overlay of prediction scores on these points revealed clusters corresponding to the labels. Finally, experienced pathologists examined the tiles in each of these clusters and summarized the general histological features in these clusters, which served as the representation of the histological features of these subgroups.

Gene set enrichment analysis

Differential Expressed Genes (DEGs) were identified using DESeq2 ([Love et al., 2014](#)) by applying the minimal pre-filtering to keep only genes that have at least 10 reads in total. We selected the genes which had $FDR \leq 0.01$ and absolute fold change larger than 2. To designate the representative pathways of immune subtypes, we selected the DEGs between the two immune subtypes and then underwent a pathway enrichment analysis of Hallmark, KEGG, and Reactome. The overrepresented pathways were selected ($FDR < 0.1$, only pathways with at least 10 genes observed in each data type are considered).

To identify significantly enriched Hallmark, KEGG, PID, and REACTOME gene sets of each immune cluster, we applied the ssGSEA on all proteins to calculate the normalized enrichment score (NES) for each gene set in each sample. Then we performed the pairwise t-test of NES among the 2 immune clusters and adjusted the p-values by FDR. We ranked gene sets by FDR and selected the top 50 gene sets (all $FDR < 0.01$) of each immune cluster.

Histone protein and acetylation calculation

Core histones H2A, H2B, H3 and H4, and linker histone H1 are encoded by multiple genes with minor changes in their sequence. Accordingly, we detected a number of peptides and acetylated peptides corresponding to either of the core histones and H1 histone.

To facilitate the interpretation of histone acetylation events, we averaged acetylation values for peptides mapped on different gene encoding practically the same histone protein.

Histone acetylation association with HATs and HDACs

To test the association between HATs/HDACs protein and acetylation levels of histone sites, we fitted Lasso regression model with HATs/HDACs and histone protein expression as independent variables and a histone acetylation site as a dependent variable. Lasso regression has been chosen because it takes expression of all enzymes into account simultaneously and is insensitive to highly correlated dependent variables. We performed 300 bootstraps with 80% training data and 20% testing data, and reported averaged coefficients returned by the model across 300 iterations.

Pathway enrichment analysis along histones H2B and H3/H4 acetylation axes

We investigated pathways from Hallmark, KEGG, Wikipathways, and REACTOME, positively or negatively aligned with averaged H2B and H3/H4 acetylation level. H2B acetylation was calculated by averaging acetylation of all H2B peptides detected. Since H3 and H4 histones are strongly correlated with each other, we averaged acetylation of histones H3 and H4 peptides together to obtain H3/H4 acetylation value.

We assumed that true biological activity of a pathway is regulated by collective changes of expression levels of majority of proteins involved in this pathway; then a difference in a pathway activity between tumors can be assessed by a difference in positioning of expression levels of proteins involved in this pathway in ranked list of expression levels of all proteins in each of tumors. Following this idea, we assessed relative positioning of pathway proteins between tumor by determining two probabilities: (1) probability of pathway proteins to occupy by random chance the observed positions in a list of tumor proteins ranked by expression level from the top to the bottom and, similarly, (2) probability to occupy by random the observed positions in a list of expression levels ranked from the bottom to the top. Then, the inferred relative activation of a given pathway across tumors was assessed as a negative logarithm of the ratio of the above “top” and “bottom” probabilities. Thus, for a pathway of a single protein, its relative activity across tumors was assessed as a negative log of ratio of two numbers: a number of proteins with expression level bigger than an expression level of given protein, and a number of proteins with expression levels less than an expression level of given protein. For pathways of multiple proteins, the “top” and “bottom” probabilities were computed as geometrically averaged P values computed for each of proteins using Fisher’s exact test, given protein’s ranks in a list of pathway proteins and in a list of ranked proteins of a tumor, a number of proteins in a pathway, and the total number of proteins with the assessed expression level in a given tumor. The thermodynamic interpretation of the inferred pathway activity scoring function is a free energy associated with deviation of the system from equilibrium either as a result of activation or suppression. Thus, the scoring function is positive, when expression levels of pathway’s proteins are overrepresented among top expressed proteins of a tumor, and it is negative, when pathway’s proteins are at the bottom of expressed proteins of a tumor; the scoring function is close to zero, when expression levels are distributed by random. Given any biological axis, e.g. histone acetylation levels in each of tumors, one can determine pathways which are significantly correlated or anti correlated with the axis.

Causative pathway interaction discovery using CausalPath

To discover the causative pathway interactions in our proteomic and phosphoproteomic data, we took the normalized expression of protein with < 10% missing values and phosphoprotein with < 25% missing values across all tumor and normal samples as the input to CausalPath (commit 7c5b934). We ran CausalPath in the mode that tests the mean values between test and control groups (value-transformation = significant-change-of-mean), where the test group being the tumors of one subtype and control group being the rest of the tumors. The pathway interaction discovery data source was Pathway Commons v9 (built-in-network-resource-selection = PC). Additionally, we enabled the causal reasoning if all the downstream targets of a gene were active or inactive (calculate-network-significance = true, use-network-significance-for-causal-reasoning = true, permutations-for-significance = 10000). The causative interactions with FDR < 0.05 were extracted and visualized (fdr-threshold-for-data-significance = 0.05 phosphoprotein, fdr-threshold-for-data-significance = 0.05 protein, fdr-threshold-for-network-significance = 0.05). Full result tables were available in [Table S5](#).

L1000 and P100 drug connectivity analysis

For mRNA abundance, RNA-seq read counts were used to perform differential expression analysis between gene-altered and WT samples using edgeR ([Robinson et al., 2010](#)). The significantly differentially expressed genes ($FC \geq 1$ and $FDR < 0.05$) were then used as input in the subsequent analysis. For protein and phosphoprotein abundance, a Wilcoxon rank sum test was performed to determine differential abundance of protein and phosphoprotein measurements between gene altered and WT samples. Protein with $FDR < 0.05$ were considered as differentially expressed.

The differentially expressed genes between gene-altered and WT samples were filtered for the 978 genes measured in the L1000 assay and then were processed using the CLUE ([Subramanian et al., 2017](#)) (summary connectivity score) and iLINCS ([Pilarczyk et al., 2019](#)) connectivity algorithms. The resulting drug connectivities were aggregated to the compound level using the summary connectivity score in CLUE and the Connected Perturbations Z-score in iLINCS. Target annotations for the ranked compounds were extracted from CLUE and iLINCS and combined in a single list.

Level 4 P100 data were downloaded from the LINCS Data Portal ([Stathias et al., 2019](#)) and were used to calculate drug connectivities on the phosphoprotein level as previously described ([Litichevskiy et al., 2018](#)). Firstly, the P100 probes were mapped to the CPTAC phosphopeptides using their respective modified peptide sequences. The spearman correlation was then calculated between the phosphoproteomic mutation signature and each drug-treated P100 experiment. The correlation was then converted to a connectivity score by comparing the distribution of the observed drug-signature correlations to a background distribution

of correlations using the two-sample Kolmogorov-Smirnov (KS) test. The directionality of the connectivity score (positive/negative) was determined by the comparison of the medians of the two distributions (The score is negative when the median of the observed distribution is lower than the one of the background distribution). Significant connectivity scores (p -value < 0.05) were then aggregated to the compound level by their mean across all tested cell lines.

Full result tables were available in [Table S6](#).

Supplemental information

Proteogenomic and metabolomic characterization of human glioblastoma

Liang-Bo Wang, Alla Karpova, Marina A. Gritsenko, Jennifer E. Kyle, Song Cao, Yize Li, Dmitry Rykunov, Antonio Colaprico, Joseph H. Rothstein, Runyu Hong, Vasileios Stathias, MacIntosh Cornwell, Francesca Petralia, Yige Wu, Boris Reva, Karsten Krug, Pietro Pugliese, Emily Kawaler, Lindsey K. Olsen, Wen-Wei Liang, Xiaoyu Song, Yongchao Dou, Michael C. Wendl, Wagama Caravan, Wenke Liu, Daniel Cui Zhou, Jiayi Ji, Chia-Feng Tsai, Vladislav A. Petyuk, Jamie Moon, Weiping Ma, Rosalie K. Chu, Karl K. Weitz, Ronald J. Moore, Matthew E. Monroe, Rui Zhao, Xiaolu Yang, Seungyeul Yoo, Azra Krek, Alexis Demopoulos, Houxiang Zhu, Matthew A. Wyczalkowski, Joshua F. McMichael, Brittany L. Henderson, Caleb M. Lindgren, Hannah Boekweg, Shuangjia Lu, Jessika Baral, Lijun Yao, Kelly G. Stratton, Lisa M. Brammer, Erika Zink, Sneha P. Couvillion, Kent J. Bloodsworth, Shankha Satpathy, Weiva Sieh, Simina M. Boca, Stephan Schürer, Feng Chen, Maciej Wiznerowicz, Karen A. Ketchum, Emily S. Boja, Christopher R. Kinsinger, Ana I. Robles, Tara Hiltke, Mathangi Thiagarajan, Alexey I. Nesvizhskii, Bing Zhang, D.R. Mani, Michele Ceccarelli, Xi S. Chen, Sandra L. Cottingham, Qing Kay Li, Albert H. Kim, David Fenyö, Kelly V. Ruggles, Henry Rodriguez, Mehdi Mesri, Samuel H. Payne, Adam C. Resnick, Pei Wang, Richard D. Smith, Antonio Iavarone, Milan G. Chheda, Jill S. Barnholtz-Sloan, Karin D. Rodland, Tao Liu, Li Ding, and Clinical Proteomic Tumor Analysis Consortium

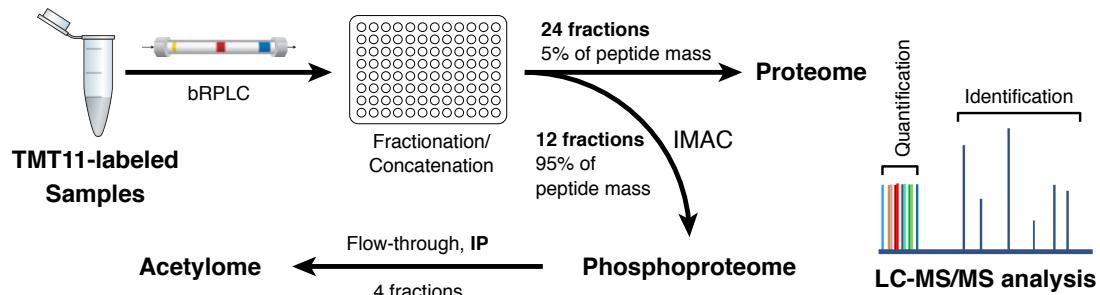
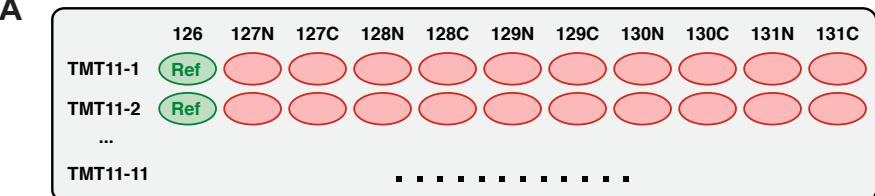
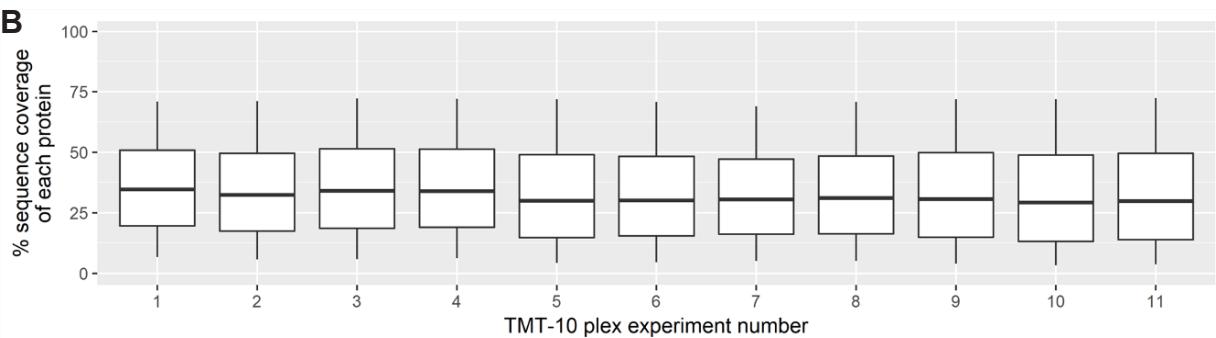
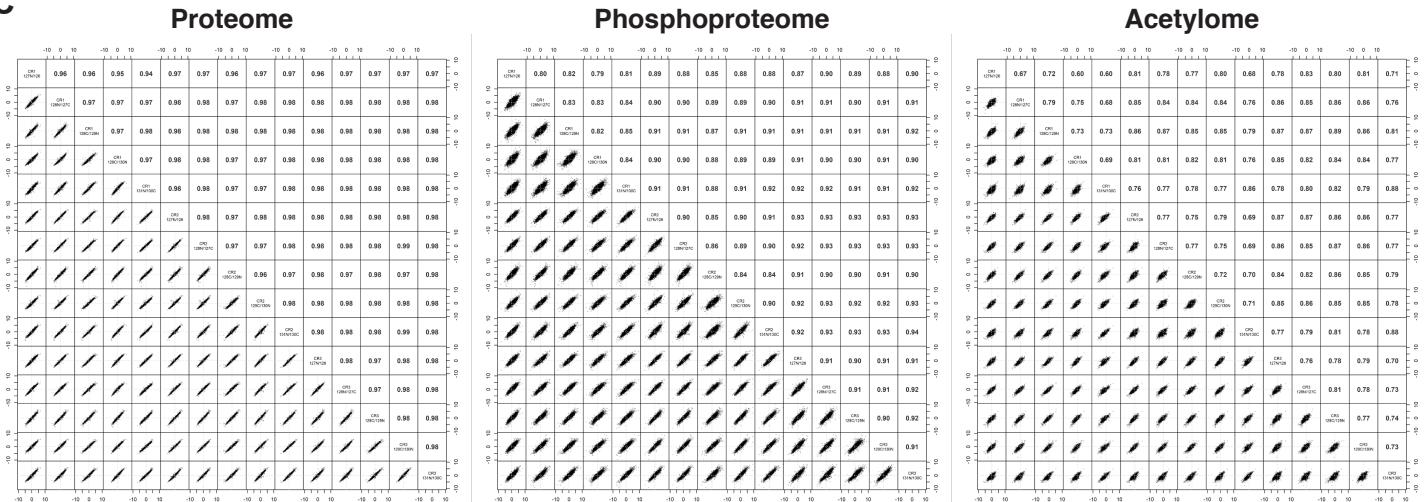
A**B****C**

Figure S1. Integrated Proteomic Workflow and Quality Control, Related to Figure 1.

A) TMT-11 based global proteome, phosphoproteome, and acetylome analysis workflow. The GBM tumors and GTEx normal tissues were analyzed in 11 TMT 11-plex experiments, each with 10 study samples and a common internal reference sample created by pooling all study samples (equal contribution). The TMT11-labeled samples were then fractionated, split (with 5% peptide mass analyzed directly for global proteome), and subjected to tandem enrichment of phosphopeptides and acetylated peptides. Peptides were detected and quantified using information from the TMT-11 MS/MS spectra.

B) Distribution of sequence coverage of the identified proteins with tryptic peptides detected by MS/MS in each TMT-11 plex; whiskers show the 5–95 percentiles. The total numbers of identified proteins in each TMT plex range from 9,992 to 10,572.

C) Robust and precise proteomics platforms. Longitudinal performance was tested by repeated proteome, phosphoproteome and acetylome analysis of aliquots of the same patient-derived xenograft QC samples in standalone TMT-11 plexes, along with the GBM study samples; scatter plots and Pearson correlations comparing individual replicate measurements are shown.

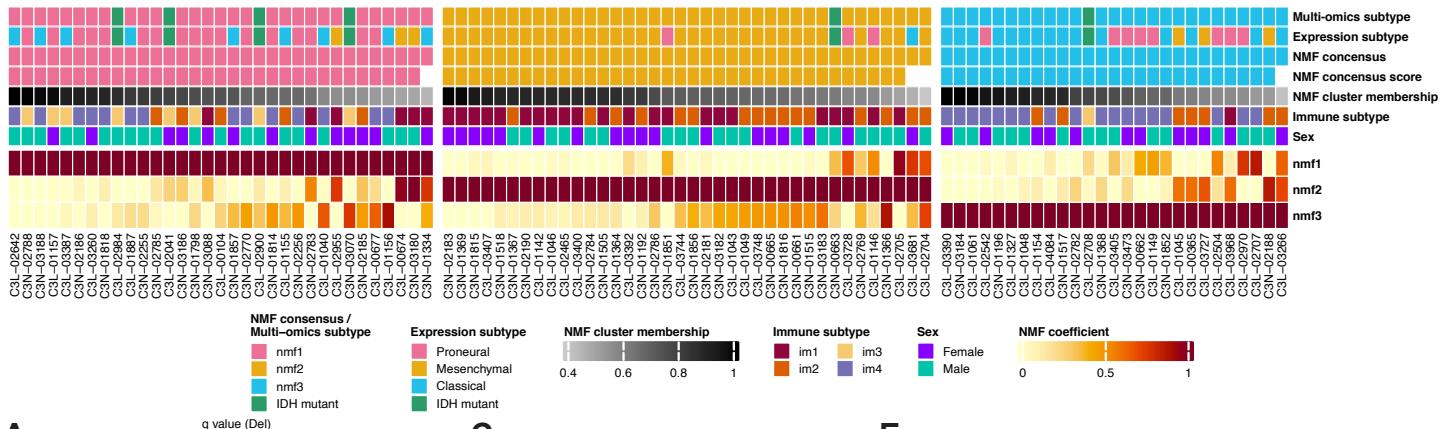
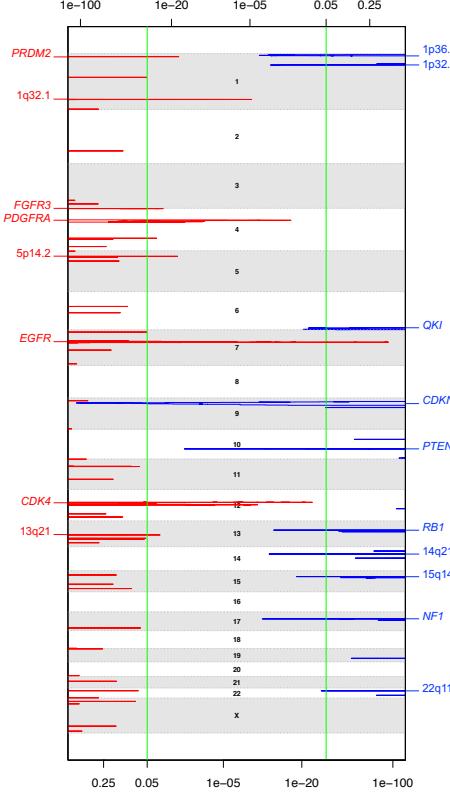
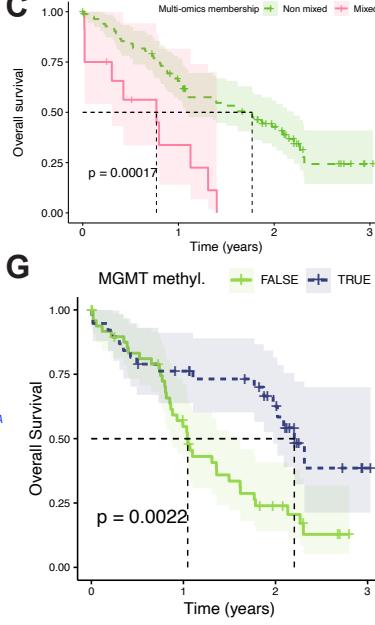
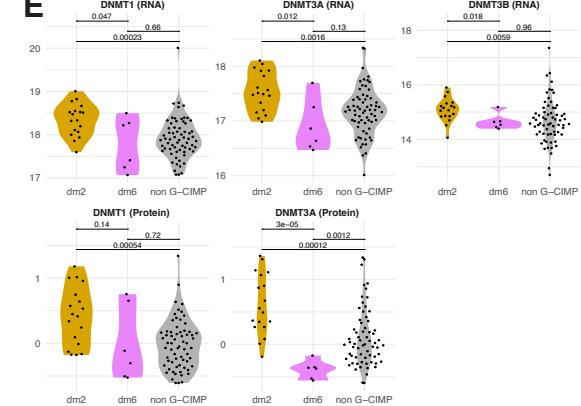
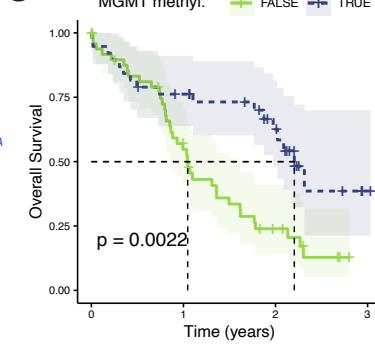
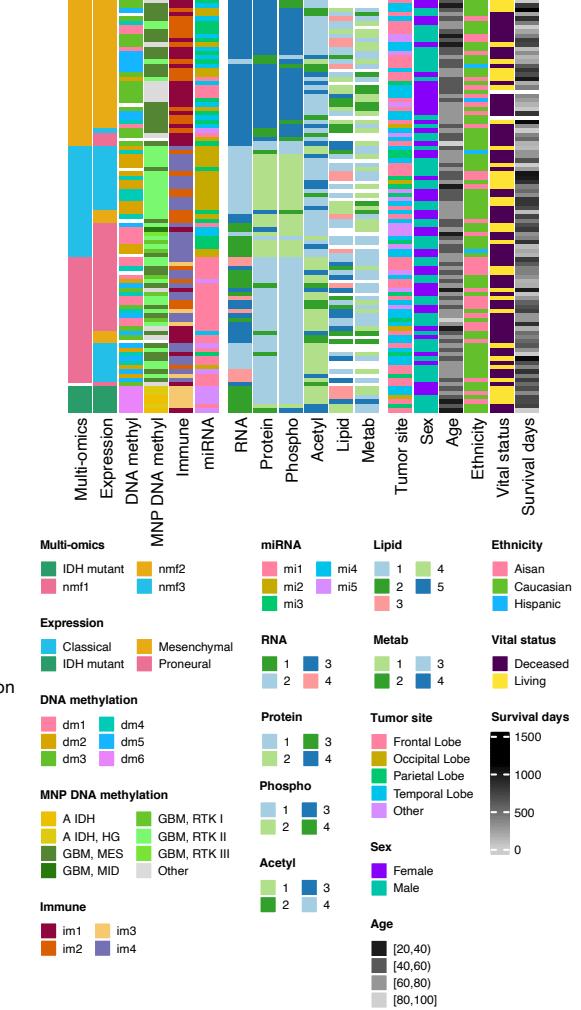
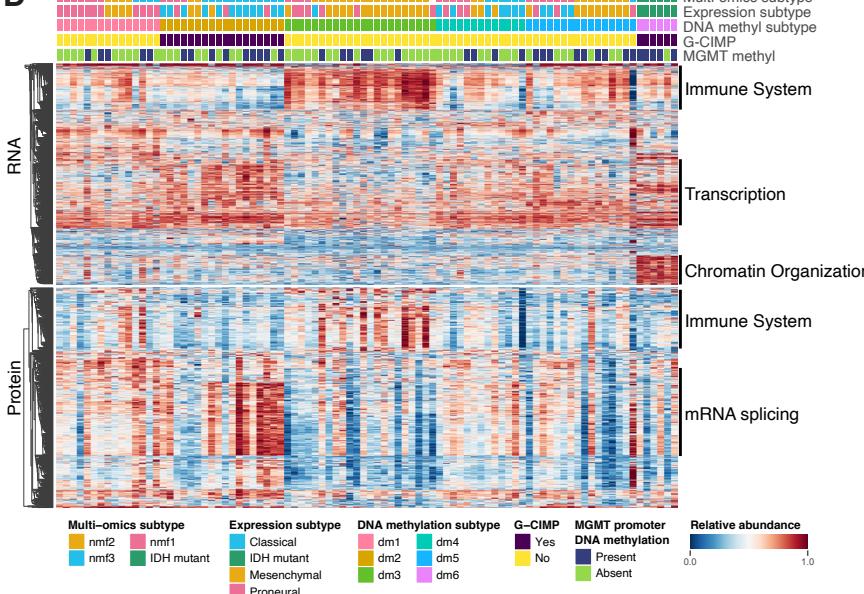
B**A****C****E****G****F****D**

Figure S2. Proteogenomic Characterization and Subtyping of GBM, Related to Figure 1.

- A) GISTIC2 copy number variations at and focal level of all tumors with WGS available.
- B) Heatmap of multi-omics membership scores of all three multi-omics subtypes for each tumor.
- C) Survival Kaplan–Meier curves of mixed-subtype tumors with low multi-omics membership score versus the rest of the tumors.
- D) Pathway enrichment analysis of gene expression with differential DNA methylation in promoter regions across all six DNA methylation subtypes.
- E) Differentially expressed DNMTs in two G-CIMP subtypes (dm2 and dm6) at both RNA and protein level.
- F) Comparison of the clustering results using individual data type, including DNA methylation, miRNAs, RNAs, proteins, phosphoproteins, acetylproteins, lipids, and metabolites.
- G) Survival Kaplan–Meier curves of tumors with different MGMT promoter DNA methylation status.

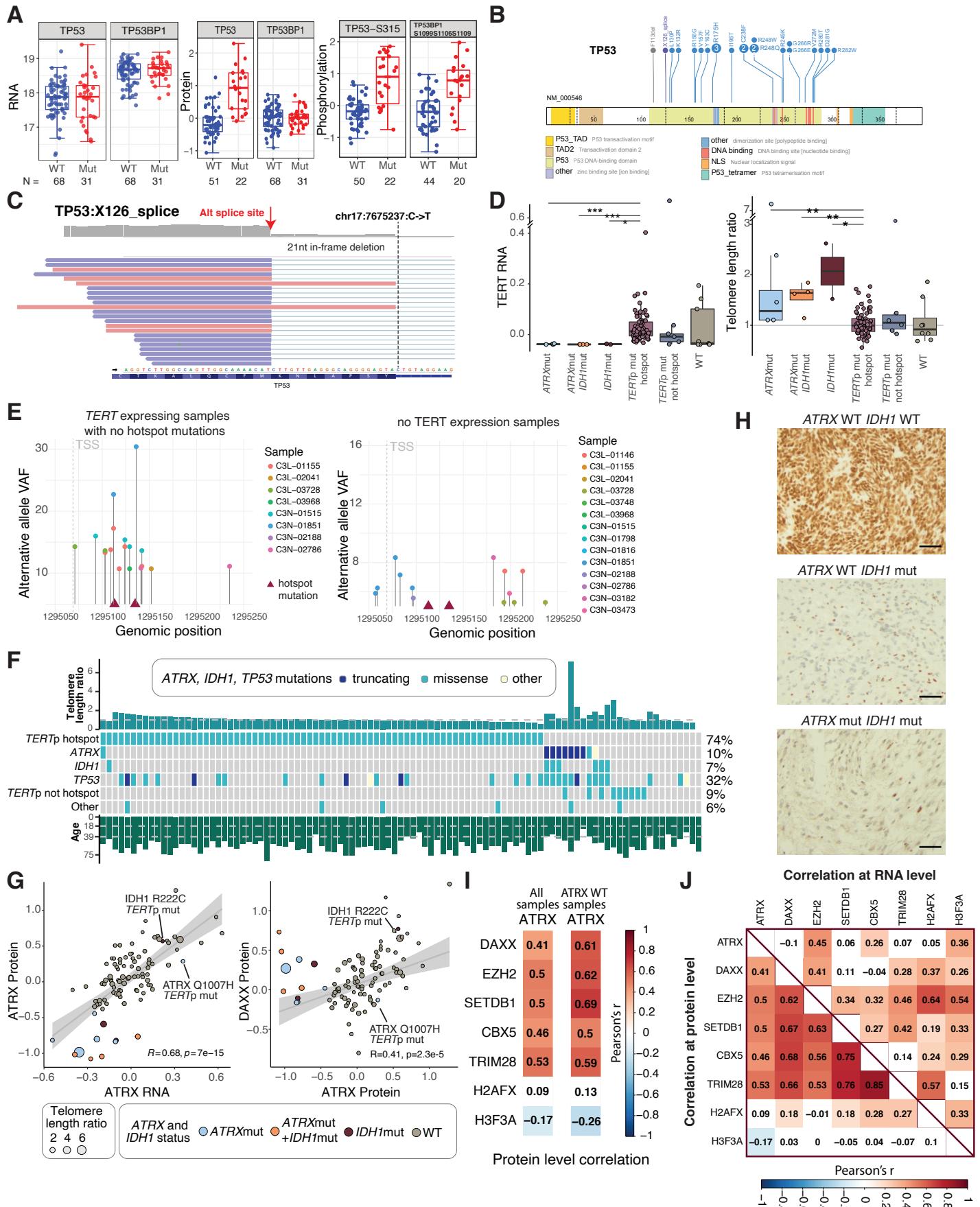


Figure S3. Effect of *TP53*, *TERTp*, *IDH1*, and *ATRX* Mutations, Related to Figure 2.

A) Boxplots showing RNA, protein and phosphosite abundance for *TP53* and *TP53BP1* in *TP53*-WT tumors (blue; n = 68) and *TP53*-mutated tumors (red; n = 31).

B) The lollipop of *TP53* mutations in **Figure 2C**.

C) IGV screenshot of RNA supporting reads for alternative splice site associated with *TP53* X126_splice mutation.

D) *TERT* RNA expression in proposed telomere genotypes. *TERTp* mutations are associated with stable *TERT* expression (left). Telomere length ratio (WGS tumor telomere length / WGS blood normal telomere length) in proposed telomere genotypes (right).

E) Variant allele frequency of mutations in *TERTp* region found in tumors not carrying *TERTp* hotspot mutations but expressing *TERT* RNA (upper panel) and in tumors not carrying *TERTp* hotspot mutations and no expressing *TERT* RNA (bottom panel). For the upper panel mutations with > 10% VAF are shown, for the bottom panel mutations with > 5% VAF are shown. Triangles indicate hotspot genomic position 1295113 and 1295135 (-124 and -146 bp from TSS labeled by the dashed line).

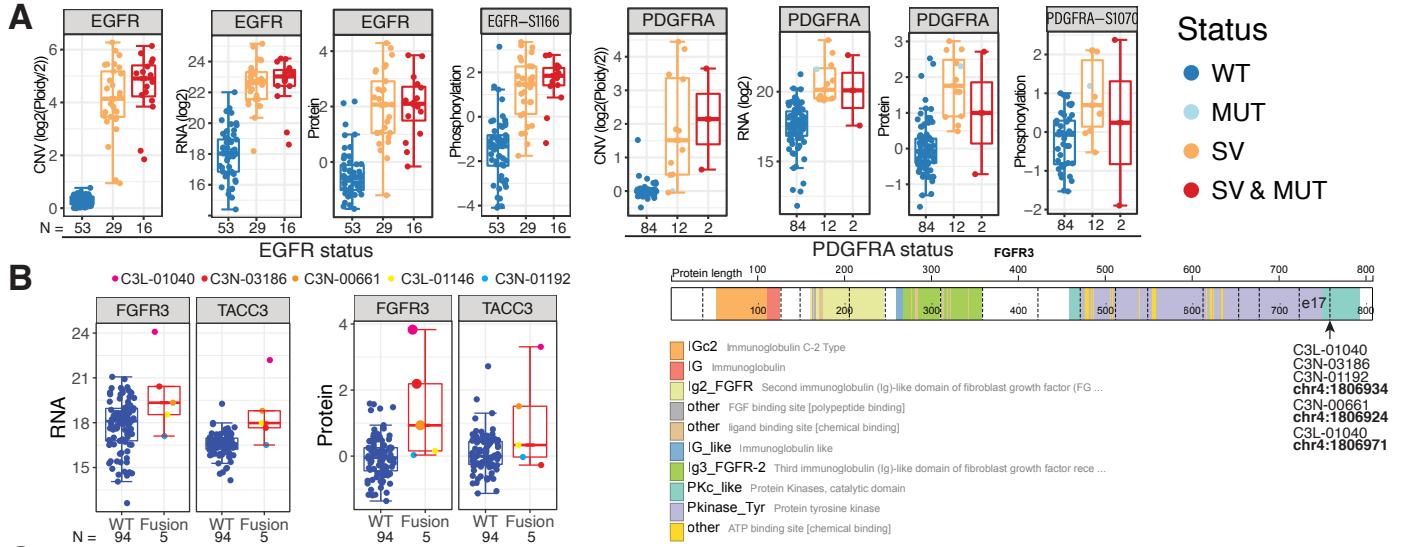
F) Overview of genetic alterations in *TERTp* hotspot and not hotspot positions, mutations in *ATRX*, *IDH1* and *TP53*. Telomere length ratio in tumor/blood normal is indicated on top, patient age is indicated in the bottom.

G) *ATRX* and *IDH1* mutational effect on *ATRX* RNA and protein levels. *IDH1* R132H mutations affect *ATRX* protein only (left). Neither *ATRX* nor *IDH1* mutations are associated with reduced DAXX protein expression (right). The missense *ATRX* mutant, Q1007H, residing outside of the SNF2 and helicase domains, had normal RNA and protein expression. We noticed that other gene mutations also led to altered *ATRX* protein levels, including significantly decreased *ATRX* protein in *IDH1* R132H mutants, even though the *ATRX* RNA level itself appeared unchanged

H) Immunohistochemistry staining for *ATRX* protein in FFPE sections of *ATRX* WT and *IDH1* WT tumor (positive), *ATRX* WT and *IDH1*mut tumor (negative), and in *ATRX*mut and *IDH1*mut tumor (negative). IHC staining confirmed low *ATRX* expression in *IDH1* mutants regardless of *ATRX* mutational status. Scale bars: 100 µm.

I) WT *ATRX* protein is better correlated with its complex partners than if WT and mutant combined together.

J) *ATRX* complex members showed high correlation of their abundance in protein but not in RNA across all tumors.



Protein length 100 200 300 400 500 600 700 800

IGc2 Immunoglobulin C-2 Type

IG Immunoglobulin

lg2_FGFR Second immunoglobulin (Ig)-like domain of fibroblast growth factor receptor

Other FGF binding site [polypeptide binding]

Other ligand binding site [chemical binding]

IG_like Immunoglobulin like

lg3_FGFR-2 Third immunoglobulin (Ig)-like domain of fibroblast growth factor receptor

PKc_like Protein Kinases, catalytic domain

Pkinase_Tyr Protein tyrosine kinase

Other ATP binding site [chemical binding]

C3L-01040

C3N-03186

C3N-01192

chr4:1806934

C3N-00661

chr4:1806924

C3L-01040

chr4:1806971

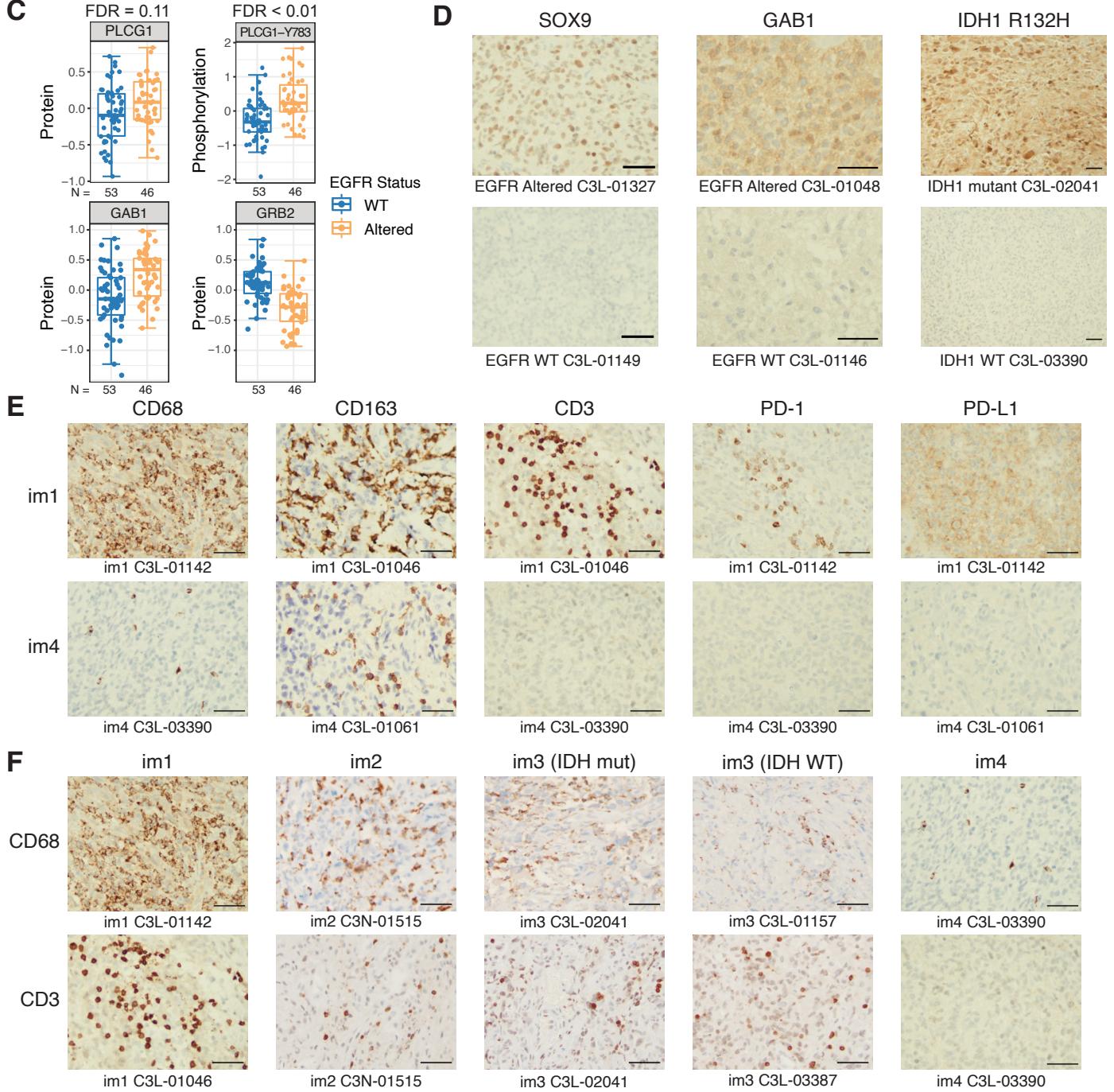


Figure S4. CNV, RNA, Protein and Phosphosite Level Change in Samples with Different RTK Alterations, and Immunohistochemistry Staining Results, Related to Figures 2, 3 and 4.

A) The comparison of CNV, RNA and protein expressions and phosphosite level between *EGFR* altered and WT samples (left panel) and *PDGFRA* altered and WT samples (right panel). We did not find expression differences between samples having a sole SV event versus those with dual mutation and SV events, suggesting EGFR upregulation in GBM is largely due to SV, as associated with CNV amplification, rather than mutationally driven, which is different from other tumor types, such as lung cancer (Cancer Genome Atlas Research Network, 2014). We also found nine samples in which *EGFR* SV co-occurred with *PDGFRA* or *FGFR3* SV, while 13 samples with either *PDGFRA* or *FGFR3* alteration did not show any alterations in *EGFR*. For *PDGFRA*, two out of three mutations overlap with SV events. Only one sample with mutation in *PDGFRA* had high *PDGFRA* RNA and protein expression. For EGFR, n = 53 (WT), 29 (SV), and 16 (SV and MUT); for PDGFRA, n = 84 (WT), 12 (SV), and 2 (SV and MUT).

B) The comparison of RNA and protein expression of FGFR3 and TACC3 between samples with and without FGFR3-TACC3 fusion and the breakpoints found in *FGFR3* from RNA-Seq data (n=94 and 5, respectively). Five samples with FGFR3-TACC3 fusions with an intact FGFR3 kinase domain. Three samples which are protein expression outliers in FGFR3 are marked by larger circles.

C) The comparison of PLCG1, GAB1, and GRB2 protein expression and PLCG1-Y783 phosphosite level between *EGFR* altered (n = 46) and WT (n = 53) samples.

D) Immunohistochemistry staining for SOX9 and GAB1 expression in *EGFR* altered and WT tumors is concordant with the mass spectrometry findings. Positive *IDH1* R132H staining of the ATRX WT and *IDH1* mutant tumor shown in **Figure 2H**. Scale bars: 100 μm.

E) Immunohistochemistry staining for CD68, CD163, CD3, PD-1, PD-L1 in tumors of different immune subtypes is concordant with the mass spectrometry and gene expression analyses. Scale bars: 100 μm.

F) Immunohistochemistry staining for CD68 and CD3 in tumors of all four immune subtypes is concordant with the mass spectrometry and gene expression analyses. Scale bars: 100 μm.

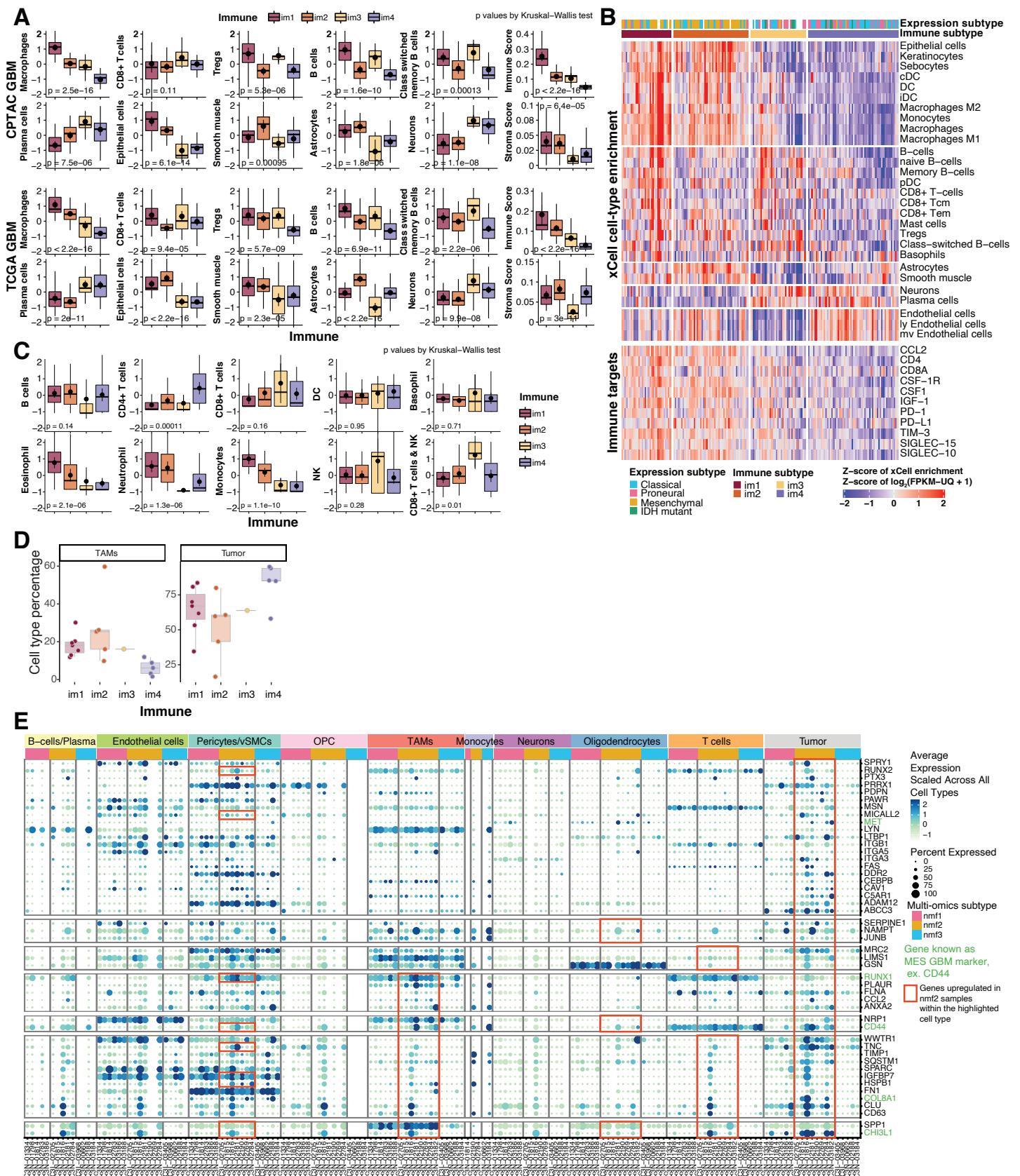


Figure S5. Distinct Immune Marker Expression Ascertains Immune Subtypes, Related to Figure 4.

- A) The distributions of xCell cell type enrichments, immune scores, and stroma scores among the 4 immune subtypes in the CPTAC GBM cohort and TCGA GBM cohort. For CPTAC, n = 30 (im1), 28 (im2), 9 (im3), and 32 (im4); for TCGA, n = 32 (im1), 48 (im2), 35 (im3), and 59 (im4).
- B) The 4 immune subtypes identified by the consensus clustering on the cell type enrichment in the TCGA GBM study.
- C) The distributions of protein-based CIBERSORT cell type enrichments among the 4 immune subtypes in the CPTAC GBM cohort. n = 30 (im1), 28 (im2), 9 (im3), and 32 (im4).
- D) snRNA-seq cell type percentage per sample in each immune subtypes.
- E) Expression of epithelial–mesenchymal transition genes from different cell types in different snRNA-seq samples. Genes upregulated (Wilcoxon FDR < 0.05 and $\log_2\text{FC} > 0.25$) in tumor cells of nmf2 samples. Genes were also tested for significant upregulation in TAMs, T-cells, oligodendrocytes and pericytes/vSMCs; the same thresholds were applied to genes highlighted in red boxes. Dot size denotes the percentage of cells expressing a gene in the given cell type.

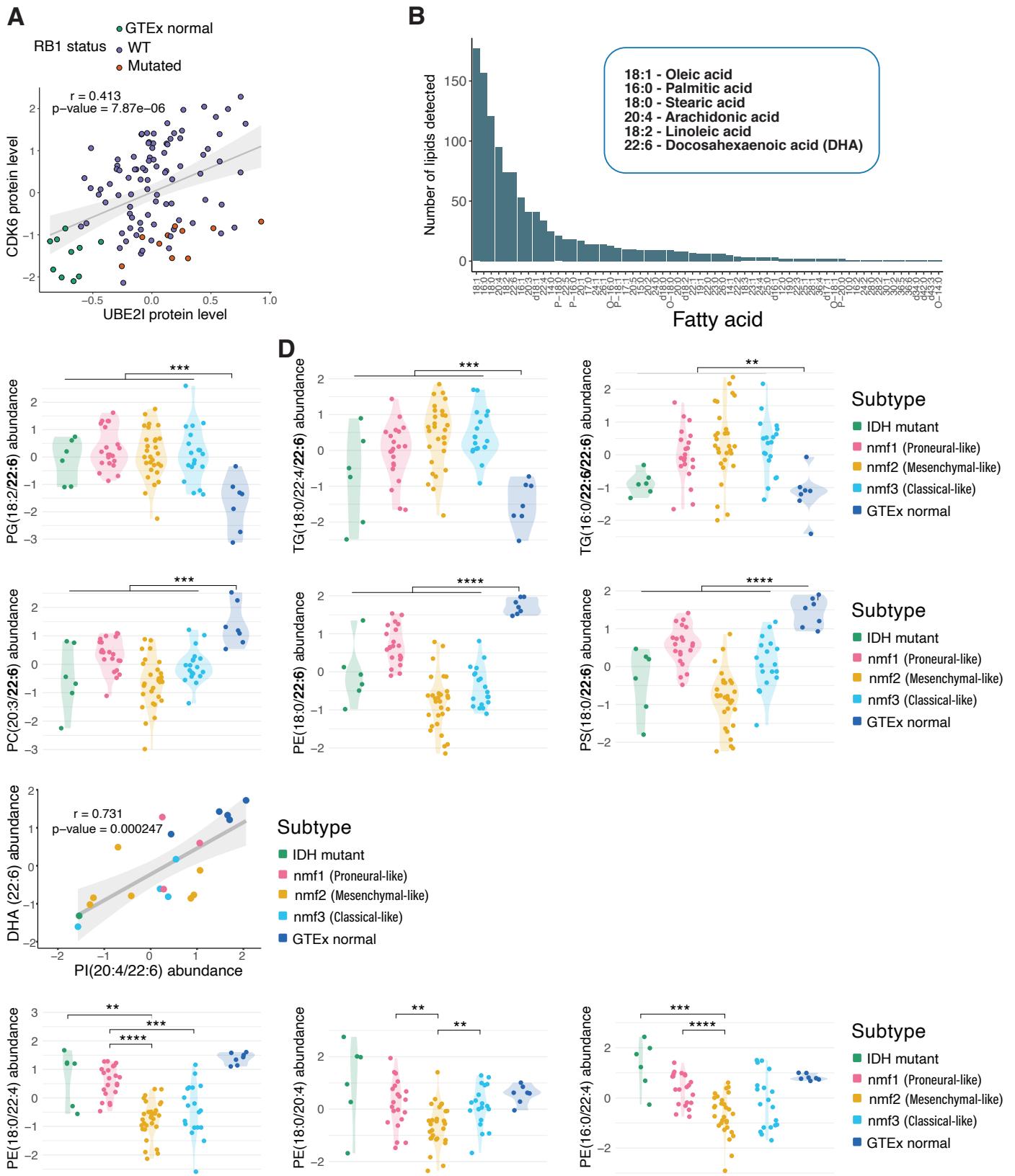


Figure S6. Relations Between H2B Acetylation and SUMOylation Pathway, Related to Figure 5; Additional Lipid Abundance Evidence, Related to Figure 6.

- A) Scatterplot showing correlation between UBE2I protein level and CDK6 protein level. Pearson's correlation value and *p*-value are shown. RB1 mutated samples are labeled in dark orange.
- B) Frequency of the side chains observed in the detected lipids.
- C) Abundance of phosphatidylglycerol PG(18:2/22:6) across multi-omics subtypes and GTEx normal samples. 22:6 is likely DHA. Wilcoxon test *p*-value is labeled.
- D) Abundance of 22:6-carrying TGs across multi-omics subtypes and GTEx normal samples. 22:6 is likely DHA. Wilcoxon test *p*-value is labeled.
- E) Abundance of 22:6-carrying PC, PE and PS across multi-omics subtypes and GTEx normal samples. 22:6 is likely DHA. Wilcoxon test *p*-value is labeled.
- F) Scatterplot showing correlation between DHA abundance and PI(20:4/22:6) abundance. Pearson's correlation value and *p*-value are shown.
- G) PEs carrying 22:4 or 20:4 fatty acids are downregulated in nmf2/mesenchymal-like subtype, while PE not containing either of the two FAs is upregulated in IDH1 mutant subtype only.

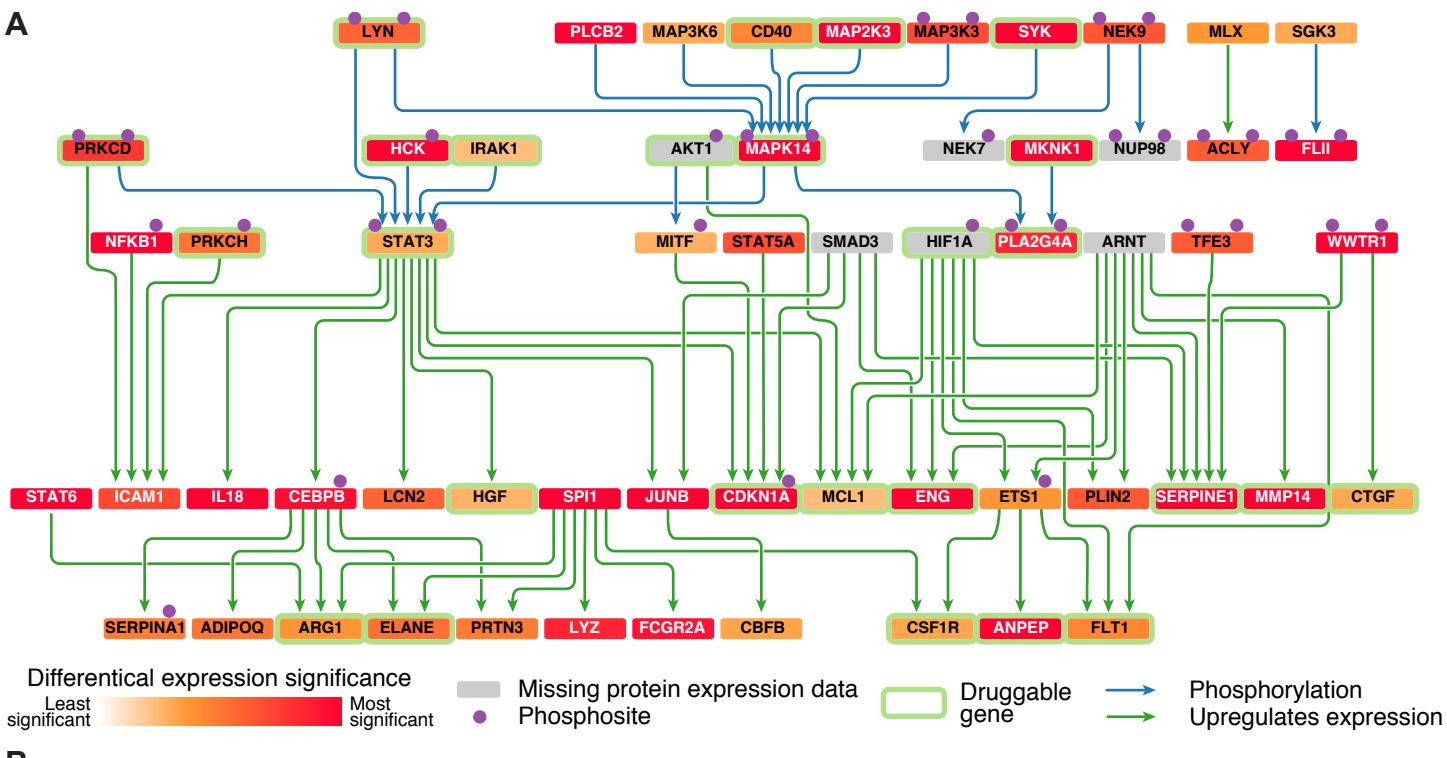
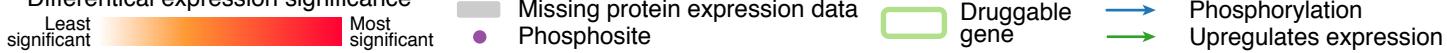
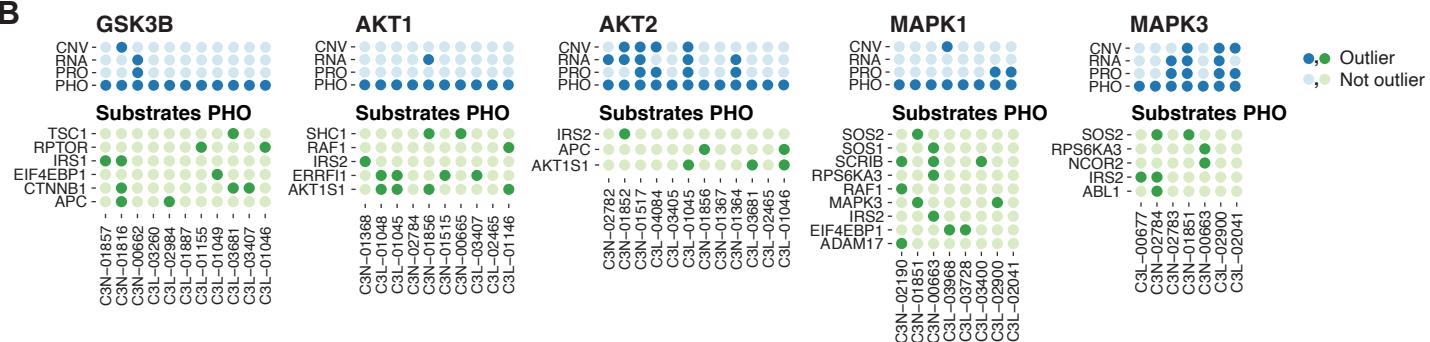
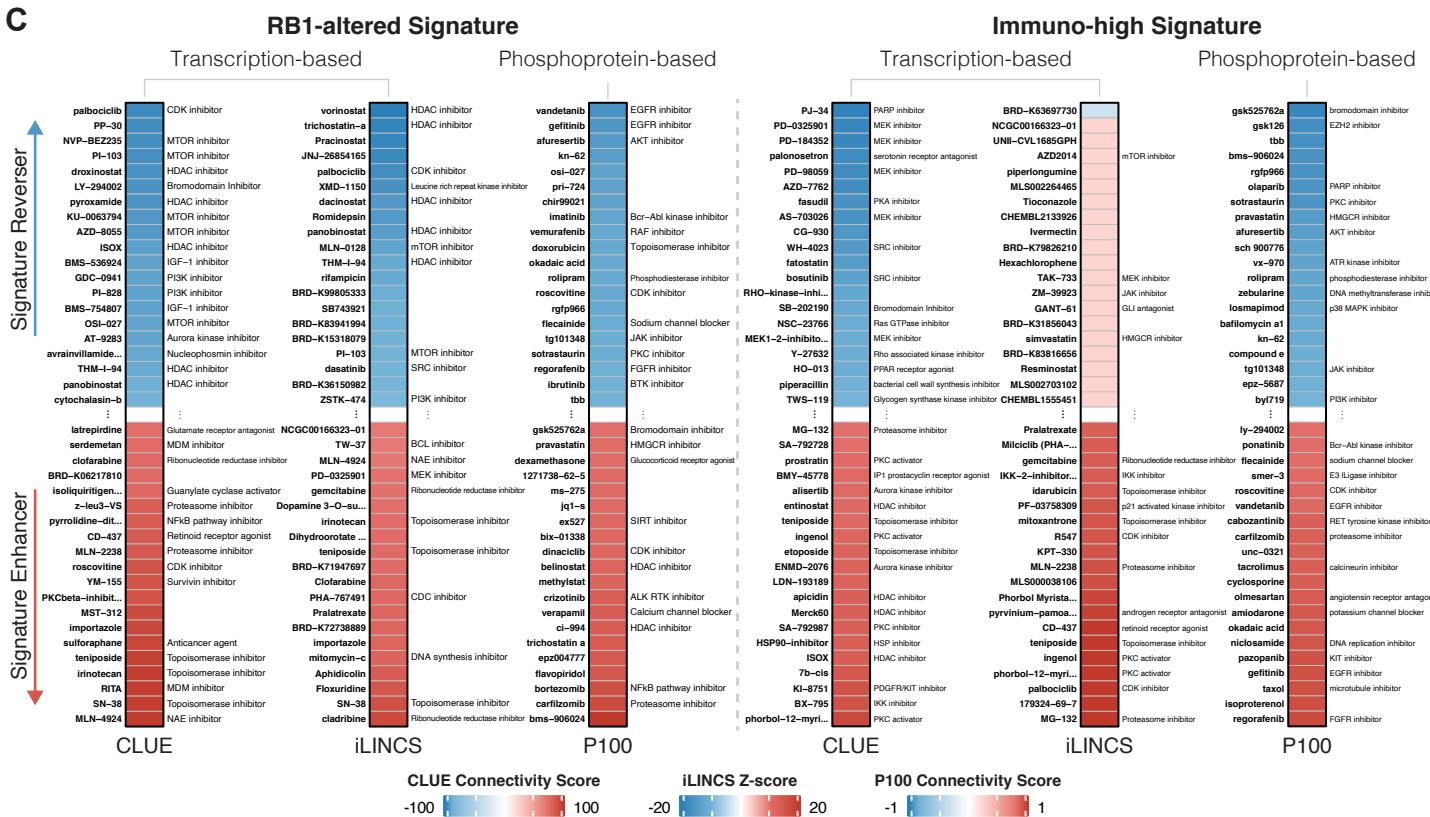
A**Differential expression significance****B****C**

Figure S7. Summary of Pathway Alterations of Mesenchymal Tumors Compared to the Other Tumors, Related to Figure 7.

A) Causal explanations for differentially expressed proteomic and phosphoproteomic profiles in mesenchymal tumors versus the rest of the tumors using CausalPath. We highlighted the immune and hypoxia-related interactions enriched in the graph. Genes controlling the immune response might be a potential therapeutic target. Complete result graph is given in **Table S5**.

B) Sample level outlier status for kinase-substrate pairs shown in **Figure 7B**. For kinases (blue) outlier status for four levels of data is shown: CNV, RNA, PRO (protein) and PHO (phosphorylation). For their substrates (green) only PHO outlier status is shown.

C) Drug connectivity analysis using alteration-specific transcriptional (CLUE and iLINCS) and phosphoproteomic (P100) signatures of RB1 altered versus WT tumors and immune-high (im1, im2, and im3) versus immune-low (im4) tumors.